U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**

NHTSA
www.nhtsa.gov

DOT HS 811 358 August 2010

Assessing the Feasibility of Vehicle-Based Sensors to Detect Alcohol Impairment

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 811 358		

4. Title and Subtitle	5. Report Date
Assessing the Feasibility of Vehicle-Based Sensors to Detect Alcohol Impairment	August 2010
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
John D. Lee, Dary Fiorentino, Michelle L. Reyes, Timothy L. Brown, Omar Ahmad, James Fell, Nic Ward, Robert Dufour	N10-005
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
National Advanced Driving Simulator, The University of Iowa 2401 Oakdale Blvd Iowa City, IA 52242	11. Contract or Grant No. DTNH22-06-D-00043
12. Sponsoring Agency Name and Address	**13. Type of Report and Period Covered**
National Highway Traffic Safety Administration 1200 New Jersey Avenue, SE, West Building Washington, DC 20590	Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Eric Traube (Task Order Manager); Dr. Michael Perel (Ret.); Dr. Eric Nadler (Volpe Center); Melanie Laverman; Sue Ellen Salisbury; Cheryl Roe; Dawn Marshall; Dan Johnsen; Meiji Zhang; Claude LaRoche; Jeff Gordon; Catherine Horak; Larry Hynes; Scott Egerton; Dr. Jane Moeckli; and Stephanie C. Binder (COTR).

16. Abstract

Despite persistent efforts at the local, state, and federal levels, alcohol-impaired driving crashes still account for 31% of all traffic fatalities. The proportion of fatally injured drivers with blood alcohol concentrations (BAC) greater than or equal to 0.08% has remained at 31-32% for the past ten years. Vehicle-based countermeasures have the potential to address this problem and save thousands of lives each year. Many of these vehicle-based countermeasures depend on developing an algorithm that uses driver performance to assess impairment. The National Advanced Driving Simulator (NADS) was used to collect data needed to develop an algorithm for detecting alcohol impairment. Data collection involved 108 drivers from three age groups (21-34, 38-51, and 55-68 years of age) driving on three types of roadways (urban, freeway, and rural) at three levels of alcohol concentration (0.00%, 0.05%, and 0.10% BAC). The scenarios used for this data collection were selected so that they were both representative of alcohol-impaired driving and sensitive to alcohol impairment. The data from these scenarios supported the development of three algorithms. One algorithm used logistic regression and standard speed and lane-keeping measures; a second used decision trees and a broad range of driving metrics that are grounded in cues NHTSA has suggested police officers use to identify alcohol-impaired drivers; a third used a support vector machines. The results demonstrate the feasibility of a vehicle-based system to detect alcohol impairment based on driver behavior. The algorithms differentiate between drivers with BAC levels at and above and below 0.08%BAC with an accuracy of 73 to 86%, comparable to the standardized field sobriety test. This accuracy can be achieved with approximately eight minutes of driving performance data. Differences between drivers and between roadway situations have a large influence on algorithm performance, which suggests the algorithms should be tailored to drivers and to road situations.

17. Key Words	18. Distribution Statement
Alcohol, Algorithm, Driver Impairment Scenario Design, National Advanced Driving Simulator	Document is available to the public from the National Technical Information Service www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified		98	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 EXECUTIVE SUMMARY

The most notable findings from this study include:

- The National Advanced Driving Simulator (NADS-1) and a low-workload scenario are sensitive to alcohol. Alcohol effects were apparent in a simulation scenario representing a typical drive home at night from an urban bar.
- These effects were evident in drivers' control of vehicle lane position and speed. Standard deviation of lane position and average speed differentiated BAC conditions most precisely.
- The most sensitive indicators of impairment are associated with continuous performance (e.g., lane keeping) rather than discrete events (e.g., response to a traffic signal or use of turn signals).
- The three algorithms detected impairment at and above the legal limit about as well as the Standardized Field Sobriety Test (SFST), with sensitivity increasing with BAC level.
- This project demonstrates the feasibility of a driving-behavior-based approach to detecting alcohol impairment in real time.

Background

Despite persistent efforts at the local, state, and federal levels, alcohol-impaired driving crashes still contribute to approximately 31% of all traffic fatalities. Although regulatory and educational approaches have helped reduce alcohol-related fatalities, other approaches merit investigation. One such approach detects alcohol impairment in real time using the increasingly sophisticated sensor and computational platform that is available on many production vehicles.

It may be possible to detect impairment using driver state (e.g., eye movements), drivers' control inputs (e.g., steering and accelerator movements), and vehicle state (e.g., speed or lane position). Once detected, this information can support interventions that discourage drivers from driving while impaired and prevent alcohol-related crashes. This study assessed how well algorithms could detect impairment in a widely applicable and timely manner.

Objectives

The long-term research objective is to use algorithms that detect impairment as feedback to drivers to discourage or prevent drinking and driving. This report describes how, individually and in combination, driver actions reveal signatures of alcohol impairment, and how well algorithms built on these signatures detect drivers with BAC levels that are over the legal limit. Specific objectives include:

- Understand how driving-related metrics reflect the impairment associated with BACs at 0.05% and 0.10% (currently, the legal limit in the United States is 0.08%)
- Determine how well these metrics apply to different roadway situations and to different drivers (i.e., determine robustness)
- Develop algorithms to detect alcohol-related impairment
- Compare robustness of metrics and algorithms

Method

Data were collected in the National Advanced Driving Simulator (NADS) from 108 moderate to heavy drinkers dosed at placebo (0.00%), 0.05%, and 0.10% BAC on three separate visits. Drivers were divided into equal groups by age (21-34, 38-51, and 35-68) and gender. The participants drove a scenario representative of a drive home from an urban bar: a nighttime trip that involved a stretch of freeway and ended on a rural gravel road. The drives started with an urban segment composed of a two-lane roadway through a city with posted speed limits of 25 to 45 mph with signal-controlled and uncontrolled intersections. An interstate segment followed that consisted of a four-lane divided expressway with a posted speed limit of 70 mph. The drives concluded with a rural segment composed of a two-lane undivided road with curves. A portion of the rural segment was gravel. Drivers' steering, accelerator, and brake inputs, vehicle lane position and speed, the driving context (whether the vehicle was in a urban, interstate, or rural environment), and driver eye and eyelid movements were captured in representative driving situations, with precise control and in great detail.

Results

The objectives were addressed with two broad sets of analyses. The first focused on whether BAC affected performance. The second focused on detection of impairment. These analyses show the simulator and scenario to be sensitive to alcohol, and that algorithms can detect alcohol-related impairment.

Driving performance measures (i.e., mean speed, standard deviation of speed, and standard deviation of lane position) indicated systematic differences between BAC conditions. No statistically reliable effects of age and gender were found for lane deviation, but BAC affected lane deviation. Normalized lane deviation for the entire drive was 46.77 at 0.00% BAC, 49.79 at 0.05% BAC, and 54.31 at 0.10% BAC. Age reliably affected average speed, with average speed increasing with increasing age. BAC also affected average speed, with a higher BAC, in general, leading to lower average speed. Neither age nor BAC reliably affected speed deviation. Surprisingly, gender reliably affected speed deviation, with speed deviation greater for males than for females. These results are notable because the alcohol effects are apparent even though all participants were moderate to heavy drinkers and the driving situation was representative of daily driving, placing relatively low demands on the driver.

Taken together, the results from the impairment analyses indicate that alcohol affected performance, and that the NADS is sensitive to those changes. The next set of analyses focused on whether it is possible to classify BAC status (BAC $< 0.08\%$ v. BAC $\geq 0.08\%$) on the basis of those changes. Three algorithms were developed to predict BAC status based on logistic regression, decision tree modeling, and support vector machines. The algorithms were assessed in terms of sensitivity, robustness, timeliness, and bias.

Sensitivity is the degree to which the algorithm can differentiate between drivers above and below the legal limit. These algorithms show sensitivity comparable to that of the SFST. Accuracy of the algorithms ranged from 73 to 86%. The logistic regression algorithm achieved an accuracy of 74.4% by combining information across the entire drive, achieving maximum sensitivity after approximately 25 minutes of driving. Decision tree and Support Vector Machine algorithms are much more sensitive and timely, identifying impairment in the situations tested with greater precision after only approximately eight minutes of driving. Greatest sensitivity was achieved by the decision tree, which combined driving performance indicators tailored to individual drivers. The most sensitive indicators of impairment were associated with continuous performance (e.g., lane keeping) rather than discrete events (e.g., response to a traffic signal or use of turn signals). Performance of the algorithms showed substantial differences in the degree to which they and their constituent measures provide robust and timely indications of impairment.

Robustness is insensitivity to confounding factors, such as different driving environments, and it applies to many factors affecting algorithm performance. An important element of robustness addressed in this study concerns the dependence of the algorithm on differences between drivers and the driving environment (i.e., urban, freeway, rural). Consistent with previous research, algorithms tailored to individuals outperformed generic algorithms by approximately 13%. Algorithm performance also depends on the driving context: different driving situations provide different measures, and these measures differ in their sensitivity. Current vehicle technology makes it quite feasible to capitalize on the benefits of tailoring algorithms to individuals by comparing a driver’s performance to his or her past performance in similar roadway situations.

Timeliness is the speed with which an algorithm is able to detect impairment. Timeliness is a critical consideration for real-time algorithms because some interventions rely on impairment detection well before the end of the drive. Timely impairment detection depends critically on the driving context with some events and variables being more sensitive than others. The most sensitive indicators of impairment involve continuous measures cumulated over time, such as the standard deviation of lane position. In addition, important signatures of alcohol impairment (straddling, weaving, and gaze concentration) are defined by behavior that evolves over a relatively long time horizon, requiring samples of driving behavior that extends over 30 seconds to several minutes. Even with such constraints, sensitivity comparable to the SFST was obtained over approximately eight minutes of driving.

Bias refers to the tendency of the algorithm to favor correctly detecting impairment at the expense of incorrectly identifying impairment when there is none. The fundamental differences in the optimization approaches between decision trees and SVM lead to differences in bias. These complementary differences can be leveraged to minimize false detection of impairment and to maximize detection of impaired drivers. Algorithms can be combined according to the benefits of detecting impairment and the costs of failing to detect impairment, so that one algorithm is used to maximize impairment detection and another is used to avoid false detection. Support Vector Machines show a tendency to outperform decision trees in maximizing detection.

Recommendations and conclusions

This study demonstrates the feasibility of vehicle-based sensors to detect alcohol-related impairment in real time: sensitivity is comparable to the SFST. This sensitivity is likely a very conservative estimate relative to sensitivity in detecting higher BAC levels. Because 66% of alcohol-related fatalities occur with BAC levels above 0.15% (compared with impaired drivers at 0.08 BAC or greater), the greatest value of a vehicle-based countermeasure may lie in detecting high BAC levels, where algorithms are likely to be very sensitive. These results suggest substantial promise in detecting other impairments, such as drowsiness, distraction, and even age-related cognitive decline.

The ultimate aim of impairment-detection algorithms is to support interventions that guide the driver to safer behavior. The desirability and feasibility of any particular algorithm depends on how it meets the particular needs of an intervention. This study demonstrates that algorithms differ substantially on these dimensions and that design must consider the inevitable tradeoffs. Algorithms become more sensitive, but less timely, as measures are integrated over time. The ultimate feasibility of impairment-detection algorithms depends on matching the performance profile of an algorithm to an intervention.

This project demonstrated the feasibility of a behavior-based approach to detecting alcohol impairment. It also identified many issues that merit further investigation. This study focused on moderate levels of alcohol (0.05% and 0.10% BAC) in people who reported to be moderate to heavy drinkers, but not problem drinkers. Given the assumption that moderate and heavy drinkers show less obvious indications of impaired driving, algorithms that are able to detect impaired driving from this population are likely to be much more effective for people who are light drinkers or for higher BAC levels, whereas the effectiveness of the algorithms for problem drinker is unknown. It would be useful to empirically assess how algorithm sensitivity differs at higher BACs, and for light and chronic drinkers.

More generally, this study identified a huge design space of sensors, measures, signatures and algorithms, algorithm parameter combinations, and meta algorithms. Assuming there are at least 10 sensors, 10 metrics for each sensor, 4 time scales, 10 algorithms, 10 implementations of each algorithm, and 4 meta algorithms, a total of more than 160,000 potential algorithms exist. This project sampled only a small region of that space. A specific challenge facing deployment of the algorithms developed in this study is the reliable measurement of lane position. Current lane-tracking technology is vulnerable to sun glare, adverse weather, and poorly maintained lane markings. Steering wheel position is not subject to these limits and further analysis of its ability to detect alcohol-impaired driving is warranted. More generally, a promising direction for algorithm development is to identify classes of drivers and classes of driving situations, and an important direction for algorithm assessment is to develop metrics that relate to the interventions the algorithm intends to support. These results also suggest that further exploration is warranted, not just for alcohol impairment detection, but also for detecting impairment associated with drowsiness, distraction, and even age-related cognitive decline.

2 INTRODUCTION AND OBJECTIVES

Despite persistent efforts at the local, state, and federal levels, alcohol-impaired driving crashes still contribute to approximately 31% of all traffic fatalities. The proportion of fatally injured drivers with blood alcohol concentrations (BAC) greater than or equal to 0.08% has remained at 31-32% for the past ten years (National Highway Traffic Safety Administration, 2009b). Although regulatory and educational approaches have helped to reduce alcohol-related fatalities, other approaches merit investigation. One such approach concerns countermeasures that capitalize on the increasingly sophisticated sensor and computational platform that is available on many production vehicles. Such vehicle-based countermeasures have the potential to address this problem and save thousands of lives each year.

Vehicle-based countermeasures use sensors that describe drivers' control inputs (e.g., steering wheel and brake pedal movement), vehicle state (e.g., accelerometer and lane position), driving context (e.g., speed zone information and proximity of surrounding vehicles), and driver state (e.g., eye movements and posture). Data from these sensors can be transformed, combined, and processed with a variety of algorithms to develop a detailed description of the driver's response to the roadway. These sensors and algorithms hold promise for identifying a range of driver impairments, including distraction, drowsiness, and even age-related cognitive decline. Alcohol represents a particularly important impairment that might be detected by vehicle-based sensors and algorithms.

This report describes the development and evaluation of algorithms to detect the behavioral signature of alcohol. Such an algorithm is a central element of any vehicle-based countermeasure for alcohol-related crashes. Algorithm development depends on collecting data from impaired and unimpaired drivers. This report describes data collection that involved drivers from three age groups (21-34, 38-51, and 55-68 years of age) driving through representative situations on three types of roadways (urban, freeway, and rural) at three levels of alcohol concentration (0.00%, 0.05%, and 0.10% BAC). The high fidelity of the National Advanced Driving Simulator (NADS) makes these data unique. Drivers' control inputs, vehicle state, driving context, and driver state were captured in representative driving situations, with precise control and in great detail. This report describes how, individually and in combination, these data reveal signatures of alcohol impairment, and how well algorithms built on these signatures detect drivers with BAC levels that are over the legal limit of 0.08%.

The overall objectives of the data collection and analysis efforts were to:

- Understand how driving-related metrics reflect the impairment associated with BAC at 0.05% and 0.10%
- Determine the robustness of these metrics with respect to individual differences such as age and gender, as well as the roadway situation
- Identify signatures of impairment and develop algorithms to detect alcohol-related impairment
- Compare robustness of metrics and algorithms

This document contains six main sections. The first describes the prevalence and consequences of drinking and driving. The second describes the sensors and metrics that are likely to capture a clear behavioral signature of alcohol impairment. The third section describes the criteria and considerations for detecting impairment. The fourth section outlines the experimental methods and experimental design, independent variables, dependent variables, and data collection procedures. The fifth section describes the general effects of three alcohol levels on driver performance. The sixth section describes the characteristics of algorithms used to detect alcohol impairment, their performance, and the validation process.

This report focuses on detecting alcohol impairment, but the issues are common to any vehicle-based system that detects driver impairment. As manufacturers work to make vehicles increasingly aware of the roadway and driver state, such systems will become important contributors to vehicle safety. Consequently, this report provides an example of assessing the limits and capabilities of vehicle-based systems that support a behavior-based estimation of driver impairment.

The long-term research objective is to use algorithms that detect impairment to provide feedback to drivers that will discourage drinking and driving. Ultimately the alcohol-impairment-detection algorithms developed in this study could support a range of vehicle-based interventions to prevent alcohol-related crashes. Such interventions could include limiting drivers' ability to drive dangerously (e.g., lockout distractions or limit speed), providing feedback to impaired drivers that may motivate them to pull over or drive more cautiously, adjusting crash warning systems to provide an earlier warnings, or providing long-term feedback that highlights dangerous driving. This approach to detecting alcohol impairment and associated countermeasures complements in-vehicle technology that prevents an alcohol-impaired driver from starting the car, such as that being pursued in the Driver Alcohol Detection System for Safety (DADSS, www.dadss.org/). This dual-prong approach is consistent with many public health programs, ranging from preventing teen pregnancy to reducing smoking-related illnesses.

Using vehicle-based technology to prevent alcohol-impaired drivers from crashing also begins to address the more general problem of preventing drivers suffering from a broad range of other impairments from crashing. The combined influence of alcohol, distraction, fatigue, and drug- and age-related cognitive impairments represent the predominant cause of crashes. Mitigating these impairments could have profound safety benefits.

3 DRINKING AND DRIVING PREVELANCE

Impaired driving is a significant public health and safety problem in the United States (U.S.). It was estimated by the National Highway Traffic Safety Administration (NHTSA) that impaired driving costs society $51 billion annually (Blincoe, et al., 2002). Although alcohol consumption is legal for U.S. citizens aged 21 or older, it is illegal in all states to drive with a blood alcohol concentration (BAC) of 0.08% or greater. In 2008, there were an estimated 11,773 traffic crash fatalities involving drivers at these illegal BAC limits (BAC=0.08% and above) (National Highway Traffic Safety Administration, 2009b).

3.1 Drinking in America

A report by the National Center for Health Statistics (NCHS) on the health behaviors of adults estimated that 6 of 10 people in the United States were drinkers of alcohol in 1999-2001 (consumed alcohol within the past year), whereas about one in four were total abstainers (Public Health Service, 2004). About 20% of the adults in that survey reported consuming five or more drinks in a day at least once in the past year. A Gallup Poll (Saad, 2003) indicated that the percentage of adult drinkers who consumed one to seven drinks in the past week has increased from 36% in 1992 to 45% in 1997 to 50% in 2001. The proportion of drinkers in that same survey who reported consuming eight or more drinks in the past week increased from 12% in 1992 and 1997 to 18% in 2001. In a separate analysis of Gallup's surveys on alcohol and drinking from 1999 to 2003, 40% of men aged 18-29, 29% of men aged 30-49, 21% of men aged 50-64, and only 16% of men aged 65 and older admitted to drinking more than they should (Carroll, 2003). Carroll also found that women of all ages were much less likely to report drinking more than they should at times: 26% of women aged 18-29; 19% of women aged 30-49; 12% of women aged 50-64; and only 5% of women aged 65 and older. Alcohol consumption represents a prevalent and increasingly common behavior.

Although moderate consumption describes the drinking patterns of many, the prevalence of excessive drinking represents an important concern. Recent studies and surveys have indicated that about 8% of the U.S. population can be classified as alcohol dependent or alcohol abusers (NIAAA, 2004). This translates to about 17.6 million American adults, most of whom drive motor vehicles. Binge drinking (defined as consuming 5 or more drinks in one session) increased in the United States between 1995 and 2001, according to a recent survey (Naimi et al., 2003). Binge-drinking episodes per person per year increased 35% during that same period, with men accounting for 81% of the episodes. Overall, 47% of binge-drinking episodes occurred among otherwise moderate drinkers, and 73% of all binge drinkers were classified as moderate drinkers in that study. Binge drinkers were 14 times more likely to drive impaired than non-binge drinkers. The patterns of moderate and excessive drinking describe a societal context in which drinking is likely to precede or accompany driving, and that the societal trends will require countervailing forces to prevent alcohol-impaired driving from reflecting similar trends.

3.2 Drinking and driving on U.S. roads

Our knowledge about the impaired-driving problem in the United States has been augmented by a series of National Roadside Survey (NRS) studies from which we can estimate the prevalence of drinking and driving over time in the contiguous 48 states by randomly selecting drivers from the road and requesting breath samples. The first NRS, sponsored by NHTSA, was conducted in 1973 (Wolfe, 1974). The second NRS was sponsored by the Insurance Institute for Highway

Safety (IIHS) in 1986 (Lund & Wolfe, 1991), and the third was jointly funded by IIHS and NHTSA in 1996 (Voas, Wells, Lestina, Williams, & Greene, 1998). The first three surveys (1973, 1986, and 1996) included a brief interview of randomly selected drivers, and a breath sample to measure the BAC. The fourth in this series of national surveys, conducted in 2007, followed the general methodology of the three prior surveys in obtaining BACs for comparison with the earlier surveys, but also incorporated several new features (Lacey, et al., 2009). These included questionnaires on drivers' drug use, interaction with the criminal justice and treatment systems, drug- and alcohol-use disorders, and collecting and analyzing oral fluid and blood to determine the presence of drugs (over-the-counter, prescription, and illegal) other than alcohol.

Figure 1 summarizes and compares the results of the four NRS studies of weekend nighttime drivers. The figure shows that the percentage of drivers in all BAC range categories has decreased in succeeding decades, with the exception of an increase in the percentage of drivers with BACs between 0.050% and 0.079% between 1986 and 1996. However, the overall percentage of positive BAC drivers decreased between 1986 and 1996 (Fell, Tippetts, & Voas, 2009).

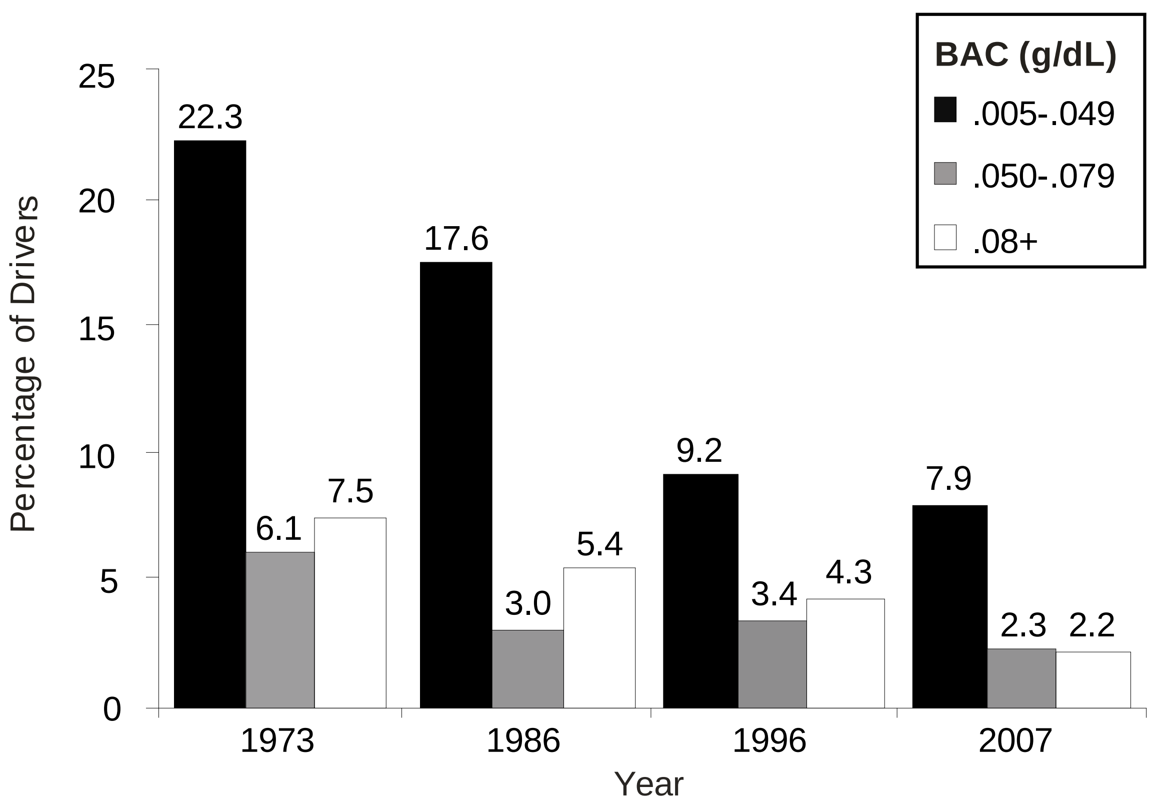

Figure 1. Percentage of nighttime drivers in three BAC categories in the four national roadside surveys.

Based upon data collected in the 2007 NRS, the best predictor of being a drinking driver on U.S. roads was reported as binge drinking (six or more drinks in a session for males; five or more drinks for females). The data showed that people who reported binge drinking were more likely to have had a positive BAC on the road. Given that binge drinkers were more likely to be impaired drivers, binge drinking was a better predictor than being classified as alcohol dependent or an alcohol abuser, as defined by the *Alcohol Use Disorders Identification Test* (AUDIT)

(Babor, de la Fuente, Saunders, & Grant, 1992; Chung, Colby, Barnett, & Monti, 2002; Conley, 2001).

Telephone surveys provide a less direct estimate of drinking and driving behavior, but one that is largely consistent with the roadside surveys. In a telephone survey of more than 6,000 people aged 16 and older in the United States in 2001, 23% reported driving within two hours of drinking alcohol in the past year (Royal, 2003). In that same survey, problem drinkers were estimated to make up 29% of the past year's drinking drivers, accounting for about 46% of all drinking-and-driving trips reported in that survey. "Problem drinkers" were defined in that survey as having two or more positive responses to the CAGE instrument (Ewing, 1984), or having consumed five or more drinks on four or more days in the past month, or having consumed nine or more drinks (eight for females) on at least one occasion in the last month. These problem drinkers accounted for 343 to 491 million drinking-and-driving trips reported in 2001.

3.3 Arrests and fatal crashes of alcohol-intoxicated drivers

In recent years, between 1.4 and 1.5 million drivers have been arrested annually for driving while intoxicated (DWI) or driving under the influence (DUI) (FBI, 2009). This is more people than are arrested each year for larceny or theft, assaults, weapons charges, or vandalism. About the same number of people is arrested each year for drug abuse violations as for DWI. In 2006, the DWI arrest rate was about one DWI arrest for every 138 licensed drivers in the United States. When combined with drinking-and-driving surveys, this amounts to one DWI arrest for every 772 reported episodes of driving after drinking, one DWI arrest for every 88 reported episodes of driving over the BAC limit, and one DWI arrest for every six stops by police for suspicion of DWI (Zador, Krawchuk & Moore, 2000).

Jones and Lacey (2000) summarized the state of knowledge on repeat DWI offenders and concluded that repeat offenders have many of the characteristics of first-time offenders. Some of these common characteristics include a high BAC at the time of arrest, alcohol dependency indications, drinking at multiple locations, and experiencing other problems related to alcohol. Kennedy et al. (1993), in an earlier review of the literature on convicted drinking drivers, reported that 80-95% of DWI offenders were males, 70-80% were aged 25-45, and 35-60% reported usually drinking 5 or more drinks at a session. In a NHTSA study, Fell (1992) showed that drivers with prior DWI convictions were overrepresented as drivers in fatal crashes by a factor of 1.8, but that only 1 out of 7 intoxicated drivers in fatal crashes had a prior DWI conviction within the past three years.

Alcohol involvement in fatal crashes (any driver with a BAC of 0.01% or greater) in 2007 was more than three times higher at night (6 p.m.–6 a.m.) than during the day (6 a.m.–6 p.m.) (62% versus 19%). Alcohol involvement was 35% during weekdays compared to 54% on weekends. Nearly one in four drivers (23%) of personal vehicles (e.g., passenger cars or light trucks) and more than one in four motorcyclists (27%) in fatal crashes were intoxicated (i.e., had a BAC equal to or greater than the 0.08% illegal limit in the United States). In contrast, only 1% of the commercial drivers of heavy trucks had BACs equal to 0.08% or higher. The 21-24 age group had the highest proportion (35%) of drivers with BACs≥0.08%, followed by the 25-34 age group (29%). The oldest and the youngest drivers had the lowest percentages of BACs ≥ 0.08%: those aged 75 or older were at 4%, and those aged 16-20 were at 17% (Fell, et al., 2009).

There were 55,681 drivers involved in fatal crashes in 2007 that resulted in 41,059 deaths. Twenty-two percent of these drivers were legally intoxicated (BAC ≥0.08%), whereas 14% had BACs exceeding 0.15% (see Figure 2). Thirteen percent of these intoxicated drivers were driving a motorcycle or an off-road vehicle in the fatal crash. Only 7.5% of these intoxicated drivers had a prior conviction for DWI within the past three years. It is believed, however, that this percentage would be much higher if the look-back period for prior offenses went beyond three years (e.g., 5 to 10 years). Over 10% of these impaired drivers were younger than 21, ages at which any drinking of alcohol and driving is illegal in every state.

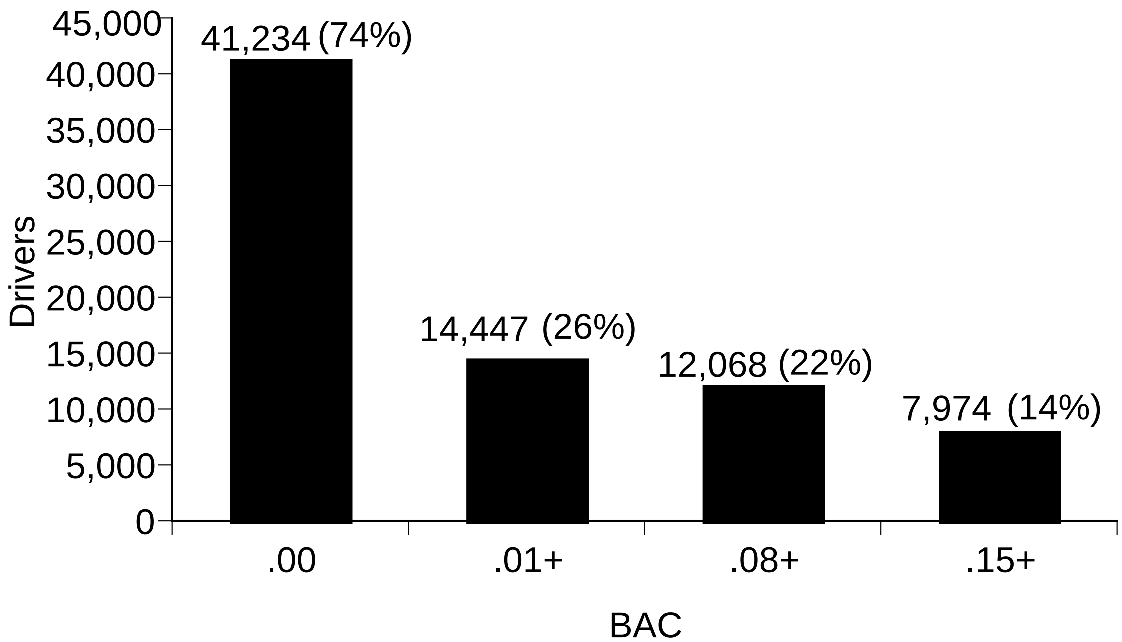

Figure 2. Drivers involved in fatal crashes by BAC level, FARS 2007 (N=55,681) (NHTSA, 2009a).

According to Voas, Romano, Tippetts & Furr-Holden (2006), one out of four drinking drivers in fatal crashes (BAC>=0.01%) are estimated to be heavy episodic drinkers, while over half are considered current normative drinkers (see Table 1).

Table 1. Percent of drinking drivers in fatal crashes in five consumption categories [adapted from Voas, et al. (2006)] .

Drinker Classifications	**Number of Drinking Drivers in Fatal Crashes** **[FARS Average from 1999-2001]**	**Percent of Drinking Drivers in Fatal Crashes**
Dependent Drinkers	1,410	11.3%
Abusive Drinkers	560	4.5%
Dependent and Abusive Drinkers	270	2.2%
Heavy Episodic (Binge) Drinkers	3,170	25.3%
Current Normative (Social) Drinkers	7,110	56.8%
ALL	12,520	100%

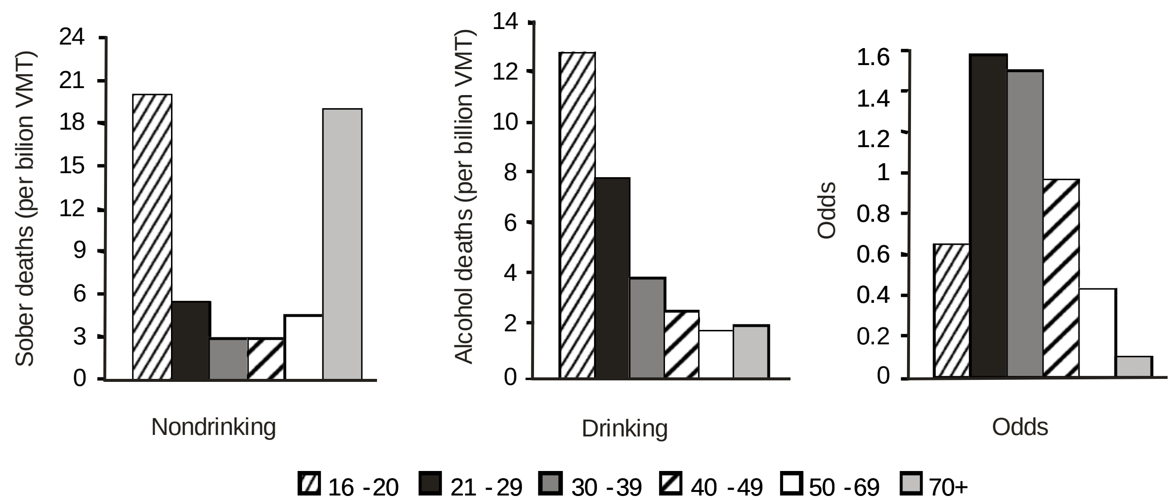

Figure 3. Drinking and nondrinking driver fatal crash rates by driver age and the odds of drinking to nondrinking driver rates, 1990-1994 [adapted from (Tippetts & Voas, 2002)].

To better understand the role of alcohol, it is useful to compare drinking and nondrinking drivers in fatal crashes, as shown in Figure 3. The first graph shows the age distribution of non-alcohol-related fatal crash involvements as a function of vehicle miles traveled (VMT) as reported by the

1990 Federal Highway Administration (FHWA) National Personal Transportation Survey. That graph takes the traditional U-shape, with underage and elderly drivers having the highest non-alcohol-related crash rates per mile driven. The common explanation for this relationship is the inexperience and risk-taking of youthful drivers and the deterioration of driving skills and, perhaps more importantly, the greater fragility of elderly drivers because of their greater susceptibility to fatal injury under certain crash conditions (Lee, 2006).

When, as in the second graph of Figure 3, alcohol-related rates per VMT are plotted, an L-shaped curve results with drinking-driver rates per mile driven being the highest among youthful drivers and gradually dropping with age, with the elderly least involved. The common explanation is that the driving skills of underage drivers are more vulnerable to alcohol. A somewhat different impression is provided when alcohol-involvement rates are considered as a ratio (drinking-driver rates/non-drinking-driver rates) by age group (third graph in Figure 3). This takes the shape of an inverted "U." Thus, when the involvement of underage drinking drivers in fatal crashes is related to mileage driven, their risk level is high. This is in part because their risk when sober is high. When the effect of their high risk when sober is accounted for by using that measure to normalize the data to compare across age groups, drivers 21 to 49 have a higher relative risk (odds) when drinking than do drivers under the age of 21.

The percentage of all traffic fatalities that were alcohol-related declined from 59.6% in 1982 to 41.5% in 2007, a 30% relative reduction. The percentage of all fatally injured drivers (where the testing for BAC is relatively high in the Fatality Analysis Reporting System [FARS]) who had a BAC higher than the current illegal limit in the United States (BAC ≥0.08%) dropped from 53% in 1982 to 35.5% in 2007, a 33% relative decline. A similar decline occurred for fatally injured drivers with extremely high BACs (≥0.20%). The ratio of drinking drivers to nondrinking drivers in fatal crashes (the crash incidence ratio (CIR) reflects the effectiveness of various impaired-driving countermeasures) declined from 69% in 1982 to 35% in 2007, a 51% decline in that measure (Figure 4). This decline occurred before 1995, and the CIR has remained relatively constant for the last 15 years.

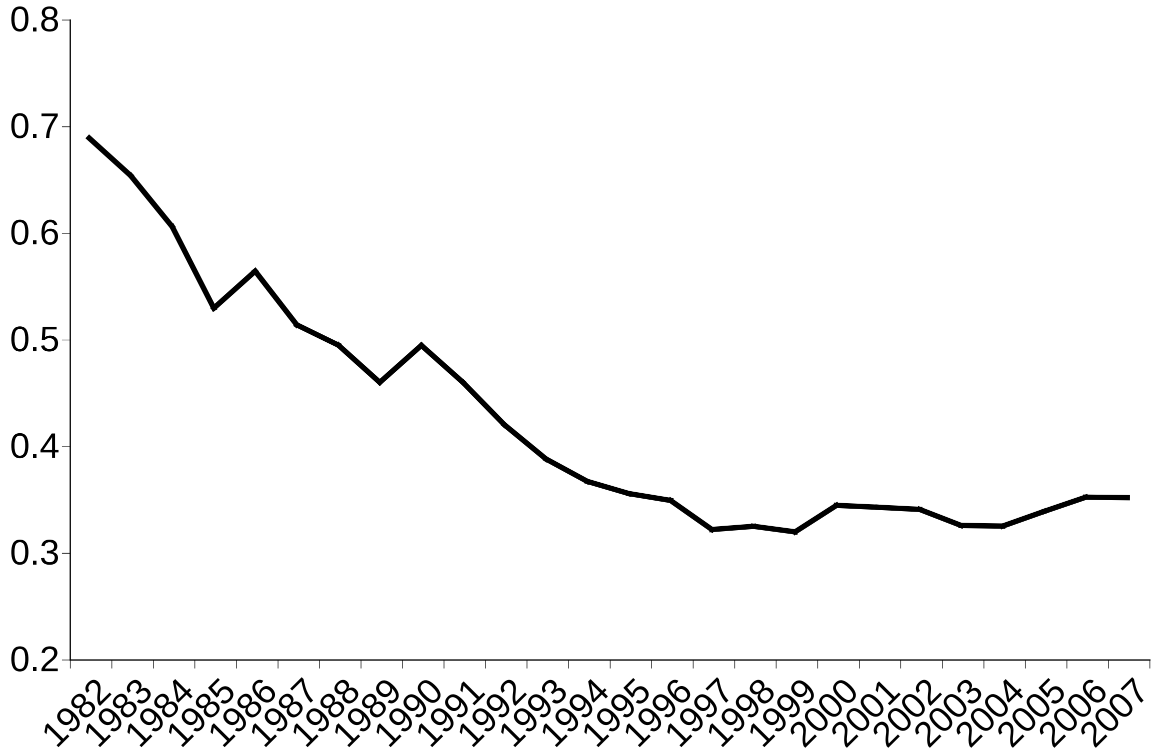

Figure 4. Ratio of drinking drivers to nondrinking drivers in fatal crashes (NHTSA, 2009a).

The relative risk of being involved as a driver in a fatal crash at various BAC levels (relative to a 0.00% BAC) depends on drivers' age and gender. Although not a typical case-control study, Zador, Krawchuk, and Voas (2000) selected crash data from the FARS that matched the weekend hours and the months in 1996 when a national survey of driver BACs was conducted. They showed that the relative risk of being killed as a driver in a single-vehicle crash varied considerably by age and gender, ranging from a low of 0.07 for drivers aged 35 and older at very low BACs (0.010%-0.019%) to 15,559.85 for 16-20 year old male drivers at very high BACs (0.15%+). Age and gender clearly interact with alcohol levels to influence crash risk.

These studies support several conclusions. First, the societal trends of increasing alcohol consumption, increasing instances of excessive drinking, and the stable rate of drinking and driving all suggest the need for countermeasures that go beyond the traditional enforcement and education. Second, because the minority of alcohol-intoxicated drivers in fatal crashes are problem drinkers (25%), compared to 50% who are normative drinkers. Third, because the best predictor of being a drinking driver on U.S. roads is binge drinking, countermeasures that target only repeat offenders will have limited efficacy. In combination, these characteristics of the drinking and driving problem all support a vehicle-based countermeasure based on detecting behavioral signatures of alcohol impairment.

4 SENSORS, MEASURES, AND METRICS FOR DETECTING ALCOHOL IMPAIRMENT

This study seeks to identify impairment using data from driver's own vehicles. Creating a successful vehicle-based system to detect alcohol impairment depends on the sensors and associated data. More and better data will support more precise detection, but the benefit of additional sensors must be weighed against their intrusiveness, reliability, and cost. Although EEG sensors might be able to detect changes in brain function associated with increased alcohol levels (Brookhuis & De Waard, 1993), such systems are not feasible for production vehicles. EEG sensors are expensive, produce noisy data, and require intrusive instrumentation that most drivers would not tolerate. Many other techniques do not require sophisticated sensors, such as cognitive task batteries that require people to perform a series of working memory, tracking, vigilance, and reaction time tasks (Kennedy, Turnage, Rugotzke, & Dunlap, 1994; Kennedy, Turnage, Wilkes, & Dunlap, 1993). Unfortunately, such techniques are quite intrusive and impractical—few drivers are willing to perform these tasks every time they start a trip. Because the sensor suite must be feasible for implementation in a production vehicle, the sensors must be low-cost and non-intrusive. In addition, the sensors must be robust to a range of environmental conditions and not sensitive to the efforts of drivers to circumvent them. Recent reviews of emerging sensor technology guided our analysis of potential sensors, measures, and metrics (Pollard, Nadler, & Stearns, 2007; Ward, 2006). This section outlines the considerations for choosing sensors, measures, and metrics to detect alcohol impairment, as well as challenges associated with developing robust metrics as input to an impairment-detection algorithm. The distinction between measures (data available from sensors) and metrics (summary statistics) emphasizes the critical translation of raw data into diagnostic behavioral signatures of impairment.

4.1 Alcohol impairment, driving performance, behavior, and safety

Figure 5 shows that alcohol has a systematic effect on driving safety, with BAC levels over 0.10% having an increasingly dramatic effect on crash risk (Blomberg, Peck, Moskowitz, Burns, & Fiorentino, 2005). This increased crash risk reflects both performance impairment and behavioral change. Performance impairment reflects diminished perceptual, motor, and cognitive capacity that accompanies increased BAC levels. Substantial evidence shows that alcohol diminishes performance, particularly in situations that require attention to be divided across multiple activities. Behavioral change reflects the increase in the willful engagement in risky behavior (e.g., speeding) with increased BAC levels. Both performance impairment and behavioral change can increase crash risk.

The degree to which performance impairment or behavioral change influences the crash risk in Figure 5 is difficult to determine, although the conditions surrounding crashes suggest that both contribute. Crash data show that alcohol-impaired drivers are more likely to be exceeding the speed limit at the time of the crash than those who are not alcohol-impaired. Similarly, crash risk for those over the age of 35 with a BAC between 0.08% and 0.10% is 11.4 times that of a sober driver, but the same BAC leads to a crash risk 51.9 times that of a sober driver for male drivers aged between 16 and 20 (Zador, Krawchuk, & Voas, 2001). The differential effect of alcohol on these two groups of drivers suggests that the influence of alcohol on crash risk depends on more than its effect on performance impairment alone.

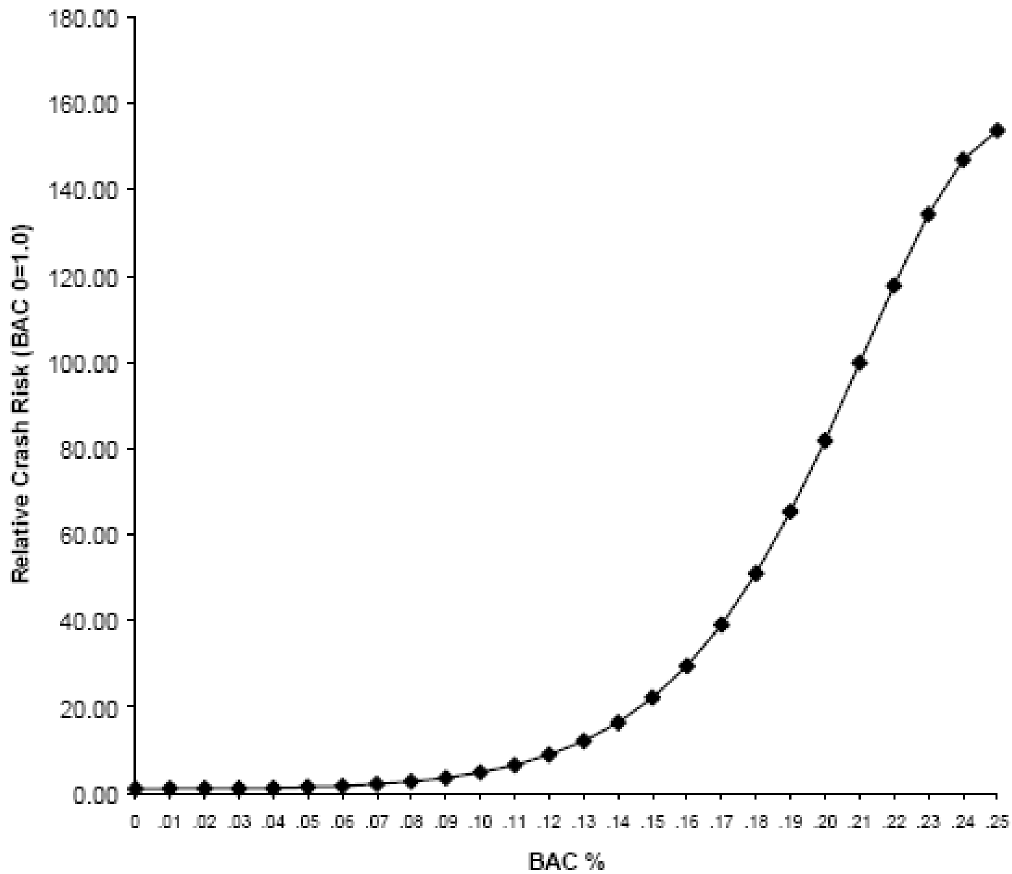

Figure 5. The effect of increasing levels of blood alcohol content (BAC) on motor vehicle crash risk (Blomberg, et al., 2005).

Hundreds of studies have explored the cognitive impairments associated with alcohol. Several major reviews document these results (Ogden & Moskowitz, 2004). First, alcohol does not always lead to a statistically reliable effect on performance, but increasing BAC levels increase the likelihood of detecting an effect. Sixty percent of all studies show an effect for BAC levels over 0.09% (Evans, 1991). The difficulty in detecting the effect of alcohol at lower BAC levels suggests that people can be quite effective in adapting to low levels of alcohol impairment in low-demand situations by investing additional resources or adjusting their response strategies. Second, alcohol can affect performance at levels as low as 0.03% BAC. Situations that demand precise and timely responses show performance impairments even at low BAC levels. Third, alcohol affects the various cognitive processes differentially: perception and vigilance tasks are much less sensitive compared to selective and divided attention tasks or those associated with high stimulus-response complexity. This differential sensitivity suggests that detecting alcohol impairment would be most effective in complex and demanding driving situations that engage selective and divided attention.

Several reviews of alcohol-related performance impairment show a consistent set of findings regarding the differential effect of alcohol on task types. For studies examining impairment due to low levels of alcohol (<0.05% BAC), 70-80% show an effect of alcohol for complex tasks compared to only 33% that show an effect for simple tasks (Holloway, 1994). Simple, automatic tasks are relatively unaffected, and complex, controlled tasks are relatively sensitive to low levels of alcohol (<0.05% BAC). Complex, controlled tasks are those that require simultaneous attention to multiple task elements and lack a consistent stimulus-response mapping. Simple, automatic tasks are those that require response to a single stimulus that is paired with a single pre-defined response that the person has extensive experience selecting. Psychomotor processes associated with balance are relatively automatic and are affected only at relatively high BAC levels. In contrast, the frequency, duration, and distribution of eye fixations tend to be quite sensitive to alcohol, reflecting the sensitivity of divided attention to alcohol impairment (Moskowitz, Ziedman, & Sharma, 1976). Similarly, simple reaction time tasks are relatively unaffected and choice reaction time tasks tend to be sensitive (Tzambazis & Stough, 2000). Overall, alcohol impairs performance most strongly in situations that require people to integrate information from complex scenes, attend to multiple elements, generate responses that are contingent on multiple conditions, and coordinate multiple responses.

This simple description of alcohol impairment has important implications for selecting measures and metrics to detect alcohol impairment. First, measures depend on the driving scenarios, so scenario development is critical to gathering sensitive measures. Scenarios that involve common driving tasks occurring in isolation (e.g., lane keeping on a straight road) are likely to be less sensitive than less-practiced tasks that require distribution of attention to multiple aspects of the driving environment (e.g., adjusting to the speed of a lead vehicle on a twisting road), or coordination of vehicle control and secondary tasks (e.g., selecting a song from a CD). This dependence on driving scenarios suggests that relating measures to driving context might be an effective way to enhance their sensitivity. The second general implication is that metrics derived from multiple measures are likely to be more sensitive than metrics derived from individual measures. Alcohol impairment is most apparent in complex tasks or tasks that demand coordination on multiple dimensions. Integrating measurements from multiple aspects of driving is likely to reflect alcohol-related impairment that might not be apparent from metrics based on individual measures, a theme we will return to in defining the algorithms.

4.2 Considerations for choosing alcohol-sensitive measures and metrics

Several critical design considerations for a vehicle-based system govern the selection of impairment metrics:

1. Sensitivity – The measure must be sensitive enough to account for a large portion of the variation in performance due to alcohol.
2. Specificity – Ideally, the measures should be indicative of alcohol impairment and not other sources of impairment, such as distraction and drowsiness, or roadway situations that might disrupt driving performance, such as work zones.
3. Availability – The measures must be accessible continuously during all driving conditions such that impairment monitoring is continuous.
4. Acceptability – The measure must appear to be specific to alcohol impairment and safety as well as not perceived as being invasive (such as speed data being used to detect impairment and enforcing speeding fines despite lack of driver impairment).

5. Practicality – The technology required for sensing the measurement data must be easily integrated into the vehicle at minimal cost and must operate robustly in all driving environments.

Of these five considerations, specificity is perhaps the most challenging to address. Specificity refers to the ability of measures to differentiate alcohol from other types of impairment. Specificity is particularly challenging because the underlying mechanisms of alcohol impairment are similar to other types of impairment. Alcohol is a central nervous system depressant, so it can induce drowsiness, which may also be manifest in people who have not had sufficient sleep. In fact, drowsiness-related driving impairment and alcohol-related driving impairment are comparable in how they affect some driving performance metrics, such as lateral control (Arnedt, Wilde, Munt, & MacLean, 2001). Furthermore, many pharmaceuticals also induce drowsiness and can impair driving performance (Lenne, Dietze, Rumbold, Redman, & Triggs, 2003; Linnoila & Mattila, 1973; Vanakoski, Mattila, & Seppala, 2000). Because drowsiness-related impairment poses a serious threat to driving safety, differentiating alcohol and drowsiness impairment may not be necessary because both might merit similar countermeasures.

The challenge of distinguishing between drowsiness and alcohol impairment was addressed in a simulator study. Peters and van Winsum (1998) report two studies involving a highway and an overtaking scenario to examine driver impairment in relation to fatigue and alcohol. In one study, data from eight drivers were used to examine the reliability of a driver state monitor to classify drowsiness. These participants drove on two separate days after working a full night shift. On the first day, they drove for 30 minutes along a simulated highway and overtook a slower lead vehicle. Data from this initial drive were used to train a classification algorithm and were also used to define the baseline condition. On the second day, the same subjects drove a longer version of the highway route with more overtaking scenarios for a total of three hours. Data from this longer drive were used to define the fatigue condition. In a separate study, data from ten participants were used to examine the reliability of the system to classify alcohol impairment. These drivers participated in three separate drives on different days. In the initial drive, they drove for 30 minutes along a simulated highway and overtook a slower lead vehicle. Data from this initial drive were used to train the classification algorithm and also were used to define the baseline condition. For the next two drives, the same participants drove the same route twice on separate days; that is, once without alcohol and once after consuming alcohol to reach a BAC of 0.05%.

For both these studies of drowsiness and alcohol impairment, the classification algorithm was derived using data from vehicle sensors that produced measures of brake pressure, lateral position (mean, standard deviation, min), steering wheel position (standard deviation), accelerator pedal position, vehicle speed (mean, standard deviation), engine RPM, eye blink activity, and time to line crossing. In addition to an analysis of the classification performance of the system, this study also examined the similarity between the patterns of driving behavior that were derived for the normal and impaired conditions. Notably, the cases of normal and fatigued conditions overlapped in 21% to 45% of cases. The degree of overlap was even greater between the normal and alcohol conditions (41% to 65%). This suggests that (1) drowsiness has a more distinct effect than alcohol on driving, and (2) alcohol-impaired driving can be similar to unimpaired driving. Such conclusions suggest that both drowsiness and alcohol impairment can be detected, and that drowsiness might produce a similar signature to that of low levels of alcohol.

Not only do impairment from drowsiness and alcohol share a resemblance, but these two sources may combine to compound impairment. As an example, Roehrs, Beare, Zorick, & Roth (1994) examined the effect of combining alcohol and drowsiness on lateral control in a driving simulator. In this study, twelve male subjects participated in four counter-balanced test conditions: (1) full sleep with eight hours of sleep and sober; (2) full sleep and alcohol to reach BAC 0.05%; (3) partial sleep deprivation with four hours sleep and sober; and (4) partial sleep and alcohol. The results indicated that alcohol consumption increased levels of drowsiness (as measured by the Multiple Sleep Latency Test) even in the full sleep condition. Moreover, alcohol reduced lateral control, as measured by increased lane departures, when combined with drowsiness in the partial sleep deprivation. Thus, this study demonstrated the potential for a synergistic effect of alcohol in combination with drowsiness such that the risk of a crash from a combination of these factors may be higher than in the presence of either factor alone.

Another study examined the combined effect of alcohol and fatigue in car-following and lane-keeping scenarios on a two-lane roadway (Fairclough & Graham, 1999). Four matched groups of male subjects (N = 64) completed two 40-minute drives along this route: (1) control group with no sleep deprivation and a placebo drink; (2) partly sleep-deprived group with four hours sleep on night before test, placebo drink; (3) full sleep-deprived group with no hours sleep on night before test and a placebo drink; and (4) alcohol group with normal night sleep before test and an alcohol drink to yield average 0.08% BAC. This study examined performance in terms of headway and lateral control measures. The measures of lateral control were most sensitive to the impairment effects of alcohol and drowsiness. Drowsiness and alcohol both resulted in significantly more lane departures than normal driving. Moreover, drowsiness resulted in significantly less steering activity compared to both the normal and alcohol conditions. This suggests that drowsiness-related performance decrements resulted from inattention and inaction, whereas the impaired lane keeping with alcohol may result from other mechanisms such as reduced safety margins associated with the willingness to take greater risks, or ineffective steering control. Such differences may point to algorithms that differentiate impairment due to alcohol and drowsiness, but are beyond the scope of the current study.

The combined evidence of these studies supports several important conclusions. First, alcohol and drowsiness can both undermine driving performance in a way that is reflected in easily measured variables, such as steering activity and lateral position. Second, the effects of drowsiness are similar to and perhaps greater than those of alcohol, making it difficult to create an algorithm specific to alcohol and making it likely that an algorithm might correctly identify impairment associated with drowsiness even when the driver has not been dosed with alcohol. Third, the combined effect of alcohol and drowsiness may be more than either alone, which diminishes the practical importance of differentiating between drowsiness and alcohol impairment. For many interventions, the need for an algorithm that is specific to alcohol may be outweighed by the benefit of detecting impairment independent of its cause.

4.3 Measures and metrics from simulator and on-road studies

When reviewing behaviors that could be assessed to support the detection of impairment, it is important to differentiate between a behavioral *measure* and an impairment *metric.* A measure refers to data that is sensed directly, such as speed or lane position. A metric refers to the summary statistic used to characterize behavior over time or space, such as median or standard deviation. Unfortunately, there is no consistency in the literature in terms of either the choice of measure or metric to quantify driver impairment associated with alcohol (Brookhuis, De Waard, & Fairclough, 2003; Ogden & Moskowitz, 2004; Tzambazis & Stough, 2000; Zador, et al., 2001). For this reason, it is difficult to use previous research to develop a coherent picture of the effect of alcohol on driving behavior and to identify the most sensitive behavioral measures of impairment.

Considering the limits of the current research base, Table 2 provides a preliminary assessment of measures of impairment. This evaluation speculates on the compliance of each measure with each proposed criteria. As a result, it is apparent that some measures are compliant (●) while others are non-compliant (○) or have an ambiguous status (⦿). For example, it is apparent that all of the measures are sensitive to impairment, although time headway (TH), stability of lateral position, and steering wheel activity have been more consistently associated with impairment effects of alcohol in the published research. However, headway measures presume the presence of a lead vehicle, which may not always be available. In contrast, speed, lateral position, and steering wheel activity are continuously available.

All of these measures would appear to have face validity in terms of intuitive relevance to impairment and safety (with perhaps the exception of steering wheel activity) and so might be understandable and acceptable to drivers. Despite this, speed may not be an acceptable measure if the public is averse to the perception that such data would be used for police enforcement of speeding fines. Finally, speed and steering wheel activity can be considered to be practical measures to the extent that effective and inexpensive sensor technologies exist. Sensors also exist to make headway and lateral position measures practical, although they tend to be more expensive and there remains some debate about the type of technology to be used for particular applications.

A range of sensors that go beyond measuring driver control inputs and vehicle state are becoming feasible for inclusion in production vehicles. These include video cameras for measuring head position and tracking eye movements. Pressure transducers in the seat can also measure movement and posture. Such sensors provide a non-intrusive, continuous, and potentially sensitive set of measures of driver impairment. For example, eyelid closure has long been recognized as a sensitive measure of drowsiness (Bergasa, Nuevo, Sotelo, Barea, & Lopez, 2006; Grace & Suski, 2001), and eye movements provide a promising indicator of distraction-related impairment (Liang, Reyes, & Lee, 2007a, 2007b) and are sensitive to alcohol (Marple-Horvat, et al., 2008; Moskowitz, et al., 1976). Eye tracking, particularly horizontal gaze nystagmus, has long been known to be sensitive to alcohol; however, eye-tracking systems that might be incorporated into production vehicles do not have the precision to estimate this metric but may support other metrics such as gaze concentration. Physiological measures are generally infeasible because they currently require a degree of intrusive instrumentation that undermines their acceptance and practicality. As vehicle automation and collision warning technology becomes more common, measures of headway and lane position will become increasingly available.

Table 2. Performance measures of driving for impairment detection.

Measure	**Sensitivity**	**Availability**	**Acceptance**	**Practicality**
Steering wheel activity	●	●	⊙	●
Vehicle lateral position	●	●	●	⊙
Eye-tracking measures	●	⊙	●	⊙
Postural stability	⊙	●	⊙	●
Time headway (TH)	●	○	●	⊙
Vehicle speed	⊙	●	○	●
Time to collision (TTC)	⊙	○	●	⊙
Physiological measures (e.g., EEG)	●	○	○	○

Filtering the potential measures according to the four criteria for real-time detection of driver impairment identifies several measures as promising candidates. Lateral control measures, including the steering wheel, and measures of longitudinal control that are not dependent on lead vehicles (e.g., speed, pedal input) may be the most viable measures from which to compute performance metrics to quantify impairment – especially when considering the need for continuous impairment detection. Lateral control measures and metrics are more promising than longitudinal control measures.

Lateral control measures refer to the vehicle state and are partially the consequence of drivers' control input. Drivers' control input can be a more sensitive indicator of impairment because it is not filtered through the vehicle dynamics. There are three particularly promising metrics of steering control:

- Steering reversals – Mean number of steering reversals computed with a 2-degree filter (Verwey & Veltman, 1996).

- Steering error – The degree to which the steering wheel movements deviate from a smooth trajectory defined by a second-order Taylor series approximation of the steering wheel movement. Steering error provides the base data for calculating steering entropy.

- Steering entropy – The predictability of the driver's steering responses, as defined in Nakayama et al. (1999); also see Boer (2001).

Whereas these measures can be sensitive to impairment, steering entropy can be problematic given that low entropy (and few reversals) can be indicative of attention to the driving task, very effective control input, or a lack of attention with insufficient control activity. In both these cases, the amount of input is minimal or absent, which gives the impression of equivalent states, although the case of diminished input associated with lack of attention is clearly an indication of impairment.

Steering entropy (SE) has been extensively developed and applied to detecting distraction-related impairment (Boer, 2001; Nakayama, et al., 1999). Other metrics of steering behavior have also been used (e.g., reversals, standard deviation of steering position, power spectral analysis). However, most of these metrics have been used primarily to characterize impairment effects of drowsiness or distraction. Naturally, since the sedative effect of alcohol does resemble some of the characteristics of drowsy driving, these alternative measures of steering activity may also be useful. Recent research suggests SE is very sensitive to distraction, but may not be any more sensitive to alcohol than simpler metrics of steering reversals (Rakauskas, et al., 2008). However, SE may be less sensitive to contextual variables (e.g., road curvature) than some of the alternative metrics. For this reason, SE may be a promising alternative to steering reversals.

4.4 Measures and metrics from naturalistic roadway observations

Substantial research has shown that behavioral cues available to police officers monitoring drivers can be indicative of alcohol-related driving impairment (Harris, 1980). These cues include those that can be observed while the car is in motion, such as weaving, and cues that are observed after a police officer has stopped a motorist, such as difficulty exiting the vehicle. A study confirmed the validity of these cues and found that the same cues used to identify drivers at 0.10% BAC were also effective in identifying those at 0.08% BAC (Stuster, 1997). Interestingly, none of these cues proved to be indicative of BAC levels below 0.08%, despite an attempt to identify such cues. Candidate cues sensitive to low BAC levels were identified through a review of relevant literature and through interviews with police officers. Table 3 shows the cues most sensitive to alcohol impairment; those that were not sensitive to alcohol impairment, such as speeding more than 10 miles over the speed limit, have not been included.

Table 3. Cues for identifying alcohol-impaired drivers from field studies (Stuster, 1997).

Element of Driving	**Behavioral Indicator** **(probability of impairment given observed behavior)**
Lane Position Maintenance	Weaving within a lane (0.52)
	Weaving across lane lines (0.54)
	Straddling a lane line (0.61)
	Swerving (0.78)
	Turning with a wide radius (0.68)
	Drifting during a curve (0.51)
	Almost striking a vehicle or other object (0.79)
Speed Control and Braking	Stopping problems (too far, too short, or too jerky) (0.69)
	Accelerating or decelerating for no apparent reason (0.70)
	Varying speed (0.49)
	Slow speed (10+ mph under limit) (0.48)
Vigilance	Driving in opposing lanes or wrong way on one-way (0.54)
	Slow response to traffic signals (0.65)
	Slow or failure to respond to officer’s signals (0.65)

Element of Driving	Behavioral Indicator (probability of impairment given observed behavior)
	Stopping in lane for no apparent reason (0.55)
	Driving without headlights at night (0.14)
	Failure to signal or signal inconsistent with action (0.18)
Judgment	Following too closely (0.37)
	Improper or unsafe lane change (0.35)
	Improper turn (too jerky, sharp, etc.) (0.50)
	Driving on other than the designated roadway (0.80)
	Stopping inappropriately in response to officer (0.69)
	Inappropriate or unusual behavior (throwing, arguing, etc.) (0.48)
	Appearing to be impaired (0.90)
	Illegal turn (0.19)

Similar to the conclusions based on simulator and on-road experiments, Table 3 suggests that lateral control may be sensitive to alcohol and that metrics such as the standard deviation of lane position might be most effective. An automated system to detect alcohol impairment was developed and found lateral control, rather than longitudinal control, to be sensitive. Data were collected over approximately 10 seconds as drivers decelerated in a 300 ft approach lane to a checkpoint (Stuster, 1999). The difference between the maximum and minimum lateral displacement and the minimum lateral displacement were both sensitive to alcohol and, when combined, they identified 67% of those with BAC levels above 0.04% in a population of drivers in which 80% had BAC levels of 0.04% to 0.12%.

4.5 Conclusion

The measurement of driver impairment often involves the formulation of metrics based on safety-relevant measures. Typically, these measures quantify behaviors associated with operating vehicle controls and the resulting vehicle state with respect to lateral and longitudinal control. The most common behavioral measures that have a demonstrated sensitivity to impairment are metrics of steering input and lateral control. Such measures are relevant, given that increased lateral variability is related to the probability of a lane departure that is itself a precursor to numerous types of crashes, including road departures. The standard deviation of lane position and steering reversals are two particularly promising metrics. These measures are safety-relevant in that they have direct implications for vehicle control. Other indicators of impairment, such as the association of eye movements and steering wheel position, may not have such a direct relationship with vehicle control, but may be diagnostic. Safety-relevant metrics may be advantageous if the algorithm produces an output that the driver must interpret: safety-relevant metrics are more likely to be understood and accepted.

The nature of alcohol impairment suggests that metrics based on integrated measures, rather than metrics based on single measures, represent a productive and underexplored approach. Alcohol impairment is most apparent in complex, multi-task situations. Consequently, metrics that reflect the joint performance and coordination of multiple driving tasks seem most likely to be sensitive to alcohol impairment. The relationship between horizontal eye position and steering wheel angle is one such example of coordinated tasks.

Overall, there is no critical mass of evidence to recommend specific metrics. Moreover, few studies have deliberately compared various measures and metrics in terms of sensitivity to alcohol in the driving context. Based on the four criteria listed above, the most promising measures and associated metrics are:

- Steering wheel activity – Reversals and steering error
- Vehicle lateral position – Standard deviation of lane position and range
- Eye movement – Gaze concentration

5 DETECTION CRITERIA AND CONSIDERATIONS

The criterion that defines a particular level of impairment represents an important design consideration for any impairment-detection system. Indeed, assessing impairment depends on specifying a classification rule rather than simply selecting a metric or behavioral signature. There are several frameworks for establishing the thresholds that define impairment: absolute criteria, relative criteria, and pattern-based criteria. Any practical algorithm is likely to contain elements of a pattern-based algorithm because alcohol impairment is most likely to be revealed in a combination of performance decrements rather than declines on one or two measures.

5.1 Absolute criteria

Absolute criteria classify behavior as impaired based on a threshold that is independent of the distribution of unimpaired behavior in a baseline condition. Absolute criteria reflect safety boundaries and human performance capabilities. Absolute criteria can also be set in terms of thresholds that are accepted by convention as having face validity for safety, such as excessive speed; for example, a speed threshold can be set at the speed limit plus a margin of 10%.

Based on these considerations, Brookhuis, De Waard, & Fairclough (2003) have proposed the "tentative" list of absolute criteria shown in Table 4. These criteria were derived from a subset of possible behavioral metrics from impairment conditions of alcohol, distraction, drowsiness, and visual occlusion[1]. Based on these criteria, a driver would be classified as impaired if the standard deviation of lane position changed by more than 4 cm (relative criterion) or was more than 25 cm (absolute criterion). Such diagnoses of impairment are considered valid because these criteria have been calibrated with the effects of alcohol at BAC 0.08%.

5.2 Relative criteria

Relative criteria classify behavior as impaired based on a threshold that is defined relative to the distribution of normal behavior in an unimpaired baseline condition. Relative criteria implicitly define safety in terms of the *change* in crash risk from the baseline condition. Such criteria are determined by testing the statistical significance of the deviation between observed behaviors and the distribution of behaviors in the baseline condition. A statistically significant deviation signifies that the observed behavior is probably not representative of the baseline condition and is, therefore, presumed to represent an impaired condition. Based on these considerations, Brookhuis, De Waard, & Fairclough (2003) have also proposed the list of relative criteria shown in Table 4. The criteria listed in Table 4 represent absolute and relative decrements in driving performance based on observed driving behavior, and the criteria listed in Table 5 are based on observations of drivers in the visual occlusion paradigm. The absolute criteria in Table 5 are absolute in the sense they do not depend on the driver, but are adjusted according to the speed of the driver.

---

[1] Visual occlusion is a research technique that blocks the driver's ability to view the driving environment except when the driver presses a button. On pressing the button, the driving environment is revealed for one second.

Table 4. Definition criteria for following too closely, straddling lanes, and driving too fast (Brookhuis, et al., 2003).

Measure	Absolute Change	Relative Change
Following too close:		
Time headway to lead vehicle (TTC)	<0.7 s	-0.3 s
Lane keeping:		
Steering standard deviation (SD)	>1.5	+0.5
Lateral deviation (SD) of the vehicle	>0.25 m	+0.04 m
Minimum time-to-line crossing (TLC) right	<1.3 s	-0.3 s
Minimum time-to-line crossing (TLC) left	<1.7 s	-0.2 s
Median TLC (right lane)	<3.1 s	-0.7 s
Median TLC (left lane)	<4.0 s	-1.4 s
Driving too fast:		
Vehicle speed[2]	limit +10%	+/- 20%

Table 5. Lane-keeping criteria based on visual occlusion (Brookhuis et al., 2003).

Measure	Speed (km/h)	Absolute Criteria
Standard deviation of lane position	>50	0.25 m
Standard deviation of steering wheel	At 60	1.7°
	>80-120	1.5°
Median TLC	60	6.0 s
	80	5.7 s
	100	5.0 s
	120	4.2 s
15% TLC	60	3.8 s
	80	3.5 s
	100	3.1 s
	120	2.9 s
Minimum TLC at different speeds		1.1 s

[2] For the purposes of this project, we considered relative change in speed associated with driving too fast only as an increase of 20%.

5.3 Pattern-based criteria

Pattern recognition criteria are based on neural nets or similar statistical approaches to generate multivariate rules that are "trained" to differentiate between normal and "impaired" states. For example, De Waard, Brookhuis, & Hernández-Gress (2001) conducted a driving simulator study with 20 drivers to evaluate the performance of an impairment classification system to detect driver distraction based on input parameters of pedal and steering control, vehicle speed, lateral position, and headway to lead vehicle.

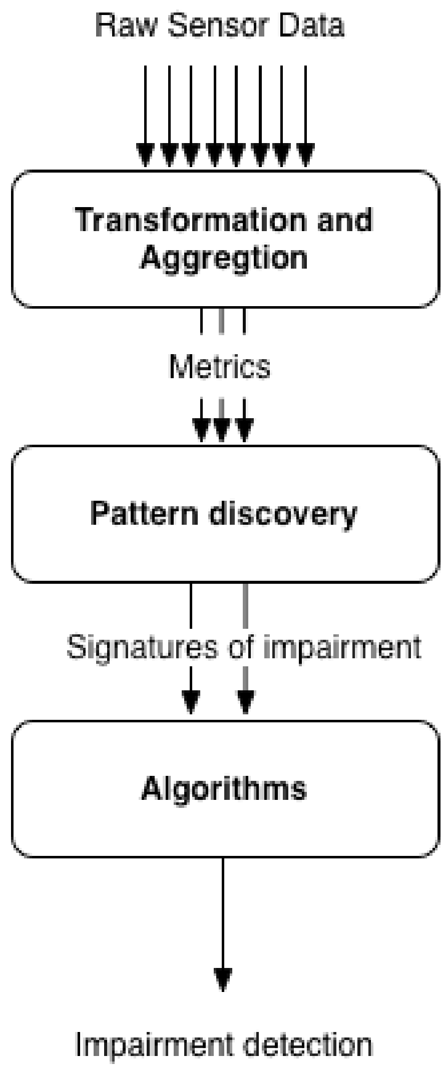

Figure 6. Representation of processes in generic classification system to classify normal and non-normal conditions.

The classification system had three components, as shown in Figure 6. First, the original set of measures from the sensor data is reduced to a number of composite variables or metrics that transform and aggregate the individual measures. Second, pattern discovery is applied to identify signatures of impairment. These signatures are combinations of metrics that differentiate normal and non-normal conditions. Third, a classification algorithm (e.g., Support Vector Machine or decision tree) recognizes patterns of measured behavior and establishes thresholds that differentiate between normal and non-normal conditions. This requires a training process in which the algorithm is told the actual condition for the assigned data so that patterns and thresholds can be learned. The algorithm is then used to recognize those learned patterns indicative of normal and impaired conditions for new data sets. Note that this system in effect uses relative criteria, but unlike other conventions that apply thresholds to individual measures (see Table 4), it adopts multidimensional thresholds for derived components that characterize patterns of behavior indicative of impairment. The result of this process is the detection of non-normal conditions, such as instances of impaired driving.

5.4 Selecting appropriate criteria: matching criteria to mitigations

Each of these general criteria has different strengths and weaknesses. Most generally, the absolute criteria are most easily understood and explained. In contrast, the relative and pattern-based criteria are likely to be much more effective in distinguishing impaired from unimpaired drivers because they make more complete use of the data. The relative importance of understandability and effectiveness depends on the particular mitigation that the algorithm will ultimately guide. If the algorithm drives immediate feedback to the driver, then driver acceptance of the information is critical and understandability would be relatively important. Drivers might not understand, and therefore ignore, an impairment warning based on how quickly their eyelids close, but would be more likely to understand a warning based on weaving or swerving. If the algorithm guides collision warning adaptation, then the effectiveness would be most critical. In addition, some algorithms, such as those based on driver state (e.g., postural stability and eye movements), do not have any well-defined absolute criteria and their output does not have benchmarks that can be easily understood by the driver. The primary purpose of this study is to develop an algorithm that can identify alcohol-impaired drivers, so the emphasis will be on pattern-based criteria in which multiple metrics are combined and evaluated to maximize correct detection and minimize false positives.

6 DATA COLLECTION METHODS

Algorithm development and evaluation requires detailed data from alcohol-impaired drivers in a controlled, yet realistic situation. Data were collected from drivers at three BAC levels experiencing representative driving scenarios in a high-fidelity driving simulator. The data collection involved 108 drivers from three age groups (21-34, 38-51, and 55-68 years of age) driving through representative situations on three types of roadways (urban, freeway, and rural) at three levels of alcohol concentration (0.00%, 0.05%, and 0.10% BAC). BAC dosing was limited to 0.10% due to practical and ethical reasons. The following sections summarize the data collection methods: participant population, simulator and sensor suite, experimental design, and dependent variables. Details are provided in appendices referenced in each section.

6.1 Participants

One hundred and eight participants completed all three drives of the study. Participants were healthy men and women aged 21 and older, with a valid driver's license, and who were moderate to heavy drinkers. All drivers had been licensed for at least two years and drove a minimum of 10,000 miles per year. Efforts were made to recruit a racially and ethnically diverse participant population. Participants were paid $250 for completing all study sessions. Pro-rated compensation was provided for participants who did not complete the study. Inclusion and exclusion criteria include:

- Possess a valid US driver's license
- Licensed driver for two or more years
- Drive at least 10,000 miles per year
- Restrictions on driver's license limited to vision
- Not currently taking illegal drugs or drugs that interact with alcohol

  All participants provided a urine specimen on each of the three experimental days and were screened for 10 types of drugs: methamphetamine, morphine, cocaine, marijuana, PCP, benzodiazepines, barbiturates, methadone, trycyclic antidepressants, and amphetamine. Those participants who tested positive were discontinued from participation in the study. The urine specimens of females were screened for hCG, a pregnancy hormone. Due to health concerns, those participants who tested positive for pregnancy or illicit drug use were discontinued from the study.

- Does not use any special equipment to drive, such as pedal extensions, hand brake or throttle, spinner wheel knobs, or other non-standard equipment that would limit interpretation of accelerator pedal, brake pedal, or steering inputs.
- Be a moderate to heavy drinker as determined by the Quantity-Frequency-Variability (QFV) scale. Appendix A contains the survey and scoring that defines the specific inclusion and exclusion criteria for this criterion.

Participants were recruited to fit specific age and gender characteristics and, as shown in Table 6.

Table 6. Number of participants reporting to visits[3].

Group	Enrolled Visit 1 (Screening)	Passed Screening	Visit 2	Visit 3	Visit 4	Completed
Young Male	35	24	21	18	18	18
Young Female	24	20	20	18	18	18
Middle Male	27	20	20	18	18	18
Middle Female	27	20	19	18	18	18
Older Male	31	26	22	20	19	18
Older Female	21	20	19	18	18	18
Total	165	130	121	110	109	108

Table 7 summarizes the population of participants showing the intended balance of age and gender. The distribution of drinking patterns based upon the QFV reflects that participants in this study were more likely to be heavy drinkers rather than moderate drinkers. The sleepiness scale reveals that, with the exception of the older females, drivers were more drowsy at the completion of their drive that they were before it started. Overall simulator sickness scores reflect the relative lack of adverse affects associated with driving in a simulated environment.

Table 7. Participant characteristics.

		Age 21-34		**Age 38-51**		**Age 55-68**	
Variable		**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Number completed		18	18	18	18	18	18
Mean age (years)		26.56	26.83	43.22	44.72	59.56	61.06
Mean height (inches)		70.65	65.53	70.61	65.35	70.14	64.76
Mean weight (pounds)		199.81	159.56	220.61	175.29	211.86	172.86
Mean body mass index		27.9	26.1	31.1	28.6	30.2	29.0
Distribution of Drinking Patterns	Moderate	2	7	4	9	6	7
	Heavy	16	11	14	9	12	11
Sleepiness scale	Pre-Drive	2.3	2.6	2.3	2.2	2.0	2.0
	Post-Drive	2.7	3.1	2.8	2.4	2.3	2.0
Simulator sickness score		8.9	18.3	19.2	21.9	17.9	15.1

[3] A total of nine participants passed the screening but were not scheduled for a second visit.

6.2 Simulator and sensor suite

The National Advanced Driving Simulator (NADS) is located at The University of Iowa's Oakdale Campus. It consists of a 24-foot dome in which an entire car is mounted. All participants drove the same vehicle—a 1996 Malibu sedan. The motion system, on which the dome is mounted, provides 400 square meters of horizontal and longitudinal travel and ±330 degrees of rotation. The driver feels acceleration, braking, and steering cues as if he or she were actually driving a real vehicle. Each of the three front projectors has a resolution of 1600 x 1200; the five rear projectors have a resolution of 1024 x 768. The edge blending between projectors is five degrees horizontal. The NADS produces a complete record of vehicle state (e.g., lane position) and driver inputs (e.g., steering wheel position), sampled at 240 Hz.

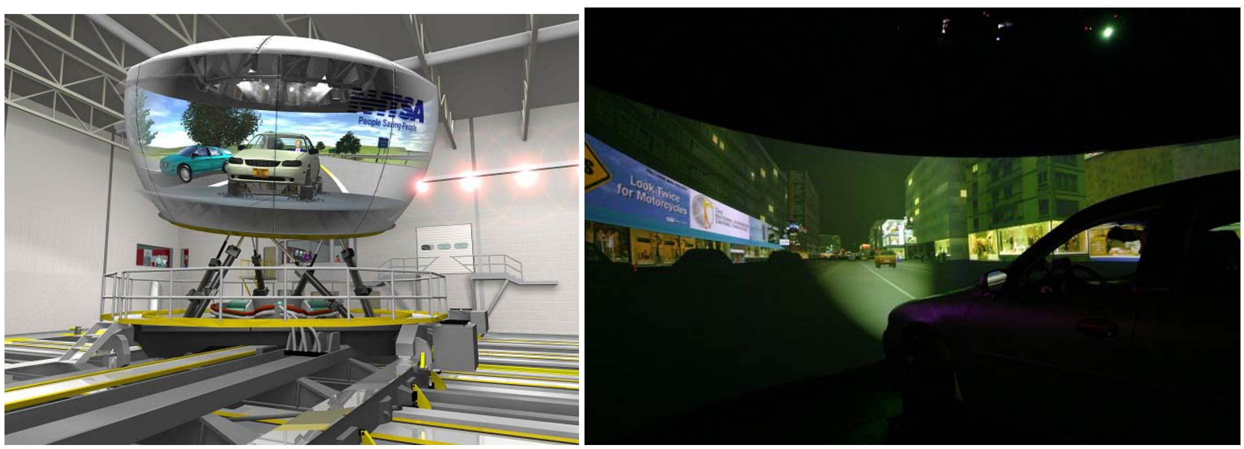

Figure 7. Representation of NADS-1 driving simulator (left) with a driving scene from inside the dome (right).

The cab is equipped with a Face Lab™ 4.0 (Seeing Machines, Canberra, Australia) eye-tracking system that is mounted on the dash in front of the driver's seat above the steering wheel. The worst-case head-pose accuracy is estimated to be about 5º. In the best case, where the head is motionless and both eyes are visible, a fixated gaze may be measured with an RMS error of 2º.

Softflex™ (Vista Medical, Winnipeg, Manitoba) force sensor array (FSA) mats were positioned on the bottom and back of the driver's seat. Each of these mats consists of 256 piezo resistive sensors distributed in a 16 x 16 array over an area of approximately 17" x 17." These mats sample data from the sensor array a rate of 35 Hz.

An Alco-Sensor IV (Intoximeters, Inc., St. Louis, MO) breath alcohol-testing instrument was used to measure participants' BAC. The hand-held sensor uses a fuel cell to determine BAC level. The system is approved by the US DOT for evidential use and exceeds the federal model specification for traffic enforcement and Omnibus Breath Alcohol Testing. The system is designed to measure BAC levels from 0.00% to 0.40% with drift of less than 0.005% BAC over several months. The system was checked at least every other day for calibration and recalibrated using an approved dry gas standard.

6.3 Driving scenarios

Each drive was composed of three nighttime driving segments. The drives started with an urban segment composed of a two-lane roadway through a city with posted speed limits of 25 to 45 mph with signal-controlled and uncontrolled intersections. An interstate segment followed that consisted of a four-lane divided expressway with a posted speed limit of 70 mph. Following a period in which drivers followed the vehicle ahead, they encountered infrequent lane changes associated with the need to pass several slower-moving trucks. The drives concluded with a rural segment composed of a two-lane undivided road with curves. A portion of the rural segment was gravel. These three segments mimicked a drive home from an urban bar to a rural home via an interstate. Events in each of the three segments combined to provide a representative trip home in which drivers encountered situations that might be encountered in a real drive. Scenario events are summarized in Table 8.

Table 8. Summary of events for each of the three segments of the drive.

Segment	**Event Name (number)**	**Description**	**Approximate Duration (seconds)**
Urban	Pull Out (101)	Pull out of parallel parking spot into traffic	30
	Urban Drive (102)	Drive on a narrow two-lane road with traffic and parked vehicles	45
	Green Light (103)	Navigate green traffic light on urban two-lane road with parked vehicles along the road, oncoming traffic, traffic behind driver	30
	Yellow Dilemma (104)	Navigate yellow light dilemma on urban two-lane with parked vehicle, oncoming traffic, traffic behind driver	75
	Left Turn (105)	Left turn at signalized intersection (no green arrow, no dedicated turn lane), oncoming traffic, variety of gaps	80
	Urban Curves (106)	Three curve segments of mixed radius of curvature	180
Freeway	Turn On Ramp (201)	Turn right onto interstate on-ramp	30
	Merge On (202)	Merge onto interstate	50
	Following (203)	Intermittent slower-moving truck traffic in the driving lane and a single slow moving passenger vehicle in the passing lane, interleaved with a CD-	180

Segment	Event Name (number)	Description	Approximate Duration (seconds)
		change distraction task	
	Merging Traffic (204)	Approach second interchange, interact with traffic merging into interstate	60
	Interstate Curves (205)	Navigate three curves on interstate	185
	Exit Ramp (206)	Take exit ramp off interstate	30
Rural	Turn Off Ramp (301)	Turn right from ramp onto rural two-lane road	30
	Lighted Rural (302)	Lighted two-lane rural road, 55 mph	90
	Transition to Dark Rural (303)	Straight roadway that transitions between lighted and unlit	20
	Dark Rural (304)	Unlit straight and curved road, segments, center and road edge marking are faded and the road surface is grayish. Data from the Hairpin Curve is not included in this.	270
	Dark Rural Hairpin Curve (304)	A hairpin turn and a vertical curve located within the Dark Rural (304). Data for Dark Rural and Hairpin Curve are mutually exlusive.	30
	Gravel Transition (305)	Transition to gravel surface on straight road	30
	Gravel Rural (306)	Gravel road (straight and curves)	90
	Driveway (307)	Pull into driveway with gravel	30

Throughout the urban section, a series of potential hazards required drivers to scan the roadside. These hazards included pedestrians, motorbikes, and cars entering and exiting the roadway. These hazards had paths that would cross the driver's path if they were to remain on their initial headings. There was an instance where a pedestrian crossed the driver's path well in front of the driver.

Because each participant drove three times, once for each BAC level, three scenarios with varied event orders were required to minimize the learning effects from one drive to the next. For each of the three scenarios, there were the same number of curves and turns, but the order of the curves varied. For example, the position of the left turn in the urban section varied so that it was located at a different position for each drive. Additionally, the order of the left and right rural curves varied between drives. The scenario specification in Appendix B provides additional details concerning the differences between the three drives and the events.

6.4 Experimental design and independent variables

A 3 x 2 x 3 between-between-within subjects design exposed six groups of participants to three BAC levels. Between-subject independent variables were age group (21-34, 38-51, and 55-68 years) and gender. The within-subject independent variable was BAC (0.00%, 0.05%, and 0.10%).

Three factors motivated the choice of the age ranges. The first factor was that only those who could legally drink in the state of Iowa would be included. Therefore, enrollment in the study was restricted to those 21 years of age or older. The second factor was that to the extent possible, the entire spectrum of adults who drink and drive should be included, which motivated including a group with maximum age of approximately 70. The third factor was that the age ranges should be uniform, with equal spacing between them. Based on these requirements, the age groups were 21-34, 38-51, and 55-68, so that each group had a range of fourteen years.

6.5 Procedure

Participants were recruited with newspaper ads, internet postings, and referrals (see Appendix C for recruitment materials). An initial telephone interview determined eligibility for the study. Applicants were screened in terms of health history, current health status, and use of alcohol and other drugs (see Appendix D). The Cahalan, Cisin, and Crossley Quantity-Frequency-Variability (QFV) scale was used to determine whether applicants were moderate drinkers or heavy drinkers and eligible for participation in the study (see Appendix A). The Audit survey was used to exclude chronic alcohol abusers (see Appendix E). Pregnancy, disease, or evidence of substance abuse resulted in exclusion from the study. Participants taking prescription medications that interact with alcohol were also excluded from the study.

Each participant participated in four sessions, the last three separated by one week. Order of target BAC levels and scenario event sequence were counterbalanced across participants, as shown in Appendix F. The time of day of each of the three sessions was the same for a given participant.

Appendix F includes the full experimental protocol. On study Visit 1 (screening), upon arrival at the NADS, each participant gave informed consent to participate in the study and received a copy of the signed informed consent form (see Appendix I). They then provided a urine sample for the drug screen and, for females, the pregnancy screen. During a five-minute period following these activities, the participant sat alone in the room where subsequent measurements of blood pressure, heart rate, height, and weight were made. Cardiovascular measures within acceptable ranges (systolic blood pressure = 120 ± 30 mmHg, diastolic blood pressure = 80 ± 20 mmHg, heart rate = 70 ± 20) confirmed eligibility for the study. If participants met study criteria, they were then administered a breath alcohol test and verbally administered the QFV (see Appendix J) and the Audit Survey (see Appendix E) to further confirm eligibility. If participants met study criteria, they completed demographic surveys. These surveys included questions related to crashes, moving violations, driver behavior, drinking, and driving history (see Appendix K). Participants viewed an orientation and training presentation (see Appendix L) that provided an overview of the simulator cab and the in-cab task they were asked to complete while driving. Participants then completed the practice drive (see Appendix G for in-cab protocol and Appendix H for control room logs) and completed surveys after their drive about and how they felt and about the realism of the simulator (see Appendices M and N). The practice drive included making a left hand turn, driving on two- and four-lane roads, and practicing the CD changing task. Date and time for Visits 2, 3, and 4 were confirmed with participants at the completion of this visit.

During Visits 2, 3, and 4 all participants completed a urine drug screen and, for females, a pregnancy screen to confirm eligibility for the study. Participants waited for five minutes following these activities during which the participant sat alone in the room where subsequent measurements of blood pressure and heart rate were obtained to determine study eligibility. If participants met study criteria, they then received a breath alcohol test, the QFV, and the Audit Survey to further confirm eligibility. If eligible to continue, the time and duration of last sleep, and time and contents of last meal were recorded (Appendix O). Age, gender, height, weight, and drinking practice were used to calculate the alcohol dose. The Sahlgrenska Formula was used to estimate body water for each participant in order to calculate the amount of alcohol and juice required to reach the target BAC (Equation 1 and Equation 2). Participants were served three equal-sized drinks at 10-minute intervals and were instructed to pace each drink evenly over the 10-minute period. NADS staff monitored the participants periodically throughout the drinking period to ensure an even pace of drinking.

Equation 1. Sahlgrenska formula for body water.

$$Body\ Water\ (liters) = \begin{cases} if\ Male, -10.759 + 0.192 \times Height(cm) + 0.312 \times Weight(Kg) - 0.078 \times Age(years) \\ if\ Female, -29.994 + 0.294 \times Height(cm) + 0.214 \times Weight(Kg) - 0.0004 \times Age(years) \end{cases}$$

Equation 2. Vodka volume formula.

$$Amount\ of\ Vodka\ (mL) = \frac{\dfrac{DesiredPeakBAC + \dfrac{Total\ Absorption\ Time}{60 \times Standard\ Clearance\ Rate}}{H_20\ in\ Blood} \times \dfrac{Sahlgrenska\ Body\ Water \times 10}{Specific\ Gravity\ of\ Alcohol}}{0.4}$$

$$\left[\begin{array}{c} \text{Specific Gravity of Alcohol} = 0.79 \\ \text{Sahlgrenska Body Water} = \text{Equation 1} \\ \text{H}_2\text{O in Blood} = 85\% \\ \text{Desired Peak BAC} = 0.115\% \text{ or } 0.065\% \\ \text{Standard Clearance Rate} = 0.017 \end{array}\right.$$

On the days when participants were dosed to achieve 0.10% and 0.05% BAC, the amount of alcohol consumed was calculated to produce a peak BAC of 0.115% or a peak BAC of 0.065%. On the 0.00% peak BAC day, the drink consisted of one part water and 1.5 parts orange juice. Each of the glasses had its rim swabbed with vodka and 10 ml of vodka was floated to produce an initial taste and odor of alcohol.

Sixteen minutes after the end of the third drink, BAC measurements were taken at two- to five-minute intervals until the target BAC (± 0.005%) was reached. BAC values were plotted to identify the peak BAC. Each data point includes the measurement error of 0.005% associated with the accuracy of the sensor. A sample plot for identifying when BAC has peaked and begun to decline to the point where the participants could go into the simulator with a BAC of 0.10% is provided in Appendix P. Prior to insertion in the simulator, a participant must have had at least two declining data points and been within the target range, or below if the target was not reached. Peak BAC was expected 30 minutes after the end of the third drink.

When the target BAC was reached, the participants drove in the NADS. All data were collected as the BAC declined to minimize extraneous variation associated with the effect of rising and falling BAC levels and to represent the most likely situation under which alcohol-impaired driving occurs. As soon as the simulator returned to the dock and the participant exited the simulator (within 5 minutes of completing the drive), a BAC measurement was obtained, followed by an SFST (see Appendix Q). The individuals conducting the SFST were trained according to NHTSA's guidelines. The primary individual leading the training and administration of the SFST was a former police officer with 20 years of experience in law enforcement. The Stanford Sleepiness scale was also administered before and after each drive (see Appendix R).

Participants were not informed of their measured BACs until their participation in the study was completed. On all experimental days, the participants were transported home after their BAC dropped below 0.03%. Measurements of BAC were taken every 20 minutes after exiting the simulator to indicate when the participant was to be transported home. In the case of the 0.00% BAC condition, participants were held for at least three measurements before being transported home. At the end of Visit 4, participants were debriefed (see Appendixes S and T) and paid $250. Pro-rated compensation was provided for participants who did not complete the study.

6.6 Dependent variables

Each drive consisted of 19 distinct events, so the specific dependent variables were not constant across the drive. The primary measures examined across the drive were standard deviation of lane position, average speed, and standard deviation of speed. Average speed over a segment was calculated as the mean velocity irrespective of the speed limit as was the standard deviation of speed. The scenario specification describes the dependent variables for each event (see Appendix B).

A critical step in developing algorithms concerns the translation of measures into alcohol-sensitive metrics. The data collection produced an archive of more than 100 hours of driving data, sampled at 240 Hz, across 19 distinct events. Metrics of driving performance are not invariant over roadway situations. The road type, traffic situations, and the particular maneuver the driver happens to be performing all make it important to consider the metrics in the context of the roadway situation. Alcohol-related metrics may differ on a segment-by-segment basis and even on a minute-by-minute basis. Lane position averaged over a trip might mask lane-straddling behavior on a freeway that did not occur on an urban arterial during the same trip. A challenge in aggregating the raw data is to avoid combining qualitatively different data for different behaviors and situations, for this reason data were aggregated over each event.

The following sections describe the algorithm development and evaluation that builds on the data reduction. The primary objectives associated with the algorithm development and analysis include:

- Understand how driving-related metrics reflect the impairment associated with BAC at 0.05% and 0.10%
- Determine the robustness of these metrics with respect to individual differences such as age, and gender, and roadway situation
- Develop algorithms to detect alcohol-related impairment
- Compare robustness and timeliness of metrics and algorithms

These sections serve two general purposes: to describe the effect of alcohol on driving and to assess how well algorithms can identify BAC levels over the legal limit.

7 ALCOHOL LEVELS AND DRIVING PERFORMANCE

This section describes participants' BAC levels and the effect of these levels on three standard driving performance measures: standard deviation of lane position, mean speed, and speed deviation. This section first describes the stability of driving performance across the three drives, indicating where drivers might have adapted to the scenario with repeated exposure to the events. Although the independent variables were counterbalanced to minimize learning effects, substantial changes in response to events over time might undermine the sensitivity and limit the generalization of the results. This section then describes how the three dependent measures (lane deviation, average speed, and speed deviation) varied as a function of BAC, age, and gender. This analysis also describes the sensitivity of driving performance during each event to the BAC levels. Overall, this section addresses the following objectives:

- Understand how driving-related metrics reflect the impairment associated with BAC at 0.05% and 0.10%
- Determine the robustness of these metrics with respect to individual differences such as age, and gender, and roadway situation

Based on previous research concerning the effect of alcohol on driving performance, the analysis focuses on three measures, which are summarized in Table 9. It is not meaningful to calculate these measures for some events, such as pulling out of a parking spot (Event 101), which is indicated in the table.

Table 9. Events for which Standard Deviation of Lane Position, Average Speed, and Speed Deviation were analyzed.

Event Name (number)	**Lane Deviation**	**Average Speed**	**Speed Deviation**
Pull Out (101)			
Urban Drive (102)	√	√	√
Green Light (103)	√	√	√
Yellow Dilemma (104)	√	√	√
Left Turn (105)		√	
Urban Curves (106)	√	√	√

Event Name (number)	Lane Deviation	Average Speed	Speed Deviation
Turn On Ramp (201)			
Merge On (202)			
Following (203)			
Merging Traffic (204)			
Interstate Curves (205)	√	√	√
Exit Ramp (206)		√	
Turn Off Ramp (301)			
Lighted Rural (302)	√	√	√
Transition to Dark Rural (303)	√	√	√
Dark Rural (304)	√	√	√
Dark Rural hair pin curve (304)	√	√	√
Gravel Transition (305)		√	√
Gravel Rural (306)	√	√	√
Driveway (307)			

7.1 Blood alcohol concentrations of participants

Table 10 shows the BACs that were obtained pre-drive, post-drive, and the pre-post drive average. The values for the conditions where drivers were not dosed were 0.00% BAC. The dosing and testing procedures effectively produced the intended experimental conditions. The median BAC for the 0.05% condition was 0.048%, and the median BAC for the 0.10% condition was 0.095%. Figure 8 shows the individual data points that underlie the data in Table 10. Points under the diagonal line represent cases where the post-drive BAC was lower than the pre-drive BAC. These points represent cases where the drivers had declining BAC during the drive, which was a goal of the dosing protocol that was achieved in 204 of the 216 cases.

Table 10. Summary of BAC levels for the two experimental conditions.

	0.05% BAC (*N* = 108)			**0.10% BAC (*N* = 108)**		
Test Time	***M***	***SD***	**Median**	***M***	***SD***	**Median**
Pre-drive	0.053	0.005	0.054	0.098	0.009	0.102
Post-drive	0.042	0.006	0.043	0.088	0.009	0.090
Mean	0.047	0.005	0.048	0.093	0.008	0.095

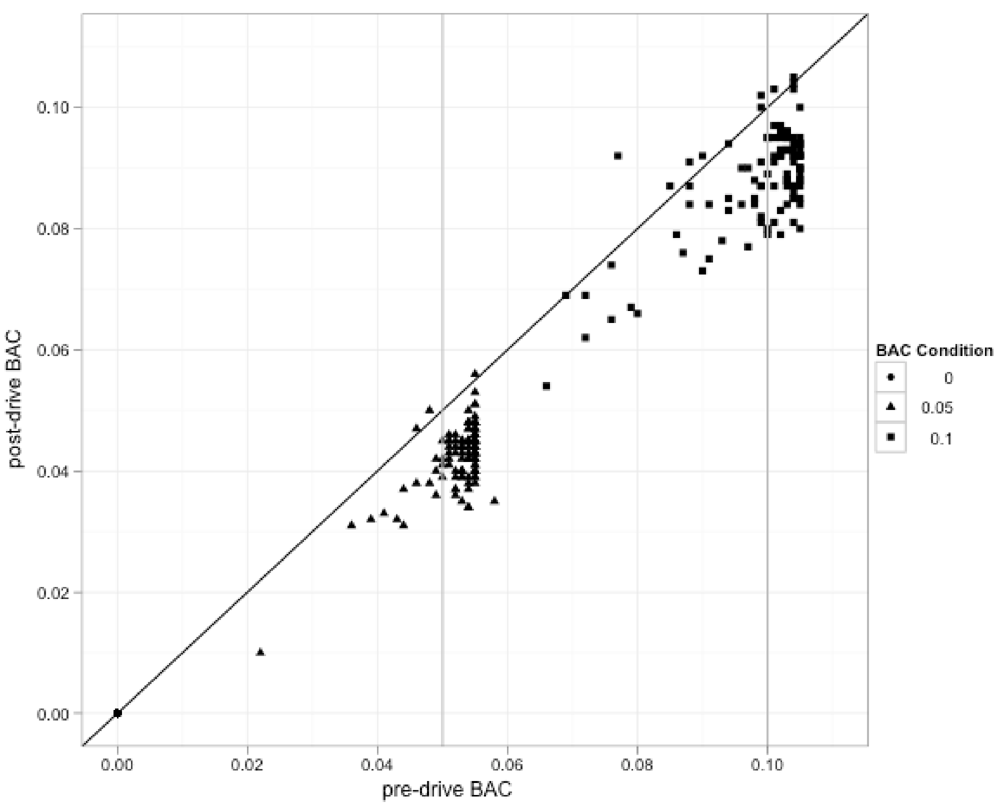

Figure 8. Pre- and post-drive BAC levels for the two experimental conditions.

7.2 Driver adaptation to events with repeated exposure

Ideally, driver response to each event would be unaffected by previous exposure to the events—their response would be similar across the three drives. The following analysis examines the extent to which the results departed from that ideal.

Figure 9 shows lane deviation, average speed, and speed deviation by drive number across scenario events. No lane deviation measures had statistically reliable differences as a function of drive number (1, 2, or 3). The variation between events is much greater than the variation across drives. The standard deviation of lane position is stable over drives.

For average speed, there were statistically significant differences as a function of drive number in Interstate Curves (205), Exit Ramp (206), Transition to Dark Rural (303), Gravel Transition (305) and Gravel Rural (306) events. Although statistically significant, the differences between drives are small relative to the differences between events and only in Gravel Rural (306) is the difference substantial. This event consisted of driving on a gravel road, which might not have been a routine driving experience for some participants, and the increase may be explained by increased comfort with the task.

For the speed deviation measures, there were statistically significant differences as a function of drive number in the Green Light (103) and Gravel Transition (306) events. Although not statistically significant, Yellow Dilemma (104) had the largest change in speed variation. This difference likely stems from more participants noticing and braking for the yellow light on subsequent drives.

Overall, these analyses show generally stable driving behavior with a minimal amount of learning or adaptation over the three drives. Several events, such as the gravel road and the several urban events show moderate changes in speed selection and modulation. However, these changes are small relative to the differences between events.

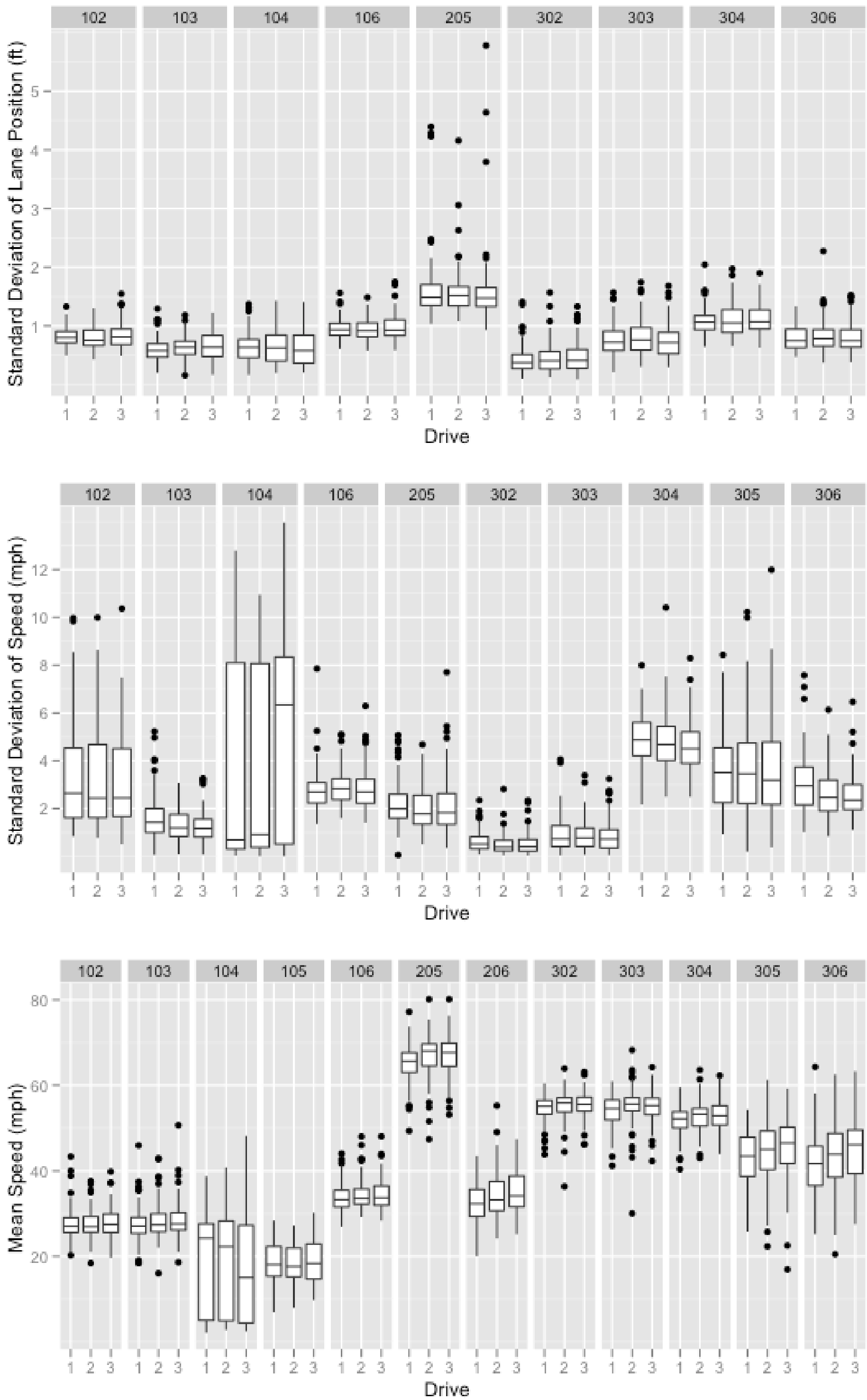

Figure 9. Standard deviation of lane position, average speed, and speed deviation by drive number across scenario events for Urban Drive (102), Green Light (103), Yellow Dilemma (104), Urban Curves (106), Interstate Curves (205), Lighted Rural (302), Transition to Dark Rural (303), Dark Rural (304), Gravel Transition (305), and Gravel Rural (306).

7.3 Effect of BAC levels on driving performance across roadway situations

The purpose of this analysis is to assess the overall sensitivity of common driving metrics to alcohol and to assess how robust they are to differences between roadway events and to differences between urban, freeway, and rural driving segments. This analysis also determines the robustness of these metrics with respect to individual differences such as age and gender.

Due to simulator restarts, there were instances in which data could not be collected for the entire drive, indicated by cases where there were less than 108 data points. No efforts were made to replace the missing data. The tables (see Table 11 to Table 13) also show the lane deviation, average speed measures, and speed deviation measures, and lane deviation measures by BAC group. Figure 10 illustrates those differences across the events. Tables showing these measures by age group and gender are included in Appendix U.

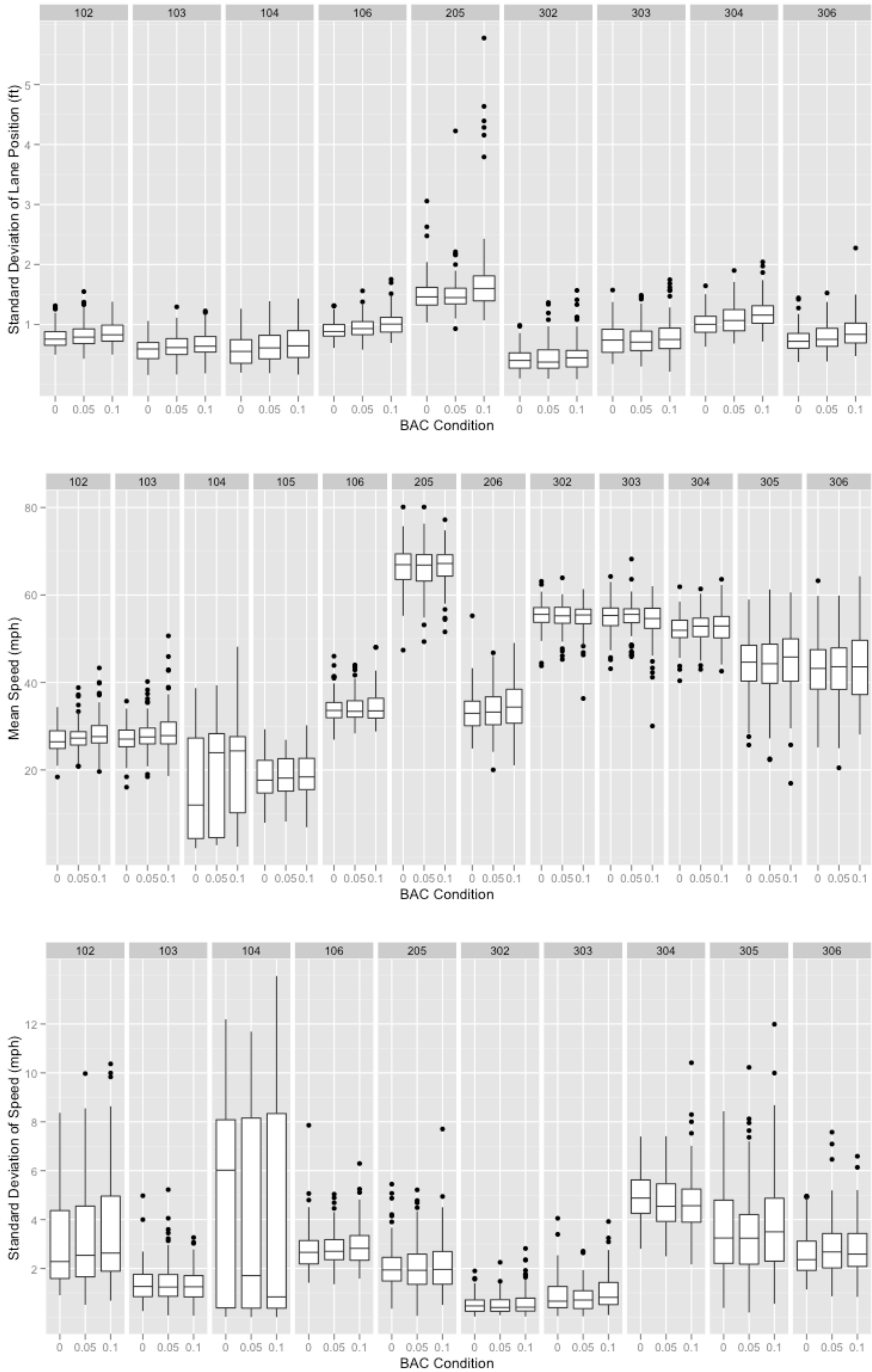

Figure 10. Effect of BAC on standard deviation of lane position, average speed, and speed deviation for Urban Drive (102), Green Light (103), Yellow Dilemma (104), Urban Curves (106), Interstate Curves (205), Lighted Rural (302), Transition to Dark Rural (303), Dark Rural (304), Gravel Transition (305), and Gravel Rural (306).

Table 11. Lane deviation (ft) by BAC group across events.

	BAC Group											
	0.00% BAC			0.05% BAC			0.10% BAC			Total		
Event	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*
Urban Drive (102)	**.78**	**108**	**.17**	**.82**	**107**	**.21**	**.85**	**108**	**.18**	**.82**	**323**	**.19**
Green Light (103)	**.58**	**108**	**.19**	**.63**	**107**	**.22**	**.66**	**108**	**.23**	**.63**	**323**	**.21**
Yellow Dilemma (104)	**.59**	**107**	**.26**	**.63**	**108**	**.26**	**.70**	**105**	**.30**	**.64**	**320**	**.28**
Urban Curves (106)	**.90**	**108**	**.16**	**.94**	**108**	**.18**	**1.03**	**108**	**.21**	**.96**	**324**	**.19**
Interstate Curves (205)	**1.51**	**107**	**.30**	**1.51**	**108**	**.35**	**1.75**	**105**	**.75**	**1.59**	**320**	**.52**
Lighted Rural (302)	.42	108	.19	.45	108	.26	.48	108	.27	.45	324	.24
Transition to Dark Rural (303)	**.75**	**108**	**.25**	**.77**	**108**	**.28**	**.79**	**108**	**.30**	**.77**	**324**	**.28**
Dark Rural (304)	**1.02**	**108**	**.20**	**1.09**	**108**	**.23**	**1.20**	**106**	**.27**	**1.10**	**322**	**.25**
Dark Rural Hairpin Curve (304)	.88	108	.27	.93	108	.32	.95	106	.34	.92	322	.31
Gravel Rural (306)	**.75**	**108**	**.20**	**.81**	**108**	**.24**	**.87**	**106**	**.27**	**.81**	**322**	**.24**
Note. BAC differences shown in bold are statistically significant at $p < 0.05$.												

Table 12. Average speed (mph) by BAC group across events.

	BAC Group											
	0.00% BAC			0.05% BAC			0.10% BAC			Total		
Event	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*
Urban Drive (102)	**26.95**	**108**	**3.14**	**27.58**	**107**	**3.33**	**28.51**	**108**	**4.00**	**27.68**	**323**	**3.56**
Green Light (103)	**27.06**	**108**	**3.34**	**27.98**	**107**	**3.67**	**29.03**	**108**	**4.97**	**28.02**	**323**	**4.12**
Yellow Dilemma (104)	**15.79**	**108**	**11.82**	**18.17**	**108**	**11.41**	**20.42**	**108**	**10.81**	**18.13**	**324**	**11.48**
Left Turn (105)	18.35	108	4.65	18.48	108	4.52	18.99	108	4.64	18.61	324	4.60
Urban Curves (106)	33.92	108	3.21	34.21	108	3.22	34.44	108	3.66	34.19	324	3.37
Interstate Curves (205)	66.31	107	4.65	66.02	108	4.96	66.56	105	4.32	66.29	320	4.64
Exit Ramp (206)	33.26	108	4.94	33.72	108	5.14	34.41	108	5.24	33.80	324	5.11
Lighted Rural (302)	55.32	108	3.00	55.13	108	3.03	54.97	108	3.50	55.14	324	3.18
Transition to Dark Rural (303)	55.00	108	3.46	55.10	108	3.36	54.22	108	4.52	54.77	324	3.82
Dark Rural (304)	52.03	108	3.21	52.58	108	3.39	52.83	107	3.62	52.48	323	3.41
Dark Rural Hairpin Curve (304)	44.86	108	4.46	45.78	108	4.09	45.50	107	4.69	45.38	323	4.42
Gravel Transition (305)	44.40	108	6.52	44.09	108	6.74	44.82	107	7.43	44.44	323	6.89
Gravel Rural (306)	43.07	108	7.06	43.20	108	7.32	43.97	106	8.14	43.41	322	7.50

Note. BAC differences shown in bold are statistically significant at $p < 0.05$.

Table 13. Speed deviation (mph) by BAC group across events.

	BAC Group											
	0.00% BAC			0.05% BAC			0.10% BAC			Total		
Event	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*
Urban Drive (102)	3.03	108	1.85	3.25	107	2.00	3.52	108	2.24	3.27	323	2.04
Green Light (103)	1.33	108	.73	1.42	107	.84	1.34	108	.67	1.36	323	.75
Yellow Dilemma (104)	4.52	108	3.88	4.21	108	4.07	4.03	108	4.38	4.25	324	4.11
Urban Curves (106)	2.75	108	.86	2.82	108	.77	2.93	108	.85	2.84	324	.83
Interstate Curves (205)	2.08	107	.94	2.08	108	1.01	2.17	105	1.11	2.11	320	1.02
Lighted Rural (302)	.53	108	.36	.51	108	.37	.60	108	.53	.55	324	.43
Transition to Dark Rural (303)	**.90**	**108**	**.72**	**.80**	**108**	**.55**	**1.04**	**108**	**.75**	**.91**	**324**	**.68**
Dark Rural (304)	4.91	108	.95	4.66	108	1.05	4.69	107	1.26	4.76	323	1.10
Dark Rural Hairpin Curve (304)	2.73	108	1.20	2.61	108	1.15	2.55	107	1.22	2.63	323	1.19
Gravel Transition (305)	3.51	108	1.80	3.45	108	1.82	3.76	107	1.95	3.57	323	1.85
Gravel Rural (306)	2.60	108	.89	2.81	108	1.22	2.81	106	1.10	2.74	322	1.08
Note. BAC differences shown in bold are statistically significant at $p < 0.05$.												

7.4 Robustness of metrics with respect of age, gender, and driver state

In contrast to the prior section where raw values were analyzed, this section will use composite score. These were chosen as the basis of the analysis because participants' performance and impairment may fluctuate across events resulting in impairment at the event level may be difficult to interpret. Composite scores for lane deviation, average speed, and speed deviation were examined to determine whether impairment was present across the entire drive. The composite scores were the *t*-scores ($M = 50$, $SD = 10$) based on the average of the *z*-scores of the measures across the events.

A 2 x 3 x 3 between-between-within ANOVA was performed on each of the three composite measures. Between-subjects independent measures were gender and age group (21-34, 38-51, 35-68). Within-subjects independent measure was (0.00%, 0.05%, and 0.10%). Because of multiple analyses, α was set to 0.01[4].

In computing the lane deviation composite score eleven participants were deleted from the analyses because of incomplete data. The analyses, therefore, were conducted with 97 subjects. The mean lane deviation composite scores by BAC, age group, and gender are shown in Appendix U. Mauchly's Test of Sphericity was not significant, indicating that no adjustment to the degrees of freedom was required.

Of the within-subjects effects, the only statistically significant effect was the main effect of BAC, $F\ (2, 182) = 34.82$, $p < 0.001$, partial $\eta^2 = 0.28$. As shown in Figure 11, lane deviation composite scores increased as a function of BAC, producing a statistically significant linear trend, $F\ (1, 91) = 60.78$, $p < 0.001$, partial $\eta^2 = 0.28$. The quadratic trend was not significant, $F\ (1, 91) = 1.49$, $p > 0.05$, partial $\eta^2 = 0.02$. No between-subjects effect or interactive effects were statistically significant.

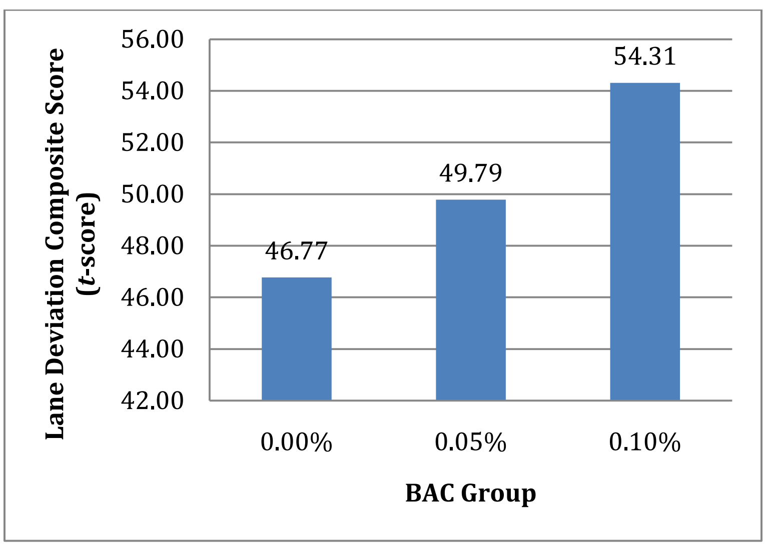

Figure 11. Lane deviation as a function of BAC group.

---

[4] Note that this was an a-priori adjustment to the alpha level to control for type I error rather than a particular statistical approach for control. Although arbitrary, we believe it is reasonable.

The average speed deviation composite scores by BAC, age group, and gender are shown Appendix U. Mauchly's Test of Sphericity was statistically significant, and the Greenhouse-Geisser adjustment was used to adjust the degrees of freedom. Of the within-subjects effects, the only statistically significant effect was the main effect of BAC, F (1.89, 192.27) = 6.59, $p < 0.01$, partial $\eta^2 = 0.06$. As shown in Figure 12, average speed composite scores decreased as a function of BAC, producing a statistically significant linear trend, F (1, 102) = 11.70, $p < 0.01$, partial $\eta^2 = 0.10$. The quadratic trend was not significant, F (1, 102) = 0.00, $p > .05$, partial $\eta^2 = 0.00$. No interactive effects were statistically significant.

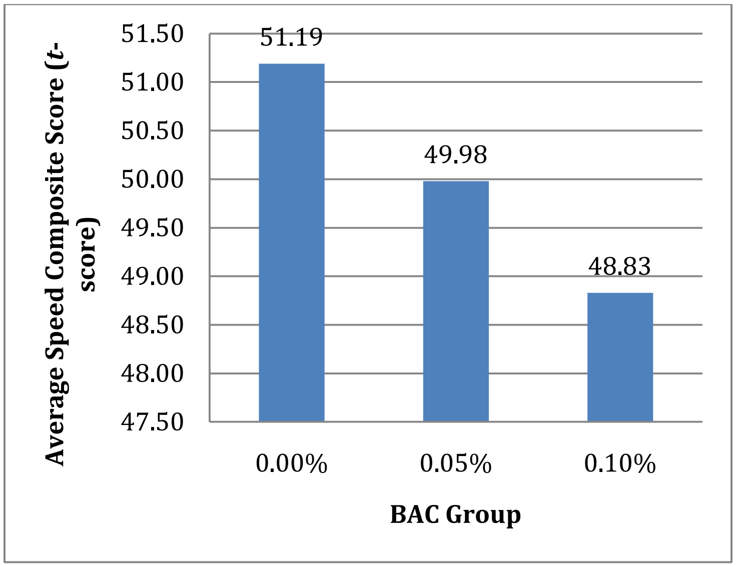

Figure 12. Average speed as a function of BAC group.

Of the within-subjects effects, the only statistically significant effect was the main effect of BAC, F (2, 102) = 8.81, $p < 0.001$, partial $\eta^2 = 0.15$. Of the between-subjects effects, the only statistically significant effect was the main effect of Gender, F (1, 102) = 8.47, $p < 0.01$, partial $\eta^2 = 0.08$. Speed deviation was greater among males (52.23) than among females (47.77). The mean speed deviation composite scores by BAC, age group, and gender are shown in Appendix U. Mauchly's Test of Sphericity was not significant, indicating that no adjustment to the degrees of freedom was required. No interactive effects were statistically significant.

The analyses were also performed with sleepiness as a covariate. There was a within-subjects interaction of BAC and sleepiness for lane deviation, F (2, 180) = 5.97, $p < 0.01$, partial $\eta^2 = 0.12$; and a within-subject interaction of BAC and sleepiness for average speed, F (1.89, 190.57) = 4.21, $p < 0.05$, partial $\eta^2 = 0.08$. Thus, sleepiness was associated with some of the effects of BAC, which is not surprising given that BAC is positively correlated with sleepiness, $r = 0.25$, $p < 0.001$.

7.5 Conclusion

Analysis of common driving metrics demonstrates the sensitivity of the drive to alcohol impairment. As expected, lane position variation was particularly sensitive and speed variation was less so. Increasing BAC levels generally affected driving performance in an orderly manner—higher BAC levels led to a linear decrease in performance. Drivers' response to the events was generally robust and unaffected by repeating the drives, with the variation between events being much greater than the variation between drives. The metrics were also robust to the effects of age and gender. No interactions affected driver performance and only the composite speed score was influenced by the main effects of age and gender. Alcohol levels did not interact with age, gender, and roadway situation, which might have otherwise undermined the association of driving metrics and alcohol impairment.

8 ALGORITHM DEVELOPMENT AND EVALUATION

The primary objectives for algorithm development and evaluation include:

- Develop algorithms to detect alcohol-related impairment based on behavioral signatures that vehicle-based sensors can measure
- Compare sensitivity, robustness, and timeliness of metrics and algorithms.

This chapter addresses these objectives by first describing the performance of a logistic regression algorithm that builds directly on an analysis of simple measures of driving performance—lane position variability, mean speed, and speed variability. To go beyond these three simple indicators of driver impairment, a decision tree algorithm fit to individual events and to the urban, freeway, and rural segments identifies behavioral signatures of alcohol impairment. These signatures provide a detailed description of alcohol impairment that supports more accurate detection that the three-variable logistic regression. The final sections of the chapter assess algorithm sensitivity, robustness, timeliness, and bias defined as:

> *Sensitivity*—The number of correctly classified cases—true positives and true negatives
> *Robustness*—Vulnerability to generalization error, context, or available data
> *Timeliness*—The amount of data aggregated over time to produce an accurate classification
> *Bias*—The tendency to favor detecting impairment at the expense of incorrectly identifying impairment when there is none.

8.1 Data and general methods of algorithm development and evaluation

The objective of the following analyses was to determine whether it is possible to distinguish between drivers with BACs at and above 0.08% and those below 0.08%. To that end, a new variable was created (BAC Status) by dichotomizing the pre- and post-drive BACs as either both being less than 0.08% or both being at or above 0.08%. The dichotomization produced 313 valid cases. Eleven cases were eliminated because the pre- and post-drive BACs were not on the same side of the 0.08% cutoff. BAC status characteristics are shown in Table 14. The median BAC for the low BAC status condition (BAC < 0.08%) was 0.037%. The median BAC for the high BAC status condition (BAC ≥ 0.08%) was 0.097%. The median differences between the conditions were 0.06%.

Table 14. Measured BAC levels associated with the two BAC classes.

	BAC < 0.08% (N = 224)			BAC ≥ 0.08% (N = 89)			Total (N = 313)		
Test Time	*M*	*SD*	Median	*M*	*SD*	Median	*M*	*SD*	Median
Pre-drive	.028	.027	.042	.101	.004	.103	.049	.041	.054
Post-drive	.023	.023	.032	.091	.006	.092	.042	.037	.043
Average	.025	.025	.037	.096	.004	.097	.045	.038	.048

Three general algorithms were developed. The first was based on logistic regression and was fit using a standard least squares regression approach using the entire dataset. The two other approaches to algorithm development used support vector machines (SVMs) and decision trees, which can often outperform linear combinations of the features (Liang, et al., 2007b). Originally developed by Vapnik (1995), SVMs have several advantages over approaches that make assumptions of linearity and normality. The SVM approach identifies a hyperplane that separates instances with different BAC levels (Saarikoski, 2008). SVMs are particularly well-suited to extract information from noisy data (Byun & Lee, 2002) and avoid overfitting by minimizing the upper bound of the generalization error (Amari & Wu, 1999). A decision tree approach, C4.5, classifies data by creating a tree that divides the data using the gini index, which weights feature influence in a linear fashion (Lim, Loh, & Shih, 2000; Quinlan, 1996). Adaptive boosting (AdaBoost) sequentially fits a series of classification algorithms, with greater emphasis on previously misclassified instances. It then combines the output of the classification algorithms by adjusting the importance of each classifier based on its error rate (Freund & Schapire, 1996). This approach is particularly valuable where a single decision tree or SVM cannot capture the complexity of the underlying relationships. Adaptive boosting was applied to both the Decision Tree and SVM, but not the logistic regression.

Three criteria are used throughout to assess algorithm sensitivity: accuracy, positive predictive performance (PPP), and area under curve (AUC). Accuracy measures the percent of cases that were correctly classified, and PPP measures the degree to which those drivers that were judged to have high BAC levels actually had high BAC levels. Performance measures such as correct detection or overall accuracy fail to provide a complete description of algorithm performance because they do not account for the baseline frequency of impairment nor differences in the decision criterion. An algorithm can correctly identify all instances of impairment simply by setting a very low decision criterion, but such an algorithm would misclassify all cases where there was no impairment. The signal detection parameter, d', avoids these problems, but its underlying assumptions include symmetry of signal and noise distributions, which are often violated. AUC is a nonparametric version of d' and represents the area under the receiver operator curve, which provides a robust performance measure that does not depend on the assumptions underlying d'. Perfect classification performance is indicated by an AUC of 1.0, and chance performance is indicated by 0.50. AUC is an unbiased measure of algorithm performance, but accuracy and PPP are more easily interpreted, so all three are used in describing the algorithms.

8.2 Logistic regression algorithm and basic driving performance

A sequential logistic regression was performed to assign each case to one of the two BAC categories (BAC < 0.08% or BAC ≥ 0.08%), first using 11 speed deviation measures, then after adding 13 average speed measures, and then after adding 10 lane deviation measures. Ten cases had missing data and were deleted, leaving 303 cases for the logistic regression.

A test of the speed deviation predictors against a constant-only model was not statistically significant, $\chi^2(11, N = 303) = 8.26$, $p > .05$. Adding 13 average speed predictors produced a statistically significant improvement, $\chi^2(13, N = 303) = 32.74$, $p < .01$. Adding the lane deviation predictors was also statistically significant, $\chi^2(10, N = 303) = 54.38$, $p < .001$. These results are summarized in Table 15. Only measures of average speed and lane deviation, therefore, were useful in predicting BAC status. Figure 13 shows the increase of overall classification accuracy as a function of the events composing the drive.

Table 15. Logistic regression for BAC status as a function of speed deviation, average speed, and lane deviation across segments.

Measures, variables, and regression parameters						95% Confidence Interval for Odds Ratio	
Measure	Variable	*B*	Wald Test	*p*	Odds Ratio	Lower	Upper
Speed Deviation	Urban Drive (102)	.254	4.460	.035	1.290	1.018	1.633
	Green Light (103)	-.323	1.305	.253	.724	.416	1.260
	Yellow Dilemma (104)	.301	8.559	.003	1.351	1.104	1.653
	Urban Curves (106)	-.167	.561	.454	.846	.546	1.311
	Interstate Curves (205)	-.195	.993	.319	.823	.561	1.207
	Lighted Rural (302)	.008	.000	.985	1.008	.455	2.229
	Transition to Dark Rural (303)	.150	.244	.622	1.162	.640	2.109
	Dark Rural (304)	-.101	.245	.620	.904	.606	1.348
	Dark Rural Hairpin Curve (304)	.008	.003	.958	1.008	.748	1.359
	Gravel Transition (305)	.123	1.231	.267	1.131	.910	1.405
	Gravel Rural (306)	-.141	.650	.420	.869	.617	1.223

Measure	Variable	*B*	Wald Test	*p*	Odds Ratio	Lower	Upper
Average Speed	Urban Drive (102)	.067	.337	.561	1.069	.853	1.339
	Green Light (103)	.075	.486	.486	1.078	.873	1.331
	Yellow Dilemma (104)	.106	7.072	.008	1.111	1.028	1.201
	Left Turn (105)	.093	2.155	.142	1.097	.969	1.242
	Urban Curves (106)	-.285	8.163	.004	.752	.618	.914
	Interstate Curves (205)	-.012	.042	.838	.988	.881	1.108
	Exit Ramp (206)	.073	2.178	.140	1.075	.977	1.184
	Lighted Rural (302)	.002	.000	.985	1.002	.791	1.270
	Transition to Dark Rural (303)	-.186	2.320	.128	.831	.654	1.055
	Dark Rural (304)	-.024	.049	.824	.976	.786	1.211
	Dark Rural Hairpin Curve (304)	-.072	1.304	.253	.930	.821	1.053
	Gravel Transition (305)	.053	1.316	.251	1.054	.963	1.154
	Gravel Rural (306)	-.009	.050	.823	.992	.920	1.069
	Urban Drive (102)	-2.196	4.190	.041	.111	.014	.911

Measure	Variable	*B*	Wald Test	*p*	Odds Ratio	Lower	Upper
Lane Deviation	Green Light (103)	.426	.234	.628	1.532	.272	8.609
	Yellow Dilemma (104)	1.899	4.670	.031	6.680	1.193	37.394
	Urban Curves (106)	3.211	6.885	.009	24.796	2.253	272.873
	Interstate Curves (205)	.758	3.828	.050	2.135	.999	4.564
	Lighted Rural (302)	-.451	.350	.554	.637	.143	2.836
	Transition to Dark Rural (303)	-.590	.808	.369	.554	.153	2.007
	Dark Rural (304)	2.638	7.582	.006	13.989	2.139	91.492
	Dark Rural Hairpin Curve (304)	-.337	.270	.603	.714	.200	2.546
	Gravel Transition (306)	1.369	3.436	.064	3.933	.924	16.734
	Constant	4.285	1.067	.302	72.613		

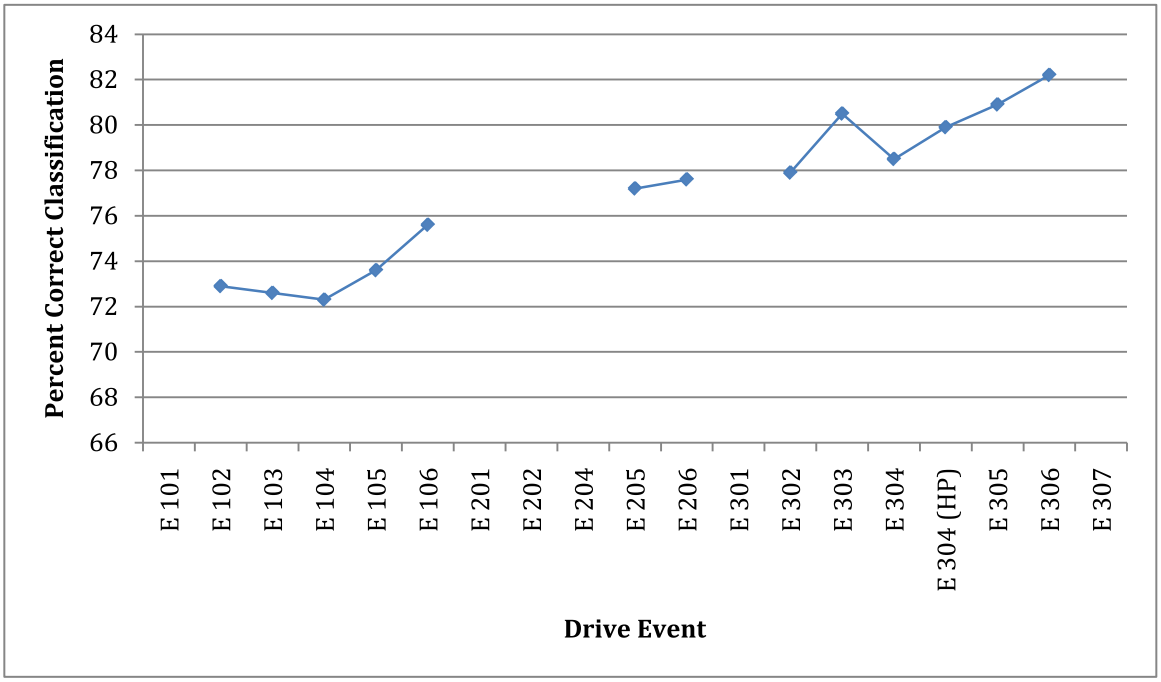

Figure 13. Correct classification for the cumulative logistic regression algorithm.

Overall classification performance was acceptable. The 11 speed deviation variables and the 13 average speed variables produced a correct classification of 95% for low BAC status and 21% for high BAC status (PPP), with an overall correct classification of 74%. The addition of the lane deviation variables resulted in a correct classification of 95% for the low BAC status and a correct classification of 49% for the high BAC status, with an overall correct classification of 82%. At the last event, the variance in BAC status accounted for by the algorithm was moderate, with Nagelkerke R^2 = .39. The AUC metric for this analysis was .720 (CI=0.073). Table 15 shows the regression coefficients, Wald statistic, odds ratios, and the 95% confidence interval for odds ratios for the 34 predictors of BAC status, based on all events.

It could be argued that logistic regression is not the correct analytic tool for these data. Logistic regression is a between-subjects strategy, which assumes the cases are unrelated to each other. In the current study, however, each subject was tested three times. The correlations of the within-subjects runs are not accounted for in the logistic regression model. To address these concerns, separate analyses were conducted, adjusting for individual differences such as age, gender, height, weight, and learning. The results of those analyses were very similar to those presented above and are not presented here. This analysis provides a baseline and point of comparison for the more complex algorithms.

8.3 Signatures of alcohol impairment

The diversity of driving situations and associated driver responses might provide behavioral signatures of impairment that are more sensitive than the simple measures of lane position variability, mean speed, and speed variability. To assess this possibility, a decision tree algorithm was fit to each event using a diverse set of variables, with the results shown in Table 16. These variables reflect the range of cues police officers use to detect impaired drivers discussed earlier and cataloged in Table 3.

The variables in Table 16 and throughout the document are abbreviated with a consistent notation. The first part of the variable indicates the measure, such as lane position (lp_), normalized lane position (lpn_), speed (sp_), and acceleration (acc_). The average, minimum, maximum, and standard deviation are shown as (_avg , _min, _max, _sd), and the initial and final values for an event are shown as _init and _end. Appendix B contains a complete list and detailed definition of each variable. Some of these variables, such as turn signal use, are relevant to only a few events, but others, such as minimum speed, are applicable to many.

Table 16 shows the variables indicative of alcohol impairment and how well a decision tree analysis can combine them to separate drivers with high and low levels of BAC. The decision tree analysis identifies a set of variables and their associated levels that best separate one condition from another. AUC values show that most events contain diagnostic information, with the exception of Merge On (202), which has an AUC of 0.50. Some events are highly sensitive to alcohol impairment, with AUCs exceeding 0.75. Dark Rural (304) is the most sensitive event with an AUC of 0.84. Similar to the other sensitive events in Table 16, this event is relatively long and places substantial demands on the driver. Dark Rural requires drivers to negotiate an unlit rural road with a sharp curve. Urban Drive (102), while very different in the details, is also quite sensitive, with an AUC of 0.78. The Urban Drive places substantial demands on the driver to scan the environment, maintain position within a relatively narrow lane, and monitor surrounding vehicles and pedestrians. This analysis shows that the behavioral signatures associated with various driving situations differ substantially in their sensitivity to alcohol. For six events, the sensitivity of the decision tree algorithm exceeded that of the logistic regression even when the logistic regression included data from entire drive.

Table 16. Decision tree algorithm applied to each event of the drive.

Segment	**Event Name (number)**	**Description (variables)**	**Approximate Duration (seconds)**	**AUC (Accuracy)**
Urban	Pull Out (101)	Pull out of parallel parking spot into traffic (gap_taken_t, turn_signal)	30	.60
	Urban Drive (102)	Drive on a narrow 2-lane road with traffic and parked vehicle (sp_avg, lpn_sd, lp_sd, sp_end, sp_min, sp_sd)	45	.78
	Green Light (103)	Navigate green traffic light on urban 2-lane road with parked vehicles along the road, oncoming traffic, traffic behind driver (sp_min, spn_avg sp_min, brake_press)	30	.66

Segment	Event Name (number)	Description (variables)	Approximate Duration (seconds)	AUC (Accuracy)
	Yellow Dilemma (104)	Navigate yellow light dilemma on urban two-lane with parked vehicle, oncoming traffic, traffic behind driver (lpn_sd, sp_max, spn_avg, spn_avg spn_avg)	75	.64
	Left Turn (105)	Left turn at signalized intersection (no green arrow, no dedicated turn lane), oncoming traffic, variety of gaps (sp_max, sp_avg, sp_min)	80	.65
	Urban Curves (106)	Three curve segments of mixed radius of curvature (lpn_sd, lp_avg, spn_sd, sp_sd)	180	.80
Freeway	Turn On Ramp (201)	Turn right onto interstate on-ramp (accel_release, sp_end, acc_avg, sp_init)	30	.61
	Merge On (202)	Merge onto interstate (acc_avg)	50	.50
	Interstate Curves (205)	Navigate three curves on interstate (lp_sd, sp_avg, sp_sd)	185	.65
	Exit Ramp (206)	Take exit ramp off interstate (acc_avg, sp_avg)	30	.65
Rural	Turn Off Ramp (301)	Turn right from ramp onto rural two-lane road (stop_pos, sp_end, acc_done_d, spe_end, acc_end, acc_done)	30	.73
	Lighted Rural (302)	Lighted two-lane rural road, 55 mph (lp_avg, spn_sd, lp_sd)	90	.68

Segment	Event Name (number)	Description (variables)	Approximate Duration (seconds)	AUC (Accuracy)
	Transition to Dark Rural (303)	Straight roadway that transitions between lighted and unlit (sp_sd, sp_avg, lp_sd, lpn_sd, sp_init)	20	.65
	Dark Rural (304)	Unlit straight and curved road, segments, center and road edge marking are faded and the road surface is grayish. A hairpin turn and a vertical curve (lp_sd, lp_sd_hp, lp_avg, sp_min, lpn_sd_hp, sp_sd)	300	.84
	Gravel Transition (305)	Transition to gravel surface on straight road (steer_sd, sp_avg, sp_sd, sp_init)	30	.73
	Gravel Rural (306)	Gravel road (straight and curves) (lp_sd, lpn_sd, sp_end, sp_init, spn_sd)	90	.76
	Driveway (307)	Pull into driveway with gravel (steer_max, sp_end, sp_init, turn_signal)	30	.63

The greater sensitivity of the decision tree algorithm relative to the logistic regression reflects, in part, different variables. Although some variables are shared with the logistic regression analysis, many are notably different. Urban Drive (102), for instance, includes the speed and lane variability measures used in the logistic regression, as well as minimum speed and standard deviation of normalized lane position). The standard deviation of normalized lane position (lpn_sd) is the variability relative to the lane center, rather than the variability relative to the mean lane position (lp_sd).

Figure 14 shows the decision trees associated with Urban Drive (102). Rounded rectangles show variables used to divide instances into high and low BAC. Numbers on the lines indicate criteria for the divisions. The rectangles represent leafs of the decision tree—the classification outcome associated with the criteria leading to the leaf. The label indicates whether the instances associated with the leaf are identified as high (TRUE) or low (FALSE) BAC levels. The width of the bar of each leaf indicates the number of instances that correspond to the conditions associated with the leaf. The clear and filled components of this bar indicate the proportion of high and low

BAC levels: Clear corresponds to low and filled to high. A perfectly accurate decision tree would have bars that are either entirely clear or entirely filled.

The trees are developed sequentially according to the ADAboost process, with the second tree being fit to those cases that the first misclassified. Because each tree represents a solution for those cases that were not fit by the others, each reveals a different behavioral signature of impairment. For Urban Drive (102), the first decision tree shows that a combination of speed and lane variability identifies alcohol impairment—those driving faster and with greater lane position variability tend to be those with high BAC levels. The second tree shows lane position as the dominant differentiating factor, followed by speed variability. The third tree uses a combination of lane position variability relative to the mean lane position, as opposed to the centerline, as in other two trees. These figures demonstrate that alcohol impairment might be detected most efficiently by a collection of variables combined in a non-linear fashion.

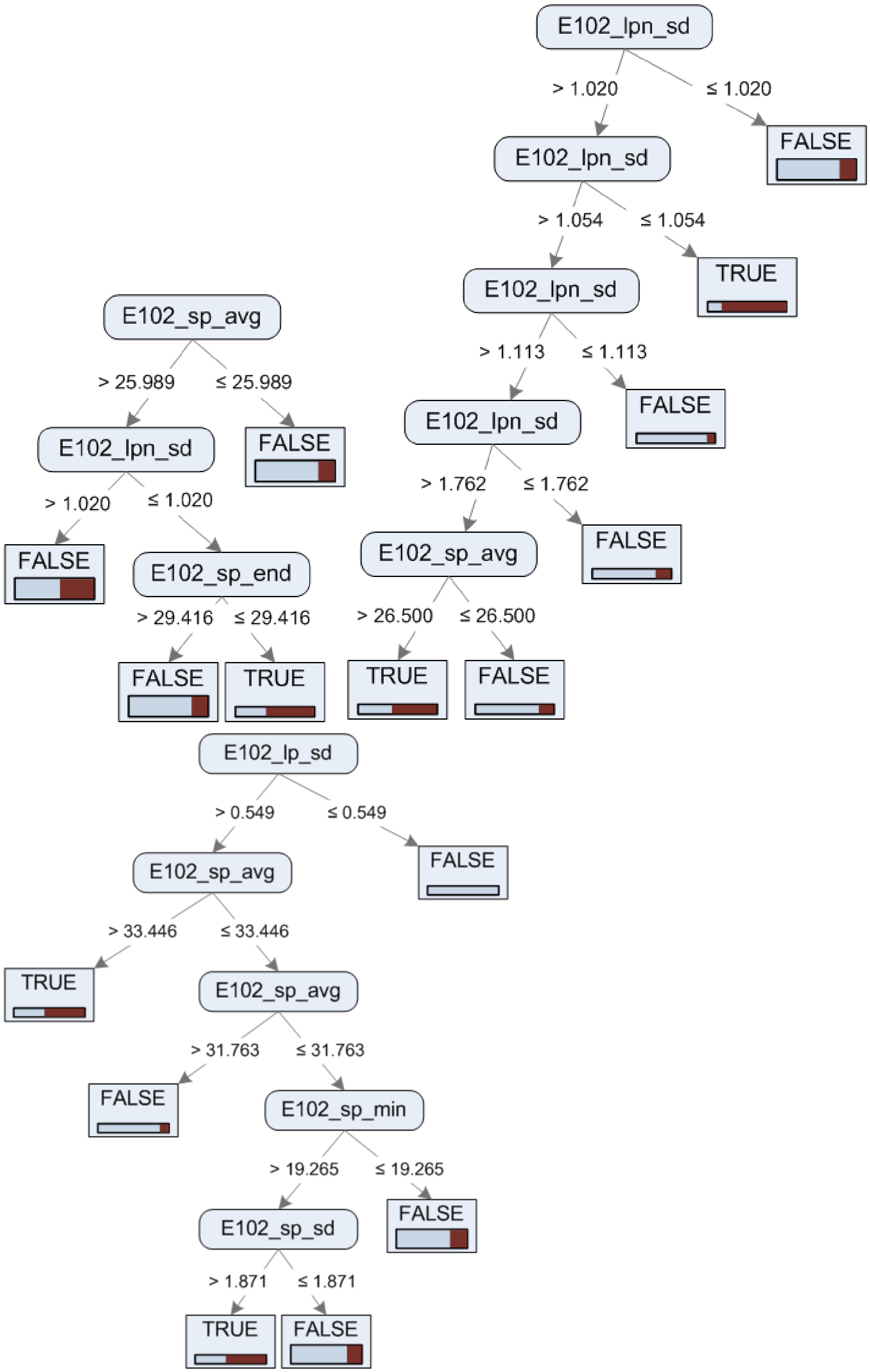

Figure 14. Decision trees for Urban Drive (102), showing the three distinct alcohol impairment signatures.

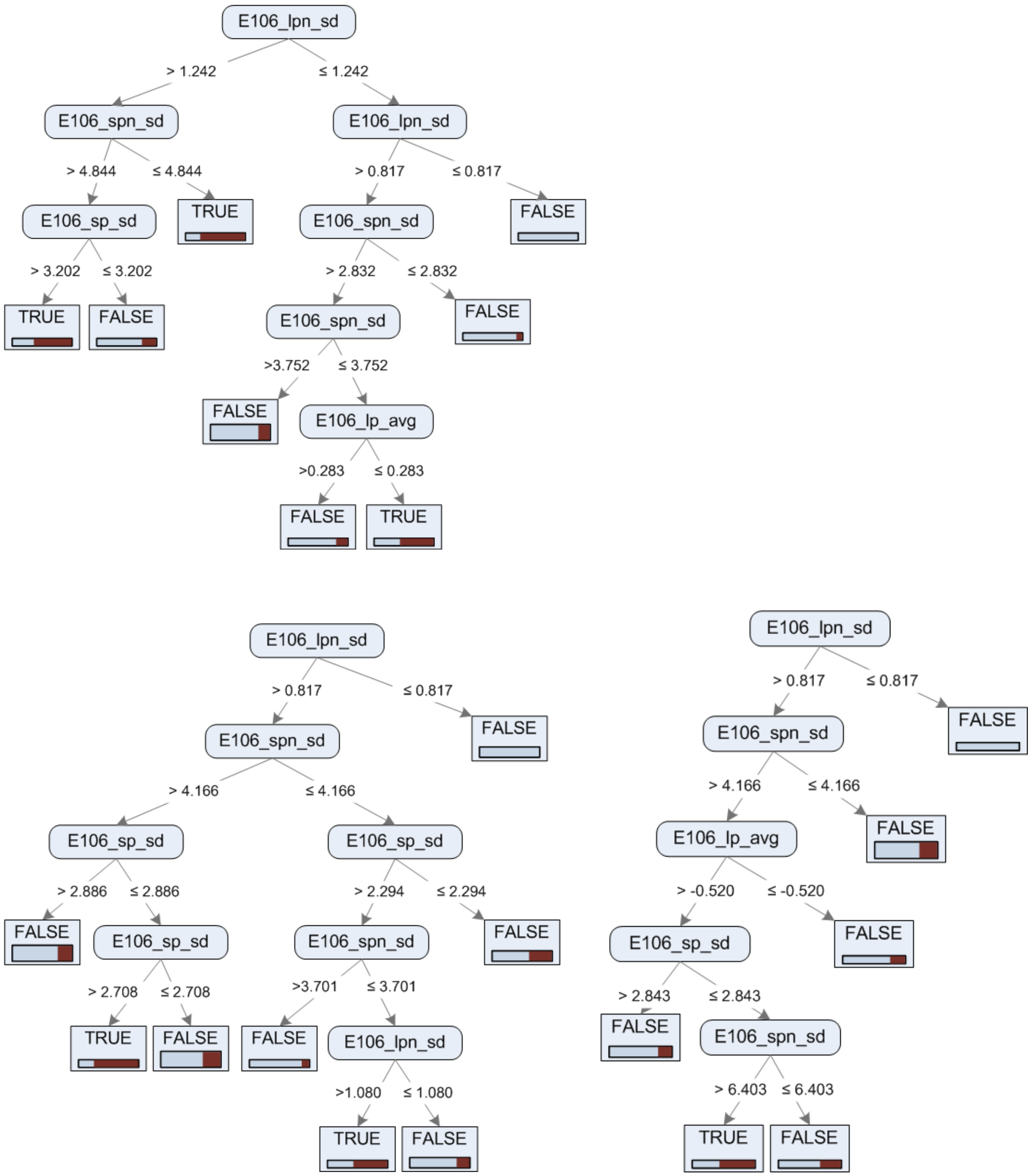

Figure 15. Decision trees for Urban Curves (106), showing the three distinct signatures of impairment.

The decision trees associated with the Urban Curves (106) in Figure 15 show signatures of alcohol impairment that share many features with those of Urban Drive (102). In both, normalized lane position variability plays a dominant role. The average lane position is also part of the behavioral signature for this event, reflecting a tendency of impaired drivers to straddle the lane in curves, but not in the dense urban environment of Urban Drive (102). Another notable difference is the role of speed variability, which was not a factor in Urban Drive (102) or the logistic regression, but emerges as in important feature in Urban Curves (106). This may reflect the diminished ability of drivers with high BAC to manage the dual tasks of lane keeping and speed control that are both demanding on the rural curves. The most striking difference between these two events is the decision tree complexity: the trees for Urban Curves (106) have 19 nodes in total, compared to 12 for Urban Drive (102). This and the variety of variables associated with the decision trees for each event support two important conclusions: alcohol has a clear and consistent effect on lane-keeping performance, and lane position variability combines with speed and other variables in a way that depends on the particular driving context.

A notable feature of both of both figures is the occasional occurrence of seeming arbitrary levels of a variable that are associated with impairment, but other levels are not—the standard deviation of lane position in the second decision tree in Urban Drive (102) as an example. Some of these reflect the complex relationship between variables, such as when impaired drivers drive more slowly, which tends to reduce their lane position variability relative to unimpaired drivers who adopt a higher speed. Other instances might represent cases where the decision trees over fit the data, so the accuracy might be lower if the decision trees were applied to data not used in their construction; this is a topic that we will return to in the analysis of algorithm robustness.

Building decision trees with data from the urban, freeway, and rural segments highlights impairment signatures that cut across several events. Figure 16 shows how lane position dominates, but that factors such as average speed combine to identify a set of unimpaired drivers, such as those that have a low standard deviation of normalized lane position on Urban Curves (106) (E106_lpn_sd) and who stopped at the yellow light dilemma (E104_sp_avg). In the second and third decision trees, the minimum speed in Urban Drive (102) (E102_sp_min) is combined with speed metrics from other events to indicate aberrant speed control in drivers whose lane position variability did not indicate high BAC.

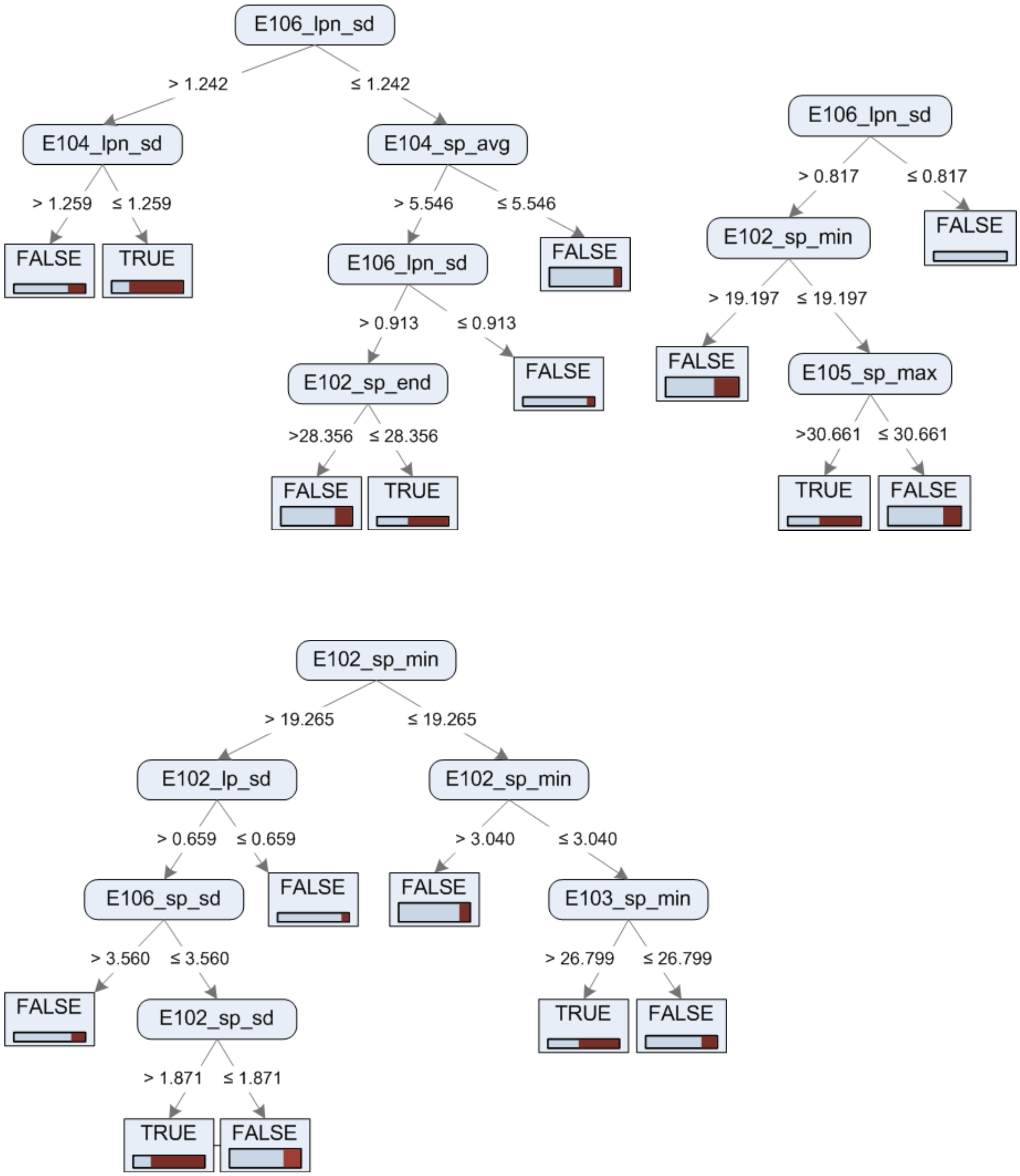

Figure 16. Decision trees from the urban segment, indicating signatures of impairment that span the Urban Drive (102), Green Light (103), Yellow Dilemma (104), Left Turn (105), and Urban Curves (106) events.

Figure 17 shows the decision trees for the freeway segment. In the first decision tree, the degree of abrupt acceleration and deceleration helps identify impaired drivers, but the standard deviation of lane position during the curves on the interstate dominate both decision trees. High BAC leads to poor lane keeping, abrupt acceleration, and slower speeds.

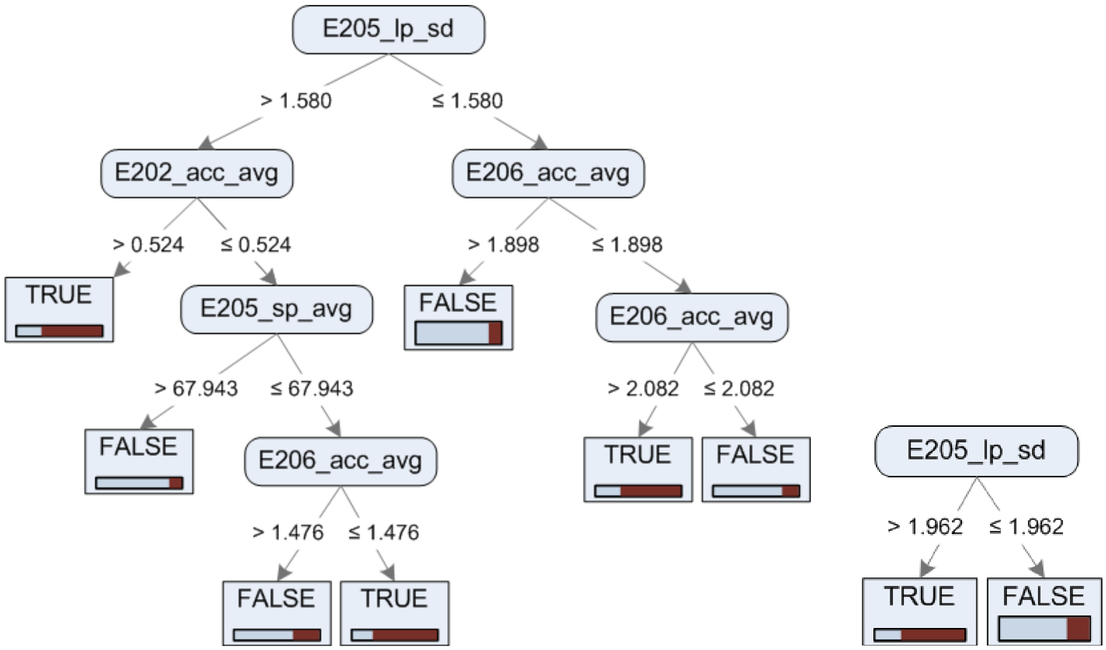

Figure 17. Decision trees from the freeway segment, indicating signatures of impairment that span the Merge On (202), Interstate Curves (205) and Exit Ramp (206) events.

Figure 18 shows the decision trees and associated signatures of impairment for the rural segment. Similar to the other segments, the first and fourth decision trees show that lane position variability strongly differentiates between impaired and unimpaired drivers. The other decision trees show that poor speed control differentiates between high and low BAC.

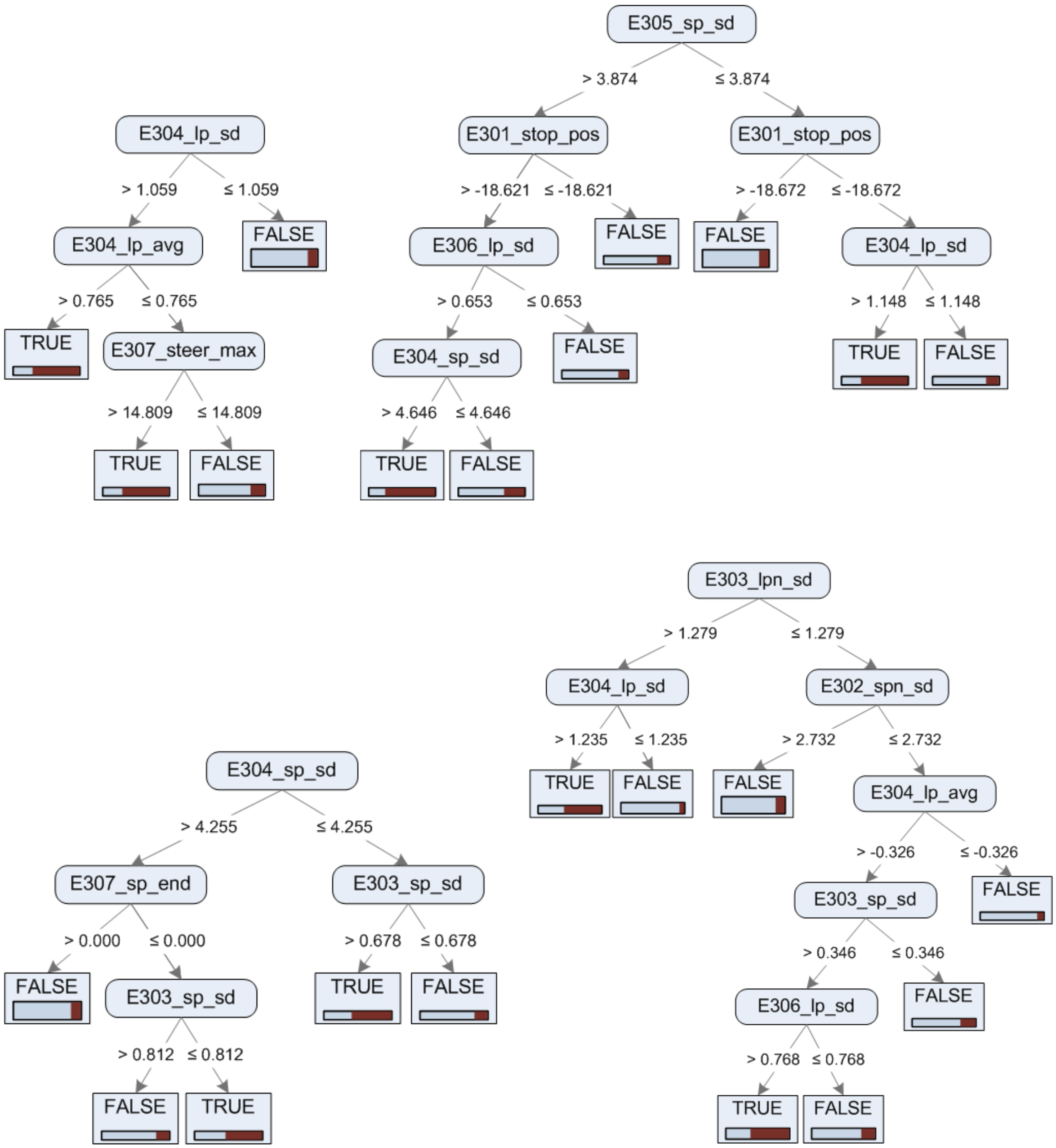

Figure 18. Decision trees from the rural segment, indicating signatures of impairment that span the Turn Off Ramp (301), Light Rural (302), Transition to Dark Rural (303), Dark Rural (304), Gravel Transition (305), Gravel Rural (306) and Driveway (307) events.

The classification performance of the decision trees that combine data over the urban, freeway, and rural segments substantially exceeds that of logistic regression, which included data from the entire drive. The logistic regression algorithm produced an AUC value of .720, compared to .824 for the decision tree algorithm applied only to the urban segment. The decision trees for the freeway and rural segments were similarly diagnostic with AUCs of .727 and .851, respectively. Positive predictive performance measures the degree to which those drivers that were judged to have high BAC levels actually had high BAC levels. The urban segment produced a PPP of

89.5%, and the freeway and rural segments produced PPP of 65.2% and 74.4%, respectively. The overall accuracy was 81.2%, 78.3%, and 77.6% for the urban, freeway, and rural segments.

The decision tree algorithm reveals important signatures of impairment not apparent in the logistic regression algorithm. Consistent with the logistic regression analysis, lane position variability emerged as a dominant indicator of impairment; however, several context-specific variables, such as maximum acceleration and minimum speed, proved to be indicative of high BAC levels, but only for specific events. These signatures of impairment strongly suggest that impairment-detection algorithms should consider the driving context.

Algorithm development ultimately aims not to identify signatures of impairment, but to identify a sensitive, robust, and timely indicator of impairment. With a mean accuracy of 79.0% for the three segments of the drive, the decision tree algorithm confirms that it is possible to create a diagnostic algorithm that is not tailored to an individual driver. The logistic regression algorithm achieved an accuracy of only 74.4% by combining information across the entire drive, achieving maximum sensitivity after approximately 25 minutes of driving. The decision tree algorithm is much more timely, identifying impairment in these situations with greater precision after only approximately 8 minutes of driving. This analysis also shows that timely impairment detection depends critically on the driving context: specific variables that reflect the demands of challenging events result in a much more timely impairment detection than generic variables.

8.4 Assessing and enhancing algorithm sensitivity, robustness, and timeliness

The following section further explores the degree to which the decision tree algorithm, by assessing how robust it is to generalization and whether other algorithms, such as one based on support vector machines, can provide a more timely and sensitive assessment of impairment. More generally, the following sections assess the degree to which decision trees and support vector machine algorithms are robust, and the degree to which they can be made more robust, timely, and sensitive. For these analyses, performance of the standardized field sobriety test provides a point of comparison.

8.4.1 Standardized field sobriety test as a baseline for algorithm sensitivity

The common application of the SFST in assessing alcohol impairment of drivers makes it a useful benchmark for assessing algorithm sensitivity—i.e., a point of comparison to see if algorithms are as good as SFST in determining impairment. Ideally, algorithms using vehicle-based sensors would exceed the capacity of the SFST to discriminate between BAC levels. This section applies three algorithms to the SFST to identify a level of accuracy that can be used to assess behavior-based algorithms later in the report.

Several studies of the SFST have demonstrated that a battery of simple visual motor tasks can discriminate between people with BAC levels above or below 0.10 with an accuracy of 83% (Burns & Moskowitz, 1977). This accuracy was based on data from five groups: Group 1 had an average BAC of 0.000% (N = 79), Group 2 had an average BAC of 0.041% (N = 20), Group 3 had an average BAC of 0.073% (N = 20), Group 4 had an average BAC of 0.120% (N = 48), and Group 5 had an average BAC of 0.156% (N = 16). The SFST battery was later reduced to a battery of three subtests: horizontal gaze nystagmus, walk and turn, and one leg stand. This reduced battery discriminated between those with a BAC of 0.10% or higher with an accuracy of 81% (Tharp Burns Moskowitz, 1981). These results were obtained in the laboratory, with relatively inexperienced testers.

More recent validation studies with more experienced testers indicate that overall accuracy is much higher in the field. In a validation study using 0.08% as the BAC criterion, officers were able to correctly identify 96% of drivers with BAC ≥ 0.08%, 93% of drivers with BAC < 0.08%, for an overall accuracy of 93% (Burns & Dioquino, 1997). These results are similar to those of a more recent study (Stuster, 2006). These results suggest the SFST can be very sensitive to BAC; however, police officers applying the SFST in the field studies were not blind to driving behavior or driver state, such as aberrant driving, drivers slurring words, and the smell of alcohol in the vehicle. Such cues may have contributed to the high classification accuracy.

This study placed drivers into three conditions (0.0%, 0.05%, and 0.10% BAC) with the aim of producing declining levels of BAC while the participants were in the simulator. The mean BACs for the three experimental conditions were 0.000% (N = 108, SD = 0.000%), 0.047% (N = 108, SD = 0.005%), and 0.093% (N = 108, SD = 0.008%), respectively. None of the drivers reached the higher BAC levels observed in the previous studies, making the discrimination task of the algorithms more difficult. Previous investigations using the SFST to classify BAC levels derived and applied decision criteria to the same sample, potentially inflating detection accuracy. A more conservative approach involves cross validation that derives decision criteria from one sample and applies them to another sample.

To identify SFST performance as a baseline for comparing the algorithm, two situations were considered: the first discriminating between the experimental conditions of 0.00% BAC and 0.10% BAC and the second between drivers above and below 0.08% BAC. Based on previous studies, one might expect greater accuracy in discriminating between people with a large difference in BAC. This analysis is based on the SFST administered immediately after the drive and scored according to the protocol in Appendix Q. Analysis of the SFST considers three techniques for discriminating between BAC levels: logistic regression, decision tree, and support vector machine.

Classification accuracy was consistent with previous studies—classification accuracy exceeded 82% for all three algorithms, with the decision tree being most accurate (84.7), followed by SVM (82.3) and logistic regression (82.0). Not surprisingly, the performance discriminating between BAC levels above and below 0.08 was somewhat worse than between the more extreme range defined by the experimental conditions of 0.00% and 0.10% BAC. Table 17 shows the accuracy ranges from approximately 80.5 to 82.5%. The failure to perfectly discriminate between BACs is not surprising given the overlap in SFST scores across BAC levels, as shown in Figure 19.

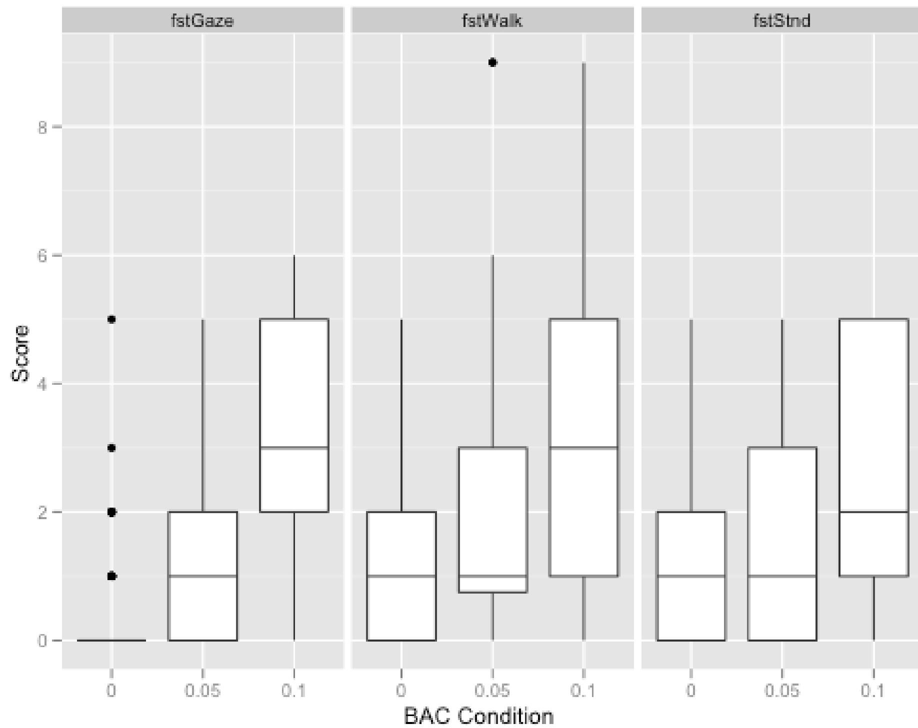

Figure 19. SFST scores show considerable overlap across BAC conditions.

Table 17. Performance of three algorithms classifying drivers with BAC above and below 0.08% using the SFST, with confidence intervals in the parentheses.

	Accuracy	**AUC**	**PPP**
Decision tree	81.8 (5.9)	.76 (0.087)	78.4 (15.5)
SVM	80.5 (6.9)	.81 (0.072)	75.6 (17.9)
Logistic regression	82.5 (5.5)	.80 (0.062)	75.9 (13.6)

Algorithms for combining elements of the SFST correctly predicted BAC levels with similar accuracy as that found in previous laboratory studies, but with lower accuracy than recent field studies. Reasons for the lower accuracy include less experienced administrators and a cross-validation statistical analysis that produces accuracy estimates that may be more representative predictions using data that are not included in estimating algorithm parameters. Another factor that contributes to algorithm performance is the BAC level being discriminated. The highest average BAC in the current study was 0.108%. This was dictated by practical and ethical reasons. The range of BACs in the current study, therefore, is narrower than any previous SFST studies, either in the laboratory or in the field. The importance of BAC range is demonstrated by the four to five percent greater accuracy when discriminating between placebo and the 0.10% BAC condition, compared to discriminating between those above and below 0.08% BAC. In addition, police officers in the recent field studies likely benefitted from a range of cues beyond those in the SFST, which were not available to those performing the SFST in this study (Rubenzer & Stevenson, 2010).

8.4.2 Algorithm sensitivity and robustness

Robustness is the degree to which algorithm performance depends on factors unrelated to driver impairment, such as road type and individual differences. This analysis considers three elements of robustness: generalization error, dependence on individual differences, and the effect of different road types.

Generalization error is the degree to which algorithm sensitivity declines when it is applied to data that were not included in its development. An algorithm that identifies impairment based on the idiosyncratic behaviors of particular drivers in a particular driving context will perform well when applied to the data used to develop the algorithm, but poorly when applied to data not used in its development. A large generalization error indicates over fitting and poor robustness.

Sensitivity of an algorithm assessed with cross validation was compared to sensitivity of an algorithm fit to the entire data set to estimate generalization error. In cross validation, one data set is withheld and the algorithm is trained on the remaining data set and then tested on the withheld data. This study used a 10-fold cross validation, in which 10 datasets were created, withholding for testing a stratified sample of 10% of the data. This produces 10 estimates of algorithm performance, which are averaged to assess the algorithm performance. Results are not typically sensitive to the number of folds in the cross validation process, but a 10-fold is most commonly used (Efron & Gong, 1983; Feng, et al., 2008).

Table 18 shows two aspects of algorithm robustness: generalization error and sensitivity to the driving context. Cross validation reveals substantial generalization error within all three roadway segments and relatively robust performance between the three segments. Accuracy was approximately 10% lower for all road segments with cross validation. Likewise AUC and PPP were substantially lower, with AUC dropping from by approximately 0.15. In contrast, the algorithm was quite robust to the differences between segments. AUC for the cross validation ranged from 0.63 to 0.65 across segments, almost an order of magnitude less than the generalization error. This difference reflects the tailoring of the algorithm to the event—signatures of alcohol impairment were derived for measures from each event rather than a single criterion applied to all events.

Table 18. Cross validation to assess robustness associated generalization error and driving context.

	Accuracy	AUC	PPP
Urban segment			
Decision tree fit to all data	81.2	0.82	89.5
Decision tree cross validation	70.3 (3.9)	0.65 (0.061)	50.2 (18.7)
Freeway			
Decision tree fit to all data	78.3	0.73	65.2%
Decision tree cross validation	68.7 (5.5)	0.63 (0.091)	43.1 (24.9)
Rural			
Decision tree fit to all data	77.6	0.85	74.4%
Decision tree cross validation	73.2 (6.5)	0.65 (0.11)	55.2 (25.1)
Field sobriety test cross validation	81.8 (5.9)	0.76 (0.087)	78.4 (15.5)

95% confidence interval in parentheses

Individualization describes how the algorithms are matched to individual drivers. A highly robust algorithm would perform well for everyone without the need to tune the algorithm to each person or class of people. Applying a z-transform to the raw data is one way to individualize the algorithms. A z-score was calculated for each measure by subtracting the mean and dividing by the standard deviation using the data for each variable, BAC level, and driver combination. Figure 20 shows the effect of this transform for three measures—the standard deviation of lane position for Urban Curves (106), Interstate Curves (205), and Gravel Rural (306) events. The left panels show raw measures and the right panel shows z-transformed variables. The benefit of this form of individualization is revealed by the greater separation of the z-transformed variables across BAC levels.

Figure 21 shows the how well the z-transformed variables separate the experimental conditions—drivers with high BAC tend to cluster in the upper right of the graph with high lane position variability for both events. Surprisingly, the lane position variability in one event is not strongly correlated with that in the other event. The bottom panel shows instances of high and low BAC as blue and red dots, overlaid by a color-coded SVM value function that classifies high and low BAC levels. The right SVM image shows the ADAboost solution addressing those cases that were not classified properly by the first. The z-transformed data from events 106 and 304 effectively identify impairment. This algorithm uses just lane position variability from two events, over total span of only 480 seconds, and classifies BAC level as effectively as the SFST, with accuracy of 80.3%, an AUC of 0.75, and a PPP of 75.0%.

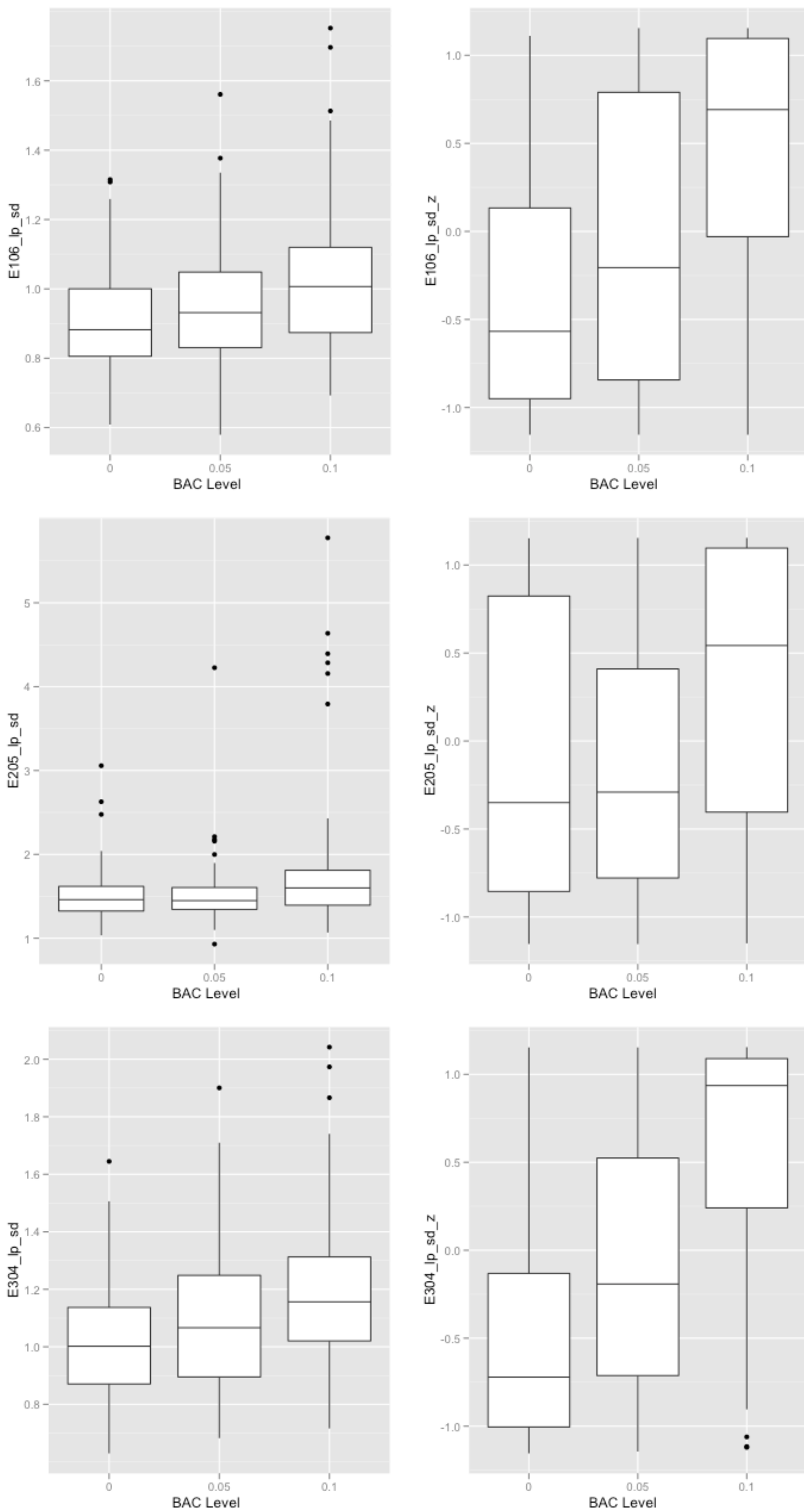

Figure 20. The z-transform of the standard deviation of lane position individualizes the model input and separates BAC levels more clearly.

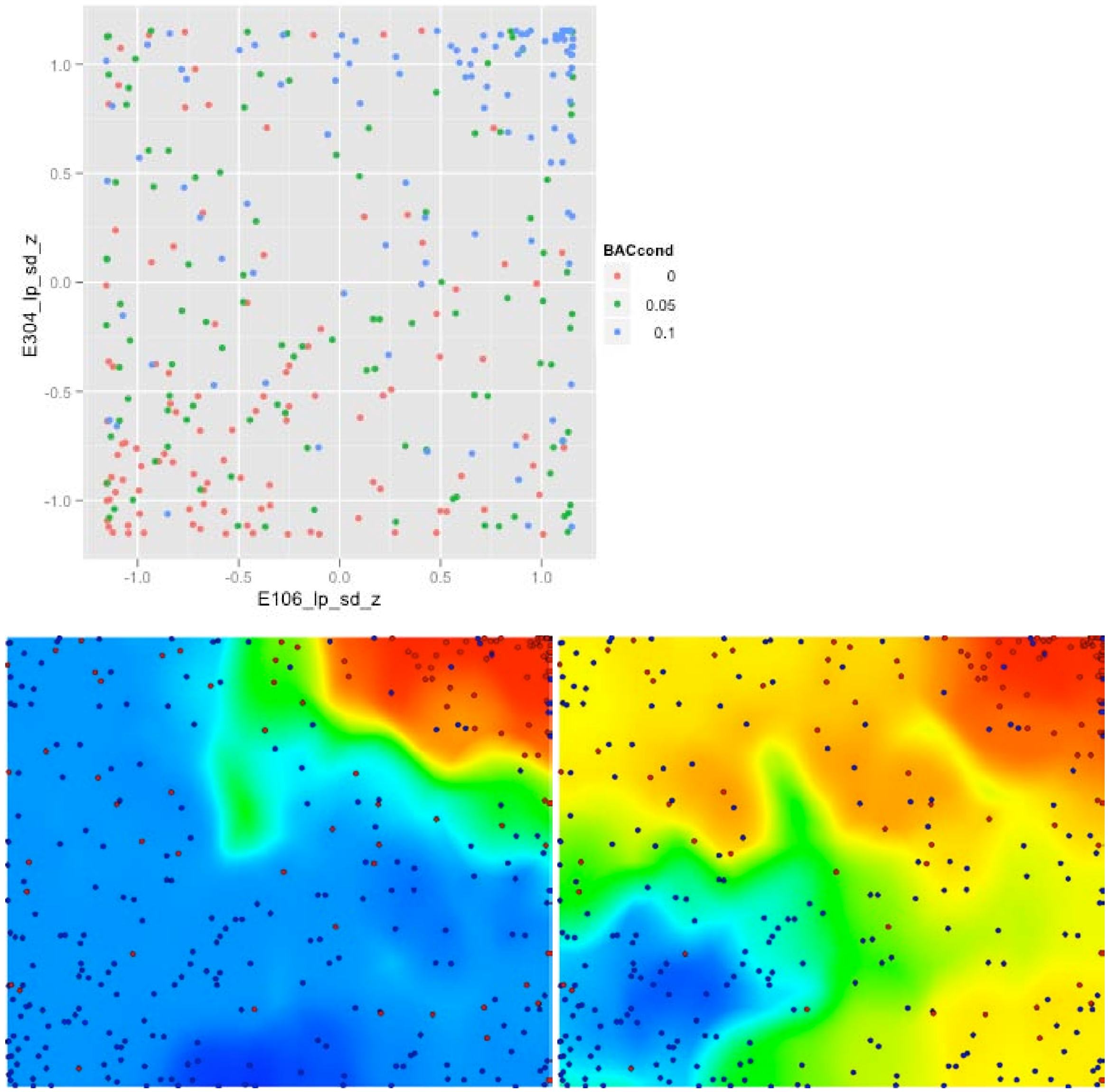

Figure 21. The joint distribution of the z-transform and the corresponding values for the support vector machine classification algorithm.

Applying the z-transform to other driving metrics provides a similar benefit to that seen with the standard deviation of lane position. Including z-transformed variables in the algorithms increased their sensitivity for all segments of the drive.

Table 19. Cross validation to assess robustness defined by generalization error.

	Accuracy	AUC	PPP
Urban			
Decision tree	78.6 (5.6)	0.79 (0.10)	68.3 (16.9)
SVM	78.7 (7.4)	0.82 (0.10)	64.5 (25.0)
Freeway			
Decision tree	72.2 (5.2)	0.71 (0.06)	52.2 (25.4)
SVM	71.6 (1.9)	0.68 (0.15)	NA
Rural			
Decision tree	77.6 (3.6)	0.81 (0.048)	68.3 (14.0)
SVM	77.4 (5.4)	0.82 (0.074)	62.8 (11.4)
Field sobriety test cross validation	81.8 (5.9)	0.76 (0.087)	78.4 (15.5)

95% confidence interval in parentheses

Another means of increasing sensitivity is to add additional variables. Eye data are a particularly promising source of information because such data reflect both executive function and attention management, as well as alcohol-induced drowsiness. Variables addressing gaze concentration, such as percent on road center and horizontal standard deviation of gaze might indicate diminished executive control and Perclos and blink duration, might reflect drowsiness. Eye movement failed to enhance the sensitivity of the algorithms in the freeway and rural segments, but the data suggests potential improvement for the urban segment, although this improvement did not achieve statistical significant, as indicated by overlapping confidence intervals. The perclos and blink duration in Event 103 helped produce an overall accuracy of 81.2 (CI=5.3) and an AUC of 0.84 (CI=0.074).

As a point of comparison, the performance of the decision tree for the urban segment, without cross validation, was very high: accuracy (85.9%), AUC (0.90), and PPP (79.2%). The difference between this performance and that in Table 18 reflects a similar generalization error as seen with the algorithms that were not individualized. Individualized algorithms are no more robust to generalization error than generic algorithms. Compared to the generic algorithm, the individualized algorithms are less robust to the types of roadway. Table 19 shows that sensitivity, measured by AUC, ranges from 0.71 to 0.81; this compares to a much narrower range of 0.63 to 0.65, in Table 18. One explanation for why individualized algorithms are more vulnerable to differences in road types is that the differences between urban and rural segments provide more opportunities for idiosyncratic behavior and driver-specific strategies. The SVM shows similar sensitivity and robustness to the decision tree, both of which performed worst in the freeway segment.

8.4.3 Algorithm robustness, driver characteristics, and alcohol levels

Another approach to robustness considers the degree to which irrelevant variables influence impairment detection. Developing a logistic regression algorithm that detects impairment by combining decision tree output for each segment assessed this possibility. Because this algorithm combines data from all three segments it is very sensitive, detecting impairment with an AUC of .96. With a perfectly robust algorithm, the misclassified instances would not be systematically related to irrelevant factors such as driver age, experience, or weight. Examining the residuals of a logistic regression can assess whether the residuals depend on various conditions. Residuals are the difference between the predicted and actual BAC condition, and the degree various factors affect algorithm performance is an indicator of algorithm robustness.

Of particular interest is the influence of age and driving experience. The first analysis considered age as a continuous variable and no difference in the distribution of residuals across age, $F(1,310)=0.435$, $p=.509$. A similar pattern is seen for weight, $F(1, 310)=1.536$, $p=.216$. A second analysis addressed the possible benefit of greater driving experience associated with age. For this comparison, drivers were divided into two groups: those below 25 years of age and those at or above 25. No statistically significant difference emerged, $F(1, 311)=.14$, $p=.707$.

Another factor that might be expected to confound alcohol detection is drowsiness. Although substantial care was taken to minimize conflicts with drivers' typical sleep schedules, some drivers drove while fatigued. Drowsiness, as measured by the Stanford Sleepiness Scale, was not strongly associated with the residuals when measured before the drive, $F(1, 311)=.955$, $p=.329$, or when measured after the drive, $F(1, 309)=1.457$, $p=.228$. Similar analyses were performed for other potentially confounding factors, such as drive number and scenario number with similar results.

Several variables were strongly related to the residuals: the gaze nystagmus score from the SFST, $F(1, 310)=65.09$, $p<.0001$, and the pre-, $F(1, 311)=121.60$, $p<.0001$, and post-drive, $F(1, 311)=131.22$, $p<.0001$, BAC levels. Figure 22 shows that the algorithm underestimates BAC levels associated with the 0.05% BAC condition and overestimates the level in the 0.10% BAC condition. This simply reflects the criterion used in training the algorithm, where the algorithm differentiated between those above and below 0.08% BAC and so classed those at 0.05% BAC as unimpaired. More interesting is the relationship between the gaze nystagmus and the residuals. This relatively strong relationship suggests that gaze nystagmus is sensitive to indicators of alcohol-related impairment that the algorithm does not capture. This has important practical consequences because combining the vehicle-based estimate of alcohol impairment with gaze nystagmus might produce a much more effective indicator of impairment than either alone. Overall, the analysis of residuals shows that the algorithm is robust to the potentially confounding effects of age, experience, and drowsiness, and is related to BAC levels as expected given the impairment criterion chosen for algorithm development.

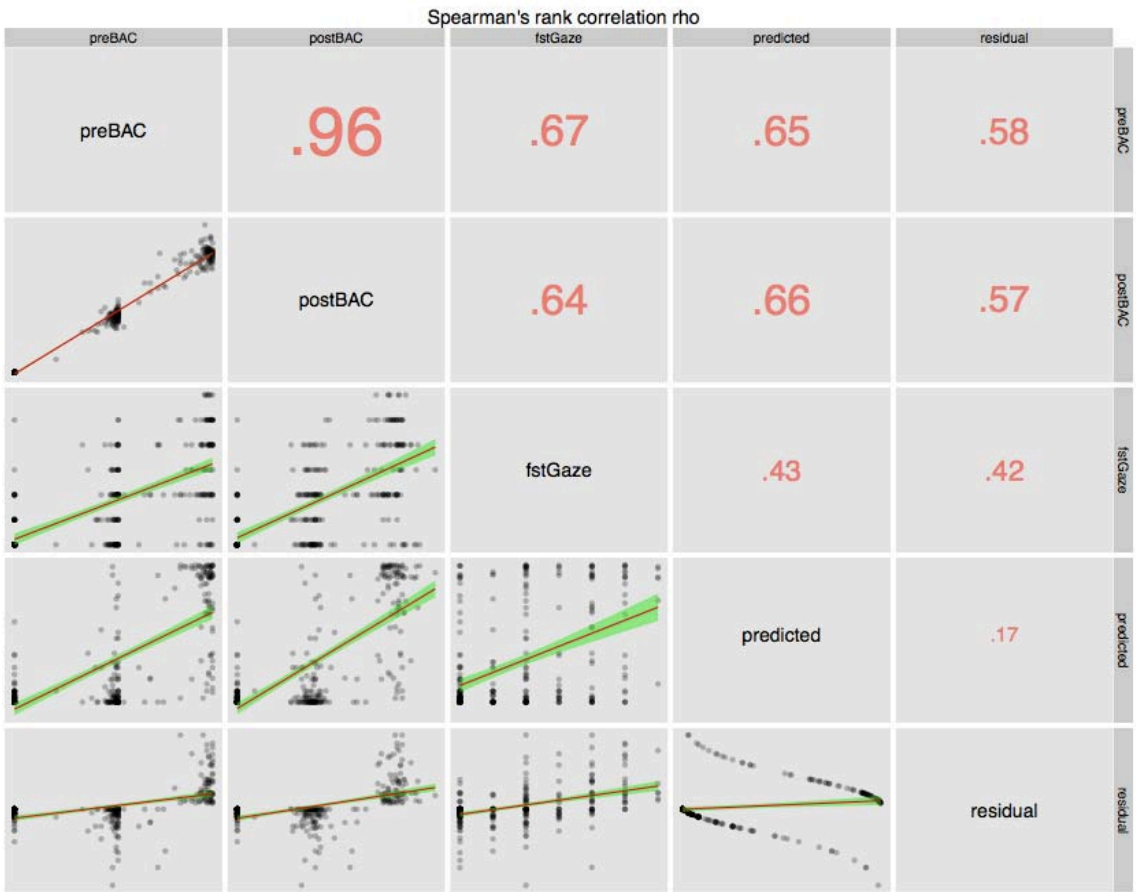

Figure 22. The relationship between algorithm predictions, residuals and related variables.

8.4.4 Algorithm sensitivity and timeliness

Timeliness refers to how quickly algorithms accumulate sufficient information to judge impairment precisely. Three fundamental considerations govern timeliness: rate of change of impairment, change in sensitivity with accumulation of evidence, and time course of impairment signature. The relationship between the rate of change of the system state relative to the time needed to identify the state is specified by the Nyquist interval, which requires a sampling rate twice that of the bandwidth of the system. Alcohol impairment changes over a time course of hours rather than seconds, so accurate state estimation can be achieved with an algorithm that estimates impairment at a time scale of 20-40 minutes. This contrasts with impairment associated with distraction, which can change much more rapidly and would require an algorithm that detects changes on a time scale of several seconds. Given the long time-constant associated with BAC, this analysis focuses on how much information must be accumulated to provide a sensitive indicator of impairment and the time scale of behavioral signatures needed for real-time interventions.

Sampling theory predicts that measurement uncertainty diminishes with the square root of the number of independent samples, suggesting that algorithm sensitivity will increase with the accumulation of information, but at a diminishing rate. The data from Table 16 plotted in Figure

23 show a general trend toward increasing sensitivity with longer events, but also indicate that longer events provide an increasing benefit. This figure also shows the substantial differences between events, with Urban Drive (102) and Urban Curves (106) being more sensitive than their duration would suggest, contrasting with Interstate Curves (205), which is less sensitive. As noted previously, highly precise impairment detection can occur in eight minutes if the driver encounters situations similar to Urban Curves (106) followed by Dark Rural (304). These results show that timely impairment detection depends on the types of events encountered by the driver, as well as the duration of information accumulation.

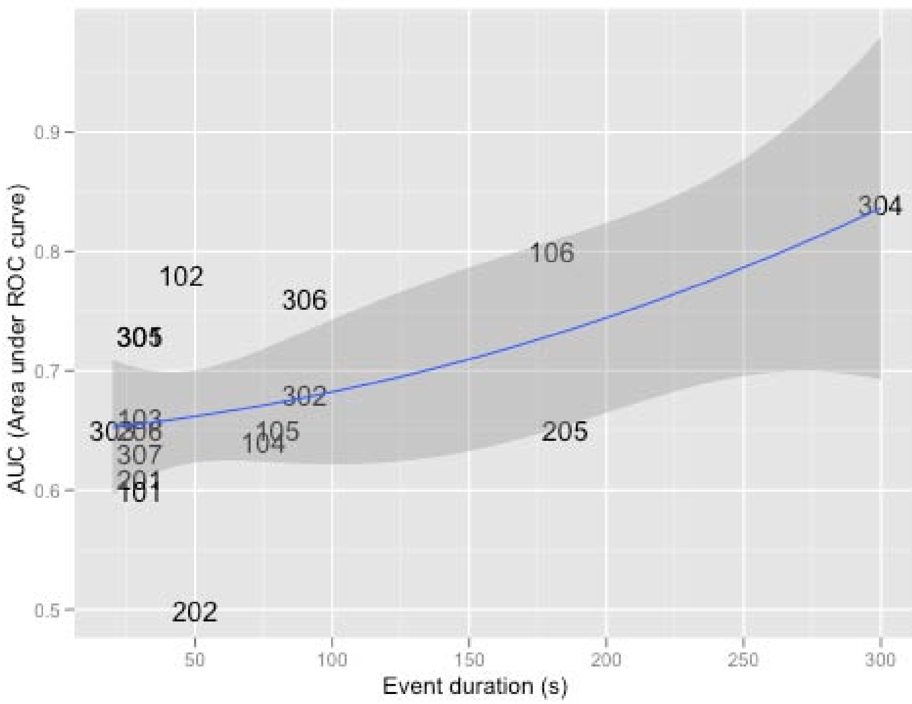

Figure 23. The sensitivity of each event as a function of its duration.

The most sensitive indicators of impairment involve continuous measures cumulated over time, such as the standard deviation of lane position. In addition, important signatures of alcohol impairment are defined by behavior that evolves over a relatively long time horizon, requiring samples of driving behavior that extend over 30 seconds to several minutes—weaving cannot be captured in 10 seconds of data. Even with such constraints, reasonable sensitivity was obtained over approximately three to nine minutes.

8.4.5 Bias and combining algorithms to minimize false alarms

Both robustness and timeliness concern factors affecting algorithm sensitivity and its overall ability to detect impaired drivers *and* avoid labeling unimpaired drivers as impaired. In contrast, bias concerns the tendency to err on the side of detecting impaired drivers *or* avoiding labeling unimpaired drivers as impaired. Generally the algorithms were biased toward correctly identifying unimpaired drivers. PPP, the probability that a driver had a high BAC when labeled as such by the algorithm, ranged between 63% and 68%. This contrasts with approximately 80% for negative predictive performance. Negative predictive performance (NPP) is the probability that a driver had a low BAC when labeled as such by the algorithm. Ideally an algorithm would have high PPP and NPP. One reason for this outcome is the unequal distribution of the data, where 71.5% of the training data were unimpaired cases and the focus was on maximizing the AUC rather than PPP. A cost-sensitive classification process could reverse this tendency.

The fundamental differences between decision trees and SVMs also suggest that they might have complementary strengths. Figure 24 shows how these complementary strengths can be leveraged to minimize false detection of impairment and maximize detection of impaired drivers. The vertical axis represents the probability of detecting an impaired driver, and the horizontal axis represents the chance of a false detection. A prefect algorithm would generate a curve that follows the upper left edge of the plot and an algorithm with no power to discriminate would be a diagonal line from the lower left to the upper right. The area under this curve is AUC. The curves represent the individualized algorithms, and the band around them represents the 95% confidence interval. This band shows that they are largely overlapping, but the blue line, representing the SVM, exceeds the red line of the decision tree when the decision threshold is low. This tendency is also reflected in Table 19, where SVMs produce higher PPP, but the confidence intervals overlap, so it does not represent a statistically significant difference. These data provide tentative evidence that a decision tree might be best when false alarms are of most concern and an SVM when misses are of most concern.

The bias of one algorithm relative to another makes it superior only in the context of the intervention it supports. An algorithm that favors hits over false alarms might be better for certain applications, such as post-drive feedback, but a poor choice for others. Bias describes an important trade-off between algorithms.

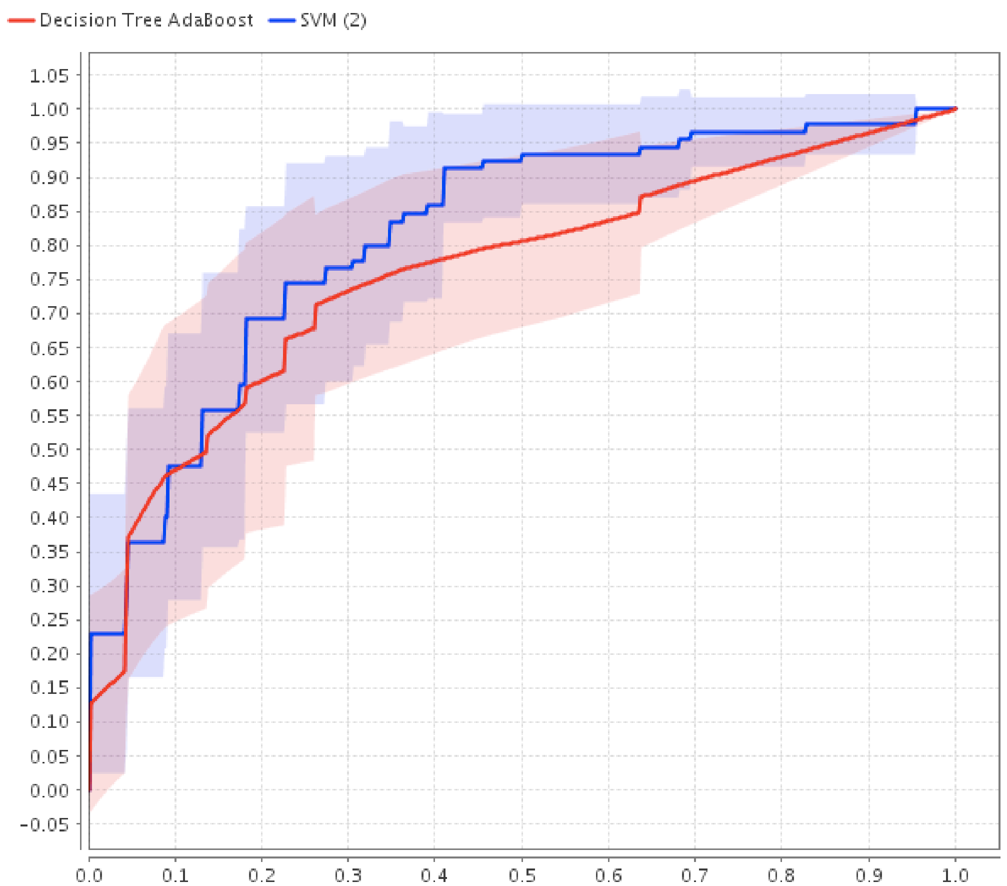

Figure 24. ROC plot for the rural segment showing that the SVM performs better than the decision tree if the decision criterion is low, favoring hits over misses.

8.5 Conclusion

The ultimate aim of this analysis was to assess the feasibility of vehicle-based sensors to detect alcohol-related impairment in real time. The results demonstrate the feasibility of such a system: the sensitivity of an individualized algorithm is comparable to that of the SFST. The behavioral signatures of alcohol impairment that form the basis of this algorithm are consistent with previous studies: diminished lateral and longitudinal control, particularly the standard deviation of lane position. The decision trees fit to individual events reveal a more complex behavioral signature of alcohol: the particular variables depend on the event, the sensitivity depends on the event, and the combination of variables is complex and depends on differences between people and events. Reflecting the complexity of the behavioral signatures, decision tree and SVM algorithms detect alcohol impairment more accurately than a traditional logistic regression using generic measures of average speed, speed variability, and lane position variability.

Analysis of algorithm sensitivity, robustness, and timeliness showed that algorithm sensitivity depends on differences between events, road segments, and drivers. The algorithms are not robust to these factors, which suggests future development must consider how to mitigate and capitalize on these effects. For example, individualizing the algorithms, even with a very simple z-transform of the raw data, substantially increases sensitivity. Similarly, using measures that best reflect alcohol impairment for each event substantially increases sensitivity. Sensitivity depends on the event and its duration. Demanding events tend to be more sensitive, as do longer events. The implication is that the amount of time an algorithm requires to accurately detect impairment depends on the sequence of roadway situations encountered by the driver. The situations included in this study show that an accuracy of 78%, comparable to that of the SFST, can be achieved in approximately eight minutes.

The ultimate aim of impairment-detection algorithms is to support interventions that guide the driver to safer behavior. The desirability and feasibility of any particular algorithm depends on how it meets the particular needs of an intervention. The ideal algorithm would be sensitive, robust, and timely. This study demonstrates that algorithms differ substantially on these dimensions and that design must consider the inevitable tradeoffs. Most importantly, algorithms become more sensitive, but less timely, as measures are integrated over time. The ultimate feasibility of impairment-detection algorithms depends on matching the performance profile of the algorithm to the nature of the intervention.

9 CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

This study demonstrated that a vehicle-based system using measures of driver behavior can differentiate between drivers with BAC levels above and below 0.08% with a sensitivity similar to the SFST. Because the indicators of alcohol impairment become much stronger at higher levels, the sensitivity would likely increase substantially if the algorithm was used to identify those with BAC levels over 0.150%. These outcomes strongly support the potential of vehicle-based systems to prevent and mitigate alcohol-related crashes.

Table 20 shows the potential of a vehicle-based indicator of impairment, particularly one that is sensitive to high BAC. Lund (2006) used the relative risk curves calculated by Zador, Krawchuk, and Voas (2000) based on the FARS and the 1996 National Roadside Survey (Voas, Wells, Lestina, Williams, & Greene, 1997) and applied those relative risk values to an estimation of the savings in driver fatalities if new vehicle "technologies" succeeded in preventing drivers at certain BACs from driving. Table 20 shows that 66% of alcohol-related fatalities occur with BAC levels above 0.150%, suggesting that the greatest value of a vehicle-based countermeasure lies in detecting high BAC levels, where algorithms are likely to be very sensitive. The exact percent of crashes associated with a given BAC level varies from year to year; therefore, a precise estimate of benefits is difficult to assess. For example, the data from NHTSA suggests a rate of 57% when comparing the number of drivers in fatal crashes with a 0.15 BAC or greater out of all drinking drivers (.01 BAC or greater). This table does not imply that the algorithm in the report would prevent all or even a majority of the crashes associated with high BAC levels, only that most fatalities occur at very high BAC levels where the algorithm is likely to be most sensitive.

Table 20. Potential lives saved in 2004 if driver BACs had been limited to <0.08% [adapted from Lund (2006)].

Driver %BAC	Fatalities	Estimated reduction
.150+	8,629	6,540
.100 – .149	3,430	1,143
.080 – .099	1,083	203
ALL	13,142	7,886

The relative contribution of alcohol to degraded performance versus risky behavior has important implications for detecting alcohol impairment with simulator data. People tend to behave better under observation, such as attempting to adhere to speed limits, than they might when they are not being monitored (Evans, 1991; Rowe, et al., 2006). As a consequence, data collected in the simulator may only capture the effects of alcohol on motivated performance and not the effects on behavior or typical driving performance. The result is that simulator data may not capture the full range of responses that an algorithm could use to identify alcohol impairment. Given that simulator data may reflect alcohol-related changes in performance and not behavior, there may be sensitive indicators of alcohol impairment, such as speeding, that might not be revealed in the simulator. Naturalistic studies of driving, such as the Strategic Highway Research Program 2 (SHRP2), could complement the data collected in driving simulators and provide a valuable platform for assessing impairment-detection algorithms.

On the basis of this research, standard deviation of lane position and average speed were shown to be reliable measures of impairment that can be feasibly captured over a number of driving situations, and appear robust enough to be useful in future vehicle-based countermeasures. Minimum speed, as well as standard deviation of lane position and speed, are useful indicators that might have particular utility in alcohol warning monitors designed to provide feedback to drivers.

The results of this study have implications for future research. One general finding is that data mining techniques, such as decision trees and support vector machines, can extract information from driving performance data that might otherwise be lost. Such information supports more timely and sensitive algorithms. Further exploration of such techniques with different types of impairment has great promise.

A second general finding is that the driving context strongly influences impairment-detection performance. Contrary to many previous simulator studies of alcohol-impaired driving, this study used a representative series of 19 events over three types of roadway situations. These events revealed that impairment detection depends on the type of event. Because driving is a statisficing rather than optimizing activity, drivers can take many paths through low-demand situations that are all satisfactory. This variety of satisfactory responses masks impairment. The variety of events also requires a greater variety of measures to capture the relevant behavior in each event. All of these findings imply that detecting alcohol-related impairment, and impairment detection more generally, depends on the driving situation. Algorithm development needs to consider roadway situations as much as it needs to consider the drivers' perceptual, motor, and decision-making response to the impairment.

A third general finding is that individual differences strongly influence driving performance indicators and the sensitivity of the algorithm. The analysis of the algorithms confirmed that it is possible to create a diagnostic algorithm that is not tailored to an individual driver. Under the conditions tested, that generalized algorithm was timely, robust, and nearly as accurate as the individualized algorithm. However, tailoring algorithms to individuals has the potential to substantially enhance algorithm sensitivity. An important consideration regarding the degree to which algorithms should consider individual differences concerns the role of algorithms in promoting safer driving. If the ultimate aim of vehicle-based systems is to identify and mitigate dangerous driving behavior, tailoring the algorithm to the individual might lead the algorithm to neglect drivers with consistently unsafe behavior. This suggests tailoring algorithms to individuals' needs to differentiate between satisfactory, but different, responses and unsafe behavior. In general, this study demonstrates that vehicle-based algorithms to detect impairment are feasible, but that their performance depends on a careful tailoring of the algorithm to the drivers and roadway situation. The success of future impairment-detection algorithms will likely depend on understanding how the impairment interacts with different types of drivers, trips, and roadway situations. An important research question concerns identifying classes of roadway situations and classes of drivers so that the algorithms do not need to be tailored to each individual driver and each individual roadway situation. Eye movement data might be particularly useful to consider in this context.

These results support the long-term research objective of using algorithms that detect impairment to provide drivers with feedback that will discourage or prevent drinking and driving. Ultimately the distraction-detection algorithms developed in this study could support a range of vehicle-based interventions to prevent alcohol-related crashes. Such interventions could include limiting drivers' ability to drive dangerously (e.g., lockout distractions or limit speed), providing feedback to impaired drivers that may motivate them to pull over or drive more cautiously, adjusting crash warning systems to provide an earlier warnings, or providing long-term feedback that highlights dangerous driving. Such interventions all depend on a reliable means of detecting alcohol impairment using driver behavior data, which this study demonstrates as being feasible.

The promising results associated with alcohol-related impairment detection suggest other types of impairment detection might also hold promise, most notably distraction and drowsiness. As such algorithms and associated interventions are developed, their joint performance must be considered. One such consideration is the definition of a false alarm. False alarms and misses need to be interpreted in the context of how the impairment detection will be used. If the ultimate goal of a system is to identify impaired behavior rather than inferring the presence of alcohol, then the meaning of false alarms and misses changes substantially. An alcohol-impairment algorithm might be sensitive to drowsiness and so could provide valuable information even if the impairment it detects is not the one it was originally designed to detect. The ultimate definition of algorithm sensitivity might need to depend on the intervention it supports, just as the required sensitivity depends on the intervention.

10 ACRONYMS AND ABBREVIATIONS

ANOVA	analysis of variance
BAC	blood alcohol concentration
CI	confidence interval
EEG	electroencephalogram
FSA	force sensor arrays
NADS	National Advanced Driving Simulator
NHTSA	National Highway Traffic Safety Administration
OEM	original equipment manufacturer
PCA	principal components analysis
PPP	positive predictive performance
QFV	Quantity-Frequency-Variability scale
ROC	receiver operating characteristic
RPM	rotations per minute
SD	standard deviation
SDLP	standard deviation of lane position
SE	steering entropy
SSS	Stanford Sleepiness Scale
SFST	standardized field sobriety test
SUV	sport-utility vehicle
SVD	singular value decomposition
TH	time headway

ACKNOWLEDGEMENTS

The project benefited from the contributions of Mark Gilmer and Tim Brewer from Intoximeter, who arranged for a loan of the Alco-Sensors for this study.

REFERENCES

Amari, S., & Wu, S. (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural Networks, 12*(6), 783-789.

Arnedt, J. T., Wilde, G. J. S., Munt, P. W., & MacLean, A. W. (2001). How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? *Accident Analysis and Prevention, 33*(3), 337-344.

Babor, T. F., de la Fuente, J. R., Saunders, J., & Grant, M. (1992). *AUDIT: The alcohol use disorders identification test: Guidelines for use in primary health care*. Geneva, Switzerland: World Health Organization.

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., & Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems, 7*(1), 63-77.

Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., et al. (2002). *The Economic Impact of Motor Vehicle Crashes, 2000* (No. DOT HS 809 446). Washington, DC: National Highway Traffic Safety Administration.

Blomberg, R. D., Peck, R. C., Moskowitz, H., Burns, M., & Fiorentino, D. (2005). *Crash risk of alcohol invovlved driving: A case control study*. Stamford, CT: Dunlap and Associates.

Boer, E. R. (2001). Behavioral entropy as a measure of driving performance. In D. V. McGehee, J. D. Lee & M. Rizzo (Eds.), *Proceedings of the International Driving Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design* (pp. 225-229). Public Policy Center, Iowa City, IA.

Brookhuis, K. A., & De Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics, 36*(9), 1099-1110.

Brookhuis, K. A., De Waard, D., & Fairclough, S. H. (2003). Criteria for driver impairment. *Ergonomics, 46*(5), 433-445.

Burns, M., & Dioquino, T. (1997). *A Florida Validation Study of the Standardized Field Sobriety Test (SFST) Battery* (No. Florida Department of Transportation Rep.No. AL-97-05-14-01). Los Angeles, CA: Southern California Research Institute.

Burns, M., & Moskowitz, H. (1977). *Psychophysical Tests for DWI Arrest* (No. US Department of Transportation Rep. No. DOT HS-802 424). Washington, DC: National Highway Traffic Safety Administration.

Byun, H., & Lee, S. W. (2002). Applications of Support Vector Machines for Pattern Recognition: A Survey. *Pattern Recognition with Support Vector Machines, First International Workshop, SVM 2000*, 213-236.

Chung, T., Colby, S. M., Barnett, N. P., & Monti, P. M. (2002). Alcohol use disorders identification test: Factor structure in an adolescent emergency department sample. *Alcoholism: Clinical and Experimental Research, 26*(2), 223-231.

Conley, T. B. (2001). Construct validity of the MAST and AUDIT with multiple offender drunk drivers. *Journal of Substance Abuse Treatment, 20*(4), 287-295.

de Waard, D., Brookhuis, K. A., & Hernandez-Gress, N. (2001). The feasibility of detecting phone-use related driver distraction. *International Journal of Vehicle Design, 26*(1), 85-95.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician, 37*(1), 36-48.

Evans, L. (1991). *Traffic Safety and the Driver*. New York: Van Nostrand Reinhold.

Fairclough, S. H., & Graham, R. (1999). Impairment of driving performance caused by sleep deprivation or alcohol: A comparative study. *Human Factors, 41*(1), 118-128.

Fell, J. C., Tippetts, A. S., & Voas, R. B. (2009). Fatal traffic crashes involving drinking drivers: what have we learned? *Annals of Advances in Automotive Medicine, 53*, 63-76.

Feng, C. X. J., Yu, Z. G. S., Emanuel, J. T., Li, P. G., Shao, X. Y., & Wang, Z. H. (2008). Threefold versus fivefold cross-validation and individual versus average data in predictive regression modelling of machining experimental data. *International Journal of Computer Integrated Manufacturing, 21*(6), 702-714.

Grace, R., & Suski, V. (2001). Improving safety for drivers and fleets: Historical and innovative approaches. In D. V. McGehee, J. D. Lee & M. Rizzo (Eds.), *Proceedings of the International Driving Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design* (pp. 345-350). Iowa City: IA: Public Policy Center.

Harris, D. H. (1980). Visual detection of driving while intoxicated. *Human Factors, 22*(6), 725-732.

Holloway, F. A. (1994). *Low-dose Alcohol Effects on Human Behavior and Performance: A Review of Post-1985 Research*. Oklahoma City, OK: Civil Aeromedical Institute, Federal Aviation Administration.

Kennedy, R. S., Turnage, J. J., Rugotzke, G. G., & Dunlap, W. P. (1994). Indexing Cognitive Tests to Alcohol Dosage and Comparison to Standardized Field Sobriety Tests. *Journal of Studies on Alcohol, 55*(5), 615-628.

Kennedy, R. S., Turnage, J. J., Wilkes, R. L., & Dunlap, W. P. (1993). Effects of Graded Dosages of Alcohol on 9 Computerized Repeated-Measures Tests. *Ergonomics, 36*(10), 1195-1222.

Lacey, J. H., Kelley-Baker, T., Furr-Holden, D., Voas, R. B., Romano, E., Torres, P., et al. (2009). *2007 National Roadside Survey of Alcohol and Drug Use by Drivers: Alcohol Results*. Washington, DC: National Highway Traffic Safety Administration.

Lenne, M. G., Dietze, P., Rumbold, G. R., Redman, J. R., & Triggs, T. J. (2003). The effects of the opioid pharmacotherapies methadone, LAAM and buprenorphine, alone and in combination with alcohol, on simulated driving. *Drug and Alcohol Dependence, 72*(3), 271-278.

Liang, Y., Reyes, M. L., & Lee, J. D. (2007a). Non-intrusive detection of driver cognitive distraction in real-time using Bayesian networks. *Transportation Research Board*(2018), 1-8.

Liang, Y., Reyes, M. L., & Lee, J. D. (2007b). Real-time detection of driver cognitive distraction using Support Vector Machines. *IEEE Intelligent Transportation Systems, 8*(2), 340-350.

Lim, T. S., Loh, W. Y., & Shih, Y. S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning, 40*(3), 203-228.

Linnoila, M., & Mattila, M. J. (1973). Drug interaction on driving skills as evaluated by laboratory tests and by a driving simulator. *Pharmakospsychiatrie Neuro-Psychopharmakologie, 6*, 127-132.

Lund, A. K. (2006). Eliminating alcohol-impaired driving: Potential effects of technology applied to the general population, Insurance Institute for Highway Safety *International Technology Symposium*. Albuquerque, NM.

Lund, A. K., & Wolfe, A. C. (1991). Changes in the incidence of alcohol-impaired driving in the United States, 1973-1986. *Journal of Studies on Alcohol, 52*(4), 293-301.

Marple-Horvat, D. E., Cooper, H. L., Gilbey, S. L., Watson, J. C., Mehta, N., Kaur-Mann, D., et al. (2008). Alcohol badly affects eye movements linked to steering, providing for automatic in-car detection of drink driving. *Neuropsychopharmacology*, *33*, 849 – 858.

Moskowitz, H., Ziedman, K., & Sharma, S. (1976). Visual search behavior while viewing driving scenes under the influence of alcohol and marihuana. *Human Factors, 18*(5), 417-431.

Nakayama, O., Futami, T., Nakamura, T., & Boer, E. R. (1999). SAE Technical Paper Series: Development of a steering entropy method for evaluating driver workload. *Human Factors in Audio Interior Systems, Driving, and Vehicle Seating, SP-1426*.

National Highway Traffic Safety Administration. (2009a). Fatality Analysis Reporting System (FARS). Retrieved March 30, 2009, from ftp://ftp.nhtsa.dot.gov/fars/

National Highway Traffic Safety Administration. (2009b). Fatality Analysis Reporting System (FARS). Retrieved June, 2009, from ftp://ftp.nhtsa.dot.gov/fars/

Ogden, E. J. D., & Moskowitz, H. (2004). Effects of alcohol and other drugs on driver performance. *Traffic injury Prevention, 5*, 185-198.

Peters, B., & van Winsum, W. (1998). *SAVE: System for effective Assessment of the driver state and Vehicle control in Emergency situations*: R&D Programme Telematics: EU.

Pollard, J. K., Nadler, E. D., & Stearns, M. D. (2007). *Review of Technology to Prevent Alcohol-Impaired Crashes (TOPIC)* (No. DOT HS 810 827). Washington, DC: US Department of Transportation, National Highway Traffic Safety Administration.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research, 4*, 77-90.

Rakauskas, M. E., Ward, N. J., Boer, E., Bernat, E. M., Cadwallader, M., & Patrick, C. J. (2008). Combined effects of alcohol and distraction on driving performance. *Accident Analysis & Prevention, 40*, 1742–1749.

Roehrs, T., Beare, D., Zorick, F., & Roth, T. (1994). Sleepiness and ethanol effects on simulated driving. *Alcoholism-Clinical and Experimental Research, 18*(1), 154-158.

Rowe, S. Y., Olewe, M. A., Kleinbaum, D. G., McGowan JR, J. E., McFarland, D. A., Rochat, R., et al. (2006). The influence of observation and setting on community health workers' practices. *International Journal for Quality in Health Care, 18*(4), 299–305.

Rubenzer, S. J., & Stevenson, S. B. (2010). Horizontal Gaze Nystagmus: A Review of Vision Science and Application Issues. *Journal of Forensic Science, 55*(2).

Saad, L. (2003). Moderate drinking on the rise: The Gallup Organization, Gallup News Service, Poll Analyses.

Stuster, J. W. (1997). *The Detection of DWI at BACs Below 0.10*. Washington, DC: US Department of Transportation, National Highway Traffic Safety Administration.

Stuster, J. W. (1999). *Development of an Automated DWI Detection System*. Washington, DC: NHTSA.

Stuster, J. W. (2006). Validation of the standardized field sobriety test battery at 0.08% blood alcohol concentration. *Human Factors, 48*, 608-614.

Tippetts, A. S., & Voas, R. B. (2002). Odds that an involved driver was drinking: Best indicator of an alcohol-related crash? In D. R. Mayhew & C. Dussault (Eds.), *Proceedings of the 16th International Conference on Alcohol, Drugs and Traffic Safety* (Vol. 1, pp. 67-71). Montreal, Canada: Société de l'assurance automobile du Québec.

Tzambazis, K., & Stough, C. (2000). Alcohol impairs speed of information processing and simple and choice reaction time and differentially impairs higher-order cognitive abilities. *Alcohol and Alcoholism, 35*(2), 197-201.

Vanakoski, J., Mattila, M. J., & Seppala, T. (2000). Driving under light and dark conditions: effects of alcohol and diazepam in young and older subjects. *European Journal of Clinical Pharmacology, 56*(6-7), 453-458.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology-Applied, 2*(3), 270-285.

Voas, R. B., Wells, J., Lestina, D., Williams, A., & Greene, M. (1997). Drinking and driving in the US: The 1996 National Roadside Survey. In C. Mercier-Guyon (Ed.), *Proceedings of the 14th International Conference on Alcohol, Drugs and Traffic Safety - T97, Annecy, 21 - 26 September 1997* (Vol. 3, pp. 1159-1166). Annecy, France: Centre d'Etudes et de Recherches en Médecine du Trafic.

Voas, R. B., Wells, J., Lestina, D., Williams, A., & Greene, M. (1998). Drinking and driving in the United States: The 1996 National Roadside Survey. *Accident Analysis and Prevention, 30*(2), 267-275.

Ward, N. J. (2006). *Alcohol, Traffic Safety, and the Conception of Vehicle-based Systems to Detect Driver Impairment*.

Wolfe, A. C. (1974). *1973 US national roadside breath testing survey: procedures and results*. Ann Arbor, MI: University of Michigan Safety Research Institute.

Zador, P. L., Krawchuk, S. A., & Voas, R. B. (2000). Alcohol-related relative risk of driver fatalities and driver involvement in fatal crashes in relation to driver age and gender: An update using 1996 data. *Journal of Studies on Alcohol, 61*(3), 387-395.

Zador, P. L., Krawchuk, S. A., & Voas, R. B. (2001). Alcohol-related relative risk of driver fatalities and driver involvement in fatal crashes in relation to driver age and gender: An update using 1996 data. *Journal of Studies on Alcohol*(May), 387-395.

DOT HS 811 358
August 2010

U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

NHTSA
www.nhtsa.gov