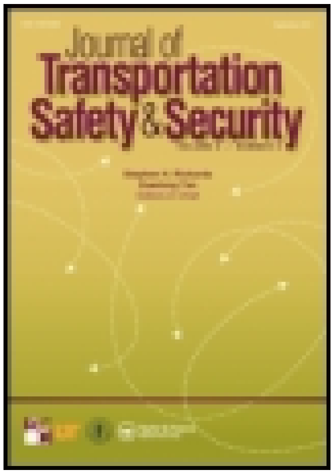

Journal of Transportation Safety & Security
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utss20

A novel approach to assessing road-curve crash severity

Andry Rakotonirainy[ab], Samantha Chen[ab], Bridie Scott-Parker[c], Seng Wai Loke[d] & Shonali Krishnaswamy[e]

[a] Centre for Accident Research and Road Safety - Queensland, Queensland University of Technology, K Block, 130 Victoria Park Road, Kelvin Grove, Queensland, 4059, Australia

[b] Institute of Health and Biomedical Innovation, Queensland University of Technology, Queensland, Australia

[c] University of the Sunshine Coast Accident Research (USCAR), Faculty of Arts and Business, University of the Sunshine Coast, Queensland, Australia

[d] Department of Computer Science and Computer Engineering, La Trobe University, Bundoora, Victoria, 3086, Australia

[e] School of Computer Science and Software Engineering, Monash University, 900 Dandenong Road, Caulfield, Victoria, 3145, Australia

Accepted author version posted online: 10 Sep 2014.

To cite this article: Andry Rakotonirainy, Samantha Chen, Bridie Scott-Parker, Seng Wai Loke & Shonali Krishnaswamy (2014): A novel approach to assessing road-curve crash severity, Journal of Transportation Safety & Security, DOI: 10.1080/19439962.2014.959585

To link to this article: http://dx.doi.org/10.1080/19439962.2014.959585

Title of Manuscript: A novel approach to assessing road-curve crash severity

Authors: Andry Rakotonirainy[1, 2]

Samantha Chen[1, 2]

Bridie Scott-Parker [3]

Seng Wai Loke [4]

Shonali Krishnaswamy[5]

[1] Centre for Accident Research and Road Safety – Queensland, Queensland University of Technology, K Block, 130 Victoria Park Road, Kelvin Grove, Queensland, 4059, Australia.

[2] Institute of Health and Biomedical Innovation, Queensland University of Technology, Queensland, Australia

[3] University of the Sunshine Coast Accident Research (USCAR), Faculty of Arts and Business, University of the Sunshine Coast, Queensland, Australia

[4] Department of Computer Science and Computer Engineering, La Trobe University, Bundoora, Victoria, 3086, Australia.

[5] School of Computer Science and Software Engineering, Monash University, 900 Dandenong Road, Caulfield, Victoria, 3145, Australia.

Corresponding Author: Dr Bridie Scott-Parker

Address: University of the Sunshine Coast Accident Research (USCAR), University of the Sunshine Coast, Locked Bag 4, Maroochydoore DC, Queensland 4558m Australia.

Email: bscottpa@usc.edu.au

Abstract

Curves are a common feature of road infrastructure; however crashes on road curves are associated with increased risk of injury and fatality to vehicle occupants. Countermeasures require the identification of contributing factors. However, current approaches to identifying contributors use traditional statistical methods and have not used self-reported narrative claim to identify factors related to the driver, vehicle and environment in a systemic way. Text mining of 3434 road-curve crash claim records filed between 1 January 2003 and 31 December 2005 at a major insurer in Queensland, Australia, was undertaken to identify risk levels and contributing factors. Rough set analysis was used on insurance claim narratives to identify significant contributing factors to crashes and their associated severity. New contributing factors unique to curve crashes were identified (e.g., tree, phone, over-steer) in addition to those previously identified via traditional statistical analysis of Police and licensing authority records. Text mining is a novel methodology to improve knowledge related to risk and contributing factors to road-curve crash severity. Future road-curve crash countermeasures should more fully consider the interrelationships between environment, the road, the driver and the vehicle, and education campaigns in particular could highlight the increased risk of crash on road-curves.

1 Introduction: Crashes on road curves

Road curves are integral in road infrastructure and can be horizontal and/or vertical in nature. In Australia, nearly one third of crashes occur on road curves (Shields et al, 2001), and the consequences of crashing on road curves are frequently more severe than crashes on straight roads. To illustrate, in Queensland, Australia, between 1 July 2008 and 30 June 2009, 39.0% of fatality crashes occurred on curves. The fatality rate of road-curve crashes was more than twice that of straight road crashes (2.3% of road-curve crashes were fatal, 1.0% of straight road crashes were fatal). In addition, 45.2% of persons injured in road-curve crashes required hospitalisation, compared to 37.3% of drivers injured on straight road crashes (Department of Transport and Main Roads (DTMR, formerly Queensland Transport), 2011).

The severity of the road-curve crash is dependent on a number of variables in the road transport system pertaining to both the crash itself and to circumstances immediately prior to the crash. Crash statistics indicate that 73% of fatal crashes that occur on road curves involve travelling at speeds in excess of the posted speed limit (Queensland Transport, 2006). Travelling in excess of posted speed limits is associated with greater severity of crashes and increased likelihood of injury and fatality (Aarts and van Schagen, 2006; Kloeden et al., 1997), and travelling at excess speed on a road curve in particular is likely to lead to loss of control of the vehicle which results in a run-off-road or roll-over crash. Given the necessity of curves in road infrastructure and the corresponding high rates of injury and fatality from crashes occurring as the driver uses these road features, it is vital that contributors to the road-curve crashes, and their severity in particular, are identified to inform countermeasure development.

1.1 Contributing factors

A number of contributing factors to road-curve crashes have been identified in the road safety literature, and these pertain to characteristics of the road, the driving environment, the driver, and the vehicle. Contributing factors have been gleaned from such sources as police reports and hospital records, and it is noteworthy that these sources require subjective assessment and interpretation of contributors to crashes, particularly if the crash-involved driver is fatally-injured. Characteristics of the road that have been found to contribute to curve-related crashes include the degree and length of the curve, sight distance, lane width, surface and side friction, super-elevation, and the location of the curve. A sharper curve, that is a curve with a smaller radius, involves greater crash potential as the level of difficulty negotiating the curve increases, principally as drivers have less time to perform corrective manoeuvres on the resulting shorter curve (Othman et al., 2010). Such curves also are frequently associated with poor visibility of the upcoming curve and driving hazards such as other vehicles. Roadside objects such as electricity poles or trees can also affect sight distance (Torbic et al., 2004). Narrow roads in a curve (Zegeer et al., 1991) encourage drivers to position their vehicles in alignment with the centre line, drivers thereby occupying more of the road space on a curve and placing themselves at greater risk of a head-on collision (Rosey et al., 2008). Road curves that have a worn surface result in lower friction and greater braking distances in the event of an emergency, particularly if the surface is wet (Das and Abdel-Aty, 2010). Super-elevation involves an inclined road-curve that capitalises upon the frictional force between the vehicle's tyres and the road surface and the weight of the vehicle to create the required centrifugal force to prevent the vehicle from sliding out of the curve. Vehicle speed is integral in the stability of the vehicle on the curve, excess speeds placing the driver at increased risk of a road-curve crash. Curves located after a long stretch of straight road are also associated with greater crash risk (Seneviratne and Islam, 1994).

Environmental factors (i.e., anything which is outside of the vehicle) that have been found to contribute to curve-related crashes include weather, time of day, animals and debris on road ways, signage, and traffic congestion. Rainy weather results in slippery roads and reduced driver visibility, compounding crash risk and increasing crash severity. Driving at night is hazardous for all drivers, and in conjunction with wet weather makes negotiating a road curve particularly risky (Caliendo et al., 2007). Animals on the road are a hazard not only for drivers who take evasive action and may steer into oncoming traffic or run off the road, but also as large animals such as moose and kangaroos can impact the vehicle, causing considerable damage to the vehicle and increasing the risk of injury to the driver (Rowden et al., 2008). Debris such as frayed tyres on the roadway can reduce the contact between the vehicle and the road-curve surface (Retting et al., 2000). Road signage (Bai et al., 2010) can also play a role in road-curve crashes, for example if the signage distracts the driver from the driving task or if the road safety instructions on signage are ambiguous. Drivers may change their driving behaviour to suit varying levels of traffic congestion (McCartt et al., 2004), for example drivers may travel at higher speed when the flow of traffic is not impeded, placing them at greater risk of injury or fatality from a road-curve crash.

Driver-related variables include the age of the driver, misjudgement, distraction and fatigue, speeding, and driving after drinking alcohol. Younger drivers are likely to be inexperienced in negotiating curves (Chen et al., 2010) as they are still novice drivers, and coupled with poorer hazard perception skills, increased sensation seeking propensity, and susceptibility to negative peer influence and risky decision making are at greater risk of crashing on road curves (Scott-Parker et al., 2009). Drivers may also under- or over-estimate the sharpness of the curve and fail to safely negotiate the road-curve. Distraction on a road-curve is particularly risky given the transient nature of the road environment, and drivers may over-correct when their attention is re-focussed upon their driving. Mobile phones have emerged in recent years as a major source of in-vehicle distraction (Ishigami and Klein, 2009). Distraction (Lam, 2002) is also more likely if the driver is fatigued, for example as a result of circadian rhythms and travelling at peak rest times in the afternoon and early morning, or through extended journey duration. Fatigued drivers take longer to detect and to react appropriately to driving hazards, and experience difficulty concentrating on the driving task and maintaining their lane position (Gnardellis et al., 2008). As noted in section 1, speeding is associated with greater risk of crash, injury and fatality, particularly on road-curves (Aarts and van Schagen, 2006), and it is noteworthy that drivers have a shorter period in which they can detect and react to hazards in a rapidly-changing road environment (Liu et al., 2005). Drivers are also at increased risk of a road-curve crash if they have been drinking alcohol (McCartt et al., 2004) which has been found to affect not only the ability of the driver to perform the motor tasks of controlling a vehicle, but also high-order processing functions such as decision making and hazard perception (Grunewald and Ponicki, 1995; Keall and Frith, 2004).

Road-curve crashes have also been found to arise from vehicle-related characteristics such as vehicle defect or failure. These are most commonly defective tyres which have insufficient tread or have been punctured and therefore do not maintain safe contact between the car and the road surface throughout the curve (Broyles et al, 2001), and inadequately maintained brakes requiring greater stopping distances (Jones and Stein, 1981). Vehicle age has also been found to be a factor (Blows et al., 2003): older and less expensive vehicles are less likely to have primary and secondary safety features which also reduce the risk of road-curve crash, such as electronic stability control.

1.2 Data mining techniques

Data mining is a process of knowledge discovery from large data sets by combining methods from statistics and artificial intelligence. Large volumes of data are analysed for patterns, trends and interrelationships (Hand et al., 2001), and data mining can be used to identify factors or predictors of crashes. Knowledge derived through data mining processes include models, patterns or relationships, and techniques such as rough set analysis, clustering algorithms and genetic algorithms have been utilised in traffic studies. Pande and Abdel-Aty (2006) used data mining techniques to examine conditions conducive to road crashes, identifying and classifying different categories of rear-end crashes in particular, using the classifications to define a prediction model and applying the model in real-time with loop detectors to assess the probability of crashing. This approach facilitated the development of a warning system that could alert drivers regarding potential rear-end crashes 5-10 minutes prior to such a collision.

Data mining can be used to identify factors or predictors of crashes, and data mining techniques have provided unique insight into the nature of car crashes, analysing data from a different perspective to discover previously unknown patterns and data correlations. For example, Kuhlmann et al. (2005) utilised data mining techniques to examine the design and crash-simulation performance of motor vehicles. Singh (2001) studied the relationships between the driver and vehicle factors that were found to contribute to crashes, such as age, gender and vehicle type using the Principal Component Analysis data mining technique, thereby identifying heretofore unrealised groupings amongst the crash data including age- and gender-based patterns. Wong and Chung (2007) utilised rough set theory to examine the contributing factors to 2,316 car crashes in Taiwan in 2003. This approach identified patterns amongst the driver, journey, behaviour, environment and crash characteristics reported in single-vehicle car crashes. Nayak et al. (2011) used data mining techniques to build a road crash proneness prediction model, demonstrating that road segments with only a few crashes (<8 crashes) have more in common with non-crash roads than roads with higher crash counts. Montella et al. (2011) used classification trees and rules discovery to provide meaningful insights for crashes involving powered two-wheelers in Italy. It is noteworthy however that existing approaches to identify contributing factors involve numerical data only, and also consider factors in isolation rather than considering the interrelationships between driver, environment and vehicle. Gao et al. (2013) used syntactic and semantic units in text, such as verbs, to improve crash classification and understanding of crash causation. Importantly, data mining techniques are capable of rapidly processing large amounts of information, and are underutilised techniques not only for identifying crash risk factors (Krishnaswamy et al., 2005) but also for identifying factors contributing to the severity of road-curve crashes in particular.

1.3 Rationale and aims

Existing countermeasures have attempted to address crash risk factors identified by government authorities. In Queensland, this has resulted in a dearth of information pertaining to contributors to road-curve severity for incidents costing less than AUD$2500. Insurance crash records can be a unique source of the contributing factors to the severity of road-curve crashes.

This is a novel approach that can potentially provide rich information regarding the crash, primarily because the driver is required to provide a text narrative description of the crash to the insurer. Importantly, this narrative description is likely to differ from the report provided to the Police for a number of reasons. This is not only because there will be a lack of punitive legal consequences (such as fines and demerits), but also because the driver is likely to provide additional information that is not pertinent to and/or collected in Police reports (which collect information regarding occurrence, witnesses, incident scene and event, crash description, unit details, towed vehicles, damaged property, person killed or injured, and versions, within a specific scope; to illustrate ‘atmospheric conditions’ are categorised as the following options only: clear, raining, smoke/dust, fog). In addition, road safety researchers traditionally have not had access to this information, relying instead on summary statistics provided by Police and licensing authorities. It is vital that contributors to road-curve crashes, particularly those road-curve crashes of increased cost (and therefore increased severity), be identified so that existing road-curve crash countermeasures can be enhanced as necessary and tailored countermeasures can be specifically designed. This paper presents a novel methodology based on text mining techniques (in which narrative text descriptions of crashes are analysed) to explore the factors contributing to the severity of road-curve crashes. The major contributing factors to road-curve crashes were identified and the relationships between these factors examined using text data mining of narratives contained in an insurance company crash database. This is a novel approach as contributors to severity of car crashes have not previously been text mined from insurance data.

2. Method

Figure 1 outlines the different data analysis steps. The research used curve-related insurance claim records stored in their crash database (*input*) pertaining to crashes occurring between January 2003 and December 2005 to identify the contributing factors to the severity of road-curve crashes. Curve-related information was retrieved from a larger dataset of all the insurance claim records (*identify factors*). The total cost incurred as property damage for all parties involved in the crash served as an indicator of the severity of the crash. It was assumed that the financial cost of the crash is related to the crash severity, and that a high (low) cost indicates a high (low) crash severity. It is also acknowledged that, whilst a variety of factors can contribute to the severity of crashes on road curves as discussed in 1.1 and 1.3, this information may not be recorded in the insurance reports.

[Insert Figure 1 here]

The output of the analytic process (Figure 1) includes the *identification of crash factors*, the *relationships between these factors*, and the *significant attributes predicting crashes*. The narrative crash description provided by the insurance company initially underwent data cleaning which involved discarding duplicate records and filtering for curve-related crashes only.

2.1 Data

The insurance crash report summarises information regarding the driver, the vehicle, and an unstructured narrative text description of the crash. Driver attributes include gender, age, and whether alcohol had been consumed prior to the journey (*yes*, *no*). The vehicle’s year of manufacture is recorded, the time, date, and type of crash (e.g., *hit*, *roll*), as is the number of parties involved, and an unstructured narrative description of the crash. A total of 6,011 curve-related crash records was further reduced to 3,434 curve-related crashes after removing negative cost values as it was unable to be determined if these were spurious entries. In addition, records with more than three missing values were removed as it was considered that insufficient data regarding the crash was available. The attributes were classified with semantic criteria based on intervals defined by DTMR (e.g., time of crash: *afternoon* = 1200-1600, *evening peak hour* = 1600-1900; age: *young* = 17-25 years, *mature-1* = 26-29 years, *mature-2* = 30-39, vehicle age: *new* = 0-5 years, *moderate* = 5.1-15 years, *old* = 15.1-25 years).

Text mining which can discover meaningful information by detecting lexical or linguistic usage patterns in natural language text was used to identify factors contributing to the severity of road-curve crashes (e.g., "*the road was slippery*"). Rough set analysis (Pawlak, 1995) which deals with classification and analysis of data tables (Pawlak, 1982, 1992) was used to identify the minimal subset of contributing factors to the severity of road-curve crashes and their inter-relationship(s) (a detailed description of rough set theory can be found in Annexe 1). Rough set theory is particularly advantageous in this case as it can find hidden patterns in data which cannot be otherwise identified with traditional statistical tools, evaluate the significance of the data, and facilitate the interpretation of results and generate sets of decision rules from the data.

2.2 Ward algorithm

The SAS Ward agglomerative hierarchical clustering algorithm which joins clusters of similar characteristics without increasing the overall heterogeneity (Cizek et al., 2005) was used to mine the text and form clusters of keywords. A list of keywords and the frequency each appeared in the crash description text was produced, and the most common keywords found in the text description were identified as the contributing factors to crash severity. To validate these factors as contributors in curve-related crashes only, the list of keywords from curve-related crashes was compared to the list of keywords from 11,058 non-curve-related crashes previously filtered from the data set, and keywords that appeared in both lists were discarded (*factors validation*). The remaining keywords were used as condition attributes and were represented in columns of the table prepared for rough set analysis whilst the table rows contained the claim records. We used the notion of reducts to generate a set of rules which represented predictive relations of the form expressed below.

Let S be the decision table and be defined in a pair as S = (U, A) where U is the non-empty, finite set of objects and A is a non-empty, finite set of attributes that represents the condition attributes. Rules generate with a decision and are defined as $S = (U, A \cup \{dec\})$ where U is the object, A represents the condition attributes, and $\{dec\}$ is a decision attribute and $\{dec\} \notin A$.

The rule is presented in the form $(a_{i_1} = v_1)$ □...□$(a_{i_m} = v_m)$ □ $dec = k)$ where: $1 \leq i_1 < ... < i_m \leq |A|$,

$v_i \in V_{a_i}$

Each $a \in A$ which corresponds to the function $a: U \rightarrow V_a$ and V_a is the value set of a, the evaluation function. Rosetta (a computational kernel and graphical user interface of rough set theory, Ohrn, 2001) used a supervised learning based genetic algorithm to obtain a minimal set of attributes capable of accurately and reliably predicting the crash severity group for each record and these were used to determine the relationships between the factors. Rules validation via statistical verification was conducted upon 80% of the data, with the results confirmed in the remaining 20%. The accuracy threshold of 80% (allowance +/- 10%) was supported.

To identify the significant contributing factors influencing the severity of crashes on road curves, the data was transformed into an attribute-relation file format that contained a list of instances sharing a set of attributes suitable for analysis in Weka, a Java data mining software program. The ClassifierSubsetEval algorithm which estimates the merits of a set of attributes with the RIDOR RIpple-Down Rule learner as the attribute evaluator was used. Attributes related to the driver and the vehicle only, in addition to the contributing factors identified via text mining, were used to classify each crash record. Crashes were clustered according to their cost (and therefore crash severity), and five clusters emerged: *lowest, low, medium, high,* and *highest*. It is noteworthy however that these costs comprise the cost to the insurer to repair the crash damage only and that the costs do not consider the expenses incurred through medical treatment or lost earning(s) potential of injured individual(s). In addition, this will be the first consideration of the factors involved in crashes costing less than AUD$2500, as these crashes are not required by law to be reported to Police and the Queensland licensing authority (DTMR) does not compile crash statistics pertaining to these crashes.

3. Results SAS and Ward algorithm uses clustering mechanisms and the frequency of keywords to identify contributing factors, which are summarised in Figure 1 as *tree, embankment, gravel, pole, gutter, lost control, wet road, dirt, kangaroo, truck, lost traction, fog*. As can be seen, crashes relate to both environmental (outside of the vehicle, e.g., *fog, wet road*) and driver (pertaining to driver behaviour/characteristics, e.g., *lost control, mobile*) variables.

Figure 1 also delineates between the road features (curve vs. non-curve related) of the crash. The most frequent curve-related words were *tree, embankment, gravel, pole, gutter, loss control, wet road, dirt, kangaroo, truck, lost traction,* and *fog,* suggesting these variables are most influential in the Insured's curve-related crashes. The factors specifically relate to curve-related crashes also include driver (e.g., *over-steer*) and environmental (e.g., *puddle*) factors. Importantly factors involved in curve-related crashes differ from those reported by the Queensland licensing authority which reports the following characteristics as contributors to road curve crashes: *wet road, inattention, inexperience, alcohol, speed, fatigue, fail to give way or stop, illegal manoeuvre* (Chen et al., 2006). Key heretofore unrecognised factors specific to road curve crashes as depicted in Figure 1 include tree, lost traction, fog, mountain, phone, slippery, puddle, and over-steer.

[Insert Figure 1 here]

The aim of the research was to identify factors pertaining to curve-related crashes only, therefore factor validation comparing the keywords obtained for curve-related crashes against the factors identified as pertaining to non-curve-related crashes. These contributors were found to differ (see Figure 1), suggesting that the data mining process was effective in identifying unique contributors specific to curve-related crashes. Whilst unsurprisingly a number of factors contributed to both curve- and non-curve-related crashes (*embankment, gravel, pole, lost control, wet road, gutter, kangaroo, truck* and *dirt*), key differences are apparent between the two crash types.

Rough set analysis designed to extract consistent and optimal decision rules was undertaken based on the cost/crash severity levels depicted in Table 2 (e.g., AUD$0.00 – AUD$2499.00 = lowest cost/severity; AUD$57,076.37 – AUD$77,216.36 = highest cost/severity). These four levels of severity were obtained with rough set clustering. More than twelve hundred rules which were generated through the rough set analysis were filtered using

Rosetta based on quality with the G2 likelihood algorithm (Smyth & Goodman, 1990). Rosetta provided a support count as a measure of the strength of the rule, and the relative strength is calculated by dividing the support count by the total attributes and multiplying by 100. As such, strong rules were deemed to have an appropriate combination of support and accuracy characteristics (Koperski and Han, 1995); and the higher the support count, the higher the strength. The strongest rules comprised a combination of the time of the crash (morning peak hour 0600-0900, morning 0900-1200, afternoon 1200-1400, evening peak hour 1400-1700, evening 1700-2400, night 2400-0600); age of the vehicle; driver age (17-25, 26-29, 30-39, 40-49, 50-59, 60-100 years); alcohol involved (yes/no), crash outcome (hit, collision, roll, skid, collide), and the risk factors identified in the text mining (wet, gravel, kangaroo, gutter).

Statistical analysis measurement supported in rough set analysis software was used to validate the accuracy of the rules, using the accuracy threshold of 70% ±10% and the coverage of the data. In contrast to traditional statistical techniques which utilise predefined thresholds of ‘statistical signficance’, and consistent with data mining validation practices, the validation was carried out by dividing randomly the dataset into 20% and 80%. As a form of validation, the rules generated from 80% of the data set are applied to the 20% to establish their accuracy. The accuracy measurement validation of rules is summarised in Table 1, and the results obtained from text mining were validated (akin to the testing of ‘statistical significance’) to verify that the factors obtained are only related to crashes on road curves. This process involved the comparison of the contributing factors obtained for curve- related crashes against the ones for non-curve-related crashes. Thus, out of 11,058 crash records, 6011 curve related records were mined and the results were compared against non-curve-related mining results. As can be seen in Table 1, notwithstanding the low coverage and accuracy rate for the greater severity crashes, the overall classification accuracy obtained was 63.3% with a 54.5% coverage, which was deemed to be acceptable consistent with standard data mining practice.

[insert Table 1]

Table 2 describes the driver age and gender, alcohol-involvement, time of day, vehicle age, and type of crash (rollover, hit object) for the strongest rules with the highest support count for each of the five crash severity categories. The lowest severity crash – and accordingly the lowest claim cost – was most likely to involve a female driver aged 30-59 years, driving a car which was less than 15 years old in the morning. Peak traffic times may mean that the higher traffic volume is travelling at a lower speed, hence the lower severity. In comparison, the highest severity crash – and therefore the greatest claim cost - was equally likely to involve a young male driver who had been drinking and who rolled their car in the late morning, and an older female driver whose vehicle hit a tree in the early afternoon, the former possibly reflecting lag effects of drinking alcohol the night before the crash-involved drive, whilst the latter possibly reflects the circadian lull in alertness and commensurate increased susceptibility to distraction. Male drivers featured predominantly in the remaining crash severities (and costs) of low, medium and high, consistent with their greater driving exposure and overrepresentation in traffic crashes and fatalities in general; and newer cars featured in medium and high crash severities (and costs) which may reflect the greater expense in repairing newer, and therefore frequently more expensive, cars.

[Insert Table 2 here]

It is noteworthy that these factors differ from the understandings of contributing factors previously reported in the contribution factors summarised by the Queensland licensing authority. In addition, the Queensland licensing authority does not consider interrelationship(s) amongst these factors. Characteristics pertaining to the environment, the road, the driver and the vehicle have been found to contribute to road-curve crashes. It is unlikely that the influence of these variables operates in isolation, and research examining the contributors to road-curve crashes typically does not consider the relationships between these factors. In Queensland, Australia, the Police are required to attend every road crash in which more than AUD$2500 damage is sustained by any vehicle, any persons involved is injured or killed, or if a vehicle requires towing from the crash scene. The Police complete a traffic incident report that documents crash circumstances regarding environmental conditions such as the weather and time of day, road features such as traffic control and vertical and horizontal road alignment, and driver and vehicle conditions. The Police could interview road users involved in a crash, when it is appropriate, however their narratives are often not properly analysed. DTMR identifies key contributors to road-curve crashes from the traffic incident report, and these contributors subsequently inform countermeasure development and implementation.

4. Discussion

The research utilised a novel approach to understanding the contributors and the relationship(s) between the contributors to the severity of road-curve crashes, and also allowed an investigation into the contributing factors influencing the road-curve crash severity. Road safety research has not previously analysed narrative text with text mining techniques such as rough set theory and the ward clustering algorithm. Current information sourced from licensing and government authorities relies upon third-party crash assessments (conducted by the Police) and does not consider the experience of the crash as narrated by the driver. In addition, the crash-involved driver may report different or biased crash circumstances to the Police (who can fine the driver for risky behaviour which contributed to the crash) and to the insurer (who will approve a monetary claim for repair or replacement of the crashed vehicle). As can be seen by the text mining, current and new contributing factors to road-curve crash severity have been identified, and of particular interest to road safety researchers is the contribution of ‘phone’ and ‘oversteer’. Further, attention to countermeasure design, implementation and evaluation can be directed towards the domain of interest, be it the frequency of particular contributors, or the severity of the crash according to the insurance cost. It is also noteworthy that the relationships between contributing factors has heretofore been unexplored, and a consideration of these can allow a more in-depth and accurate understanding of road-curve crashes. Accordingly the efficient and cost-effective research has implications for traffic rule enforcement, countermeasure development and implementation, insurers, and road safety researchers alike.

Traffic rule enforcement could consider targeting speeding behaviour of drivers, particularly at entries to and on road-curves. A minimal subset of contributing factors and their combination are particularly suited to further application in road-curve crash countermeasures such as in-vehicle monitoring devices, and the verified contributors can generate more accurate predictions of crash likelihood. Identifying significant contributing factors to road-curve crashes and their interrelationships also appears to influence crash severity and indicated by crash costs, and this information is of interest to insurance companies particularly in assessing crash claims and in determining policy premiums of customers. Road safety researchers could consider developing and trialling countermeasures targeting driver error such as warning signs and education campaigns highlighting the crash risks associated with road-curves, in addition to highlighting the factors found to increase the severity of road-curve crashes.

It is noteworthy however that only limited exposure information is gathered in the insurance claim (for example, the characteristics of the curve and or any line markings where the crash occurred), therefore the potential contribution of the type of curve is unknown. In the instance of severe crashes, Police collect such road information however this is not customarily made available to insurers. In addition it is unlikely that the claimant driver will know the curve characteristics, let alone report these to the insurer as part of their crash narrative. Whilst privacy considerations prevented the linkage of driver crash records in the current research, future research could operationalise data linkage and data mine the characteristics of the road curve collected by Police and the insurer which could be used not only to validate the current research findings but to further clarify the nature and mechanisms of the contributors to road curve crashes. In addition, whilst the frequency count of the identified contributors suggests their importance in curve-related crashes, analyses could augment data mining by using statistical techniques like odds ratios to elucidate the magnitude of the influential variables.

Further, approximately half of crash-related claims were removed prior to analyses as the veracity of the data could not be established. Whilst as researchers we acknowledge that narratives may not be collected, stored nor maintained by insurance companies for research purposes, however we suggest that insurers may need to improve their quality control measures to assist in crash-related research which may ultimately guide insurer policy and practices. Future research should source larger data files from other insurance providers in Queensland and

Australia to further inform our understanding of contributors to road curve crashes. The crash cost is a composite measure and therefore the cost incurred by any individual involved in a multiple vehicle crash is unable to be determined. As noted earlier, these costs also do not consider expenses incurred through medical treatment or lost earning(s) potential of injured individual(s). In addition, narrative crash descriptions are provided by the insured claimant, therefore self-presentation biases may be evident in their representation of the crash circumstances. In particular, claimants may minimise – or may not even report – their behaviour(s) which may have contributed to the crash, such as driver fatigue, alcohol consumption, and speeding. In addition, a further potential bias is that the crash-involved driver may be likely to tell a story to their insurer that maximises the likelihood that they will have their claim approved. However such reporting is not as constraining as reports to Police (with potential consequences such as incarceration).It is noteworthy also that the actual role of the attribute (such as 'tree' or 'ditch') is unclear in the analyses. That is, whether the attribute was actively involved in the crash, for example did the driver lose control of the vehicle and collide with the tree, or whether the attribute was simply involved as the end point of the crash, for example the driver and their vehicle came to rest in a ditch. Whilst the role of the attribute will also be of interest for intervention, the fact that the attribute was identified as an attribute in a curve-related crash – and that some attributes actively increased the severity of the crash (as indicated by crash cost) – suggests that broad interventions be developed which recognise their role in curve-related crashes (for example, where possible removal of trees in the vicinity of road curves, particularly if a high volume of crashes are identified, may correspond to a reduction in crash severity). Interestingly the crash type (e.g., rear-end) did not emerge as a significant variable. Future research targeting specific crash types may select these crashes only, identifying influential variables which can be validated by comparing the emergent list with variables identified through data mining all other crash types.

Moreover, no or limited crash information is available for fatal crashes, and arguably injury and fatality crashes are of most interest to road safety researchers, licensing authorities and insurers alike. Data linkage between police, the fatally-injured insured driver, and narratives provided by any witnesses could also enhance our understanding of curve-related crashes. Using past crash data to determine crash severity requires continual re-evaluation of contributing factors as new vehicle technology, enhanced curve engineering and targeted countermeasures are developed and implemented. The rough set analysis produced multiple rules for each crash severity, similar to findings of other data mining research (e.g., Wong and Chung, 2007); however validation facilitated the ranking of these rules by support counts.

It is also noteworthy that generally in Australia, crashes are categorised as *fatal, hospitalisation, medical treatment, minor injury*, and *property damage only*. Whilst arguably fatal crashes ‘cost’ the most (particularly when the costs of emergency services personnel and lost wages are considered), it is important to note that it is not unheard of for a driver to emerge relatively unscathed from a crash in which their vehicle is written off (and therefore there is a high cost for the Insurer). Conversely, serious injuries can be incurred in crashes which result in very little vehicle damage (and therefore low cost to the Insurer). Thus the relationship between crash severity and crash cost, for the Insurer, is not necessarily linear. Future research projects may wish to apply data mining to narratives only for crashes in which the driver and/or other vehicle occupants reportedly was injured (of course notwithstanding that the fatally-injured driver would be unable to contribute a narrative), the comparison of which to non-injury curve-related crashes potentially allowing identification of factors associated with curve-related crashes considered more serious.

A wealth of potential future research emerges from the current research. The research explored the narrative crash claims of one insurer in Queensland only, and the findings could be further strengthened by sourcing additional claim information from other insurers. There is however no reason to suspect that the circumstances of the road-curve crash claimed through this insurer are unique. The curve characteristics could be examined for each claim, particularly as insurers frequently require the claimant to provide the address of the incident. Curves at crash locations could be assessed for features such as sharpness and elevation. This would allow consideration of the influence of curve characteristics, and potential interrelationships with other identified contributing factors. Research could also focus on individual curves or curve type(s) which have an extraordinarily high volume of crashes. Specific details regarding the identified keywords could be further examined to determine the nature of their contribution to curve crash severity and the development of countermeasures; for example, identifying persistent trends in the location drivers ‘lost traction’ in the curve may result in improved road design. Text data mining techniques could be applied to insurance claims for other crash types, such as at roundabouts and in identified black-spots in which a high number of crashes occurred within a comparatively short period of time. In addition, techniques such as multinomial or ordered models used in conjunction with text data mining may produce more robust contributions to improving road-curve safety.

5. Conclusions

The application of existing and robust text mining techniques to the claim narratives of 3,434 road-curve crashes in Queensland of one major insurer has revealed heretofore undiscovered contributing factors to crash severity of considerable importance to road engineers, road safety researchers, and policy-makers alike (e.g., phone, oversteer). To our knowledge, this is the first time that Rough set theory has been used to mine crash data. It is unlikely the new factors identified by our text mining approach could have been identified using traditional statistical tools. Importantly, the relationship between these factors and crash severity as indicated by cost of the claim was also investigated. The research approach proved to be valuable for the purpose of discovering both knowledge and relationships from a road crash database, and the results have great potential in assisting the understanding of safety issues related specifically to road curves when used in conjunction with other traditional forms of crash data analysis. Further, the techniques could be applied within other safety domains and may reveal heretofore unrealised contributors to incidents and accidents. A number of limitations were associated with the data, including a lack of information regarding the curve at the site of the crash and possible self-presentation biases by the claimant, and a wealth of future research including improved quality of and linkage between data sources is suggested by the research findings and application of the unique methodology.

References

Aarts, L., van Schagen, I. 2006. Driving speed and risk of road crashes: A review. Accident Analysis and Prevention, 38, 215-224.

Aldridge, C. H. 2001. A rough set based methodology for geographic knowledge discovery. In Proceedings of the 6th International Conference on GeoComputation, Brisbane Australia, 24-26 September 2001.

Bai, Y., Finger, K., Li, Y. 2010. Analyzing motorists' responses to temporary signage in highway work zones. Safety Science, 48, 215-221.

Blows, S., Ivers, R. Q., Woodward, M., Connor, J., Ameratunga, A., Norton, R. 2003. Vehicle year and the risk of car crash injury. Injury Prevention, 9, 353-356.

Broyles, R. W., Clarke, S. R., Narine, L., Baker, D. R. 2001. Factors contributing to the amount of vehicular damage resulting from collisions between four-wheel drive vehicles and passenger cars. Accident Analysis and Prevention, 33, 673-678.

Caliendo, C.,Guida, M., Parisi, A. 2007.A crash-prediction model for multilane roads. Accident Analysis and Prevention, 39, 657-670.

Chen, S. H. 2010. Mining patterns and factors contributing to crash severity on road curves. PhD thesis, Queensland University of Technology, http://eprints.qut.edu.au/31711/

Chen, H. Y., Ivers, R. Q., Martiniuk, A. L., Boufous, A., Senserrick, T., Woodward, M., Stephenson, M., Williamson, A., Norton, R. 2010. Socioeconomic status and risk of car crash injury, independent of place of residence and driving exposure: Results from the DRIVE study. Journal of Epidemiology and Community Health, 64, 998-1003.

Chen, S., Rakotonirainy, A., Sheehan, M., Krishnaswamy, S., Loke, S. W. 2006. Assessing Crash Risks on Curves. In *Proceedings Australian Road Safety Research, Policing and Education Conference*, Gold Coast, Queensland.

Cizek, P., Hrdle, W., Weron, R. 2005.Statistical tools for finance and insurance: Cluster algorithms. Retrieved 30 January 2009 from http://fedc.wiwi.hu-berlin.de/xplore/ebooks/html/stf/.

Das, A., Abdel-Aty, M. 2010. A genetic programming approach to explore the crash severity on multi-lane roads. Accident Analysis and Prevention, 42, 548-557.

Department of Transport and Main Roads (DTMR). 2011. Crashes within Queensland by crash severity by horizontal alignment, 1 July 2008 to 30 June 2009. DTMR: Brisbane.

Dey, L., Ahmad, A., Kumar, S. 2005. Finding interesting rules exploiting rough memberships. Lecture Notes in Computer Science, 3776/2005, 732-737.

Duntsch, I., Gediga, G. 2000. Rough set data analysis: A road to non-invasive knowledge discovery. Retrieved 1 January 2008 from http://bib.tiera.ru/dvd56/Duntsch%20I.,%20Gediga%20G.%20%20Rough%20set%20data%20analysis.%20A%20road%20to%20noninvasive%20knowledge%20discovery(2000)(107).pdf

Gao, L., Wu, H., 2013 Verb Based Text Mining of road crash report. Proceedings 92nd Annual Meeting of the Transportation Research Board. Washington 2013.

Gnardellis, C., Tzamalouka, G., Papadakaki, M., Chliaoutakis, J. E. 2008. An investigation of the effect of sleepiness, drowsy driving, and lifestyle on vehicle crashes. Transportation Research Part F: Traffic Psychology and Behaviour, 11, 270-281.

Grunewald, P. Ponicki, E. 1995. The relationship of the retail availability of alcohol and alcohol sales to alcohol-related traffic crashes. Accident Analysis and Prevention, 27(2), 249-259.

Hand, D. J., Manilla, H., Smyth, P. 2001. Principles of data mining. MIT Press: Cambridge, Mass.

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. In Proceedings of SIAM International Conference on Data Mining, California.

Ishigami, Y., Klein, R. M. 2009. Is a hands-free phone safer than a handheld phone? Journal of Safety Research, 40, 157-164.

Jones, I. S., Stein, H. S. 1981. Defective equipment and tractor-trailer crash involvement. Accident Analysis and Prevention, 21, 469-481.

Keall, M., Frith, W. 2004. Older driver crash types in relation to type and quantity of travel. Traffic Injury Prevention, 5, 26-36.

Kloeden, C. N., Ponte, G., McLean, A. J. 2001. Travelling speed and the risk of crash involvement on rural roads. The University of Adelaide Report CR 204.

Komorowski, J. Pawlak, Z. Polkowski, L. Skowron, A. 1998. Rough sets: A tutorial. Rough fuzzy hybridization: A new trend in decision making. Eds: S. K. Pal and A. Skowron, Springer Verlag, pp. 3-98.

Koperski, K., Han, J. 1995. Discovery of spatial association rules in geographic information databases. In Proceedings of the Symposium on Large Spatial Database – SSD, 6-9 August, Portland, ME.

Krishnaswamy, S., Loke, S. W., Rakotonirainy, A., Horovitz, O., Gaber, M. M. 2005.Towards situation-awareness and ubiquitous data mining for road safety: Rationale and architecture for a compelling application. In Proceedings of Conference on Intelligent Vehicles and Road Infrastructure, University of Melbourne.

Kuhlmann, S., Ralf-Michael, V., Lubbing, C., Clemens-August, T. 2005.Data mining on crash simulation data. Machine Learning and Data Mining in Pattern Recognition, 3587/2005, 558-269.

Lam, L. T. 2002. Distractions and the risk of car crash injury: The effect of drivers' age. Journal of Safety Research, 33, 411-419.

Liu, C., Chen, C. L., Subramanian, R., Utter, D. 2005. Analysis of speeding-related fatal motor vehicle traffic crashes. NHTSA Technical Report DOT HS 809 839. NHTSA. Washington.

Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F. 2011.Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery, Accident Analysis and Prevention, In Press, Corrected Proof.

McCartt, A. T., Northrup, V. S., Retting, R. A. 2004.Types and characteristics of ramp-related crashes on urban interstate roadways in Northern Virginia. Insurance Institute for Highway Safety: Arlington.

Nayak, R., Emerson, D., Weligamage, J., Piyatrapoomi, N. 2011. Road crash proneness prediction using data mining. In A. Ailamaki, S. Emer-Yahia (Eds.), Proceedings of the 14th International Conference on Extending Database Technology, Association for Computing Machinery (ACM), 21-25 March 2011, Uppsala, Sweden., 521-526.

Nguyen, S. H., Nguyen, H. S. 2003. Analysis of STULONG data by rough set exploration system (RSES). Technical Report. PKDD/ECML Discovery Challenge.

Ohrn, A. 2001. ROSETTA Technical Reference Manual. Trondheim, Norway: Norwegian University of Science and Technology.

Othman, S., Thomson, R., Lanner, G. 2010. Are driving and overtaking on right curves more dangerous than on left curves? Annals of Advances in Automotive Medicine, 54, 253-264.

Pande, A., Abdel-Aty, M. 2006. Assessment of freeway traffic parameters leading to lane-change related collisions. Accident Analysis and Prevention, 38, 936-948.

Pawlak, Z. 1982. Rough sets. International Journal of Computer and Information Sciences, 11, 413-433.

Pawlak, Z. 1992. Rough sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, London, UK.

Pawlak, Z. 1995.Rough sets. In ACM Conference on Computer Science, 262-264.

Pawlak, Z.,Grzymala-Busse. J.,Slowinski. R., Ziarko, W. 1995. Rough sets: An emerging technology November 1995/Vol. 38, No. 11 Communications of the ACM.

Queensland Transport. 2006.Webcrash 2.3. Queensland Transport, Brisbane.

Retting, R. A., Williams, J., Schwartz, S. I. 2000. Motor vehicle crashes on bridges and countermeasure opportunities. Journal of Safety Research, 31, 203-210.

Rosey, F., Auberlet, J. M., Bertrand, J., Plainchault, P. 2008. Impact of perceptual treatments on lateral control during driving on crest vertical curves: A driving simulator study. Accident Analysis and Prevention, 40, 1513-1523.

Rowden, P., Steinhardt, D., Sheehan, M. 2008. Road crashes involving animals in Australia. Accident Analysis and Prevention, 40, 1865-1871.

Salim, F. D., Krishnaswamy, S.,Loke, S. W., Rakotonirainy, A. 2005. Context-aware ubiquitous data mining based agent model for intersection safety. Embedded and Ubiquitous Computing – EUC, 61-70.

Scott-Parker, B., Watson, B., King, M. J. 2009.Understanding the influence of parents and peers upon the risky behaviour of young drivers. Transportation Research Part F, 12, 470-482.

Seneviratne, P. N., Islam, M. N. 1994. Optimum curvature for simple horizontal curves. Journal of Transportation Engineering, 120, 773-786.

Shields, B., Morris, A., Barnes, J. S., Fildes, B. 2001. Australia's national crash in-depth study progress report. In Road Safety Research, Policing and Education Conference, 19-20 November 2001, Melbourne Australia.

Singh, S. 2001. Identification of driver and vehicle characteristics through data mining the highway crash data. Retrieved 12 April 2011 from http://www.fcsm.gov/03papers/Singh8c.pdf

Skowron, A. Rauszer, C. 1992. The discernibility matrices and functions in information systems. In Intelligent decision support - Handbook of applications and advances of Rough Set Theory. Ed. R. Slowinski. Kluwer Academic Publishers, pp. 331-362.

Slowinski, R. 1992. Intelligent decision support - Handbook of applications and advances of the Rough Sets Theory. Kluwer Academic Publishers.

Slowinski, R., Stefanowski, J. 1993. Special issue on rough sets state of the art and perspectives. Foundations of Computing and Decision Sciences, 18, 3-4.

Smyth, P., Goodman, R. 1990. Rule induction using information theory. In G. Piarersky and W. Frawley (Eds.), *Knowledge discovery in databases (Chapter Nine)* .MIT Press.

Torbic, D. J., Harwood, D. W., Gilmore, D. K., Pfefer, R., Newman, T. R., Slack, K. L., Hardy, K. K. 2004.A guide for reducing collisions on horizontal curves. Technical Report NCHRP Report 500, Vol 7. National Cooperative Highway Research Program, Texas.

Wong, J. T., Chung, Y. S. 2007. Rough set approach for accident chains exploration. Accident Analysis and Prevention, 39, 629-637.

Zegeer, C., Stewart, R., Reinfurt, D., Council, F., Newman, E., Hamilton, E., Miller, T., Hunter, W. 1991. Cost-effectiveness of improvements for safety upgrading of horizontal curves. Report No. FHWA-RD-90-021.Federal Highway Administration.

Annexe 1

Rough sets are particularly suitable for handling uncertainty in data which may be caused by missing or noisy data or due to ambiguity in the semantics of data. When handling such data, rough sets produce an inexact or "rough" classification. The concept of approximation space provides the boundaries for classifying objects. Rough sets use two concepts known as *Upper Approximation Space* and *Lower Approximation Space.* The lower approximation of a concept (or class) consists of all objects that definitely belong to that concept and the upper approximation consists of all objects that possibly belong to the concept in question. The objects that fall between the upper and lower approximation spaces (which is also called the *boundary region*) are in the area of uncertainty or rough classification. Rough sets have been widely used in several application domains (Pawlak, 1992, Slowinski, 1992; Slowinski and Stefanowski, 1993) for rule generation, attribute reduction and prediction. A distinctive feature of rough sets is that it is known to be very effective in identifying features of data that are important for predictive accuracy, as well as dealing with noise/missing data (which is a known characteristic for most real-world data).

A.1 Rough Sets Terminology We use terminology and notation that is consistent with the seminal literature in rough sets (Pawlack, 1992, Komorowski et al., 1998; Skowron and Rauszer, 1992).

A.1.1 Information System (IS). A data set in rough sets is represented as a table called *Information System (IS)*, where each row is an object and each column is an attribute. The attributes are typically divided into *condition attributes* (the determinant attributes used to make the predictions), and the *decision attributes* (the attributes that need to be predicted). Each object of the Universe has attributes and attribute values associated with it. The attributes remain the same for all objects but the attribute values may differ.

A.1.2. Indiscernibility Relation. Any two objects in an IS are said to be 'indiscernible' if, for a given set of attributes, they have the same values. That is, the objects are not distinguishable on the basis of the values in the attributes under consideration. Thus, objects 1 and 5 are indiscernible with respect to the attribute set {a, b, c} and belong to one equivalence class. The set of all equivalence classes is termed as the *indiscernibility relation*.

A.1.3. Dispensability. For a given IS, an attribute *a* is said to be dispensable or superfluous if, after removal of that attribute, we are still able to get the same set of equivalence classes (i.e. *indiscernibility relation)* without the attribute under consideration. Thus, the particular attribute under consideration is not essential for determining the relationship between the condition and decision attributes., and those attributes are said to be dispensable or superfluous.

A.1.4. Approximation Space. This is a central concept in the rough set model and is the basis for its ability to deal with vagueness and uncertainty. The concept of approximation space provides the boundaries for classifying objects. The lower approximation comprises those objects that can be classified with certainty as elements that fall within a set of attributes under consideration. Correspondingly, the upper approximation comprises those objects that can be classified as possibly being elements of the sub-set of attributes under consideration (i.e. they can neither be accepted nor rejected with certainty). An object is a strong member if it is part of the lower approximation and a weak member if it is part of the boundary region.

A.1.5. Reduct. A reduct of is a *minimal* subset of attributes such that all attributes in the reduct are indispensable. Thus, if one attribute is removed from this subset, it will change the equivalence classes and the indiscernibility relations between the attributes being studied and/ or considered. Therefore, the reduct of a set of attributes contains only non-superfluous attributes and further maintains the indiscernibility relation between the original attribute subset and itself (i.e. the reduct). There can be several reducts for a given subset B. While it is relatively simple to compute a single reduct, the general solution for finding all reducts is NP-complex.

A.1.6 Core. The core is a set of attributes that is common to all the reducts of an IS and can be computed in a straightforward manner using a tabular representation concept developed by Skowron and Rauszer (1992) known as the *Discernibility Matrix.* Both the core and the reducts in an information system can be computed relative to a given set of attributes.

A.1.7 Degree of Dependency. Rough Sets also provides mathematical constructs to measure the degree of dependency between two sets of attributes, like the condition and decision attributes. The degree of dependency is represented as a value between [0, 1], and the higher the value of k, the greater is the dependency between the two sets of attributes.

A.1.8 Significance of Attributes. The significance of an individual attribute *a* is computed using the degree of dependency. By measuring the change in the degree of dependency between the condition and decision attributes, through the inclusion and removal of the attribute whose significance is being measured, it is possible to determine the importance or significance of that particular attribute.

Figure 1

Summary of crash factors contributing to crashes reported to a Queensland, Australia, insurance company between 1 January 2003 and 31 December 2005.

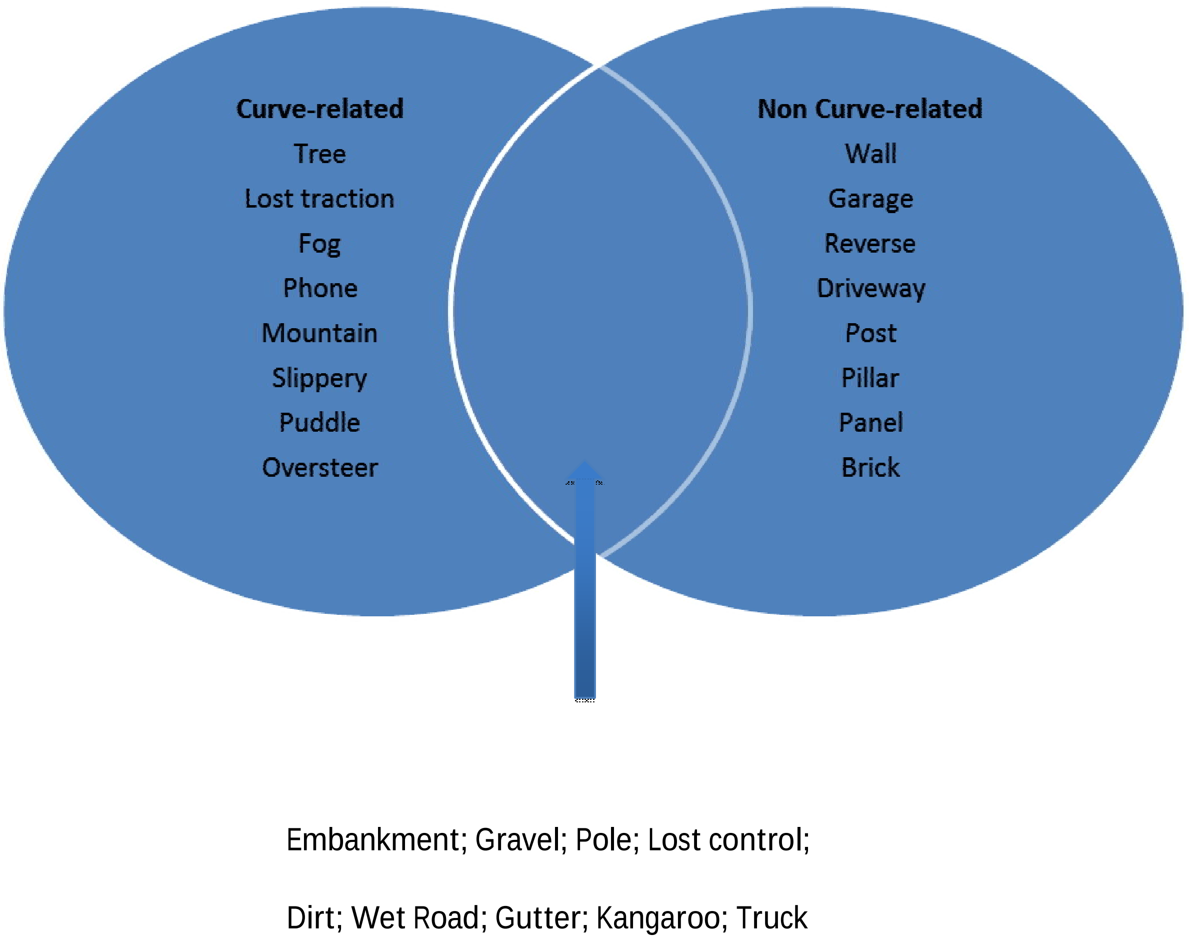

Table 1

Accuracy measurement validation of rules

	Predicted		
Severity	No. of objects	Accuracy	Coverage
Very low	933	0.601	0.546
Low	1634	0.674	0.558
Medium	153	0.437	0.474
High	25	0.286	0.280
Very high	0	0	0

Total number of objects: 2747

Total accuracy: 0.636

Total coverage: 0.545

Table 2

Summary of crash severity and contributors in curve-related road crashes, Queensland, Australia, 1 January 2003 – 31 December 2005

Crash severity	Driver	Driver	Alcohol	Time	Vehicle
Rollover	Hit	Collision			
	Age	Gender	Involved	of Day	Age
Crash	Object				
Lowest (<$2,500)	30-59	Female	No	0600-1200	1-15 years
No	No	Yes			
Low ($2,500.00					
– $16762.46)	26-29, 50-59	Male	No	0600-1200	1-15 years
No	Yes	Yes			
Medium ($16762.47					
– $38606.94)	26-39	Male	No	0900-1200	0-5 years
Yes	No	Yes			
High ($38606.95					
–$57076.36)	30-39	Male	Yes	0600-0900	0-5 years
No	Yes	Yes			
				1200-1600	
Highest ($57076.37					

–$77216.36)	17-25	Male	Yes	0900-1200	1-15 years
No	Tree	Yes			
	50-59	Female	No	1200-1600	0-5 years
Yes	No	Yes			