

Attitudes to road safety and its implications for public policy

Chirag Pathak¹, Daniel Gregg², Dr. David Gargett³

^{1,2,3}Department of Infrastructure and Transport, BITRE

Email for correspondence: Daniel.Gregg@infrastructure.gov.au or
David.Gargett@infrastructure.gov.au

Abstract

This paper provides an analysis of past survey data on community attitudes to safety, and investigates the following question: Can public policy impact community attitudes to safety? This is done by analysing the attitudes to mobile phone usage among a range of demographics (for example age, sex, and education) over a number of years and identifying high-risk groups. The conclusions should enable policy makers to understand how policy can better target high-risk groups. These high-risk groups are then recommended as a target for policy makers with a range of possible policy solutions. A unique approach of this paper is the use of two different modelling methods for data analysis. A multinomial regression model and an ensemble classifier based on regression trees are used to identify high-risk groups. Both approaches provide unique insights into the data. The comparison of the two models may be of particular interest to researchers, since it demonstrates their relative advantages in predictive capacity and in data handling.

1. Introduction

In 2011 the Australian Transport Council released the National Road Safety Strategy (NRSS) for the period 2011-2020. The annual cost of road crashes in Australia is high, estimated at 27 billion dollars, with over 32 thousand people admitted to the hospital each year. The goal of the policy is a reduction of road crash fatalities and serious injuries by 30% each by the end of 2020. NRSS describes mobile phone usage as a widely recognised form of distraction and states that there is a range of evidence linking mobile phone usage to increased casualty crash risk this is further confirmed by numerous independent studies (Adolph, 2010 and Horrey, Lesch, & Garabet, 2008). Currently the number of crashes due to the use of mobile phone while driving is not very high when compared with the overall crash statistics. However the uptake in mobile usage is increasing and this may result in an increase in mobile-related road fatalities. It is difficult to estimate the exact number of road crashes in Australia that are due to the use of mobile phones while driving. However, it is an area that will be targeted by policy makers in future. The NRSS states:

“Further investigation is required to fully understand the safety impacts of mobile phones and other potentially distracting devices, and to inform the development of appropriate countermeasures. Any consideration of changes to existing mobile phone laws would require a thorough analysis of the potential safety benefits and other impacts on the community” (Australian Transport Council, 2011)

The results of this paper may contribute some insights and suggestions for policy options and also help in identifying sections in the community which are most vulnerable to distracted driving due to the use of mobile phones.

The paper identifies high-risk groups in the community by using a combination of statistical methods: data mining analysis (using Random Forest) and multinomial regression analysis. The two models complement each other and also serve for the purposes of cross checking

results. The research finds that the age groups of 25-39 are the highest risk group and within that age group, trades people with full-time employment are even more likely to use their mobile phone while driving. To a lesser extent, but also in the high-risk group, are people aged 15-24. Not surprisingly, the 60+ age group came under the low risk category.

The paper also looks at general attitudes towards risk-taking in sections of the community, by examining their attitudes towards issues like speeding. This analysis is carried out to determine whether certain sections of the community are “using a mobile phone while driving because they have a higher propensity for risky behaviour” or whether it is “due to a lack of understanding of the dangers of using mobile phones while driving”. We find that after controlling for risky behaviour the results of the regression are still significant. This is good news for policy makers because it informs us that many users are unaware of the dangers of using a mobile phone while driving and will benefit from strong policy action in this area (since they don’t have a general tendency for risky behaviour).

Road safety researchers may also be interested in the comparison of the two modelling methods and a description of how the two methods complement each other and their respective advantages. For example, Random Forest can handle multiple variables and large data sets with missing values while regression provides us with a general (global model) that captures the information in a single predictive formula that is assumed to hold for the entire data set. These results may encourage road safety researchers to use techniques from data mining and experiment with multiple models. The paper concludes by outlining possible policy options which are derived from the analysis.

2. Background

Distracted driving and mobile phones

The use of technology in vehicles is increasing and this is leading to an increase in distracted driving. Driver distractions can broadly be classified into external and in-car distractions. Computer controlled, highly visual advertisements and billboards are an example of external distractions. In-car distractions range from the use of mobile phones to highly sophisticated vehicle controllers and GPS navigation systems.

There is no universally agreed definition of driver distraction, but there are some generally accepted rules:

A distraction is not related to impairment from alcohol or drugs.

A distraction requires a trigger event; it is not the same as inattention.

The American Automobile Foundation defines driver distraction as occurring when “a driver is delayed in the recognition of information needed to safely accomplish the driving task because some event, activity, object or person within or outside the vehicle compelled or tended to induce the driver’s shifting attention away from the driving task”. (Adolph, 2010)

The International Standards Organisation (ISO) defines driver distractions as “attention given to a non-driving related activity, typically to the detriment of driving performance”. This is the definition we use for our paper. (Pettitt, Burnett, & Stevens, 2005)

Why focus on mobile phones?

Mobile phone usage has grown rapidly in the past decade and this trend is set to continue. Survey figures from Community Attitudes towards Road Safety (CARS) indicate that 92% of people owned or used a mobile phone in 2011 (Community Attitudes to Road Safety, 2011).

This makes mobile phones one of the most recognizable types of potential driver distraction. There is a general perception that certain groups within the community may be particularly susceptible to using mobile phones while driving. We try to identify and analyse those groups in this research, which may help policy makers concentrate their efforts.

Mobile phone use while driving is an area that is also gaining interest amongst safety researchers, with numerous studies having conclusively shown that the use of mobile phones while driving results in a significant distraction for the driver. In particular it leads to visual, physical and cognitive impairment as drivers can take their eyes off the road, hands off the steering wheel or mind off the task of driving. Numerous studies have focused on particular impacts of distracted driving:

- Failure to detect traffic signals (Strayer & Johnston, 2001)
- Negative impact on drivers ability to stop the vehicle in time (Consiglio, Driscoll, Witte, & Berg, 2003)
- Increase in mental workload (McKnight & McKnight, 1993)
- Increased risk of serious accidents (through Epidemiological studies e.g. McEvoy, Stevenson, McCartt, Woodward, Haworth, & Palamara, 2005)
- General reduction in vehicle control (Stelling & Hagenzieker, 2012)

Current policy on mobile phones and driving

This is a policy area that is in a state to flux and rapidly trying to adapt to the changing environment. Current Australian laws on mobile phone usage are comparable to current best practice in the world, but there is a move towards stronger policy action to target drivers who use mobile phones while driving. Australians are not allowed to use a hand-held mobile phone while driving and use of a hands-free phone is also restricted in many states. However, research shows that even a hands-free phone can cause significant driver distraction (National Safety Council, 2010) and policy makers across the world are considering amending laws to also ban the use of hands-free phones while driving. A snapshot of current legislation in selected countries around the world is presented in Table 1.

Table 1 Driving and mobile phone use legislation across the world (OECD/International Transport Forum, 2011)

Country	Hands-free phone use	Handheld phone use
Argentina	Allowed	Allowed
Australia	In some states allowed, not allowed for learners and P-plate drivers	Not allowed
Austria	Allowed	Not allowed
Canada	Allowed	Allowed in some provinces
Denmark	Allowed	Not allowed (even for cyclists)
Finland	Allowed	Not allowed
France	Allowed	Not allowed
Germany	Allowed	Not allowed
UK	Allowed, but driver can be prosecuted if he/she is distracted while using a hands-free phone	Not allowed
US	Allowed	Not allowed in 8 states

There has been a general worldwide trend towards stricter legislation on mobile phone use while driving. Recent laws in Finland now allow for the suspension of driver's licence if they are caught using their mobile phones three times in one year, or four times in two years. Denmark is currently seeking to change laws to prevent any form of mobile phone use while driving, whether it be hand-held or hands-free. Mobile phone penetration has grown rapidly in Australia and figures from the CARS survey indicate that 92% of people surveyed owned or used a mobile phone during the 2011 period (Community Attitudes to Road Safety, 2011). Overall penetration rate for mobile phones is increasing rapidly in the Asia-Pacific and heading towards saturation at 70-90 percent of the eligible population (Figure 1).

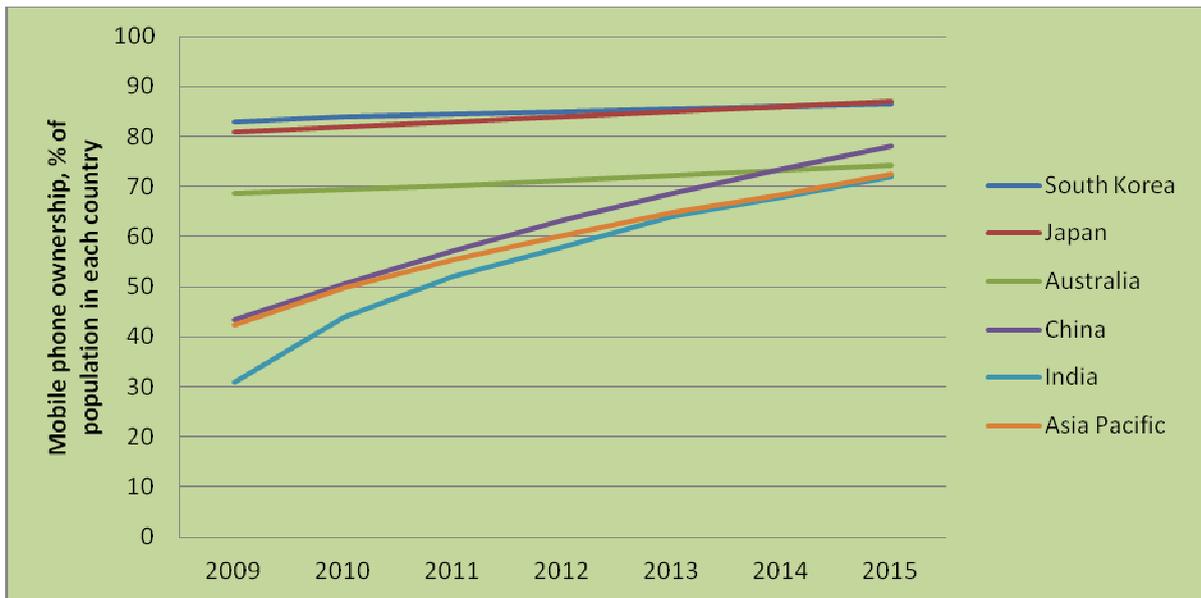


Figure 1 Mobile phone penetration rates in key Asia-Pacific countries

Australia's overall penetration rate was 69% in 2009 and is projected to reach over 74% by 2015 (eGovernment Resource Centre, 2011) In the past three surveys, approximately 60% of drivers admitted using a mobile phone while driving and in 2011 nine in ten active drivers stated that they owned a mobile phone. A recent study by the Accident Research Centre at Monash University found that 5% of drivers were observed using a mobile phone while stopped at traffic lights (Monash University Accident Research Centre, 2010).

Thus, the case for policy changes in this area is strong. This paper looks at the issue from an Australian perspective, providing insights on driver risk-profiles which may prove useful for policy makers.

3. Modelling and analysis

We used data from the Commonwealth Attitudes to Road Safety (CARS) surveys for this study. The CARS surveys are commissioned by the Department of Infrastructure and Transport and date back to the 1980's. However, data on mobile phone usage has only been collected in five surveys since 2005. These five surveys 2005, 2006, 2008, 2009 and 2011, form the basis for this study. The surveys are commissioned to monitor community attitudes to a variety of road safety issues and to suggest new areas for intervention and identify significant differences between jurisdictions.

The in-scope population for the surveys is persons 15 years and older and approximately 1500 interviews are conducted for each survey. This is done using computer assisted telephone interviews and a random digit dialling sampling frame. The average interview

length is 15 minutes and a disproportionate stratified sampling methodology is utilised to ensure adequate coverage of the population by age, sex, state/territory and capital city/other location. The response rate for the surveys has consistently been above 50%, ranging between 50 and 75%.

Each survey contains about 255 questions on a variety of topics and the focus of these surveys is not on mobile phone usage. However, we have used the data on “mobile phone usage while driving” from these surveys for our analysis. The approach taken in this research paper is very different to the goals of the CARS surveys, which present a descriptive snapshot of each year’s data. This research is a statistical analysis of the data, focusing on the question of mobile phone usage while driving. This study derives inferential statistics from multinomial regression analysis and Random Forest analysis (which is a statistical method from data mining). The paper then provides possible policy options based on the analysis results. Modelling using two different methods and across various years has meant that we have had to clean the data and standardise it, to suit our modelling and analysis goals. We have not added any new information to the original data that distorts the results and details on data cleaning are provided in the section below.

Data cleaning

We used two different models to analyse the CARS data and made certain changes to the data in order to ensure consistency and accuracy in analysis. Many questions in the original data set had too many possible responses or responses which differed from one year to another for the same variable/questions. There were also cases where the data set had possible responses which were undesirable for our analysis and we labelled those as missing. For example, if the original *age* variable had labels for “under 15” we removed those entries to keep the analysis in scope (driving age).

Some variables had too many response values (categories) as possible answers. For example, the question on whether you answer your mobile phone while driving had 8 possible response values, ranging from ‘*yes always*’ to ‘*no response*’. For the research we experimented with various scenarios for collapsing (aggregating) the data set. For example we collapsed the responses *always* and *very often* into one response category and analysed the results. Finally, we found that there is a clear difference between a definite answer of *always* and very little difference between the responses *very often* and *fairly often*. Therefore, we decided to keep categories that are very definite, separate from the indefinite categories. The collapsing as decided upon is shown in table 2.

Table 2 User responses collapsed into four categories always, often, rarely and never.

Old values						
Always =1	Very often =2	Fairly often=3	Just occ=4	Rarely=5	Never=6	Ignore=7-8
New values						
Always =1	Often=2		Rarely=3		Never=4	Ignore=7-8

Empty cells were a big problem for some variables, and this limited the number of variables we could study. In order to keep the project within scope we only used those variables that had good response labels. We conducted some analysis on attitudes to risk across categories (like attitudes to mobile phone use with regards to speeding) However, we could not do this for many other questions due to problems with missing data.

As a form of cross validation, we used mathematical techniques like series mean, to replace missing data and the results were similar to those using original data. In the main study, we only used original data and ignored missing values.

Multinomial Regression

The data provided in the CARS analysis only gives descriptive statistical analysis, providing cross tab analysis of the raw data among the different variables. This study uses inferential statistics using two different models. The first method used is the multinomial logistic regression model. This model is used when each individual has more than two alternative responses to choose from and the researcher wishes to relate these to a set of explanatory variables. (Hill, Griffiths, & Lim, 2008) In this case we had four response choices set against five explanatory variables.

Do you answer a mobile phone while driving?

As highlighted in the introduction, there are many different forms of distraction that arise with the use of a mobile phone while driving. One of the key questions from the CARS surveys asks, '*do you answer your mobile phone while driving?*' The options are, 'always', 'often', 'rarely' and 'never', which is used as the reference category.

The key category we have focussed on are those people who have responded 'always' to the question above. Using a multinomial regression model, we were able to determine which variables, representing different characteristics of individuals surveyed, would most likely influence a person to say that they 'always' answer a call while driving.

In 2005, the age category 15 to 24 years of age was found to have a significant influence on whether a respondent said that they would always answer the phone when it rang while they were driving. Those in this age group were 3.5 times more likely to respond with 'always' than those in the reference group aged 60 years and over. Similar results occurred for those aged between 25-39 years of age. However the odds ratio was approximately half that of the younger age group (1.7).

Six years later, the results varied significantly. The age bracket 15 to 24 was found not to significantly affect whether a person 'always' answers their phone while driving. Those aged 24 to 39 were likely to answer 'always' and were 6.5 times more likely to answer this way than respondents over 60 years, which is a significant increase on the results found in 2005. Another major difference was that in 2011, respondents aged 40-59 years, were found to have been more likely to answer 'always' than in prior surveys.

By conducting pooled regression analysis we are able to add time as an explanatory variable to the regression. The results show that the variables Time 3 (2008) and Time 4 (2009) were statistically significant with p-values of 0.001 and 0.043 respectively. Those surveyed were more likely to say that they always answer their phones while driving in 2008 and 2009, when compared to the reference year of 2011. Interestingly, respondents were not likely to respond 'always' in the earlier years (2005, 2006) of the CARS survey, possibly indicating the increased prevalence of mobile phone use in recent years.

When analysed by location, there are varying results between the States and Territories. The variable Western Australia consistently showed statistically significant results, indicating that by coming from Western Australia, a person has a higher likelihood of answering their phone while driving compared to the reference group from the ACT.

Pooled regression analysis over the five period data set shows that respondents from three states, New South Wales, Victoria and Western Australia, were more likely to 'always'

answer their mobile phone while driving than those from other States or Territories. Those in New South Wales were 2 times more likely to answer 'always', while respondents from Victoria and Western Australia were 1.6 and 1.5 times more likely, again using ACT used as the reference group.

General propensity for risky behaviour

Many young people enjoy risk-taking and have a higher threshold for risk than most of the older population. Therefore, a key control question to answer in a study like this is whether the responses to the question on mobile phone usage while driving is part of a general propensity for risky behaviour, or is it something to do with mobile phones?

For studying risky behaviour we had a choice of various variables ranging from attitudes towards drink driving and speeding, to the use of seatbelts. However, due to the missing data values we focused on two questions about speeding to control for risky behaviour:

We also carried out two separate regressions, to ascertain whether there was a similarity between driver's attitudes towards mobile phone use and their answers to these two questions regarding speeding.

Survey participants were asked, '*do you believe fines for speeding are mainly intended to raise revenue*'.

When compared to the reference group education level (set at diploma/degree), a level of education of 'trade/certificate' was found to be significant in every survey in determining whether respondents agreed strongly with the idea that fines for speeding are mainly intended to raise revenue. This same grouping was found to have had a significant influence on whether a person would always answer their mobile while driving in every year of the survey except 2011.

Being male was found to significantly influence whether a respondent always answers their mobile while driving and whether they strongly agree that speeding fines are issued in order to raise revenue.

Participants in the survey were also asked, '*Is it ok to exceed the speed limit if you are driving safely?*'

As with the question on speeding fines, an education level of 'trade/certificate' was a significant factor in determining whether a respondent strongly agreed that it is ok to exceed the speed limit if you are driving safely. This regression again used the group 'diploma/degree' as a reference group.

As in the previous question, being male was found to significantly influence whether a respondent always answers their mobile while driving and whether they strongly agree that it is ok to exceed the speed limit if you are driving safely, when compared to females.

Random Forest

A key objective of this paper is to analyse categorical responses from survey data and obtain a classification of the form "high-risk" or "low-risk". This problem may be solved using many mathematical models like decision trees, multinomial regression or various other machine learning methods. We have chosen Random Forest as a second analysis methodology because Random Forest provides a number of advantages over other similar techniques (Table 3).

Table 3 Key advantages of RF

Non-parametric regression model, allows for the construction of the model using information derived from the data rather than taking a predetermined form. RF has proved to be very accurate for a number of classification and regression problems.
Size of data set or variables is not a limitation and can run efficiently on very large data sets.
Open source implementation is available and no commercial software is required.
Tested successfully on various real data sets.
Non-trivial extraction of novel, implicit and actionable knowledge from large databases in a timely manner.
Provides an estimate of the importance of variables for dealing with incomplete data sets.
Runs faster than most other data mining methods.

Random Forest (RF) was first introduced by Breiman in 2001 and has since been successfully used in a number of fields (Breiman, 2001). RF works well for both regression or for classification and since our work deals primarily with categorical data we will focus on the classification problem (Liaw & Wiener, 2002). In a 2009 paper, researchers from the Louisiana State University demonstrated how RF can be used for classifying and detecting terrorists (Xu, Chen, & Li, 2009). Researchers from UCLA used data from California Department of Corrections and RF to classify and predict “high-risk” prison inmates (Berk & Baek, 2003). They were able to successfully identify inmates that were likely to engage in serious misconduct. These examples demonstrate their use as a good classifier and their ability to work with complex data sets.

RF forms part of a general set of techniques in machine learning called ensemble methods – they are algorithmic methods that generate a host of classifiers and aggregate their results. Classifiers can be of many types but the simplest classifier is a tree. A Random Forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \theta_k), k = 1 \dots\}$ where the $\{\theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x . (Breiman, 2001)

A comprehensive introduction to the RF method is out of scope for our paper, a brief introduction to the technique and concepts is described in this section for completeness. For a detailed introduction to RF we recommend Breiman’s original paper (Breiman, 2001) and the following sources (Berk R. A., 2011) (Hastie, Tibshirani, & Friedman, 2009). It is easier to understand RFs by studying its algorithm.

The standard Random Forest algorithm

For a data set consisting of C cases (labels), M classifier variables and a training data subset N : For b=1 to B (number of trees):

1. Take a sample of size N from the training data by choosing N samples with replacement from the training data (i.e. take a bootstrap sample). The remaining cases which are not in the training set will be used for testing purposes and will help determine the error rate for the RF.

2. Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
 - a. Select m variables at random from the M variables
 - b. Select the best split among the m
 - c. Split the node into two daughter nodes
3. Output the ensemble of trees $\{T_b\}_1^B$

To make a prediction at point x :

Regression: $f_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$

Classification: Let $\hat{C}_b(x)$ be the class prediction of the b th Random Forest tree. Then $\hat{C}_{rf}^B(x) = \text{majority vote } \{\hat{C}_b(x)\}_1^B$ (predict new data by averaging the predictions of the B trees) (Hastie, Tibshirani, & Friedman, 2009, pp. 587-589).

The error rate is estimated by using the generated RF for predicting and using the samples that were not in the training set to estimate the error.

It is easier to conceptualise the RF by looking at the visual representation of a tree, a RF tree may be represented as shown in Figure 2.

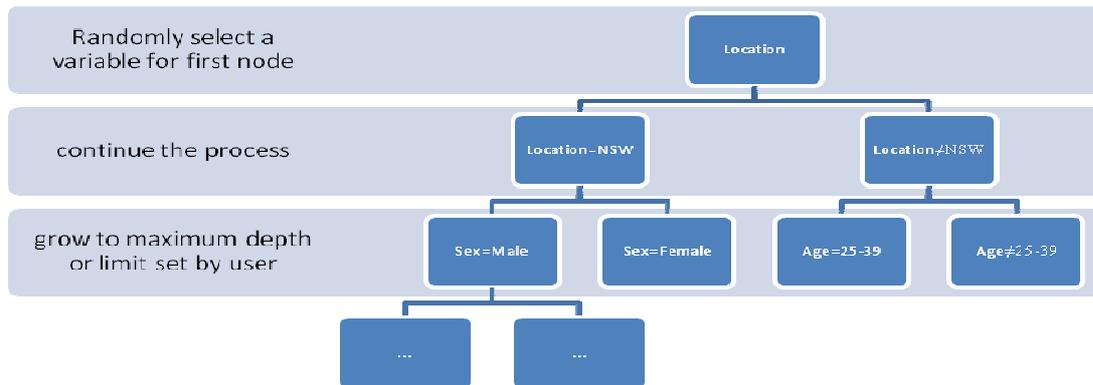


Figure 2. Representation of a randomly generated tree in RF

There are a large number of user specified parameters in the RF algorithm (Liaw & Wiener, 2002) that can have an effect on the final outcome. This section outlines some of the key parameters and how they were used in the research.

We experimented with a range of 500 to 5000 trees, set using the command *nTree*. We gradually increased the number of trees along with the variables and stopped when the results were stable. *Maxnode* and *nodesize* describes the maximum and minimum size of terminal nodes and can cause smaller or larger trees to be grown. We used the default values for these parameters. The number of variables randomly sampled as candidates at each split (*mtry*) also has to be specified. We experimented with values from 2 to 5. We found the value *mtry*=2 to be the most stable and useful in ensuring the robustness of our model. We used *xtest* to specify the training data. Most models feature a host of functions and parameters that can impact the results. The general rule of thumb used in our research, was to tweak the default parameter only if they offered a reduction in error rates and also guaranteed model stability.

Results of Random Forest modelling

The Random Forest modelling was done using the ‘randomForest’ package for R. Many different experiments were conducted with error rates ranging from 10% to 40%. In order to maintain consistency and ease of comparison with the results of the multinomial model we will focus on the experiments that used the exact data set (same number of variables and same type of data aggregation).

We carried out two series of experiments on the data, first by using data from individual years without considering the effect of time, and secondly by adding time as an additional variable. We randomly selected 80% of the data for training and 20% for testing. Each of our data files had approximately 1600 data labels and after filtering out the missing data values we obtained approximately 800-1000 data labels for training and 200 to 400 for testing.

We got error rates of around 15%, Table 4 shows a confusion matrix from one of our experiments. The values presented in this confusion matrix relate to survey responses for the question “do you answer your mobile phone while driving” – 0 corresponds to the survey response “never” and 1 corresponds to “always”. We defined high-risk as those people who always use a mobile phone while driving and low-risk as those people who never use their mobile phone while driving, so 0 is low-risk and 1 is high-risk.

Table 4 Confusion matrix from one RF experiment

Type of Random Forest :Classification			
Number of Trees 1000; No. of variables tried at each split: 2			
OOB estimate of error rate: 15.09%			
	0	1	Class Error
0	858	13	0.014
1	140	3	0.97
Test set error rate: 14.17%			
	0	1	Class Error
0	218	3	0.013
1	33	0	1.000

The error rates related to Table 4 can be viewed visually by using the ‘plot.randomForest’ function. Figure 3 shows the error rates from one experiment. The Out-of-Bag error rate is represented by the black curve and the green and red lines represent the response classes 0 and 1 for the question on mobile phone use while driving.

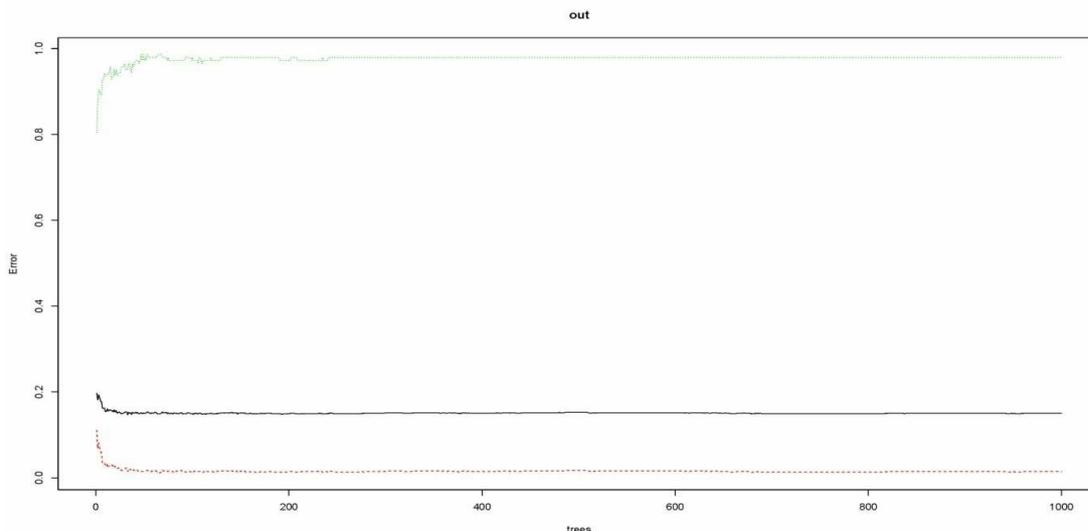


Figure 3 Error rates for mobile phone usage from an RF experiment

Using a number of experiments and varying key parameters like the number of trees and the initialisation condition (the random seed) we were able to test the robustness and stability of the RF model. This gave us confidence in our results.

The RF model outputs a score for each data record. The absolute value of this score can change from one experiment to another and therefore is not as important as the relative value. The score rating in conjunction with a low error rate provides a good classification. We obtained the following results from our experiments:

Table 5 Risk rating output for mobile phone usage from the RF model ordered relative to score highest to lowest

Age Group	Sex	Education	Score (risk rating)
25-39	Male	Trade/Certificate	Highest
25-39	Male	Degree/Higher	Highest
15-24	Male	Trade/Certificate	Highest
25-39	Female	Degree/Higher	Highest
15-24	Female	Trade/Certificate	Highest
40-59	Male	--	Lowest
40-59	Female	Trade/Certificate	Lowest
40-59	Female	Year 11/less	Lowest
60+	Male	--	Lowest
60+	Male	--	Lowest
60+	Female	-- ¹	Lowest
60+	Female	Trade/Certificate	Lowest
60+	Female	Year 11/less	Lowest

The results show very clearly that the age group 25-39 is most likely to use a mobile phone while driving. The scores in Table 5 have been aggregated by age and by risk rating, the RF model gives individual scores for each of the hundreds of data points and while this may be a useful exercise in determining credit score for a particular individual, it is not useful for our broad classification purposes. We have aggregated the groups that were given the highest and lowest risk scores in Table 5, the aggregation is done by using their relative scores and is ordered by rank. So even though both 25-39 Male and 15-24 Female are categorised as highest risk, between them the one with a higher ranking in the table is at a higher risk. We also found that the highest risk group, 25-39, tended to be people with trade/certificates who had full time employment. Employment details were only interesting for the highest risk group and therefore have been omitted from Table 5.

The second group that was determined to be at high-risk was also the youngest group in the survey, people aged 15-24, with both males and females in the high-risk category.

The model predicted that 60+ Females, who had education qualifications for Year 11 or less were the least likely to use a mobile phone while driving and therefore the least risky. The 60+ age group in general was the lowest risk group. The age group 40-59 was found to be the second lowest risk after the 60+.

¹ This category did not have a notable effect to the low risk group

Comparison of results – multinomial regression and Random Forest

Both the methods used in the research have their unique advantages, however it is interesting to look at their relative strengths and weaknesses and how one method can inform and feed information for the second method.

Regression provides us with a general (global) model that captures the information in a single predictive formula that is assumed to hold for the entire data set. In highly complex data sets with a large number of variables, compiling a global model is difficult and we face the problem of variable selection. This is an area where RF can provide the data mining insights. RF has an inbuilt function called variable importance that determines the ranking of a variables' importance relative to other variables. RF does this by removing a particular variable from the list and running permutations then averaging over all the trees in the forest and calculating the difference of the model accuracy before and after. For our work, the number of possible variables was large and RF helped us determine that the variable representing employment status was an important variable while it confirmed that the variable for location was not very relevant.

Random Forest results and the regression results were consistent with each other: regression told us that certain age groups were more likely to use the phone while driving and RF modelling confirmed this by providing relative scores for each individual which could then be aggregated into classes.

An interesting aspect of using RF is that it is able to handle as many variables as possible while regression can suffer from multicollinearity. Multicollinearity in logistic regression models results from strong correlations between independent variables and this can inflate variances of the parameter estimates. Detecting multicollinearity in regression is not an easy exercise and requires a number of different diagnostic tests. Whereas RF can be much less susceptible to multicollinearity and variables with complex interactions because of the RF functions that enable us to estimate variable importance and error rates.

A big problem in most research work is missing data. RF is naturally better at dealing with incomplete data, whereas regression requires that we exclude the missing values or we use methods like series mean to estimate the missing values. In order to ensure consistency for the results presented in this paper, we chose not to use any of the available techniques to compensate for missing data. We did experiment with some techniques and found the results were not significantly altered and this was done just as an additional test for robustness.

4. Possible Policy options

The results above have shown that policies deterring mobile phone use while driving should be targeted at those aged between 15-24 and 25-39 years of age. These are considered high-risk groups and it is vital that policies directly target these age groups. In the past, anti drink driving and speeding advertising campaigns have successfully targeted these young driver age groups. (New Zealand Transport Agency, 2011 and New South Wales Government Transport Roads & Maritime Services, 2011) A key issue facing policy makers is how to directly target these two high-risk groups in order to combat the use of mobile phones while driving.

Current laws state that using a hand-held phone while driving is illegal in all Australian States and Territories. The National Road Safety Strategy 2011–2020 has put forward a plan

to examine and extend current laws banning some novice drivers (eg, those on 'P1' licences) from using mobile phones (including hands-free); it is suggested that such restrictions could be extended to drivers in the P2 stage of their license, or to all drivers under the age of 26. (Australian Transport Council, 2011) However these policy aspirations from the State and Federal Governments do not directly address those high-risk drivers identified, aged between 25-39 years old. The following policy suggestions look to directly target these two risk groups identified.

Technology-enabled restriction on mobile usage while driving

Various forms of technology are being created to help curb driver distractions and to increase driver safety.

The National Highway Traffic Safety Administration in the USA recently put forward a set of guidelines to American manufacturers, which they believe would help curb driver distractions, such as mobile phone use while driving. The first round of non-binding recommendations provided by the NHTSA focused specifically on non-driving related distractions such as navigation systems, entertainment and communication devices (specifically mobile phones). A key recommendation suggested that in-car controls be designed to prevent the car from starting if these devices are running, which would include sending and receiving text messages. (Tison, Chaudhary, & Cosgrove, 2011). Another recommendation is what the NHTSA calls the '*2 second test*'. A device would be set up in the car to measure how often a driver's eyes are diverted away from the road. The device will allow for drivers to complete tasks that only allow drivers to take their eyes off the road for two seconds or less, for a total time of 12 seconds. (Tison, Chaudhary & Cosgrove)

Similar proposals to those recommended by the NHTSA are currently in place to curb serial drink driving offenders. Alcohol interlock programs are implemented or are being trialled in six Australian states and territories, with the exceptions being Western Australia and the ACT. An alcohol interlock device is fitted in cars of serious drink driving offenders, requiring the driver to take a breath test before they are able to start the car. If the driver fails the breath test, the car will not start and the driver will have to wait between 5 to 30 minutes before they are able to retake the test. (New South Wales Centre for Road Safety, 2011)

In Victoria, a driver must have an alcohol interlock system installed if they have two or more drink driving offences, have had a drink driving conviction where their blood alcohol concentration was over 0.15 or is either a probationary driver or under 26 years old with one conviction where their BAC was over 0.07. (Vic Roads, 2002)

The main aim of the program is to reduce drink driving among repeat offenders. A study conducted by Coben and Larkin found that in five of the six studies they performed, the alcohol interlock system was effective in reducing drink driving recidivism among offenders. (Coben & Larkin, 1999)

One recommendation is that the government, in consultation with car manufacturers, look towards investing in technology that prevents a car from starting if electronic devices are running on the dashboard. This idea, coupled with technology similar to the eye scanning technology recommended by the NHTSA, would significantly reduce driver's distractions and keep their focus on the road.

Specifically targeted advertising campaigns for high-risk groups

The results above show that mobile phone use as a driver distraction is significant for those aged between 15 and 24 years of age. In 2011, 98% of respondents surveyed, aged between 15-24 years of age, said they owned a mobile phone. Although this group

statistically were not seen to be as high-risk as the 25-39 age group, their risk for using mobile phones while driving is still high.

As outlined previously, the Federal and State governments have policy aspirations that specifically target young drivers. However better advertising campaigns could be utilised to specifically target young drivers of the dangers of using a mobile phone while driving.

Advertising against drink driving can be targeted directly at young drivers. Drink driving campaigns overseas have been successful in targeting specific demographics of society through advertising. Via various media outlets, they have been able to specifically target young drivers with their anti drink driving message. In 2010, the New Zealand Transport Agency produced a series of anti drink driving advertisements which were directly aimed at young drivers. It focuses on the idea that young people, in particular young male drivers, may feel like they are outcast if they try to convince their friends not to drink and drive. (New Zealand Transport Agency, 2011)

Advertising campaigns against speeding have been effective in targeting young drivers. Research from the New South Wales RTA found that graphic imagery was not effective in deterring young drivers. The best example of a campaign specifically targeted at young drivers was 'the pinkie campaign'. The main aim of the campaign was to highlight the social unacceptability of speeding and to challenge the mindset that speeding is 'cool', which was underlined with the tagline 'Speeding. No one thinks big of you'. The campaign also featured New South Wales and Australian cricketers in the campaign. A survey commissioned by the RTA showed that 60% of young males aged between 17-25 years were more likely to make a comment on another's driving after seeing the advertisement. (New South Wales Government Transport Roads & Maritime Services, 2008; 2011)

We believe that State governments need to advertise to a specific target market. In this case, advertising needs to be tailored specifically towards the two high-risk groups identified, those aged between 15-24 years of age and 25-39 years. This has been done successfully with speeding and anti-drink driving advertising strategies and should be implemented in a similar way to deter mobile phone use while driving.

Stricter laws on hands-free mobile phone use

Laws are in place in all Australian States and Territories that prevent drivers from using hand-held mobile phone devices (Table 6). A driver may use a hands-free device to receive calls, so long as they are not physically touching the phone in any way, which includes resting the phone on your lap or chin. (Strike, 2009) An exception to these laws is in Western Australia, where changes to laws implemented in March 2011 now allow drivers to make or receive calls by touching their phones, so long as they are mounted to a device in the car. Creating or receiving a text message, video message or Email is still considered illegal. (Government of Western Australia Office of Road Safety, 2011)

Table 6 Current laws across the Australian states

NSW	For Learner and Provisional 1 Drivers there is a total mobile phone ban while driving.
VIC	For Learner and Provisional 1 Drivers there is a total mobile phone ban while driving.
QLD	Learner and Provisional 1 Drivers under the age of 25 have total mobile phone ban
SA	For Learner and Provisional 1 Drivers there is a total mobile phone ban while driving.
WA	Hand-held mobile phone ban only.
TAS	Hand-held mobile phone ban only.

NT	For Learner and Provisional drivers there is a total mobile phone ban while driving.
ACT	Hand-held mobile phone ban only.

Various studies have shown that using a hands-free mobile device can significantly reduce drivers' concentration levels and may not be any less distracting to drivers than using hand-held phones. Analysis from academics Ishigami and Klein (2009) find that, against common public perception, driving performances are not improved by using hands-free comparatively to hand-held phones. Strayer and Johnston (2001) found that, regardless of whether a driver used hand-held or hands-free mobile device, their reaction time in being able to detect signals was severely diminished.

Consistency is needed across Australian states and territories in banning the use of hands-free mobile devices for any driver on their Learner's license, Provisional 1 or Provisional 2 license. This would specifically target the high-risk group of drivers aged 15-24 and possibly work as a deterrent in reducing mobile use among this age group. Evidence indicates that regardless of whether a driver uses a hands-free or hand-held mobile, they both distract drivers. Future policy should be aimed at eliminating all forms of mobile phone use while driving.

Reduce ambiguity in the definition of mobile phone use while driving

One of the problems identified by the Western Australian Government is that the introduction of smart phones has now produced more distractions to drivers than simply making and receiving calls. The definition of the term 'use' when referring to mobile phones is now ambiguous with the introduction of smart phones and subsequently, the Western Australian government made changes to their previous laws. As the functions that can be performed on mobile phones become more and more complex, a clear definition of what constitutes 'using a mobile phone' needs to be established in order for there to be consistency across the states.

5. Conclusion

The paper has identified high-risk groups within the community and provided an analysis using two modelling methods; Random Forest and multinomial regression. The paper has shown how public policy can identify and target high-risk groups within the community for maximum impact. However, we have not shown whether public policy does impact attitudes. This is an interesting issue that remains and could be looked at in future studies. In particular, an analysis is needed of the link between public attitudes to safety and changes in public policy. This work requires comparing enforcement data from the police on fines for mobile phone use while driving and the CARS survey data. Such a study will reveal how the changes in policy have resulted in changes in attitudes.

Bibliography

Australian Transport Council, 2011, *National Road Safety Strategy 2011-2020*, Canberra.

Berk, R. A. (2011). An Introduction to Ensemble Methods for Data Analysis. *Department of Statistics Papers, University of California Los Angeles* , 1-37.

Berk, R. A., & Baek, J.-H. (2003). Ensemble Procedures for Finding High-risk Prison Inmates. *Department of Statistics, University of California Las Angeles* , 1-21.

Breiman, L. (2001). Random Forests. *Machine Learning* , 5-32.

Coben, J. H., & Larkin, G. L. (1999). 'Effectiveness of ignition interlock devices in reducing drunk driving recidivism.' *American Journal of Preventive Medicine* , 81-87.

Consiglio, W., Driscoll, . P., Witte, M., & Berg, W. (2003). 'Effect of cellular telephone conversations and other potential interference on reaction time in a braking response.' *Accident Analysis and Prevention*, 495-500.

Department of Transport Western Australia, 2012, *Novice drivers (includes L and P plate drivers)*, viewed 16 April 2012 <from: <http://www.transport.wa.gov.au/licensing/20424.asp>>.

Department of Transportation National Highway Traffic Safety Authority, 2012, *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Devices*, Washington D.C.

eGovernment Resource Centre, 2011. *Mobile Phone Penetration in Australia* . viewed 2 April 2012, <<http://www.egov.vic.gov.au/focus-on-countries/australia/trends-and-issues-australia/telecommunications-australia/mobile-phone-penetration-in-australia-2009-2015.html>>.

Government of South Australia Department of Planning Transport and Infrastructure, 2012, *My Licence - L Plater*, viewed 15 April 2012, <http://mylicence.sa.gov.au/l_plater>.

Government of South Australia Department of Planning Transport and Infrastructure, 2012, *My License - P1 Plater*, viewed 16 April, 2012, <http://mylicence.sa.gov.au/p1_plater>.

Government of Western Australia Office of Road Safety, 2011, *Changes to the Laws Relating to the Use of Mobile Phones and Visual Display Units While Driving in Western Australia; Frequently Asked Questions*. Perth.

Government of Western Australia Office of Road Safety, 2011, *Office of Road Safety - Road Rules & Penalties* viewed 29 March 2012, <<http://ors.wa.gov.au/Demographic-Pages/New-to-Driving/Road-Rules-And-Penalties.aspx>>.

Hands-Free Info 2011, *European cell phone, texting roundup*, viewed 11 April 2012, <<http://handsfreeinfo.com/european-cell-phone-texting-roundup>>.

Hands-Free Info, 2011, *Australia, New Zealand distracted driving news*, viewed 11 April 2012, <http://handsfreeinfo.com/category/international-cell-phone-laws/australia-new-zealand>>.

Hastie, T., Tibshirani, R., & Friedman, J. 2009. *The Elements of Statistical Learning*, Springer.

Hill, R. C., Griffiths, W. E., & Lim, G. C. 2008, *Principle of Econometrics*, Hoboken, New Jersey: John Wiley & Sons, Inc.

Horrey, W. J., Lesch, M. F., & Garabet, A. 2008, 'Assessing the awareness of performance decrements in distracted drivers', *Accident Analysis & Prevention*, 375-382.

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

Justice and Community Safety Directorate, ACT Government, 2012, *ACT Road Rules Handbook*. Canberra.

Adolph, M. (2010). *Decreasing DriverDistraction*. ITU-T Technology Watch.

- Liaw, A., & Wiener, M. (2002). 'Classification and Regression by Random Forest.' *The R Journal*, 18-22.
- McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., & Palamara, P. 2005. 'Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study', *BMJ*, 331-448.
- McKnight, J. A., & McKnight, A. S. 1993, *The effect of cellular phone use upon driver attention*. *Accident Analysis and Prevention* Vol 25 Issue 3, 259-265.
- Monash University Accident Research Centre, 2010, *Look who's talking! A roadside survey of drivers' cell phone use*, *Traffic Injury Prevention*, Volume 11 Issue 6, 555-560.
- Musselwhite, C., Avineri, E., Fulcher, E., & Susilo, Y. 2010, *Understanding Public Attitudes to Road-User Safety - Literature Review: Final Report*. Bristol: Centre for Transport and Society, University of the West of England.
- National Safety Council, 2010, *Understanding the distracted brain; Why driving while using hands-free cell phones is risky behavior*, <distracteddriving.nsc.org>.
- New South Wales Centre for Road Safety, 2011 *Alcohol Interlock Program; Information for Participants*, New South Wales Government Transport Roads & Maritime Services.
- New South Wales Government Transport Roads & Maritime Services 2011, *The Pinkie Campaign*, viewed 30 March 2012, <<http://www.rta.nsw.gov.au/roadsafety/speedandspeedcameras/campaigns/pinkie.html>>.
- New South Wales Government Transport Roads & Maritime Services, 2011, *Special conditions for L and P drivers and riders*, viewed 3 April 2012, <http://www.rta.nsw.gov.au/licensing/gettingallicence/200707_restrictions.html>.
- New South Wales Transport Roads and Maritime Services, 2008, *Cricketers Star in New 'Pinkie' Campaign*, viewed 30 March 2012, <http://www.rta.nsw.gov.au/newsevents/2008/2008_10_speedblitz.html>.
- Northern Territory Government Department of Lands and Planning 2009, *L1 – How to Obtain a class "C" or class "R(r)" Learner Licence*. Darwin.
- NZ Transport Agency 2011, *Drink driving advertisements*. viewed April 3 2012, <<http://www.nzta.govt.nz/about/advertising/drink-driving/legend.html>>.
- OECD/International Transport Forum. (2011). *Road Safety Annual Report*. International Transport Forum.
- Pettitt, M., Burnett, G., & Stevens, A., 2005, *Defining driver distraction*, to be presented in the World Congress on Intelligent Transport Systems.
- Queensland Government Department of Transport and Main Roads, 2012, *Young Drivers*, viewed 3 April 2012, <<http://www.tmr.qld.gov.au/Licensing/Learning-to-drive/Young-drivers.aspx>>.
- Shih, S., 2011, *Random Forests for Classification Trees and Categorical Dependent Variables*, Stanford University, University of California, Stanford, California.
- Stelling, A., & Hagenzieker, M. P., 2012, *Afleiding in het verkeer*. Een overzicht van de literatuur.

Strayer, D. L., & Johnston, W. A., 2001, *Driven to Distraction: dual-task studies of simulated driving and conversing on a cellular telephone*. Department of Psychology, University of Utah, Salt Lake City.

Strike 2009, *Mobile Phone Driving Laws*, viewed 4 April, 2012, <<http://www.strike.com.au/s.nl/ctype.KB/it.l/id.9523/KB.34513/f>>.

Tasmanian Government Department of Infrastructure, Energy & Resources, 2010, *Using Mobile Phones and Visual Display Units*, viewed 3 April 2012, <http://www.transport.tas.gov.au/licence_information/tasmanian_road_rules/road_safety_rules/using_mobile_phones_and_visual_display_units>.

The Social Research Centre, 2005, *Community Attitudes to Road Safety, Community Attitudes Survey, Wave 18, 2005*, North Melbourne, Victoria.

The Social Research Centre, 2006, *Community Attitudes to Road Safety, Community Attitudes Survey, Wave 19, 2006*, North Melbourne, Victoria.

The Social Research Centre, 2008, *Community Attitudes to Road Safety, Community Attitudes Survey, Wave 20, 2008*, North Melbourne, Victoria.

The Social Research Centre, 2009, *Community Attitudes to Road Safety, Community Attitudes Survey, Wave 21, 2009*, North Melbourne, Victoria.

The Social Research Centre, 2011, *Community Attitudes to Road Safety, Community Attitudes Survey, Wave 22, 2011*, North Melbourne, Victoria.

Tison, J., Chaudhary, N., & Cosgrove, L., 2011, *National Phone Survey on Distracted Driving Attitudes and Behaviours (Report No. DOT HS 811 555)*. Washington, DC: National Highway Traffic Safety Administration.

Vic Roads, 2010, *Mobile phones and visual display units*. Viewed, 3 April 2012, from: <<http://www.vicroads.vic.gov.au/Home/SafetyAndRules/RoadRules/MobilePhonesandvisualdisplayunits.htm>>.

Vic Roads, *Victoria's Alcohol Interlock Program information for participating drivers*. Melbourne, Victoria.

Xu, J., Chen, J., & Li, B., 2009, 'Random Forest for Relational Classification with Application to Terrorist Profiling.' Eighteenth International Symposium on Methodologies for Intelligent Systems (ISMIS'09).

Young, K., Regan, M., & Hammer, M. (2003). 'Driver Distraction: A Review of the Literature', Melbourne, Victoria: Monash University Accident Research Centre.