

# Predictive Models of BAC among Nighttime Drivers

Wilson, RJ; Chen, GG

Insurance Corporation of British Columbia, PO Box 3750, Victoria, BC

**Keywords:** roadside surveys, drinking driving, driver characteristics

## Abstract

The purpose of the study was to determine the combination of factors which explain and classify the BAC of drivers encountered in a nighttime roadside survey conducted in the Spring and Fall of 1998 in British Columbia. Implications for detecting impaired drivers on the road are discussed.

## Introduction

Roadside surveys, which measure the blood alcohol concentration (BAC) of nighttime drivers, provide the most representative, objective and valid measures of drinking and driving. The surveys are primarily used to measure a change in the level of drinking-driving over time, usually over a period of years, but occasionally after a few months (Beirness, Foss and Mercer, 1997). These surveys have also been useful in identifying the characteristics of nighttime drinking drivers. They have consistently demonstrated several factors associated with higher BACs. These factors include: driver sex (male), driver age (highest in the mid-range, 21-45, but varies by country and region), time of night (increases with advancing hour), trip origin (bar or tavern), trip length (shorter distance), seatbelt non-use, vehicle type (pick-up truck), and alcohol consumption (greater quantity and/or frequency) (Beirness, Foss and des Groseilliers, 1999; Foss et al., 1993; Kruger et al, 1995; Lawson and Stewart, 1987; Paquette, 1991; Perrine, Meyers and Yu, 1995; Quaye, 1994; Voas et al., 1997). In general, previous studies have taken a uni-variate approach to assessing the relationship between driver, vehicle and environmental factors driver BAC. One exception is a study by Perrine, Myers and Yu which used a stepwise regression approach to predict roadside driver BAC from a set of variables which included time of interview, alcohol consumption, an alcohol problem scale, self-reported DUI and actual driving record of alcohol and traffic related offenses. Despite the use of highly relevant variables, the total  $R^2$  of the model was very modest (.23) indicating a large proportion of variance unexplained. The present study represents a second attempt to predict the roadside BAC of drivers from a set of independent variables. The purpose of the exercise is twofold: First, a model which can explain drinking driving as a function of driver, attitudinal, situational, social and environmental influences has theoretical value in explaining the behavior. Secondly, to the extent that a subset of those variables are directly observable in an enforcement setting, knowledge of the combination of factors most highly associated with impairment may enhance existing detection and enforcement practices. The usefulness of such a model depends on its power to predict BAC and its applicability to an enforcement context. This study used both a logistic regression approach and a discriminant approach to explain and predict the BAC of nighttime drivers. Logistic regression identifies the unique contribution of predictor variables to explaining drivers' BAC (as a bivariate criterion). A discriminant model was developed to identify the latent dimensions underlying differences among drivers from three BAC levels. A classification function was then used to classify those drivers into the 3 groups. Finally the

discriminant model was applied to drivers who refused to provide a breath sample, to determine the distribution of their predicted BAC group.

## **Methods**

**Survey methods.** The roadside survey methods are described in detail by Beirness, Foss, Wilson and Mercer (these proceedings). The surveys took place in three communities in British Columbia: Vancouver, Saanich and Kamloops, in the Spring and again in the Fall of 1998. The surveys took place from Wednesday through Saturday evenings from 9:00 pm to 1:30 am. Uniformed police officers waved drivers into the survey site, where civilian interviewers requested their participation in the survey. Those drivers agreeing to participate were asked a short series of questions then asked to provide a breath sample into a small hand-held device (Intoxilyzer S-D2). A total of 5,341 vehicle drivers were selected from the traffic stream of whom 4,841 (90.6%) provided breath samples.

**Statistical design and procedures.** Preliminary screening was conducted to examine the sample distributions, presence of outliers and intercorrelations among the variables. Violation of the assumption of multivariate normal distribution was not considered serious because of the large sample size. The non-linearity of variables was dealt with by creating dummy or categorical variables. All intercorrelations among independent variables were low.

A logistic regression was performed using the SAS GENMOD and LOGISTIC procedures. A logit link function was employed with a binomial error distribution assumption. Maximum likelihood procedure was used for the model estimation. The model provides the unique contribution of each predictor, controlling for all other predictors in the model.

The criterion variable is BAC. Because the distribution of BAC is heavily skewed to the low end (86.2% with 0 mg%), it was reduced to a binary variable with the values: BAC 0-49 mg% and BAC 50 mg% and above. All relevant main effects and potential interactions were pre-tested in the model. The nonsignificant variables and interactions were eliminated from the final model. The final variable set included driver age, sex and education; trip origin, passenger configuration and time of night. Measures of driver seat belt use, vehicle type and model year, prior experience going through drinking-driving check stops, opinion on the effectiveness of drinking-driving check stops, knowledge of recent drinking-driving laws were initially included, but dropped from the final model because they failed to be significant at the 0.05 level.

All significant independent variables from the logistic regression were used in subsequent discriminant analyses, using procedures from SAS PROC DISCRIM. For these analyses drivers' BACs were classified as either Zero (0-5 mg%), mid-range (5-80 mg%) or impaired (over 80 mg%). Canonical discriminant analysis was applied to summarize the variation as latent factors or dimensions underlying the predictors. While the canonical functions are used to explain the data structure, separate classification functions were derived by PROC DISCRIM and used to test the classification performance of the discriminant model. In order to validate the classification model, the total data set was randomly divided into two equal size groups (50% probability samples) one for model estimation, and the other for model validation. Thus the classification model was developed from the estimation data set, and tested on the validation data set.

Finally, the coefficients derived from the classification model were applied to the drivers who provided an interview but refused, for various stated reasons, to provide a breath test. For this group of 194 drivers, the predictor variables were available but the criterion variable was unknown. Based on their weighted discriminant scores, derived from the model, they were assigned to one of the three BAC groups.

## Results

**Logistic Regression.** The logistic regression model was tested at three levels: the overall combined model, the parameter model and the individual levels of parameter model. The results of the latter analysis are presented in Table 1. The final combined logistic regression model fits the data well. The deviance of the model is 889, substantially smaller than the 2115 degrees of freedom. The chi-square of the deviance is nonsignificant at 0.05 level, suggesting that there is no reason to reject the model.

Table 1 Logistic Regression Analysis Results

Parameter	Level Description	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Odds Ratio
INTERCEPT		1	-6.4634	0.5793	124.4945	0.0001	
SEX	male	1	0.5101	0.1858	7.5366	0.006	1.6655
SEX	female	0	0	0			
AGE	16-18	1	-0.5649	0.649	0.7575	0.3841	0.5684
AGE	19-25	1	0.815	0.4842	2.8336	0.0923	2.2592
AGE	26-35	1	1.2008	0.4813	6.2238	0.0126	3.3228
AGE	36-45	1	1.429	0.4862	8.6384	0.0033	4.1745
AGE	46-55	1	1.0027	0.5165	3.7687	0.0522	2.7256
AGE	56+	0	0	0			
CONFIG	family	1	0.0447	0.6094	0.0054	0.9416	1.0457
CONFIG	1 passenger, diff. sex	1	0.3933	0.199	3.907	0.0481	1.4819
CONFIG	1 passenger, same sex	1	-0.4494	0.2862	2.4645	0.1164	0.6380
CONFIG	group, diff. Sex	1	0.2964	0.3011	0.9691	0.3249	1.3450
CONFIG	group, same Sex	1	1.1418	0.2947	15.0072	0.0001	3.1324
CONFIG	driver only	0	0	0			
TRIPORIGIN	home	1	0.1276	0.2852	0.2	0.6547	1.1361
TRIPORIGIN	friend/relative	1	0.4893	0.2364	4.2851	0.0384	1.6312
TRIPORIGIN	restaurant	1	0.2741	0.3216	0.7268	0.3939	1.3153
TRIPORIGIN	bar/tavern	1	1.0605	0.2855	13.7919	0.0002	2.8878
TRIPORIGIN	movie	1	-0.8834	1.0334	0.7308	0.3926	0.4134
TRIPORIGIN	sport/rec.	1	0.0705	0.3245	0.0472	0.8281	1.0730
TRIPORIGIN	other	1	-0.5208	0.3785	1.8933	0.1688	0.5940
TRIPORIGIN	work	0	0	0			
EDUC	less than high school	1	0.7483	0.2783	7.2313	0.0072	2.1134
EDUC	high school	1	0.8639	0.2181	15.6868	0.0001	2.3724
EDUC	some college	1	0.2884	0.2407	1.4349	0.231	1.3343
EDUC	college grad.	0	0	0			
HOUR	survey start time	1	0.4233	0.0784	29.1239	0.0001	1.5270

At the parameter level, all selected parameters were significant at the 0.05 level, using the likelihood ratio test. For each parameter, one category was selected as the reference group from which to compute odds ratios for all other categories. The reference group was usually, but not

always the one with lowest value. The Wald test was used to test the significance for individual categories.

Table 1 indicates that driver's sex, age and education correlate significantly with the likelihood of drinking and driving. As shown in Table 1, male drivers on average demonstrated 1.67 times of the odds, compared with female drivers. This means that male drivers are 67% higher in odds than female drivers of being impaired. Driver age is also, significantly associated with the level of BAC. The chances of drinking-driving in the 26-35 and 36-45 age groups are substantially higher than for other groups. The predicted odds for a 36 to 45-year old driver having a BAC over 0.05 is between three to four times of that of a 56 year or older driver. The younger drivers showed lower tendency for drinking-driving, as has been reported in other studies, although the relationship is not statistically significant. Education was also correlated with drinking-driving. The odds ratio for drivers with high school or less is double that of college graduates.

In line with previous research, the origin of the trip contributes significantly to the odds of drinking and driving. Table 1 suggests that trips originated from bars or taverns increase the chance of drinking and driving. The group of drivers coming out of bars and taverns showed 2.89 times the odds of drinking and driving compared with the reference group, drivers coming from work. Furthermore, the passenger configuration of a same sex group (more than one passenger of same sex as driver) increased the odds of being a drinker by more than 3 times over that of a lone driver. A single opposite sex passenger is associated with an increase in odds of about 50%.

Finally, the analysis confirms the association of increasing hour with likelihood of drinking-driving. The odds of drinking and driving increased by 100 percent for each 1.5 hour interval into the night, commencing from 2100 hours and ending at 0300 hours.

**Discriminant Analysis.** Two canonical dimensions underlying the variation in the explanatory variables were extracted. The standardized coefficients or loadings of each independent variable on the canonical functions are presented in Table 2. The correlations of the two functions with the BAC criterion are 0.26 and 0.09 respectively, indicating the total explanatory power of the model is fairly poor.

Table 2. Standardized Canonical Discriminant Coefficients

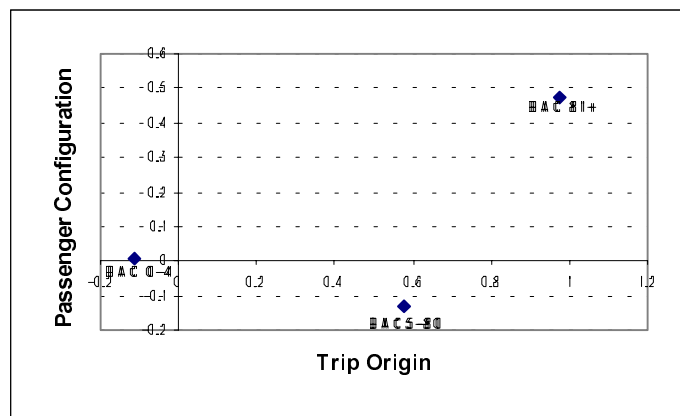
Predictor	Canonical1	Canonical2	Predictor Description
AGE16_18	-0.3082	0.0367	driver age 16-18
AGE19_55	0.3941	-0.0402	driver age 19-55
AGE56	-0.1931	0.0034	driver age 56 and over
SEX	0.2210	-0.1150	sex of driver - male
GRADE	0.0029	-0.2700	education level
CONFIG	0.2224	0.8398	group of passengers, same sex
TRIPORIGIN	0.7653	-0.2797	trip origin - bar
HOUR	0.6355	0.1696	survey start time

The first dimension appears to represent a factor of travelling from a bar late at night. Being of the mid-range age group (19-55), and to a smaller extent, being male is also associated with this

factor. Conversely, female drivers and the youngest and oldest age groups are least likely to be coming from bars late at night. The second factor has only one variable with substantial loading - passenger configuration. This indicates that the configuration of a same sex group is predictive of BAC group, independent of, and in addition to, those variables loading on the first canonical factor.

The separation of the BAC groups along the two canonical dimensions is illustrated in Figure 1. The first dimension provides the greatest separation of the three BAC groups. The impaired group (BAC=81+) has a mean score of +1 on the first (trip origin/driver age) dimension while the zero BAC drivers have a mean score of -0.1. The moderate drinking group is positioned just beyond mid-way at 0.57. The second dimension provided a further basis for separation, although small in magnitude. This dimension, characterized by same sex multi-passenger group, separates the highest BAC group, but not the zero and moderate BAC groups.

Figure 1. Position of BAC groups on two discriminant dimensions



The classification results for the validation group by the discriminant classification model are as follows: The model correctly classified 72% of the zero BAC group, 16% of the 5-80 mg% BAC group and 60% of the 81+ mg% group. These results must be compared with expected chance classification, based on knowledge of each group's proportion in the total sample. The chance level of accuracy is 85%, 13% and 2% for each of the three BAC groups respectively. Thus the discriminant model reduced the accuracy of chance classification for the zero and most frequently encountered BAC group, barely increased it for the mid-range BAC group but substantially increased it for the impaired BAC group.

When the estimated classification model was applied to the refusal group, the results indicated a distribution that was not identical to that of the participating drivers. The model predicts that 63% of the refusers had a zero BAC, 11% had a BAC between 5 and 80 mg% and 26% had a BAC over 80mg%. Based on knowledge of the discriminating variables in the model, this group appears to have a higher proportion of drinking drivers than the rest of the nighttime driving population. Unfortunately, there is no way to validate these results.

## **Discussion**

This study demonstrates that a combination of demographic and situational variables differentiate nighttime drivers on the basis of their BAC. The demographic predictors include driver age, sex and education level. The situational variables are coming from a bar, late night hour and a passenger configuration consisting of a group of passengers, of the same sex as the driver. However, like previous studies of this type the predictive model left much variance unexplained. This may be partly due to the fact that only a limited number of predictor variables were available. However, it is also likely that much of the variance in BAC is not easily measured by conventional means (eg, driver's developmental history, personality, mood, social influences and various situational antecedents).

In spite of the low overall explanatory power of both the logistic and discriminant models, the discriminant model substantially improved the accuracy of BAC prediction for the impaired (over 80 mg%) BAC group – from 2% to 60%. This means that the model has the ability to detect 60% of these impaired drivers, but would fail to detect 40%. This improvement for the impaired group was offset by increasing the rate of false positives for the zero BAC drivers. The mid-range group appears to share characteristics with the groups on either side of it. For this reason, they are not well classified.

The usefulness of these findings to an enforcement context is constrained by a number of concerns. First, the nighttime driving population is dominated by zero or low BAC drivers. Therefore the high rate of false positive identifications would pose a problem. A second limitation is that not all the predictor variables associated with BAC in this study are directly observable by a police officer. Third, the rate of false negatives for the impaired group may still be unacceptably high at 40%. Nonetheless the information (for example, on passenger configuration) could be used by police to corroborate decisions to further investigate, where other more direct evidence is also available. The information could also assist police in conducting road checks. Where detection rather than maximum exposure is the objective, road checks would be most effective carried out late at night on routes leading from bars. However, this study makes it clear that the easily observable factors on driver age, sex and passengers, even when supplemented by knowledge of trip origin and time of night are not sufficient to make reliable decisions about driver BAC.

## **Acknowledgements**

The roadside survey data were collected by the Traffic Injury Research Foundation under contract to the Insurance Corporation of British Columbia, managed by Impaired Driving Strategies.

## **References**

1. Beirness, DJ, Foss, RD, Mercer, GW. Roadside breathtesting surveys to assess the impact of an enhanced DWI enforcement campaign in British Columbia. In: Mercier-Guyon, C (Ed). Proceedings of the 14<sup>th</sup> International Conference on Alcohol, Drugs and Traffic Safety, Annecy, France, CERMT, Annecy, 1997, pp. 955-961.
2. Beirness, DJ, Foss, RD, des Groseilliers, M. ICBC roadside breathtesting survey: Final report. Traffic Research, Ottawa, Canada, 1999.

3. Foss, R, Beirness, D, Voas, R, Sprattler, K. Drinking and driving in Minnesota. In: Utzelmann, H-D, Berghaus, G. Kroj, G. (Eds). Proceedings of the 12<sup>th</sup> International Conference on Alcohol, Drugs and Traffic Safety, Cologne, Germany, Verlag TÜV Rheinland, Cologne, 1993, pp. 902-907.
4. Kruger, H-P, Reis, J, Hilsenbeck, T, Kazenwadel, et al. The German roadside survey 1992-1994: Aims, methods and results. In: Kloeden, CN, McLean, AJ (Eds). Proceedings of the 13<sup>th</sup> International Conference on Alcohol, Drugs and Traffic Safety, Adelaide, Australia, NHMRC Road Accident Research Unit, University of Adelaide, Adelaide, 1995, pp.167-178.
5. Lawson, JJ, Stewart, DE. Consumption of alcohol in night-time driving by young drivers in Canada. In: Benjamin, T (Ed). Young drivers impaired by alcohol and other drugs. International congress and symposium series, no. 116, Royal Society of Medicine Services, London, 1987, 55-63.
6. Paquette, P. Enquête sur la conduite automobile avec les facultés affaiblies par l'alcool – 1991 Québec. Société de l'assurance automobile du Québec, Québec, 1991.
7. Perrine, MW, Meyers, AR and Yu, J. Roadside BAC, alcohol use and driver record. . In: Kloeden, CN, McLean, AJ (Eds). Proceedings of the 13<sup>th</sup> International Conference on Alcohol, Drugs and Traffic Safety, Adelaide, Australia, NHMRC Road Accident Research Unit, University of Adelaide, Adelaide, 1995, pp. 185-190.
8. Quaye, K. Saskatchewan's 1993 night-time driver survey. Saskatchewan Government Insurance, Regina, Canada, 1994.
9. Voas, R. Drinking and driving in the US; the 1996 national roadside survey. In: Kloeden, CN, McLean, AJ (Eds). Proceedings of the 13<sup>th</sup> International Conference on Alcohol, Drugs and Traffic Safety, NHMRC Road Accident Research Unit, University of Adelaide, Adelaide, Australia, 1995, 1159-1166.