Rural Crash Prediction Models - The Next Generation

Dr Shane Turner, Technical Director (Transport), Beca Infrastructure Ltd BE (Hons), PhD, MIPENZ, CPEng, IntPE, shane.turner@beca.com

Dr Alistair Smith, Transport Researcher, Beca Infrastructure Ltd alistair.smith@beca.com

Dr Ian Appleton, Safety Advisor, Land Transport NZ Ian.Appleton@landtransport.govt.nz

Dr Graham Wood, Professor of Statistics, Macquarie University, Sydney gwood@efs.mq.edu.au

ABSTRACT

Over recent years a number of rural crash prediction models have been developed in New Zealand for specific purposes. However, all models to date are deficient as they do not contain all the known important predictor variables, and generally are based on limited sample sizes. As a result New Zealand does not have suitable rural crash prediction models for the identification and analysis of road safety problems, policy development and testing and project appraisal (or evaluation). New Zealand makes a significant investment in engineering improvements on rural roads each year and better crash models are expected to result in more focused investment and a reduction in fatal and injury crashes. Since such models would improve our understanding of crash causing factors, the combination of such factors and the benefits of safety countermeasures, a research team has begun work to develop the much needed models.

A three stage process is being followed by the research team, consisting of a scoping, pilot and main study. To date the team has completed the scoping study and is now progressing through the pilot study. This technical note outlines the findings of the scoping study, the data that has been collected in the pilot study, both electronically and by technicians in the field, the analysis of this data and the next steps in the process. Since the initial process was developed there have been a number of advancements in the automated collection of rural road data, which may eliminate the need to collect data on some of the variables manually in the field. Automated data collection allows capture of a larger set of variables, including road-side information; appropriate selection from such a set is expected to lead to improved models. The technical note will discuss the merits of collecting more data for rural roads and setting up a comprehensive database on rural road features. The technical note will discuss a number of associated projects and studies that have been undertaken, and which add to the knowledge that is being applied to this study. For example a number of methods have been developed for recording roadside hazards and horizontal consistency.

INTRODUCTION

The research outlined in this paper stems from a workshop with key Government Agencies and "Industry" representatives, active in road safety in 2005. That workshop identified that while progress has been made in the development of crash prediction models for rural roads in New Zealand, there are some serious "gaps" in knowledge that are impacting on policy direction and implementation. The workshop supported in principal a plan for addressing these shortcomings, through the development of the next generation of crash prediction models for rural roads.

A number of studies have developed detailed rural road crash prediction models in New Zealand, typically for one-off evaluations of specific features or policies including:

- Chadfield (1993), sought to apply Australian relationships for the impact of lane and shoulder width and shoulder slope on crash rate;
- Jackett (1990) studied the relationship between crash rates and curve radii;
- Koorey and Tate (1997), investigated how alignment consistency and speed impacted on crash rate and severity;
- Turner, (2001) developed crash prediction models for a wide variety of intersection and link types using traffic volumes and crash rates;
- Cenek et al., (2004) assessed the impact on crash rates of road surface, and alignment;
- Turner, (2004) studied the crash rate implications of roadside hazards; and
- Tate et al (2005) investigated current state highway shoulder standards and the relationship between sealed shoulder width and crash risk.

While each of the above studies has, in general, answered the question at hand, the individual models only contain a small subset (typically three to five variables) of the important predictor variables and some are only based on traffic volume. As a result the models cannot be readily or reliably used to evaluate the crash rate resulting from a combination of variables.

The overall objective of the "Rural Crash Prediction Models – A New Generation" study is to develop a new, more comprehensive set of rural crash prediction models for New Zealand, drawing on international research and experience, and New Zealand research efforts to date. The study has been broken down into three stages; Scoping, Pilot and Main Studies. The Scoping Study was completed in July 2006. The pilot study is currently underway and has progressed through data collection to preliminary data analysis. This technical note reports on the findings of the scoping study and some preliminary results from the pilot study. It also looks forward to the main study and how data might be efficiently captured for this study and others.

SCOPING STUDY STAGE

The scoping study examined the models for two-lane-two-way rural roads previously developed in New Zealand and overseas to identify all the key variables that should be included in the next generation of rural road crash prediction models. The main objectives of this study were to:

- Investigate current crash prediction models to determine which variables have found to be important and to identify existing model deficiencies;
- Identify which road and traffic related features could potentially be included in the resulting model, to prioritise these, and identify what variable sets may be available to quantify these features;
- For each possible variable develop a definition and identify whether the data is readily accessible, (accepting that for some features and variables it may be necessary to develop or modify an existing collection method);
- Develop a Data Collection Methodology that can be used by surveyors to collect field data (some data is already available from other sources); and
- Develop sampling framework for the pilot and preliminary sampling framework for the main study, suitable for budget allocation.

Important Predictor Variables

The features (variables) for investigation in the study were determined by reviewing previous research findings and following discussions between the study team and other road safety experts. The variables that were sufficiently important were; Traffic Volume, Access Density, Horizontal Geometry, Horizontal Geometry Consistency, Seal Width, Shoulder Environment, Roadside Hazards, Region and SCRIM Coefficient, in no particular order.

Data sources (Electronic and Manual)

There are a number of ways to measure and define each of the important variables, and so it was necessary in the Scoping Stage to develop definitions for each predictor variable group. Where possible data was obtained from electronic databases, including the CAS database for crash data, the RAMM database for seal width and traffic volume and the State Highway high speed database for SCRIM and horizontal geometry (and from that consistency). Other data is generally not available electronically, included shoulder environment, access density and roadside hazards.
It was expected, at least for the Pilot Study, that access density, shoulder environment and roadside hazard data would need to be collected manually via field surveys, as follows:

Access Density - Access density has previously not been collected in any known New Zealand studies. It is not readily available and needs to be collected for each section in the field. A simple collection method that provides an adequate level of detail while reducing data collection costs was required. Farm and residential accesses are to be lumped into a single category with accesses on either side of the road being counted but not further classified. Where accesses are shared, the number of adjoining properties would be counted rather than the access itself. Commercial accesses would be identified separately and for each access of this type the access would be categorised as low, medium or high.

Shoulder Environment - There are a number of measures that could be used for defining the shoulder environment. Many need to be used in unison with other measures. Examples of different shoulder environment measures that were considered include; sealed shoulder width, unsealed shoulder width, shoulder slope, width of berm, clear zone width, recoverable and traversable width. The study team concluded that the following features should be noted for each side of the road ; unsealed shoulder width, total width of recoverable slope (≥ 1:6) from edge of seal, total width of traversable slope (≥ 1:3) from edge of seal and location of continuous severe hazards, like ditches and cliffs. These measures are shown graphically in Figure 1 and 2.

Road-side Hazards - The assessment of roadside hazards mainly needs to be simple and allow for efficient data collection as an inefficient measurement technique could make the collection of roadside hazard data very labour intensive and costly compared to other variables. For this study, a limited hazard inventory is proposed. Unlike a detailed hazard inventory only a selection of roadside hazard data will be collected. Firstly, as the traversable slope measured when assessing the shoulder environment represents the distance to severe continuous hazards (such as steep up and down slopes, upright banks, deep drains and closely spaced trees), only hazards within this distance will be measured. So if the traversable width is four metres, then only hazards within four metres from the edge of seal will be included. The second restriction on the data collection is limiting the hazard collection to severe discrete hazards. Severe discrete hazards have been classified in Turner (2004). This definition is to be used in this study

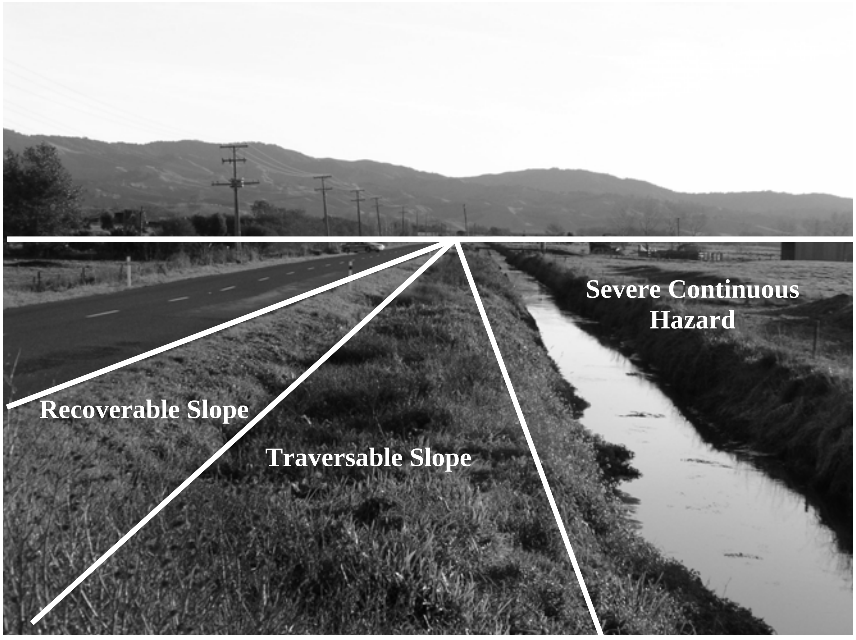


Figure 1 – Shoulder environment without unsealed shoulder but with recoverable & traversable slopes

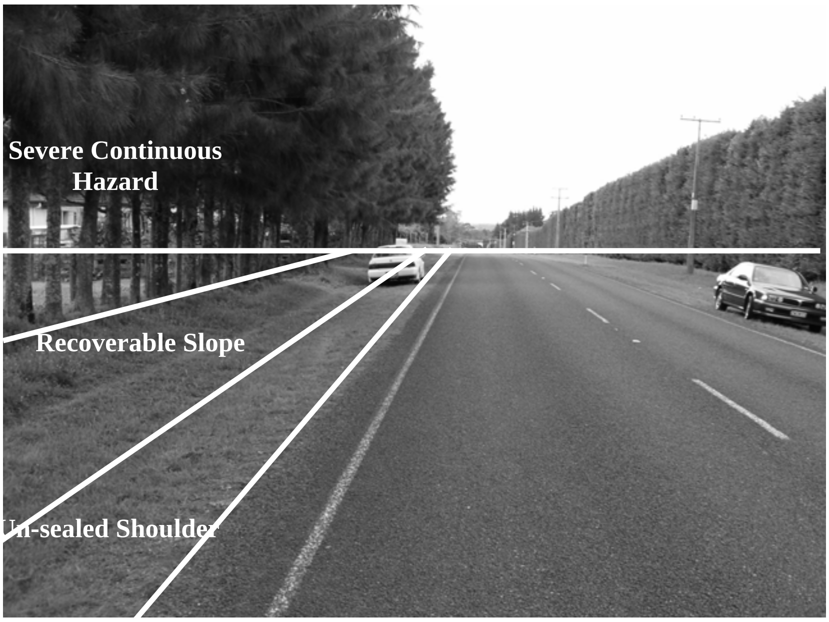


Figure 2 – Shoulder environment with unsealed shoulder and recoverable slope

Sampling Framework

In the sample framework the minimum amount of data that needs to be collected to build satisfactory models that explain the relationships between crashes and each of the variables has been estimated by a statistician (further detail on this process can be found in Turner et. al., 2006). Two sampling frameworks were proposed. The first is an ideal and the second an intelligent compromise, considering cost. The 'ideal' sampling framework would see data collected for 200 × 200-metre sections of road in each of the 12 Transit NZ regions. The sections would be, randomly sampled from the State Highway and Local Road Network. The resulting, sample size of 200 sections per region and 2400 sections for the entire county would appear satisfactory.

The alternative, 'compromise' sampling framework aims to reduce costs, by sensibly clustering Transit Regions (into 'Super Regions') so that within the clusters the regions are as uniform as possible, with respect to variables (such as weather and socio-economic effects). The clustering into 'Super Regions' was undertaken by grouping regions based on the similarity in terms of open road 85^{th} percentile speed, regional under-reporting of serious crashes, percentage of SH crashes in dry weather, percentage of SH mid-block 100km/hr alcohol related crashes, percentage of dry SH mid-block 100km/hr crashes, percentage of SH mid-block crashes in dark and percentage of mid-block 100km/hr cornering types.

In such a clustering, Auckland and the West Coast stand out as distinct regions. For this reason it was deemed they should be treated as separate regions. For the remaining regions, three clusters were identified. For this alternative sampling framework there are a total of 200 × 200-metre sections in each of the five regions, with an overall sample size of 1000 sections. This was the preferred option due to lower data collection costs. The pilot study is to focus on a single region. Based on the pilot study results it will be possible to refine the sampling framework.

PILOT STUDY STAGE

In the Pilot Study (Stage 2 of the study) data has been collected electronically and manually for 200 x 400m sections on State Highways in the Waikato. This differs from the 200m length sections specified in the scoping study. The length of the sections was extended for the pilot only so that the team had a richer data set from which to examine the relationship between crashes and road alignment. We wanted to test whether uniform sections with particular combinations of curves and straights were better or worse than having non-homogenous sections based on single curves and straights. Both approaches have been used by researchers in the literature, with no clear direction on the correct approach. In December 2007, 29 sections were surveyed, with the remaining 171 in February 2008.

The 400m road sections were segmented into eight subsections of 100m, four on each side of the road. The variables information in Table 1 was collected for each of the 100m sections.

Table 1: Variable Types

Type	Variable
Shoulder Environment	Seal Width (m)
	Unsealed Shoulder Width (m)
	Recoverable Slope Width (m)
	Traversable Slope Width (m)
Point Hazards	Wood Pole >200mm (no)
	Light Column <300mm (no)
	Concrete Pole – usually 'I' section (no)
	Heavy Street Pole >300mm without slip base (no)
	Signs Supports >120mm without slip base (no)
	Trees - trunk >100mm diameter (no)
	Culverts - road side (no)
	Culverts - road with non-traversable headwall (no)
	High impact roadside furniture (no)
	Non-traversable slope / perpendicular deep drain (m)
	End concrete barrier / bridge parallel to road (m)
	Concrete fence/barrier perpendicular to road (m)
Accesses	Farm / Residential (no)
	Commercial (no)
	Usage - Low/Medium/High (no)

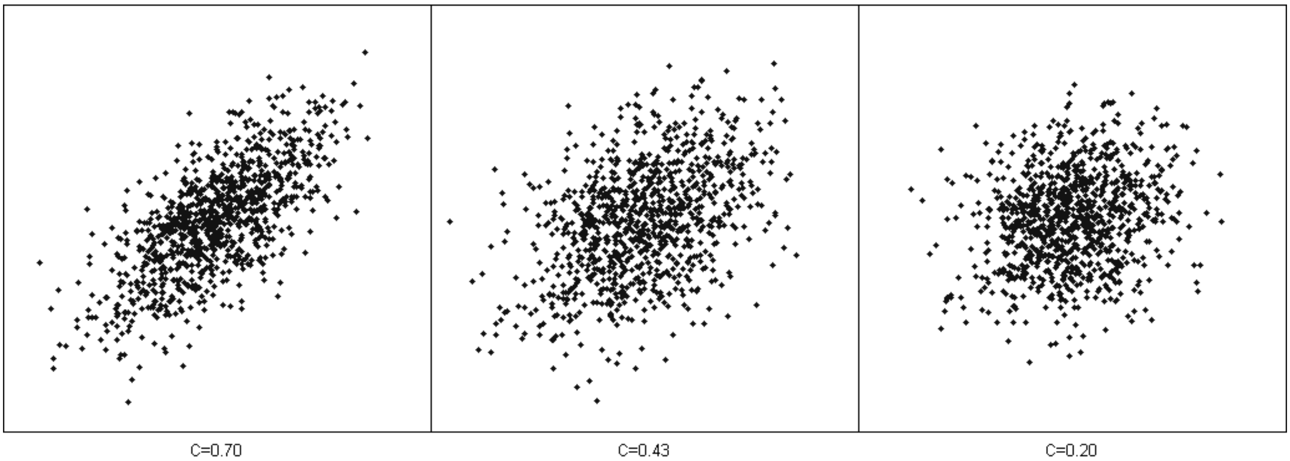
The large number of predictor variables in the rural road crash models adds to the complexity of the modelling task, in particular through cross-correlation of variables. Of the 324 combinations of variables, the following 10 variable pairs produced a correlation with absolute value greater than 0.2. These are listed in Table 2.

Table 2: Variable Cross-Correlation for Roadside Data

Variable A	Variable B	Correlation
Recoverable Slope	Traversable Slope	0.43
Traversable Slope	Wood Pole	0.43
Culverts – road side	Farm / Residential	0.42
Traversable Slope	Non-traversable Slope	-0.31
Traversable Slope	Concrete Pole	0.30
Wood Pole	Concrete Pole	0.27
Wood Pole	Non-traversable Slope	-0.27
Traversable Slope	Trees	0.26
Unsealed Shoulder	Recoverable Slope	0.23
Wood Pole	Farm / Residential	0.22

These results may be compared with example X-Y plots illustrating a range of correlation values, as shown in Figure 3. These plots indicate that the maximum correlation in our survey data, 0.43 is not particularly high. The correlation value would need to be closer to 0.7 before we would have concerns and need to consider exclusions of one of the variables from the models.

Figure 3: Example plots illustrating correlation



The next stage in the pilot study is to build some basic generalised linear models for the data collected in the Waikato. The field data on shoulder environment, access density and road-side hazards is to be combined with electronic data for the remaining variables.

LOOKING FORWARD TO THE MAIN STUDY

The major cost in the project will occur in the third stage, which will involve the collection of data for in excess of 800 road sections spread throughout the country. Even with an overall sample set of over 1000x200m sections (220 section from the pilot to be included) it is likely to be difficult to understand the relationships between all the variables and crashes. Ideally a much larger sample set of the State Highway network and also of the local authority road network would be included in the study, but this is not likely to be cost effective for a manual data collection exercise.

With emerging technologies, such as those being used by ARRB, i.e. Gypscam, information like road-side hazards, shoulder environment and access density, can be collected from video images. The quality of this data is improving over time and fast approaching the level of accuracy required for this type of research. Such technology is being utilised for the international Road Assessment Program (iRAP). With New Zealand embracing the RAP approach in the form of KiwiRAP there is a great opportunity to collect a much broader data-set of rural road features than is current available as part of this project. If this data is collected in the right format then there are a number of applications beyond KiwiRAP, to utilise the data, including the development of crash prediction models for rural roads.

REFERENCES

CENEK, P. D, DAVIES, R. B. and HENDERSON, R. J (2004), *The Effect of Skid Resistance and Texture on Crash Risk*, Sustainable Transport Conference Proceedings, Wellington, NZ

CHADFIELD, E. (1993), *Review of Cross-section Guidelines for Two Lane Rural Roads*. Transit NZ. Wellington. New Zealand.

JACKETT, M. J. (1990) *Accident Prediction: Rural Curves*. LTSA. Wellington. New Zealand.

KOOREY, G. F. and TATE, F. N. (1997), *Review of Accident Analysis Procedures for Project Evaluation Manual*. Transfund NZ Research Report

LTSA (2003) *Accident Monitoring Investigation Analysis,* Standards NZ, Wellington, NZ

TATE, F. N, ANDERSON, D. MUIRSON, M. SIZEMORE, A and WANTY, D. K (2005) *Review of State Highway Shoulders* Draft Report to Transit New Zealand

TURNER, S. A (2001) *Accident Prediction Model*, Transfund NZ Research Report No 192, Standards NZ, Wellington, NZ

TURNER, S. (2004) A*ssessing the Crash Risk Implications of Roadside Hazards*, *Unpublished Land Transport Safety Authority Report*, NZTA, Wellington, NZ.

TURNER, S., ROOZENBURG, A., TATE, F. and WOOD, G. (2006) *Rural Crash Prediction Models – A Next Generation: Stage 1 Report,* Unpublished Land Transport NZ Research Report.