COVER SHEET

Assessing Crash Risks on Curves

Samantha Chen[1], Andry Rakotonirainy[1], Mary Sheehan[1],
Shonali Krishnaswamy[2], Seng Wai Loke[3]

[1]Centre for Accident Research and Road Safety Queensland, Queensland University of Technology (Carseldine Campus), Beams Road, Carseldine, QLD, 4034, Australia.
{s2.chen, r.andry, m.sheehan}@qut.edu.au

[2]School of Computer Science & Software Engineering
Monash University
(Caulfield Campus)
900, Dandenong Road, Caulfield, Victoria-3145, Australia
Shonali.Krishnaswamy@infotech.monash.edu.au

[3]Department of Computer Science and Computer Engineering
Beth Gleeson Building, BG 215, La Trobe University, Bundoora, VIC 3086, Australia.
s.loke@latrobe.edu.au

Abstract

In Queensland, curve related crashes contributed to 63.44% of fatalities, and 25.17% required hospitalisation. In addition, 51.1% of run-off-road crashes occurred on obscured or open-view road curves (Queensland Transport, 2006). This paper presents a conceptual framework for an in-vehicle system, which assesses crash risk when a driver is manoeuvring on a curve. Our approach consists of using Intelligent Transport Systems (ITS) to collect information about the driving context. The driving context corresponds to information about the environment, driver, and vehicle gathered from sensor technology. Sensors are useful to detect drivers' high-risk situations such as curves, fogs, drivers' fatigue or slippery roads. However, sensors can be unreliable, and therefore the information gathered from them can be incomplete or inaccurate. In order to improve the accuracy, a system is built to perform information fusion from past and current driving information. The integrated information is analysed using ubiquitous data mining techniques and the results are later used in a Coupled Hidden Markov Model (CHMM), to learn and classify the information into different risk categories. CHMM is used to predict the probability of crash on curves. Based on the risk assessment, our system provides appropriate intervention to the driver. This approach could allow the driver to have sufficient time to react promptly. Hence, this could potentially promote safe driving and decrease curve related injuries and fatalities.

Keywords: Road curves, casualties, driving context, sensors, hidden Markov model, information fusion, classify, predict, interventions.

1 Introduction

Road curves are an important feature of our road infrastructure. However, road curves are serious crash risk to vehicle occupants. In Australia, 30% of crashes occur on road curves (Shields et al., 2001). Crashes on road curves frequently result in fatal injuries or casualties. Curve related crashes contributed to 63.44% of fatalities , and 54% are due to off-path crashes on curved roads (Queensland Transport, 2006). In addition, the likelihood of surviving crashes on curved roads is approximately 17% lower than on straight roads (Queensland Transport, 2006). The major types of crashes that occurred on curved roads are run-off-road, head-on, rollover, and hitting roadside objects. In Queensland, the most common type of crash on curved roads is run-off-road crashes (54%) which involve 10% more fatal crashes than on straight roads (Queensland Transport, 2006). The major contributing factors for crashes on road curves are speeding, the degree of road curve, and misjudgement. Other factors include a wet road surface, driving under the influence of alcohol, the condition of the tyres or brakes, driver fatigue, distraction, and visibility.

The Organisation for Economic Co-operation and Development (OECD) has estimated that Intelligent Transport Systems (ITS) could reduce fatalities and injuries in OECD countries by 40%, thereby saving over US$ 270 billion per year (OECD, 2003). The Australian Transport

Safety Bureau claims that ITS should bring benefits with a total of at least $14.5 billion by 2012. Of this amount, $3.8 billion is estimated to be saving due to safety improvements (Australian Government, 2004). Therefore, a potentially promising approach to improving safety on road curves is to utilize technology in combination with existing engineering interventions.

The rest of this paper is structured as follows. Sections 2 and 3 discuss existing road engineering and Intelligent Transport Systems interventions. Section 4 explains our approach, which is a novel statistical model that incorporates a driver behaviour model and ubiquitous data mining techniques for vehicles to assess road crash risks in curves. Then we present the methodology to achieve the initial stage of the project: Text mining of past crash data in Section 5, and the preliminary results obtained in Section 6. Section 7 discusses future work and concludes the paper.

2 Existing Road Engineering Interventions

Road authorities have implemented different countermeasures to prevent crashes from occurring and to minimize the consequences of a crash at horizontal curves. The primary methods are pavement markings, warning signs and delineations. The countermeasures used to reduce speeding, and provide guidance when entering a road curve, are transverse pavement markings, chevron alignment signs, horizontal alignment and advisory speed warning signs, and post-mounted delineators (PMD).

2.1 Pavement Markings and Warning signs

The transverse pavement markings used in horizontal road curves can provide drivers with the perception that the lane is narrower or that they are driver faster than their current speed. As a result, drivers will slow down. Horizontal alignment signs accompanied by an advisory speed warning sign alert drivers of a change of alignment ahead and suggest a safe speed to drive in a curve. Chevron alignment signs installed along curves provide additional guidance where there is a change in horizontal alignment (Carlson et al., 2004).

2.2 Delineators

Post-mounted delineators (PMD) are one kind of delineator used to mark an unexpected alignment. PMD provide good guidance at night as they are reflective and are installed at a height comparable to the headlights of the vehicles (Vest et al., 2005). Guideposts are another alignment delineator, used to indicate and enhance the visibility of the edge of the road (DIER, 2003). They are placed on narrow roads with insufficient road width to allow marking of the centre line and sometimes are accompanied with retro-reflective delineators to provide cues for the curve. These interventions can reduce crashes. However, they are ineffective if when drivers choose to ignore them or miss them. Hence, it is possible that road engineering interventions can be complemented by the use of intelligent transport system.

3 Existing Intelligent Transport Systems (ITS)

ITS can improve road safety by reducing the occurrence of crashes, and thereby the associated injuries and the amount of risk drivers are exposed to (Australian Government, 2004). The focus of ITS is to monitor on-road situations and real-time decision making in order to reduce accidents and fatalities. For example, smart cruise controllers can monitor the environment and adjust vehicle speed accordingly. The rest of this section presents a few of the existing in-vehicle ITS applications for curved roads situations. The section is organized according to the data type captured: environment, vehicle or/and driver information.

3.1 Information about the Vehicle

Information from the vehicle provides an idea of the driver's actions. Examples of the types of information which can be collected from the vehicle include: steering wheel movement, speed, and lane position.

One such application for curved roads is a Lane Departure Warning System (LDWS) that captures lane position information for the vehicle. It also helps to avoid run-off-road crashes and ensures that the vehicle stays within the lane. LDWS is meant to assist drivers who experience involuntary lane departure caused by fatigue, distractions from passengers, mobile phones, and radio. The system uses video cameras to keep track of the road ahead and provide 'virtual rumble strip' warnings when the vehicle is drifting out of the lane. However, the images from the video sensors can be confusing and information from maps when used may be unreliable. In Switzerland, Gontran *et al*. (Gontran et al., 2005) have proposed a new system, photobus, which uses data from cameras, sensors, satellite, road database, and a mobile mapping system, to improve LDWS.

3.2 Information about the Environment

Information about the surroundings provides an idea of the driving conditions and their effect on driver. An application that collects information on the curvature of the road is the Curve Speed Warning system (CSW) (Bishop, 2005). CSW warns drivers when they are travelling too fast to safely negotiate an upcoming curve. The system estimates a safe speed to negotiate a curve using data from a Global Position System (GPS) and an on-board map database, which determines the current vehicle position and road geometry information. Once the actual vehicle speed exceeds the recommended speed, CSW either issues a 'reduce speed' alert to the driver or reduces the speed automatically.

Another application is the Adaptive Front Lighting system (AFS), which illuminates the road ahead and the side of the vehicle path in order to optimize visibility for the driver at night. A basic system takes speed into account to create desired illumination for the driver. A more advanced system considers steering angle data along with speed and uses a swivelling lamp to automatically illuminate a wider angle of the path ahead. In addition, the next generation AFS will utilize data from GPS and digital maps to take into account any upcoming road curves. This enhanced AFS can provide proper illumination before entering and when driving through a road curve. Overall, AFS addresses night time visibility by providing a 90% improvement in the driver's view ahead and to the side. Other than that, enhanced AFS can be helpful in road curves, including sharp curves.

3.3 Information about the Driver

Information about driver can be used to detect variations in driver behaviour in different environments. Oliver & Pentland (2000) used a statistical model, the Hidden Markov Model (HMM), to model and predict driver behaviour. The Hidden Markov Model consists of a finite set of states and each state is associated with a probability distribution. An output can be generated at a particular state according to the probabilities. A HMM has both observable and hidden parameters. An example of an observable parameter in the driving context is the driver's gaze while an example of a hidden parameter is the driver's manoeuvre. HMM is a single state model, however, real-life problems consist of multiple interactions, so the model is extended to become a Coupled Hidden Markov Model (CHMM), which is capable of representing real-life problems. Brand *et al*. (1997) used a Coupled Hidden Markov Model to model seven different driver manoeuvres. CHMM has the ability to learn from given data and handle time-varying signals (Brand et al., 1997). In addition, it is able to model interacting processes at the same time, such as other vehicles. Oliver & Pentland (2000) found that a CHMM could predict driver behaviour one second before the actual driver manoeuvre. Therefore, they concluded that CHMM should be an essential feature operating between the driver and driver assistance systems to prevent potential hazard situations. On top of this, the model can be used to provide a more realistic automated vehicle in simulators (Oliver & Pentland, 2000).

There are several other ITS applications for road curves, such as electronic stability control, and a curve overshooting prevention support system. However, they will not be discussed in this paper. It can be seen from consideration of the review above that most of the applications do not use information of the environment, vehicle dynamics and driver behaviour together during data analysis process. Our approach, which is discussed below, will address the use of combined data using a ubiquitous data mining technique.

4 Our Approach

As shown in previous sections, there is a great diversity of engineering and ITS interventions to improve safety on road curves. However, research so far has not considered the combined use of environment, vehicle and driver information to improve road safety on curves. Therefore, address this shortcoming, we are proposing a model called Ubiquitous Situation Awareness Risk Prediction model for road Safety (UbiSARPS). Figure 1 provides an overview of our approach.

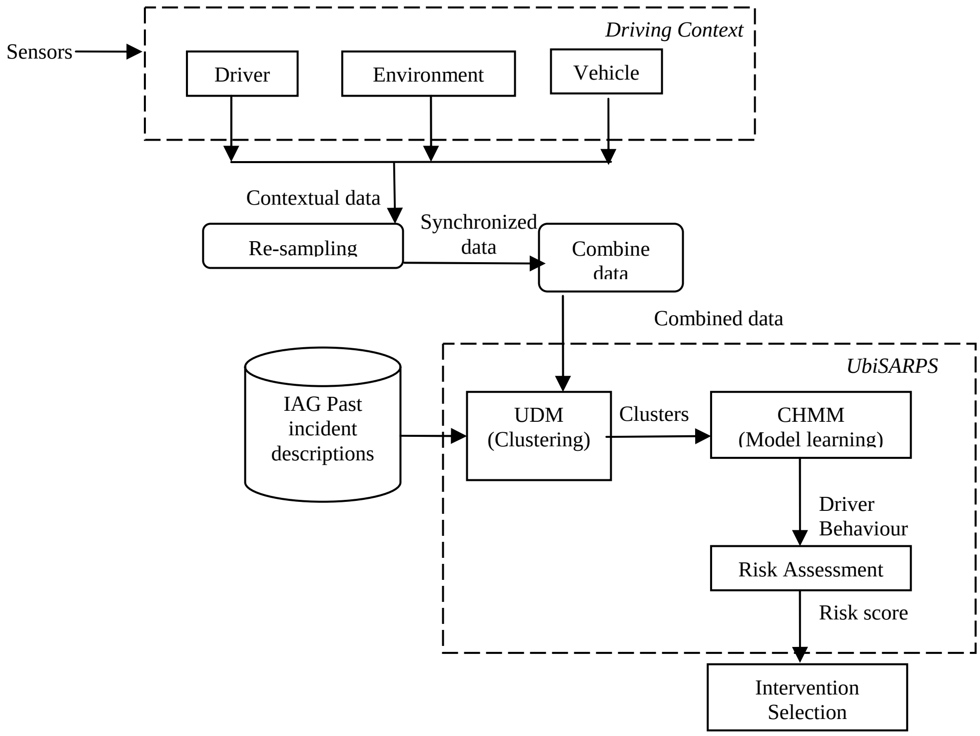

Figure 1. A brief overview of our approach.

4.1 Innovation

The main innovative contribution of our approach (UbiSARPS) is that it integrates information about driver behaviour, vehicle dynamics, environment factors, and crash history from an insurance company to assess and monitor crash risk. In addition, UbiSARPS will be personalized for each driver based on individual records provided, in order to reduce complacency towards anonymous risks.

Another innovation is that we are using a driver behaviour model, Coupled Hidden Markov Model and ubiquitous data mining algorithms to detect high-risk crash situations. Data mining is a process that extracts knowledge by analysing data to discover hidden patterns and dependencies in the database. Thus, ubiquitous data mining (UDM) is a process that analyses data from distributed and mobile devices. UDM is considered as a part of ITS, and hence has the potential to monitor the driving context and make real-time decisions. The two major UDM applications for vehicles are the Vehicle Electronics Data Acquisition System (VEDAS) and Situation Awareness with Ubiquitous data mining for Road safety (SAWUR). VEDAS (Hillol et al., 2004) is a system that monitors fleet vehicle health and driver characteristics. An alert is sent to the fleet manager whenever abnormal driving patterns arises. SAWUR (Salim et al., 2005) determines the current driving context and provides appropriate actions to avoid danger. The information is communicated in real-time using a wireless network. However, neither application considers driver behaviour as a data analysis factor.

4.2 Approach description

This subsection describes the approach in reference to the components in Figure 1.

4.2.1 Driving context

We collect contextual information about the environment, the vehicle and the driver to assess the risk of drivers when they manoeuvre on curves. We use a wide range of in-vehicle technologies and sensors to acquire the contextual factors relevant to driving on curves. Contextual factors such as driver performances, road geometry, curve type and speed are gathered to assess risks. We define performance metrics specific to curves and tailored to each driver. The collected data are composed of elements which occur at different frequencies; hence, we have to synchronize the data to the same frequencies based on the time. After synchronizing, the data are combined together for further data analysis in UbiSARPS.

4.2.2 UbiSARPS

UbiSARPS performs real-time data analysis of the combined synchronized data on vehicles/mobile devices to detect possible high-risk situations. The detection is assessed by a trained CHMM, which used the patterns learnt from past crash data and current contextual factors. For example, a driver had a crash at a particular curved road in the past so if the driver is approaching the same location or similar location, the model is able to determine that the driver is facing a high risk situation from the real-time analysis and the ubiquitous data mining process. In essence, UbiSARPS consists of Ubiquitous Data Mining (UDM), a Coupled Hidden Markov Model (CHMM), and a risk assessment process.

First, we perform data mining offline to train the CHMM, and later train using a model in real time to recognize patterns. To begin, we extract patterns from the information about past incidents occurring on road curves from an insurance company, IAG. Patterns from the past incidents can provide a further insight into the causes of road crashes on curves. Hence, we perform text mining to extract patterns as the descriptions are in free text format. Our aim is to extract possible contributing factors that lead to a crash and classify them into pattern clusters. The clustering algorithm used is a distance-based algorithm called Ward. The clusters are then used as inputs for the self-learning CHMM to learn the patterns and calculate crash risk based on the probability distribution. For example, if we obtain a cluster with keywords such as ‘wet road’, ‘skid’, and ‘lost control’, these keywords are considered as a set and act as an input for CHMM to learn. Hence, in the future, a warning is issued when the model detects a similar situation.

4.2.3 Intervention Selection

UbiSARPS can determine an intervention according to the risk assessment score obtained from the risk assessment process. The final appropriate intervention will be transferred to the technologies in the vehicle to warn the driver or correct the mistake automatically for the driver. However, we will not be designing the intervention in this research. After providing an overview of our approach, we are going to explain the methodology in the next section.

5 Methodology

In this paper, we are only presenting the initial state of the project: the text mining method and the results. This section briefly explains the aim, apparatus used, and procedures of the text mining process. In the following subsections, we discuss the apparatus used and the procedure carried out to obtain crash pattern from past crash data.

The aim is to determine the contributing factors on road curves, and identify high-risk situations and obtain crash patterns from the text mining analysis process.

5.1 Apparatus

For the text mining analysis, a text miner analysis tool within SAS is used. SAS is a software system which can be used to perform data mining (SAS, 2006).

5.2 Text Mining Procedure

We perform text mining with the information from the insurance incident records. Prior to text analysis, we filter the data to curve-related incidents only. Then, we use the SAS software to perform text mining to discover patterns within the description. Traditional data mining is ideal when dealing with numbers, but is not feasible for mining text descriptions, hence the use of text mining.

SAS uses the Ward algorithm to create pattern clusters, which consists of keywords of contributing factors. The patterns we are searching for are based on the following elements:

- The age distribution in the highest number of crashes.
- Other contributing factors not mentioned in Queensland Transport or Road Transport Authority (RTA).

- The contributing factors for crashes that cost below AUD$2,500. The cost is the monetary amount to repair property (mainly vehicles) damaged in the crash.

The discovered patterns or factors are used to create a risk behaviour scale with a weight for each factor. The patterns are also used as an input for the CHMM to learn.

6 Preliminary Results

We are in the preliminary stage of this project, and have started to perform text mining analysis. Therefore, this section discusses results of the tests performed on the text mining. We present the contributing factors for crashes at road curves from Queensland Transport, and IAG. The factors are based on the highest number of crashes and the amount being claimed in insurance for crashes on road curves.

6.2 Age distribution

For crashes that cost less than $2,500, the highest crash frequency has claims that range from $2,126 to $2,250. The highest crash frequency for crashes that cost more than $2,500 has claims that range from $2,626 to 2,760 and $2,876 to $3,000. Figure 2 shows the frequency and the claim amount distribution of the crash database.

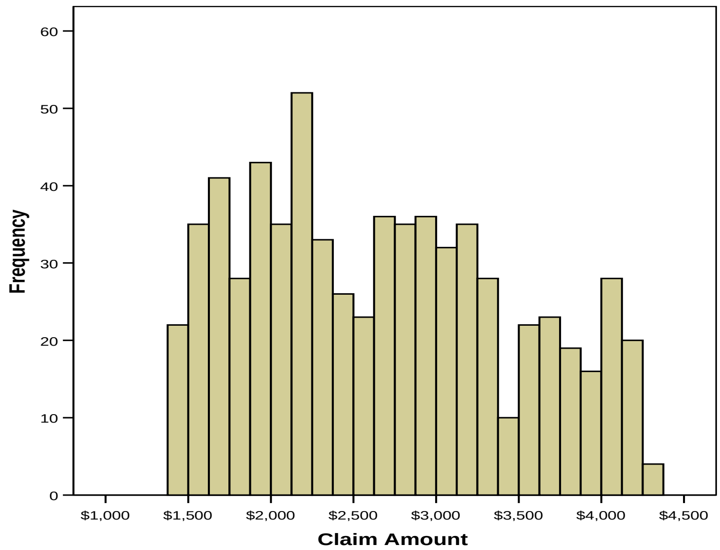

Figure 2. A histogram that displays the frequency of crashes and the claim cost.

Then using the highest frequency, we examine the contributing factors for each age group distribution: young, mature and older drivers. Young drivers of age 21 years old or less account for the highest number of crashes where crashes cost less than $2,500. Mature drivers aged between 30 and 52 years old are involved more in crashes that cost more than $2,500.

6.2 Contributing factors

Crashes that cost more than $2.500 must be reported to Queensland Transport and have their contributing factors identified. We identify contributing factors to crashes that cost less than $2,500 from the insurance records due to the following reasons:

- Crash records for these low cost crashes are not available through Queensland Transport's Webcrash (Queensland Transport, 2006).
- To determine the contributing factors that lead to crashes on road curves

- To determine the difference between the contributing factors of crashes that cost more than $2,500 and those that cost less.

Table 1 shows a list of the factors of crashes on road curves in Queensland which involve male drivers of different age group (young, mature, and older drivers). The contributing factors are divided into two categories: crashes that cost more $2,500 and crashes that cost less than $2,500.

<table>
<tr><th></th><th>Male Age Group</th><th>Queensland Transport Webcrash</th><th>IAG crash database</th></tr>
<tr><td rowspan="3">Crashes more than $2,500</td><td>Young driver</td><td rowspan="3">• Wet road
• Inattention
• Inexperience
• Alcohol related
• Speed related
• Fatigue related
• Fail to give way or stop
• Illegal manoeuvre</td><td>• Right hand bend
• Sweeping bend
• Skidded on dirt road
• Swerved to avoid
• Unable to stop</td></tr>
<tr><td>Mature driver</td><td>• Lost control
• Hit animal
• Speeding
• Side swiped
• Turn too wide
• Wet road
• Wrong side of road</td></tr>
<tr><td>Older driver</td><td>• No marked lane
• Obscured view curve.</td></tr>
<tr><td rowspan="3">Crashes less than $2,500</td><td>Young driver</td><td rowspan="3"></td><td>• Lost control
• Avoid kangaroo
• Slide off</td></tr>
<tr><td>Mature driver</td><td>• Wet road
• Right hand bend
• Wrong side of road
• Distracted
• Sharp bend
• Lost traction
• Narrow road</td></tr>
<tr><td>Older driver</td><td>• Hit parked vehicle
• Right hand bend
• No dividing line on road</td></tr>
</table>

Table 1. Contributing factors by source (Queensland Transport and IAG), age category and cost.

6.3 Analysis & Discussion

The results showed that young male drivers age 21 tend to be involved in crashes that cost less than $2,500 while mature drivers age 30 and 52 have crashes which cost more than

$2,500. This could be due to driver attitude and the age of the vehicle used. From the analysis of the age of vehicle used, we discovered that older drivers use older vehicles. This could be one of the reasons for the cost of the crash.

For the contributing factors obtained from text mining, the incident description terms are shown in Table 1. The interesting contributing factors discovered that are not available from Webcrash (Queensland Transport, 2006) are: right hand bend, no lane marking, narrow road, lost traction, sweeping bend, side swiped and avoiding kangaroo.

7 Future work & Conclusion

Road crashes on curves represent a high number of fatalities. Road authorities have suggested several engineering interventions to prevent crashes from occurring at road curves. However, the number of crashes which could be prevented with these interventions is not high. A brief review of existing ITS applications related to road curves indicates which issues should be addressed. We have begun analysing past incident descriptions and extracting patterns through the text mining process. As the next step in our research, we will build a Hidden Markov Model and then extend it to a Coupled Hidden Markov Model and train it with the inputs from the data mining processes so that our completed model can recognize and predict high risk situations. We will use a simulator to verify the performance of our model. Then we will create a risk assessment scale, which will be used to determine the risk score. This score will be obtained using the judgment of driver trainers and existing driver risk assessment software. The score will be used as an input for the learning model to learn and assess future crash risks for similar situations. We hope to obtain significant results from this research to improve road safety.

8 References

Australian Government, A. T. S. B. (2004). *Road Safety in Australia, A Publication Commemorating World Health Day 2004*. Canberra, Australia: Paragon Printers Australasia.

Bishop, R. (2005). Lateral/Side sensing and Control Systems, Chapter 6. In *Intelligent Vehicle Technology and Trends*: Artech House.

Brand, M., Oliver, N., & Pentland, A. (1997). *Coupled Hidden Markov Model for Complex Action Recognition*.

Carlson, P., Rose, E., Chrysler, S., & Bischoff, A. (2004). *Simplifying delineator and chevron applications for horizontal curves* (No. FHWA/TX-04/0-4052-1). Texas: Texas Transportation Institute.

DIER. (2003). *Road Hazard Management Guide*. Tasmina: ARRB Transport.

Gontran, H., Gilliéron, P. Y., & Skaloud, J. (2005, 9-11 March 2005). *Precise Road Geometry for Integrated Transport Safety Systems.* Paper presented at the Swiss Transport Research Conference.

Hillol, K., Ruchita, B., Kun, L., Michael, P., Patrick, B., Samuel, B., et al. (2004). *VEDAS: A Mobile and Distributed Data Stream Mining System for Real-Time Vehicle Monitoring.* Paper presented at the Proceedings of SIAM International Conference on Data Mining 2004, California.

OECD. (2003). *Road Safety Impact of New Technologies*: OECD Publishing.

Oliver, N., & Pentland, A. P. (2000). *Graphical Models for Driver Behavior Recognitioni in a SmartCar*.

Queensland Transport, Q. (2006). Webcrash 2.3: Queensland Goverment.
Salim, F. D., Krishnaswamy, S., Loke, S. W., & Rakotonirainy, A. (2005). *Context-Aware Ubiquitous Data Mining Based Agent Model for Intersection Safety*.
SAS. (2006). (Version 9.1).
Shields, B., Morris, A., Jo., B., & Fildes, B. (2001). *Australia's National Crash In-depth Study Progress Report*: Monash University, Accident Research Centre.
Vest, A., Stamatiadis, N., Clayton, A., & Pigman, J. (2005). *Effects of Warning Signs on Curve Operating Speeds* (No. KTC-05-20/SPR-259-03-1F): University of Kentucky.