The next generation of rural road crash prediction models: final report

December 2012

Shane Turner, Rohit Singh and Gary Nates
Beca Infrastructure Ltd

NZ Transport Agency research report 509

ISBN 978-0-478-40706-8 (electronic)
ISSN 1173-3764 (electronic)

NZ Transport Agency
Private Bag 6995, Wellington 6141, New Zealand
Telephone 64 4 894 5400; facsimile 64 4 894 6100
research@nzta.govt.nz
www.nzta.govt.nz

Turner, S, R Singh and G Nates (2012) The next generation of rural road crash prediction models: final report. *NZ Transport Agency research report 509*. 98pp.

Beca Infrastructure Ltd PO Box 13960, Christchurch 8141. Tel: +64-3 363 3461 E:mail:shane.turner@beca.com

Keywords: crash prediction models, New Zealand, regional factors, road safety, rural roads

An important note for the reader

The NZ Transport Agency is a Crown entity established under the Land Transport Management Act 2003. The objective of the Agency is to undertake its functions in a way that contributes to an affordable, integrated, safe, responsive and sustainable land transport system. Each year, the NZ Transport Agency funds innovative and relevant research that contributes to this objective.

The views expressed in research reports are the outcomes of the independent research, and should not be regarded as being the opinion or responsibility of the NZ Transport Agency. The material contained in the reports should not be construed in any way as policy adopted by the NZ Transport Agency or indeed any agency of the NZ Government. The reports may, however, be used by NZ Government agencies as a reference in the development of policy.

While research reports are believed to be correct at the time of their preparation, the NZ Transport Agency and agents involved in their preparation and publication do not accept any liability for use of the research. People using the research, whether directly or indirectly, should apply and rely on their own skill and judgement. They should not rely on the contents of the research reports in isolation from other sources of advice and information. If necessary, they should seek appropriate legal or other expert advice.

Acknowledgements

The research team would like to acknowledge the following individuals for their specialist advice and assistance during the course of the main study stage of this research.

Bhagwant Persaud and Craig Lyon, for helping finalise the modelling methodology and developing the crash models.

Professor Graham Wood and Zamira Hasanah, for developing a methodology and software to enable model goodness of fit testing.

Fergus Tate for his valuable guidance and inputs into the methodology and data analysis at each stage.

The Steering Group members we would like to thank are:

- Glen Koorey – University of Canterbury
- Sandy Fong – NZTA
- Colin Brodie – NZTA
- Ian Appleton – NZTA
- Fergus Tate – now NZTA
- Glen Bunting – NZTA
- James Hughes – NZTA

In particular, we would like to thank Mike Jackett and Colin Brodie for acting as peer reviewers on the project. Their advice and continued support during this research has been much appreciated.

Abbreviations and acronyms

AADT	average annual daily traffic
ADT	average daily traffic
AIC	Akaike Information Criterion
AMF	accident modification factor
BIC	Bayesian Information Criterion
CAS	crash analysis system
IHSDM	Interactive Highway Safety Design Model
ISAT	Interchange Safety Analysis Tool
MTD	mean texture depth
NB	negative binomial
RAMM	road assessment maintenance management database
SCRIM	sideways-force coefficient routine investigation machine
SH	state highway

Contents

Executive summary

The majority of rural roads in New Zealand are two-lane undivided roads, even on the state highway network. A high proportion of fatal and serious crashes in New Zealand occur on these two-lane rural roads. The New Zealand government is addressing road safety on rural roads by investing in delineation improvements, roadside clearance programmes and strategic road realignment projects. However, there are limited road safety tools available to evaluate the value for money and safety benefits of remedial schemes, especially at an individual project level.

Traditional crash reduction studies using historic crash patterns are not always sufficient to identify safety deficiencies, particularly on low-volume roads and where recorded crash data does not adequately reflect the safety of individual sections of road. Crash prediction models (specific to New Zealand conditions) can be used to assess the number of crashes by type that might occur on a road segment given various road features. They can also be used to assess the change in crashes that might result from changes to these road features, as might occur in an improvement project. They provide a much more robust methodology for evaluating the safety benefits of road safety remedial schemes than the use of traditional crash reduction studies (based on historical crash data).

A three-stage process was used to develop the models for this project. In the scoping study (stage 1) the research team identified the key variables/data collection methodology and sample size required for developing the model. The pilot study (stage 2) involved testing the data collection methods identified during the scoping stage and refinement of the data set and collection methods based on the findings. This report provides the outcomes of the final stage (stage 3) of the process, which focused on the development of crash prediction models for New Zealand rural roads.

Suitable two-lane rural roads were selected from the entire New Zealand state highway network. Sections of rural highway that had a speed limit below 100km/h and included features such as narrow bridges, railway crossings and passing lanes were excluded from the final sample set. Once the filtering process was completed, a total of 6829km of state highway was selected for the modelling process. The state highways were then separated into curved and straight elements and entered into a relational database.

The prediction variables in the relational database were collected from four key data sources. The road asset maintenance and management (RAMM) tool was the source of the majority of the variables, including annual average daily traffic flow, seal width, minimum curve speed, approach speed, region and mean texture depth. The KiwiRAP dataset was used for a roadside hazard rating. Cleaned crash data and approach speed was gathered from a study by Koorey (2009a). Finally, access way data was collected manually by analysing NZ Transport Agency videos of a number of state highways.

Generalised linear crash prediction models were developed for key crash types including head-on, loss-of-control and driveway crashes. For head-on and loss-of-control crashes, an analysis of straight, curved and all sections was undertaken. For driveways the analysis was for both curved and straight sections combined. New Zealand was divided into five regional groupings, including Auckland, the West Coast and three super regions.

The model results indicated there was a higher number of loss-of-control crashes on straight sections of road with high traffic volumes. Furthermore, crashes occurred on wide, straight sections of road, where drivers were often travelling at higher speed and might be undertaking risky passing manoeuvres. Seal width and gradient had a less pronounced effect on crash rates on curved sections of road compared with straight sections. The key element affecting crash risk on curves was minimum curve radius. The highest occurrence of driveway crashes was in the Auckland region.

The model equations can be applied to assess the impact of safety measures on specific road sections. An example of this application is provided below, where the cumulative safety benefits of improving the road surface and mitigating the presence and severity of roadside hazards are calculated for a straight section of road.

	Base scenario		**Improve road surfacing**		**Mitigate roadside hazards**	
Model Parameters	**Straights – Loss-Of-Control**	**Straights – Head-On**	**Straights – Loss-Of-Control**	**Straights – Head-On**	**Straights – Loss-Of-Control**	**Straights – Head-On**
AADT	4000	4000	4000	4000	4000	4000
Length	500	500	500	500	500	500
Seal width	7	7	7	7	7	7
Gradient	0.02	0.02	0.02	0.02	0.02	0.02
KiwiRAP weighted severity rating	2.8		2.8		0.7	
% time SCRIM below threshold	0.6	0.6	0	0	0	0
% time MTD below threshold	0.6		0		0	
Injury crashes per year	**0.618**	**0.069**	**0.206**	**0.025**	**0.180**	**0.025**
			Crash reduction compared to Base		Further reduction in crashes	
			-67%	-64%	-13%	0%

The results indicate that substantial reductions in loss-of-control and head-on crashes can be achieved by improving the road surface. Further benefits can also be gained by reducing roadside hazards, but only for loss-of-control crashes. The results provide an insight into the benefits of particular measures at a project site level.

Part of this study was to assess how crash prediction models could be utilised within a safety evaluation tool. The team considered using the Interactive Highway Safety Design Module (IHSDM) suite of software from the USA. However, discussions between the study team and the US Federal Highways Administration (FHWA) identified significant barriers to updating the in-built models currently used by IHSDM to those produced in this study, thereby limiting its widespread applicability in New Zealand.

A key recommendation of this work is that a safety analysis software tool be developed to include the rural road crash prediction models. Such a tool needs to provide an easy-to-use interface for engineering practitioners. Investigations by the study team have led to the conclusion that this is best accomplished through spreadsheet-based software similar in structure to ISAT, a USA crash prediction tool for motorway interchanges.

This study identified further research topics to enhance the crash prediction models. These include:

- Undertake predictions, using the models, across the entire network (at overlapping 100km sections), move some sections that are outliers into other regions, or add new region-like categories and then rerun the models with roads reclassified into these new regions. This should ideally improve the fit of the models.
- Look at developing shorter approach speed profiles to better understand horizontal consistency, and test these in the models.

- Investigate the relationship between seal width and traffic volume and consider developing separate models for different flow bands.
- Look at special cases of curves, such as reverse and compound curves and develop crash prediction models for these.
- Look at the effect of delineation and how it differs within each of the regions. This may explain some of the inner-region variability and improve the fit of the model.

Abstract

The majority of fatal and serious crashes in New Zealand occur on rural two-lane roads. Data on historic crash patterns is not always sufficient to enable a suitable diagnosis of the safety deficiencies of various sections of this rural road network. It also cannot readily identify safety issues on low-volume roads and shorter sections of highway, where the relative scarcity of crashes may mask the considerable potential for proactive safety improvements.

This report presents the third and final stage of a study that aims to develop crash prediction models for two-lane rural roads using data from almost 7000km of the rural state highway network. The report builds upon the findings of stage 1 (scoping study) and stage 2 (pilot study) to determine the most important parameters affecting safety on rural roads in New Zealand. The models have quantified the mathematical relationship between crashes and traffic volumes, road geometry, cross-section, road surfacing, roadside hazards and driveway density. These crash prediction models have enabled a better understanding of how safety is impacted by these factors and allow an understanding of how they interact with each other. They can also be used to determine which improvements are best to reduce crashes, for example whether to realign a road, widen the shoulder or remove roadside hazards.

1 Introduction

1.1 Background

The majority of fatal and serious crashes in New Zealand occur on rural two-lane roads. The country has an extensive two-lane rural road network, much of which has relatively low volumes of traffic. Data on historic crash patterns is not always sufficient to enable a suitable diagnosis of the safety deficiencies of various sections of this rural road network. It also cannot readily identify safety issues on low-volume roads and shorter sections of highway, where the relative scarcity of crashes may mask the considerable potential for safety improvement.

Crash prediction models have been increasingly used in the identification and evaluation of road safety issues since the mid-1990s, when initial forms of crash prediction models were first developed for the majority of road intersections and links in urban and rural areas. Crash prediction models are useful tools for evaluating the crash risk of existing small road sections, and also for evaluating the benefits of changes to each road network. These benefits are recognised internationally and many countries have developed comprehensive crash prediction models for just these purposes.

However, the majority of crash model research and development in New Zealand and internationally has focused on one-off studies to investigate specific features, such as the impact on the safety of roadside hazards, curves, carriageway width and surface friction. There are no comprehensive models that contain all the salient features and in doing so allow various options or strategies to be compared.

1.2 Purpose of the research

The overall purpose of the research was to quantify the impact of all key road features on the safety of two-lane rural roads and develop robust crash prediction models incorporating these features. A three-phase approach or plan was proposed to achieve this purpose, as described below.

1. A scoping study, aimed at identifying the salient variables and data collection and initial sampling requirements
2. A pilot study, to test alternative data collection methods and to refine the sample size and budget requirements
3. The main study, to build the necessary models.

Land Transport NZ (now NZTA) funded the scoping study (stage 1) in 2006. The main objectives of the scoping study were to:

- Investigate current crash prediction models to determine which variables have been identified as important.
- Identify which road and traffic-related features could potentially be included in the rural road crash prediction model and identify what existing (electronic) variable sets might be available to quantify these features.

- Determine what data would need to be collected in the field to provide a full set of variables for the model.
- Develop a definition and data collection methodology for each variable.
- Develop a sampling framework for the pilot study and a preliminary sampling framework for the main study.

A scoping report was produced in July 2006 which addressed each of the objectives.

The Land Transport NZ research fund then funded the pilot study (stage 2) of this project in 2008/09. The main objectives of the pilot study were to:

- Collect data for each of the key road features specified in the scoping report for 500m x 200m rural road sections in the Waikato. New data (at the time not available electronically) was collected manually in the field.
- Collect ‘new’ data for a sample of the rural road sections using video data capturing techniques and then determine if this method was accurate enough to replace the manual data collection process in the main study.
- Develop preliminary crash prediction models for rural roads, for the main crash types.
- Estimate the sample size that would be required in the main study to produce national and regional rural road crash models.

The pilot study research report was produced in March 2011 (Turner et al 2011) and contained the preliminary models that were developed during this stage.

1.3 Research objectives

Building on the two previous studies (scoping and pilot studies), the research described in this report covered the final phase of the project. The purpose of this research was to develop the next generation of rural crash prediction models for two-lane rural roads.

The objectives were to:

- Update crash prediction modelling methods to align with recent developments overseas (for example the use of homogenous rural road sections rather than sections of a fixed length.
- Develop a relationship database containing all the key variables for rural roads. This would build on the database developed by Koorey (2009a) and create a database that could be used by university students and New Zealand researchers for further analysis.
- Develop the next generation of rural crash prediction models by crash type, which would include key road features that could be changed by engineers.
- Undertake some preliminary analysis on the effects of access density on the crash rate on higher volume rural roads.
- Investigate and recommend how these new models might be included in a New Zealand version of the Interactive Highway Safety Design Model (IHSDM) safety analysis software (by talking directly to

FHWA). Alternatively to recommend the form of a new software tool that could assist in the use of crash models in New Zealand (eg an Excel spreadsheet tool).

1.4 Report structure

This report is structured as follows:

Chapter 2 presents a detailed summary of the national and international literature review carried out as part of this study.

Chapter 3 outlines study methodology.

Chapter 4 describes the adopted sample selection procedures.

Chapter 5 discusses the data collection undertaken for this study and examines data distributions.

Chapter 6 discusses the data collection of the crash data.

Chapter 7 provides the modelling methodology and explains the various models developed.

Chapter 8 presents the key conclusions of the study and discusses future research needs.

Chapter 9 contains a list of publications cited in this report.

2 Literature review

2.1 Introduction

A review of national and international research documents on crash prediction models for rural roads was undertaken to identify the key variables used in overseas studies and to identify how these variables were collected or compiled and which modelling methods were used. In addition a review of the two key approaches to the sectioning of rural roads, namely segregation into homogenous road sections (eg 100m long) or into road elements (eg curve and tangents), was carried out. The first section of this chapter outlines research on rural roads undertaken in New Zealand. The second section of this chapter outlines a selection of international studies on rural roads.

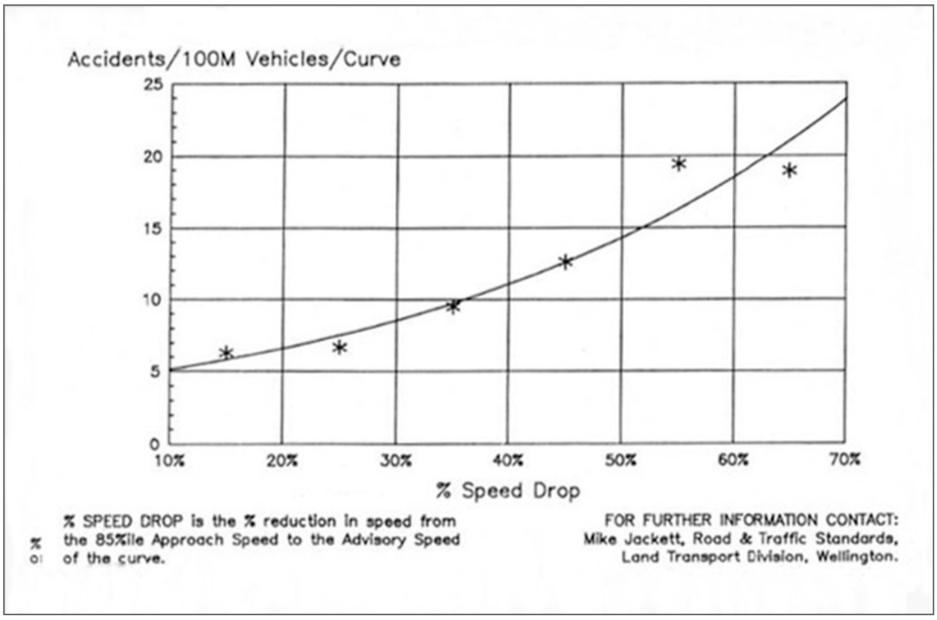
2.2 Rural crash models and relationships

2.2.1 Jackett (1992)

Using data from 990 rural curves on the following New Zealand state highways (SH) – SH1, SH2, SH6 and SH60 – Jackett (1992) developed a graph for relating loss-of-control and head-on curve crashes to advisory speed. The graph used the '%speed drop' from the estimated 85%ile approach speed to the advisory speed of the curve, as specified in equation 2.1, to predict the crash rate at the curve. This showed the number of crashes increased as the difference between the approach and advisory speed increased.

$$\text{\% speed drop} = \frac{\text{Approach Speed - Advisory Speed}}{\text{Approach Speed}} \qquad \text{(Equation 2.1)}$$

Figure 2.1 Crash prediction graph for curves (Jackett 1992)



2.2.2 Chadfield (1992)

In a study aimed at producing optimal cross-section guidelines for application on New Zealand roads, Chadfield (1992) investigated the impact of seal width, shoulder width and side slope by using the results of a VicRoads study (which actually used the results of a US study) and adjusting these for New Zealand conditions. No research was undertaken using New Zealand data. Instead the results from the VicRoads study were adjusted for differences in seal widths in Victoria and New Zealand (see table 2.1). Chadfield used crash rates per 100 million vehicle kilometres of travel and concluded that the difference in rates was attributable to variation in width and not traffic volume. No explanation for this conclusion was given. The same relationship has been used to adjust the rural crash prediction models in the NZTA's (2010b) *Economic evaluation manual* (EEM), volume 1.

Table 2.1 Base crash rate for average seal and lane widths (Chadfield 1992)

Traffic volume (vpd)	Seal width (m)	Lane width (m)	Sealed shoulder (m)	Crash rate (crashes/100 million vehicle km travelled)
<500	6.6	3.25	0	24.1
500–2500	7.5	3.5	0.25	21.5
2500–12,000	9.0	3.5	1.00	18.7

The graphs that Chadfield produced using his adjustment factors are shown in figures 2.2 and 2.3. These show a decrease in the crash rate with increased shoulder width and lane widths.

Figure 2.2 Sealed and unsealed shoulder widths and crash rates

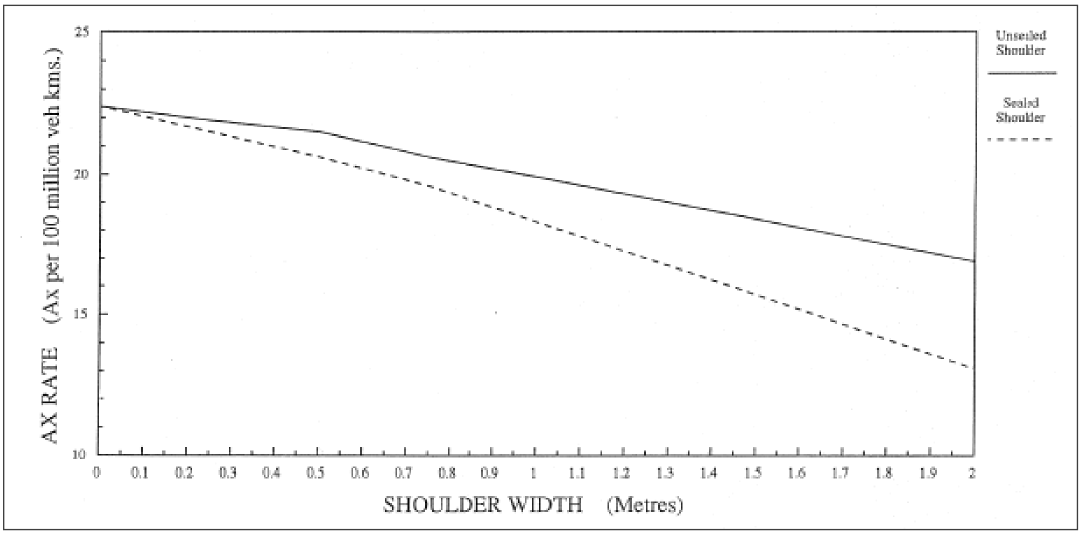
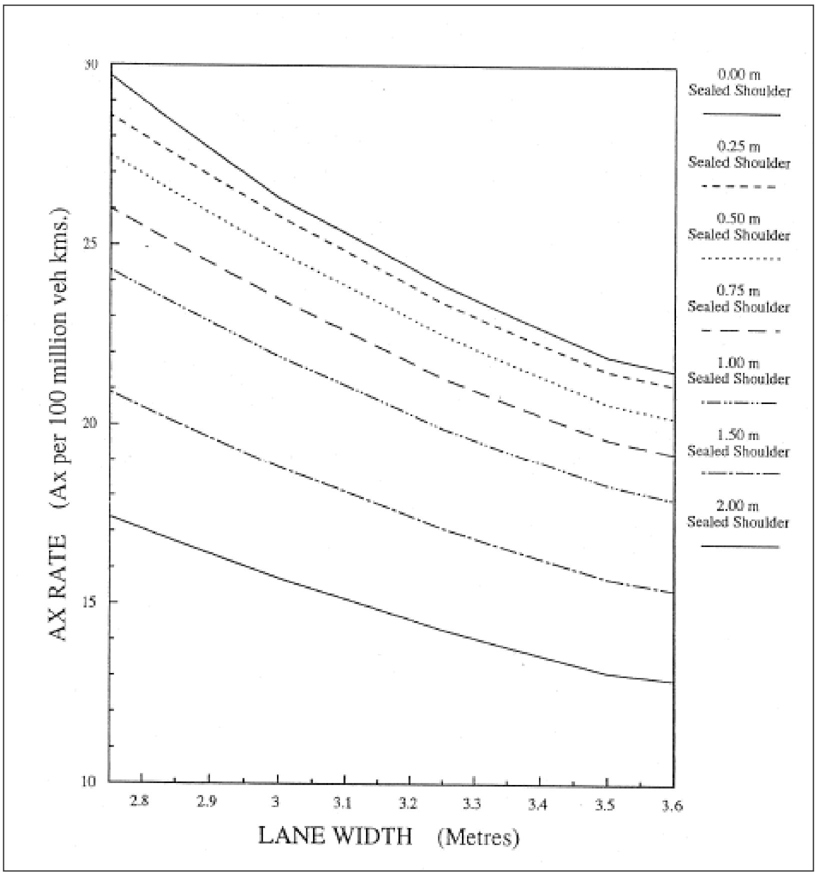


Figure 2.3 Lane and sealed shoulder widths vs crash rate (Chadfield 1992)



2.2.3 Koorey and Tate (1997)

As part of an update of the *Project evaluation manual* (now the *Economic evaluation manual* (NZTA 2010b)), Koorey and Tate (1997) investigated:

- the increased severity of crashes resulting from increases in mean speed
- the changes in crash rates resulting from improvements to horizontal curves.

For this study Koorey and Tate obtained highway geometry data for all sealed state highways, collected at 10m intervals. The data was then aggregated into 200m road segments. Crashes within 20m from an intersection were excluded from the dataset.

Using crash costs specified in the *Project evaluation manual* and the proportions of fatal, serious and minor injury crashes in the dataset for given mean speeds, Koorey and Tate produced graphs of average crash costs for various mean speeds. Relationships were then fitted to these graphs as shown in figure 2.4 and these found that average crash costs increased as the mean speed increased.

Koorey and Tate also developed a prediction model for the crash rate on 'curves' based on the reduction in mean speeds (see figure 2.5). They considered the crash rate on a particular road segment as a function of the difference between the approach speed environment and the likely speed on the segment under consideration. Although similar to the work by Jackett (1993), differences between the two studies were attributed to Jackett's measure being curves and Koorey and Tate's measure being 200m segments; and that Jackett used a subjective assessment of the approach speed. It was found that crashes on curves increased as mean speed reductions increased.

Figure 2.4 Average crash cost vs mean speed for 200m sections

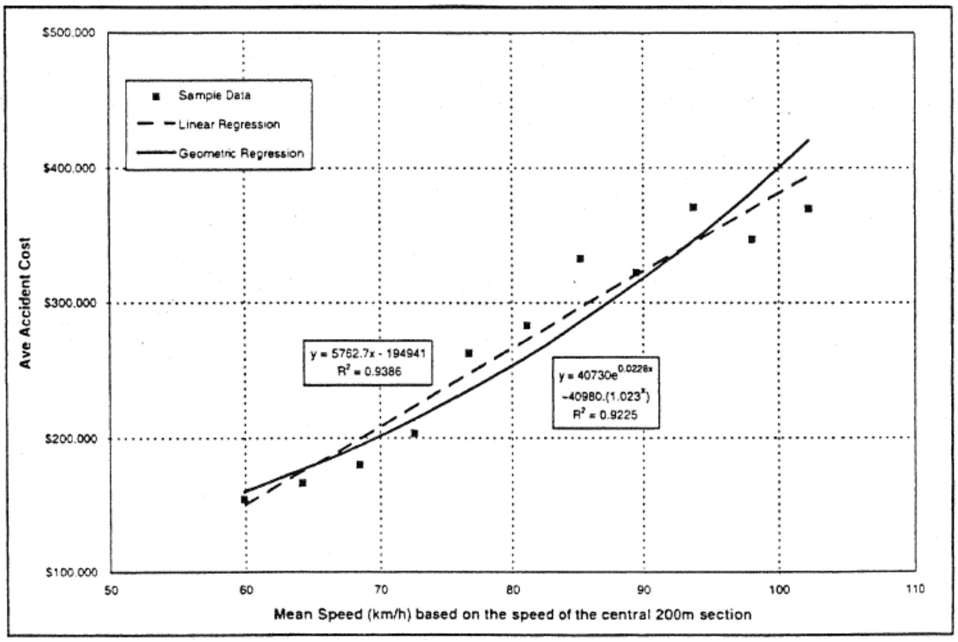
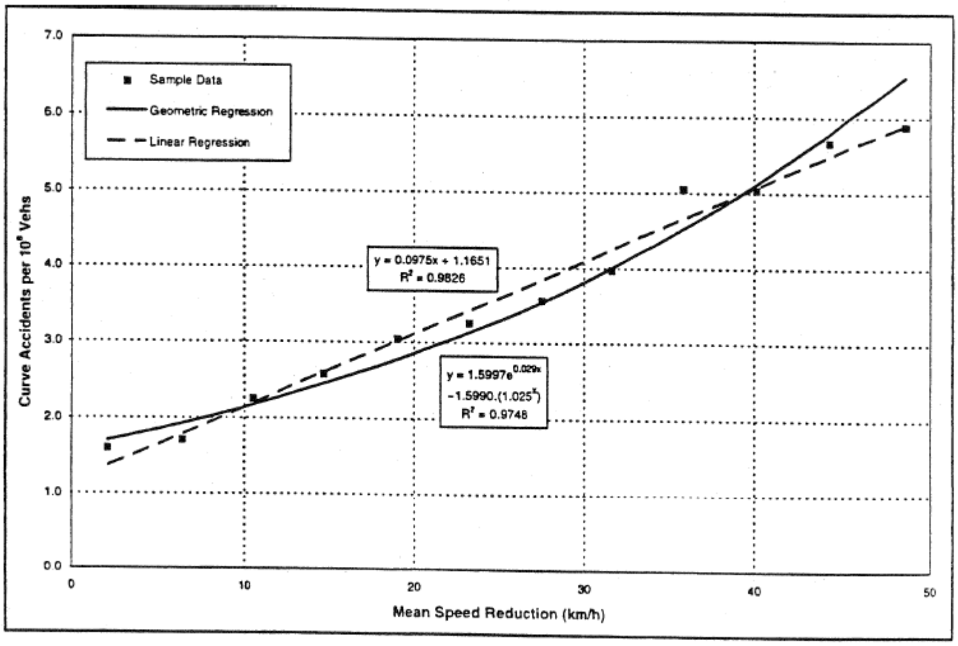


Figure 2.5 Curve crash rate vs reduction in mean speed



2.2.4 Turner (2000)

Turner (2000) developed crash prediction models for rural and urban intersections and links using generalised linear modelling techniques. For rural highways, excluding motorways and expressways, models for the mean number of mid-block injury crashes were developed.

Because the number of rural highway mid-block crashes is correlated with the terrain type (flat, rolling and mountainous) a terrain factor was included as a covariate in the crash prediction models, with sites being classified as flat or rolling. Insufficient data was available for highways in mountainous terrain.

Table 2.2 shows the model forms and crash types for the models Turner developed. These are typically of the following form, and have a Poisson or negative binomial (NB) error structure (equation 2.2).

$$A = b_0 Q_T^{b_1} \qquad \text{(Equation 2.2)}$$

where:

A annual number of reported injury and fatal crashes

b_0, b_1 parameters of the model

Q_T two-way traffic volume – average annual daily traffic (AADT)

Table 2.2 Rural highway mid-block crash prediction equations

Crash type	Crash codes	Equation (crashes per km)
Head-on	B	$A = b_0 \times Q_T^{b1}$
Overtaking	A	$A = b_0 \times Q_T^{b1}$
Rear-end (both straight)	FA to FF	$A = b_0 \times Q_T^{b1}$
Rear-end (one turning right)	GC to GE	$A = b_0 \times Q_T^{b1}$
Loss-of-control	C & D	$A = b_0 \times Q_T^{b1}$
Manoeuvring and hit object	M & E	$A = b_0 \times Q_T^{b1}$
Other		$A = b_0 \times Q_T^{b1}$
Total	All	$A = b_0 \times Q_T^{b1}$

Table 2.3 shows the model parameters by crash type. The exponents for the traffic flow variable (b_1) are distinctly different for each crash type, illustrating the importance of models for different crash types.

Table 2.3 Rural highway prediction model parameters

Crash type	b_0		b_1	Error structure
	Flat	Rolling		
Head-on	1.21×10^{-4}	1.57×10^{-4}	0.659	Poisson
Overtaking	4.16×10^{-7}	1.28×10^{-7}	1.308	NB (K = 2.2)
Rear-end (both straight)	2.18×10^{-8}	1.98×10^{-9}	1.724	NB (K = 1.7)
Rear-end (one turning right)	8.49×10^{-5}	1.25×10^{-5}	0.778	NB (K = 1.4)
Loss-of-control	5.67×10^{-3}	3.64×10^{-3}	0.483	NB (K = 1.2)
Manoeuvring and hit object	8.00×10^{-4}	2.14×10^{-4}	0.518	Poisson
Other	3.35×10^{-5}	2.45×10^{-6}	0.841	Poisson
Total	5.83×10^{-4}	2.40×10^{-4}	0.834	NB (K = 1.5)

These models show that crashes are not linearly correlated with traffic flow. Where the exponent (b_1) is less than one, crashes will increase at a diminished rate as traffic flow increases. This means that these crash types are less affected by variation in traffic flow. The opposite can be said where the exponent is greater than one.

2.2.5 Cenek, Davies and Henderson (2004)

Cenek et al (2004) developed crash prediction models for state highways using 11,000km of RAMM data (over 22,000 lane-km). Variables that Cenek et al investigated included:

- traffic volume
- road geometry (horizontal curvature, gradient and cross-fall)
- road surface condition (roughness, rut depth, texture depth and skid resistance)
- carriageway characteristics (Transit NZ (now NZTA) region, urban/rural environment).

The data on road geometry and road surface conditions has been collected in New Zealand by an instrumented vehicle since 1997. Cenek et al investigated four subsets of road crashes:

- all reported injury and fatal crashes
- selected injury and fatal crashes, covering loss of control events
- reported injury and fatal crashes occurring in wet conditions
- selected injury and fatal crashes occurring in wet conditions.

Cenek et al produced tables of crash rates for each of the key variables. Tables 2.4, 2.5 and 2.6 compare crashes with AADT, horizontal curvature and sideways coefficient routine investigation machine (SCRIM) coefficient. These tables generally show decreasing crash rates with increased AADT, horizontal curvature and SCRIM coefficient.

Table 2.4 All crashes classified by AADT

ADT Range	Road Length (km)	Number of Crashes between 1997 & 2002	Total Traffic Exposure (10^6 v-km)	Crash Rate (10^8 vkt)
ADT < 200	68	14	26	55
200 ≤ ADT < 500	650	111	517	21
500 ≤ ADT < 1000	2103	862	3268	26
1000 ≤ ADT < 2000	2646	1817	8323	22
2000 ≤ ADT < 5000	2538	3374	18144	19
5000 ≤ ADT < 10000	1485	3672	21548	17
10000 ≤ ADT < 20000	503	2329	14941	16
20000 ≤ ADT < 50000	109	660	5579	12
ADT ≥ 50000	0	0	158	0

Table 2.5 All crashes classified by horizontal curvature

Horizontal Curvature, R, (m) Range	Road Length (km)	Number of Crashes between 1997 & 2002	Total Traffic Exposure (10^6 v-km)	Crash Rate (10^8 vkt)
$10 \leq R < 100$	125	262	518	51
$100 \leq R < 1000$	2845	4277	17457	25
$1000 \leq R < 10000$	5273	6290	39620	16
$10000 \leq R < 100000$	1835	1973	14663	13
$R \geq 100000$	20	28	179	16

Table 2.6 All crashes classified by SCRIM coefficient

SCRIM Coefficient (SC)	Road Length (km)	Number of Crashes between 1997 & 2002	Total Traffic Exposure (10^6 v-km)	Crash Rate (10^8 vkt)
$SC < 0.3$	18	40	150	27
$0.3 \leq SC < 0.4$	294	730	3125	23
$0.4 \leq SC < 0.5$	2610	5144	28048	18
$0.5 \leq SC < 0.6$	4953	5421	32649	17
$0.6 \leq SC < 0.7$	2046	1287	7637	17
$SC \geq 0.7$	116	62	372	17

The tables indicate that crash rates are highest at low traffic volumes and imply a non-linear relationship, with a decreasing rate of crashes per vehicle as volumes increase. Crashes appear to be highest on roads with low values of SCRIM (ie low levels of skid resistance) and on low radius curves.

Cenek et al also developed generalised linear crash prediction models with a Poisson error structure. The main limitation with Poisson regression models was that overdispersion of crashes (variance greater than the mean) was not taken into account. Overdispersion can be taken into account by assuming a negative binomial error structure. Equation 2.3 shows Cenek et al's model form.

$$A = Q_T e^L \qquad \text{(Equation 2.3)}$$

where:

A annual number of reported injury and fatal crashes

Q_T two-way traffic volume (AADT)

L weighted sum of the values of various road characteristics

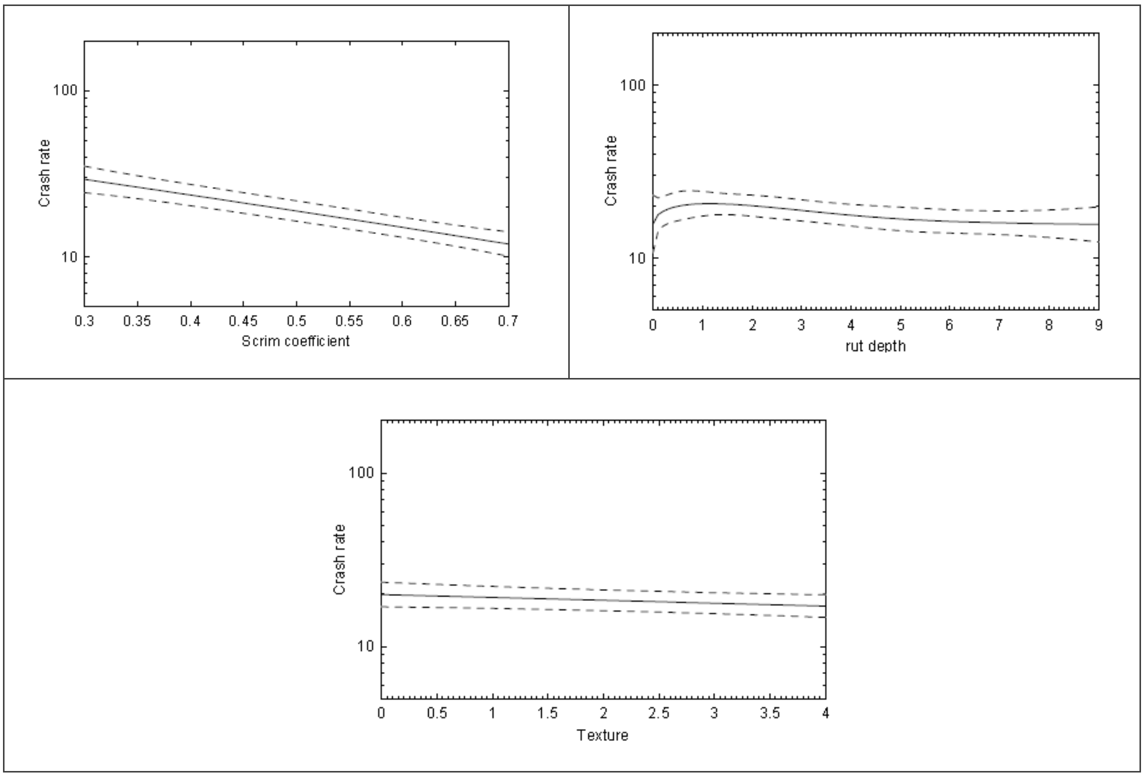
L is a function of the following variables:

- absolute gradient
- horizontal curvature
- cross-fall
- T/10 skid-site category (higher thresholds for some elements, eg railway crossings)
- skid resistance (SCRIM)

- log 10 (annual daily traffic (ADT))
- year
- Transit NZ (now NZTA) administration area
- urban/rural classification.

The models were developed for six years (1997 to 2002) and for seven Transit NZ regions. Figure 2.6 shows some of the relationships developed for road surface characteristics.

Figure 2.6 Relationship between crashes and surfacing characteristics



As with table 2.6 the crash prediction models show that the crash rate reduces as the surface friction (SCRIM) and texture improves. The relationship with rut depth is more complex. While not shown in figure 1.6, Cenek et al also found that crash rates increased as the road width increased. This same conclusion has been found by others. The relationship between crashes and seal width is a complex one, as it is affected by the correlation between traffic volume and seal width and also by the impact seal width has on driver speed, with higher speeds raising the risk of injury crashes.

2.2.6 Turner, Dixon and Wood (2004)

Turner et al (2004) investigated the impact of roadside hazards on the occurrence and severity of rural single vehicle crashes. A detailed roadside hazard inventory was undertaken on 850km of rural roads. Stratified sampling was carried out using a ‘hazard hit profile’ and six traffic flow groups.

The important causal variables were investigated using crash prediction models. Models were developed for three subsets of crashes: total single-vehicle hazard crashes, fatal and serious hazard crashes, and crashes where a specific hazard was involved. Equation 2.4 presents Turner et al's preferred model for all single-vehicle crashes involving a roadside hazard.

$$A = 1.01 \times 10^{-5} \times Q^{1.08} \times WDH^{0.65} \times ALM^{0.13} \times HVL_3^{0.16} \quad \text{(Equation 2.4)}$$

where:

A	reported injury and fatal crashes in five years
Q	two-way traffic volume (AADT)
WDH	sealed road width (m)
ALM	weighted length (m) of serious, moderate and easy curves per km
Weightings:	serious = 1
	moderate = 0.5
	easy = 0.25
HVL_3	length of severe hazards within 3m of the edge line or edge of seal per km

Equation 2.4 indicates that traffic volume, the location and type of roadside hazards, the consistency of the horizontal alignment and seal width are important variables in predicting crashes involving roadside hazards.

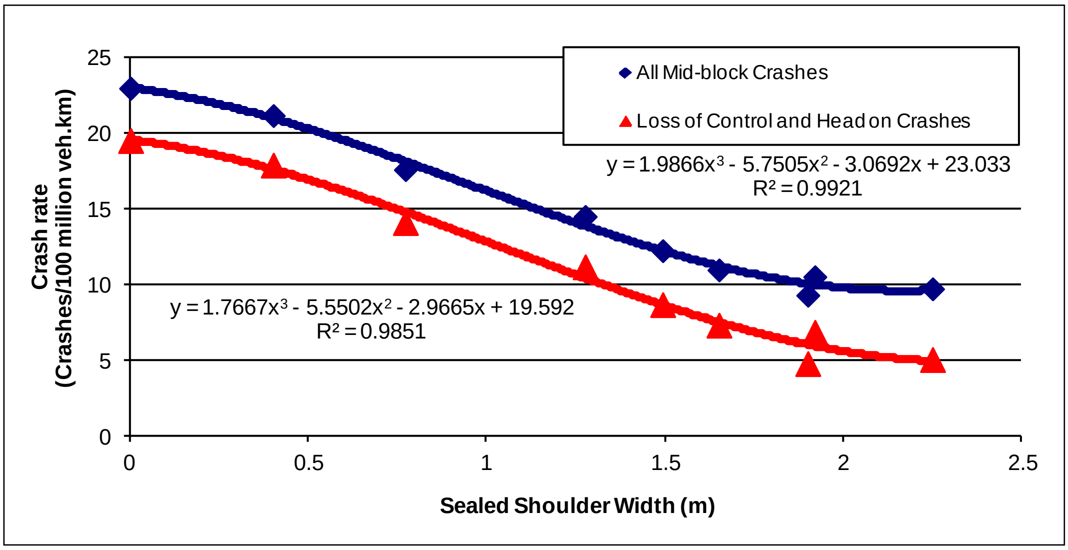
2.2.7 Tate and Turner (2007)

Tate and Turner (2007) sought to investigate the impact of shoulder width on rural mid-block crashes on two-lane state highway segments. A combination of the following was used to develop relationships between shoulder width and the risk of all mid-block crashes, and for a target set of loss-of-control and head-on crashes:

- traffic volume and crash occurrence per road centre-line vector from the crash analysis system (CAS), and
- traffic volume and sealed carriageway width, from which sealed shoulder width was estimated as: sealed shoulder width = (sealed carriageway width – 7.0m)/2.

While the relationship was similar in form to those found in international literature, concerns remained regarding the causal relationship and the fact that while roads with higher traffic volumes did have wider sealed shoulders, they also had a higher standard of horizontal alignment.

Figure 2.7 Crash rate as a function of sealed shoulder width (adapted from Tate and Turner 2007)



2.2.8 Landon-Lane (2007)

Landon-Lane examined how drivers adjusted their speed at low radius and low visibility curves. Six low radius and low visibility curves were selected for the study on state highways in Hawke's Bay. Topographical data was collected from each site to develop a digital terrain model. Road design software was used to extract information from the digital terrain model on road geometry variables and to calculate the theoretical sight distance and safe stopping sight distance. Theoretical models were used to predict the approach and curve negotiating speeds. Traffic counters were used to collect speed data and the classified traffic volume.

Landon-Lane found that observed speeds were closer to the design speed of the curve than the safe stopping speed, and that for both cars and trucks drivers generally travelled the curve well in excess of the safe stopping speed. There was no clear relationship established between exceeding safe stopping sight distance and crashes. This is perhaps understandable given the low sample size and the lack of sites with good to medium sight distance that could be used to undertake a comparison.

2.2.9 Turner and Tate (2009)

Turner and Tate (2009) examined the relationship between road geometry, observed travel speed and crashes on six 20km rural state highway sections around New Zealand. Data collected included speed profiles (of predominately young drivers) in both directions, a number of point speeds (of all drivers), traffic volume, radius of curvature, gradient, pavement cross-fall and seal width. The total data set included 488 curves and a total of 89 curve-related crashes (in five years).

The data was used to examine the relationship between road geometry and the speed choices made by the sample drivers. Using the point speed data the investigations were extended to include an assessment of the expected 85th percentile speed choice of the wider population. It was found that the speed chosen to negotiate a particular curve was more strongly related to the radius of the curve than to the design speed. In general the radius did not begin to effect negotiation speed until the curve radii fell below 300m.

The best model for predicting the negotiation speed of a particular curve was based on the radius of the curve and the speed environment measured over the preceding 500m. Two suitable models were developed for predicting the approach speed environment (over 500m). These were the bendiness ratio (degrees/km) and the mean advisory speed (a synthetic estimate derived from the radius of curvature and super-elevation).

A highly significant positive correlation was found between curve-related crash rates and the difference in the negotiating speed (85th percentile speed) and the design speed. However, due to the relatively small sample set, it was not possible to develop a robust crash prediction model relating curve crashes to the speed differential over the full range of observations, ie the model fit was poor over a section of the speed range.

2.2.10 Koorey (2009a)

The PhD thesis conducted by Koorey (2009a) focused on calibrating overseas rural road crash prediction models to New Zealand conditions. After reviewing available models and software platforms he used the US crash prediction models provided in the software tool, IHSDM, in his study. The advantage of IHSDM is that it has a built-in calibration tool that allows the various US states to calibrate the software for their jurisdiction. This same tool can be used to calibrate IHSDM to New Zealand conditions.

IHSDM is a tool that was developed by the USA Federal Highway Administration (FHWA) for assessing the safety impacts of geometric design decisions. It is used by engineers and planners to maximise the safety benefits of highway projects within the constraints of a site. It is made up of a number of modules (six in total), with the crash prediction module being the most relevant for this study.

The key variables included in the IHSDM crash prediction module are:

- traffic volume
- state covariate (0 for Minnesota and 1 for Washington)
- average lane width
- average shoulder width
- roadside hazard rating (seven-point scale based on Zegeer 1987)
- driveway density
- degree of horizontal curvature (degrees per 100 feet)
- absolute value of grade (percent)
- vertical curve rate (grade change/curve length).

To make a crash prediction, the layout and traffic operation conditions of each highway improvement option are entered into the software. Koorey has developed a number of routines to enter New Zealand-based data into IHSDM. The software predicts the crash rate at 10m intervals through each of the various road elements (eg curves and straights).

The main limitation of the IHSDM crash prediction module is that it does not include a design consistency variable. The crash prediction module therefore does not take into account the context of various road elements, eg whether a low radius curve is within a series of such curves or at the end of a long straight. As highlighted in other New Zealand research (Jackett 1992; Koorey and Tate 1997; Turner and Tate 2009) the consistency of the road design, and in turn speed environment, is an important factor. While it is not included in the crash prediction module, the IHSDM does have a separate design consistency module. Indeed the importance of this element on design is highlighted in having a separate module. However, the fact that it is not considered in the crash prediction module does limit the use of IHSDM on existing roads where there is often a problem with design consistency.

Koorey utilised the detailed road geometry and surfacing data collected by the SCRIM instrumented vehicles, RAMM data, traffic volume data from Transit NZ (now NZTA) and crash data from the CAS system to do the calibration. Once calibrated, he demonstrated the accuracy of the crash predictions from the software on several test highway sections, by comparing the observed crash rate with the pure IHSDM crash prediction and also with an adjusted IHSDM prediction that took into account the site's crash history (using the Empirical Bayes method). He found that many of the crash predictions for the test sections were well above or below the actual crash rates observed. Even with the local crash data, many of the sites had predictions significantly different from what was observed. This result really highlights the importance of developing New Zealand-specific crash models which include a design consistency factor.

2.2.11 Cenek and Henderson (2009)

Cenek and Henderson (2009) updated Cenek's crash prediction models (from 2003) following more detailed analysis of the geometric data from the instrumented vehicle and further 'cleaning' of the crash data. Over recent times, with the utilisation of GPS location data, the geometric data from the instrumented SCRIM vehicle has become more accurate than that used in the previous 2003 study. In addition, further work was undertaken in this research in terms of defining the location and length of road elements, such as curves, from the data collected at 10m intervals. In addition algorithms were developed to reposition crashes onto each highway element. There is inaccuracy in the recording of crash locations, which can be corrected in some cases using logic. For example, if a curve-related crash is positioned on a straight it is likely the crash actually occurred on a nearby curve. In this way some of the crash data was better located.

In this study models were developed for curves, as the focus of the work was on out-of-context curves. The key variables included in the model were:

- advisory speed
- curve length
- skid resistance (SCRIM)
- out-of-context curve effect (approach speed less curve speed)
- log 10 (ADT)
- year
- region.

The key relationships between each of the variables and crashes are shown in figures 2.8 to 2.11, where the rate of crashes is per 10^8 vehicle-km.

Figure 2.8 Crashes versus AADT

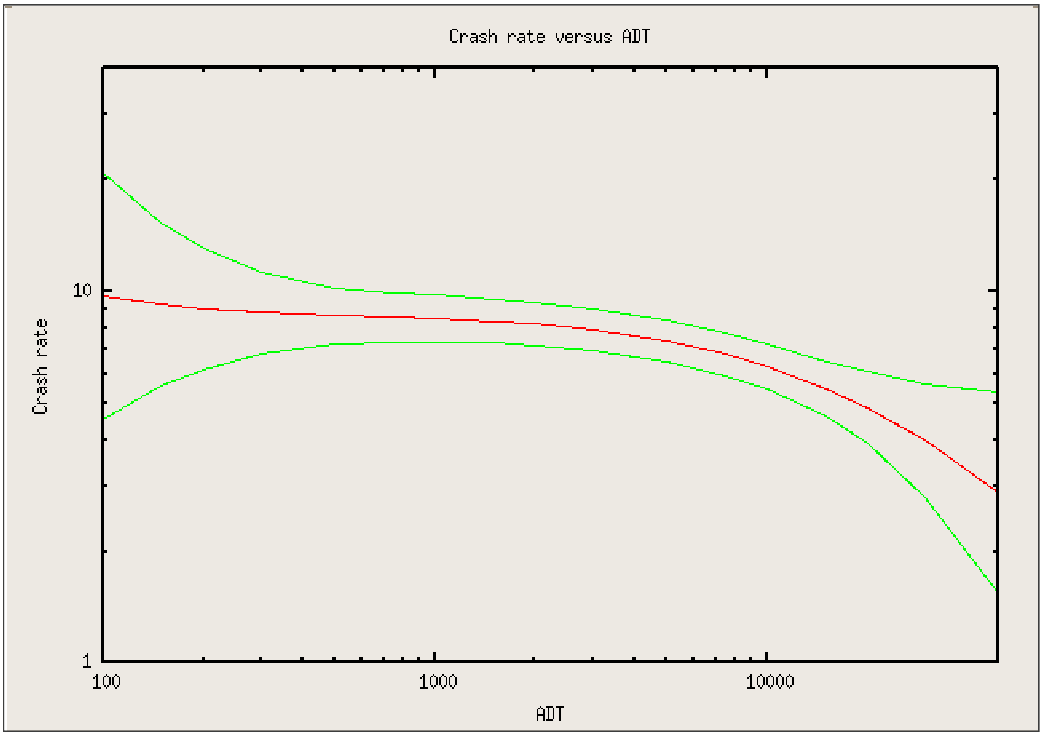


Figure 2.9 Crashes versus SCRIM (skid resistance)

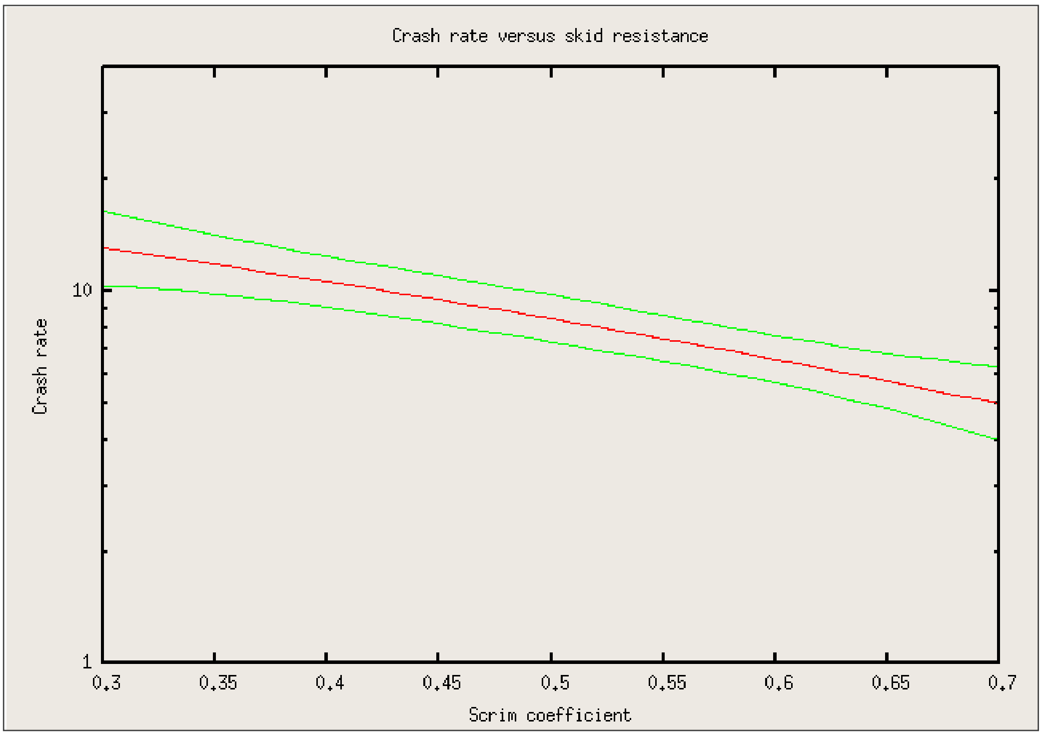


Figure 2.10 Crashes versus curve length

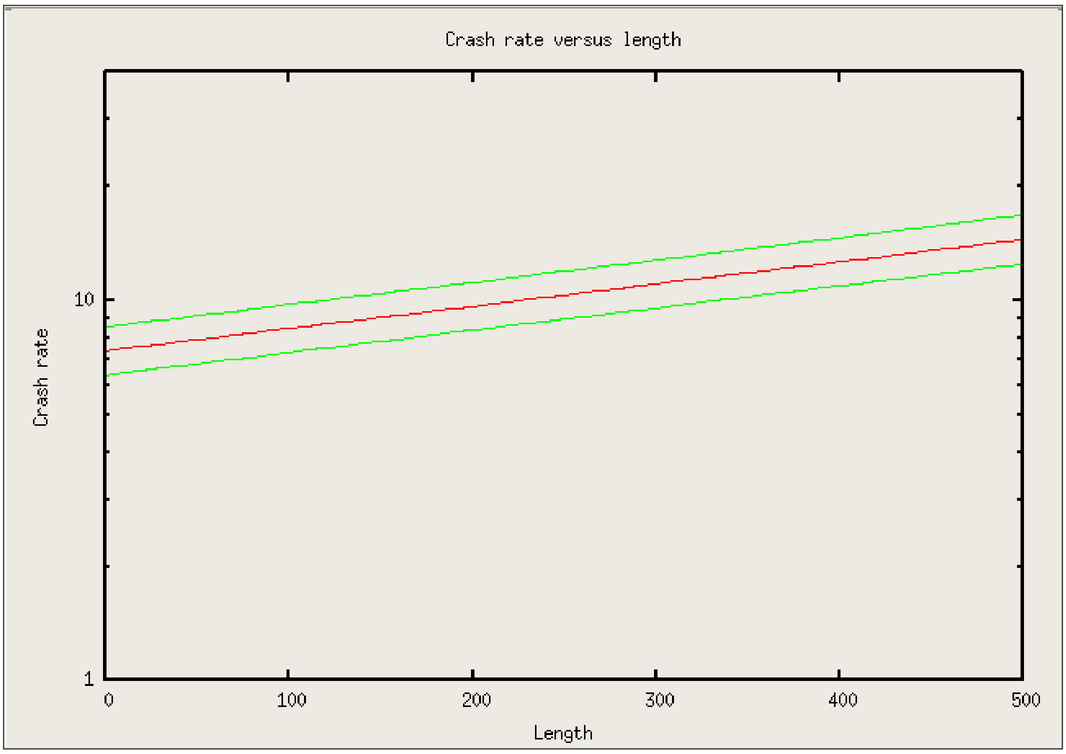
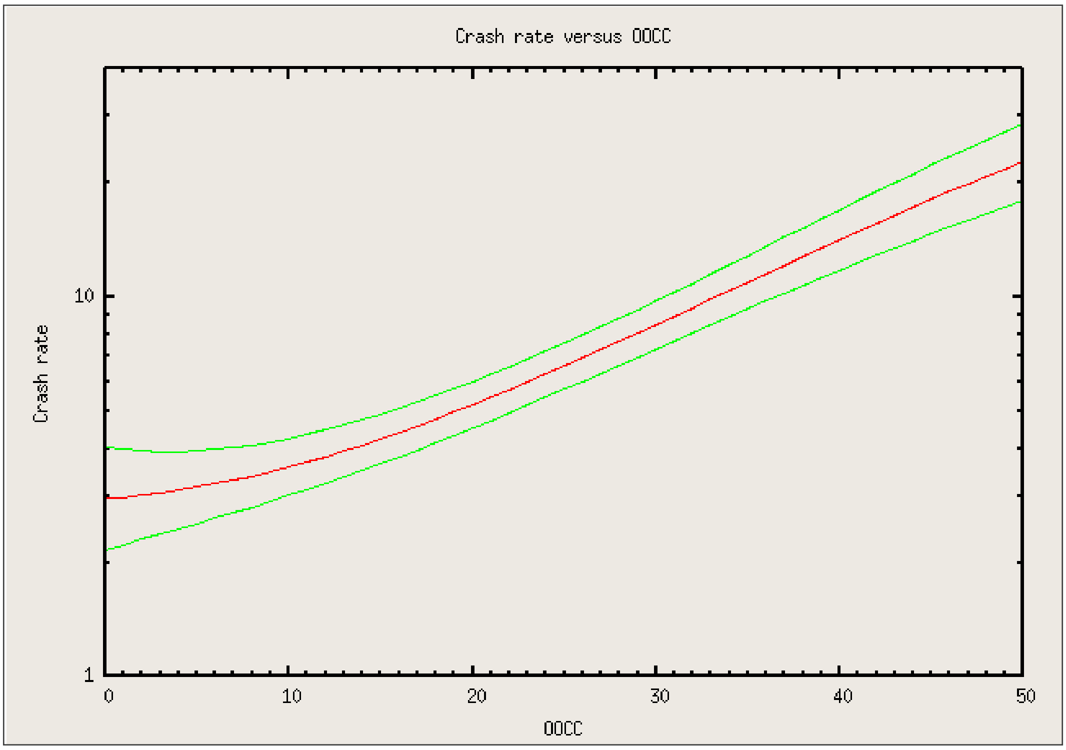


Figure 2.11 Crashes versus out-of-context curve effect



Figures 2.8 and 2.9 show that crash rates per vehicle-km reduce as traffic volumes increase and the skid resistance improves. They also show that crash rates increase for longer curves and where the approach speed is much higher than the (safe) curve speed (the out-of-context curve effect).

2.3 International rural crash prediction models

Numerous crash prediction models have been developed for rural roads internationally. The following studies were reviewed to determine the key variables that have previously been identified as being important, and areas are noted where we require further evidence to inform this research exercise.

2.3.1 Zegeer et al (1988 and 1992)

Charlie Zegeer and his colleagues produced pioneering research on crash models for rural road in North America. Two papers are of particular interest: *Accident effects of sideslope and other roadside features on two-lane roads* (Zegeer et al 1988) and *Safety effects of geometric improvements on horizontal curves* (Zegeer et al 1992). Both papers focused on what could be done to roads to improve safety. Zegeer developed crash prediction models and crash reduction factors from almost 5000 miles of rural roads and 10,000 curves across seven US states.

In the first study, on roadsides, the purpose was to a) develop a method for quantifying road-side hazards, b) determine what factors influence run-off-road crashes and c) estimate the benefits of various roadside improvements. The following variables were selected for the crash prediction models:

- traffic volume (ADT)
- horizontal curvature of each section (degrees of curvature)
- vertical curvature of each section (percentage gradient)
- side-slope length and ratio (eg 2:1)
- width of lanes and shoulders and shoulder type
- number of bridges, intersections, overpasses and railway crossings
- number of driveways (types, commercial, industrial, residential and recreational)
- type of delineation and on-street parking
- roadside hazard rating (see below).

A seven-point roadside hazard rating system was developed based on the likelihood of an object being a contributing factor in a crash and the severity of the crash. This scale was represented by a series of photos showing increased ratings as the likelihood and severity of the hazards increased. In addition to this subjective roadside rating an additional variable, roadside recovery distance, was also included. This is basically a flat, unobstructed and smooth area, from the edge of the shoulder (or edge of seal) to a side-slope or hazard. Two models were developed, one with the roadside hazard rating (equation 2.5) and the other with the roadside recovery distance (equation 2.6), as follows:

$$A = 0.0019\,(ADT)^{0.8824}\,(0.8786)^{Wft}\,(0.9192)^{PA}\,(0.9316)^{UP}\,(1.2365)^{H}\,(0.8822)^{TER1}\,(1.3221)^{TER2} \qquad \text{(Equation 2.5)}$$

$$A = 0.0076\,(ADT)^{0.8545}\,(0.8867)^{Wft}\,(0.8927)^{PA}\,(0.9098)^{UP}\,(0.9715)^{RECC}\,(0.8182)^{TER1}\,(1.2770)^{TER2} \qquad \text{(Equation 2.6)}$$

where:

A	selected crashes (single vehicle, head-on, opposite direction sideswipe, and same direction sideswipe) per mile per year
ADT	average daily traffic
Wft	lane width (in feet)
PA	average paved shoulder width
UP	average unpaved (gravel, earth or grass) shoulder width
H	median roadside hazard rating
TER1	1 is flat terrain, zero otherwise
TER2	1 is mountainous terrain, zero otherwise
RECC	average roadside recovery distance as measures from the outside edge of the shoulder.

Both models had a similar fit to the crash data, explaining just under 50% of the observed variation in the data. Using the models, Zegeer et al quantified the average reduction in crashes as a result of reducing the roadside hazard rating (up to 65% when the rating is improved by five levels) and the roadside recovery distance is increased (up to 44% when it is extended out 20 feet). They also showed the average benefit of increasing the lane width and shoulder widths, with and without changes to the roadside recovery distance. A separate model was developed for side-slopes.

The second study by Zegeer et al (1992) focused on the safety effects of geometric improvements to horizontal curves. A crash prediction model was developed for curves (equation 2.7), as follows:

$$A = [1.55(L) + 0.014(D) - 0.012(S)]\,(V)\,(0.978)^{(W-30)} \qquad \text{(Equation 2.7)}$$

where:

A	number of crashes on the curve per five years
L	length of curve
D	degree of curve
S	1 if spiral transitions present, zero otherwise
V	volume of vehicles through curve in million vehicles per five years
W	width of roadway (includes surface, inner and outer shoulder width)

This model showed the best fit to 10,900 curves in the state of Washington, with a R^2 value of 0.35. Zegeer et al showed that curve flattening could reduce crashes by up to 80%; lane widening, paved and unpaved shoulder widening could reduce crashes by up to 21%, 33% and 29% respectively; and the addition of a spiral transition could reduce crashes by approximately 5%. It was also shown that superelevation improvements could reduce crashes significantly.

2.3.2 Fitzpatrick, Park and Schneider (2008)

Fitzpatrick et al (2008) investigated how driveway density on two-lane and four-lane rural highways impacted on crash rates, known as accident modification factors (AMF). The *Highway safety manual* (ASSHTO 2010) identified driveway density as a highly influential variable when predicting crashes. Data was collected from 2756 miles in Texas which included the following key variables:

- driveway density
- lane width
- shoulder width
- segment length
- average daily traffic (ADT)
- median type (for four-lane highways).

The authors used a negative binomial regression model to develop the AMFs as it relates to driveway crashes and driveway densities. The generalised model for rural two-lane highways is as follows:

$$AMF = \exp[\beta\ (DD - base)] \qquad \text{(Equation 2.8)}$$

where:

β the regression coefficient for driveway density

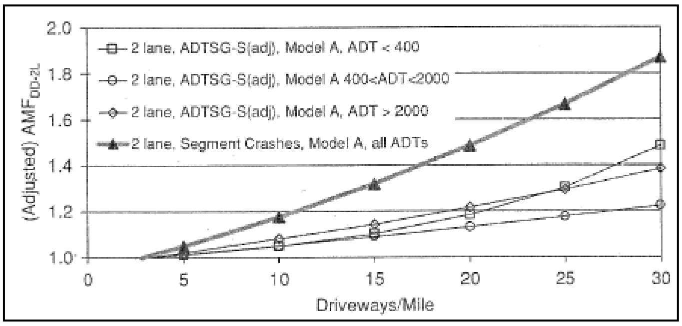
DD number of driveways per mile (driveway density)

base the base number of driveways.

The author investigated how the AMF varied with a single ADT versus discretising the ADT. A plot shown in figure 2.12 shows the model with a fixed ADT and an ADT that has been discretised into three sets:

- ADT < 400
- 400 < ADT < 2000
- ADT > 2000

Figure 2.12 AMF versus driveway density



This research showed that on a two-lane highway, the greater the number of driveways, the greater the risk of a crash occurring. The difference observed in the relationship for all ADTs and each ADT band was a result of differences in the crashes that were included in the various underlying crash prediction models, with some using all segment crashes and others based on crashes that occurred only at driveways.

2.3.3 Rengarasu, Hagiwara and Hirasawa (2009)

Rengarasu et al (2009) investigated how road segment segregation affected the accuracy of crash prediction models of rural roads. Crash data was collected from 1989 to 2005 from six national roads in Hokkaaido, Japan. The authors compared results from homogenous road segments with 1km fixed segments. The variables used were categorised by the following classifications:

- road geometry
- road cross section
- road structure
- traffic flow
- built-up area
- season.

Negative binomial regression models were used to develop the crash prediction modes. Table 2.7 outlines all the significant variables and identifies whether the variable increases or decreases the number of crashes as the variable increases or is included. Blank fields signify that the variable was not significant.

Table 2.7 Effect of variables on road crashes

Variable	Homogenous road segments	1km road segments
Bendiness	Increase	Increase
Hilliness	Decrease	Decrease
Maximum grade	–	Increase
Tunnels	Increase	–
Bridges	Increase	–
Maximum shoulder width	Increase	–
Average lane width	Increase	–
Number of lanes	Decrease	Increase
AADT	Increase	Increase
Winter	Decrease	Increase
Density inhabited districts	Increase	Increase
Railway station	Increase	–

It was found that the model with homogenous road segments had a greater accuracy than the 1km road segments. It was also found that the homogenous road segments had more significant variables.

2.3.4 Garber, Haas and Gosse (2010)

Garber et al (2010) investigated the safety performance function of the SafetyAnalyst tool (developed by the Federal Highway Administration) and developed functions specifically calibrated to the state of Virginia. SafetyAnalyst presents six safety performance functions that were calibrated with data from Ohio, North Carolina and Minnesota.

SafetyAnalyst models crashes as a function of AADT and segment length, as outlined in equation 2.9.

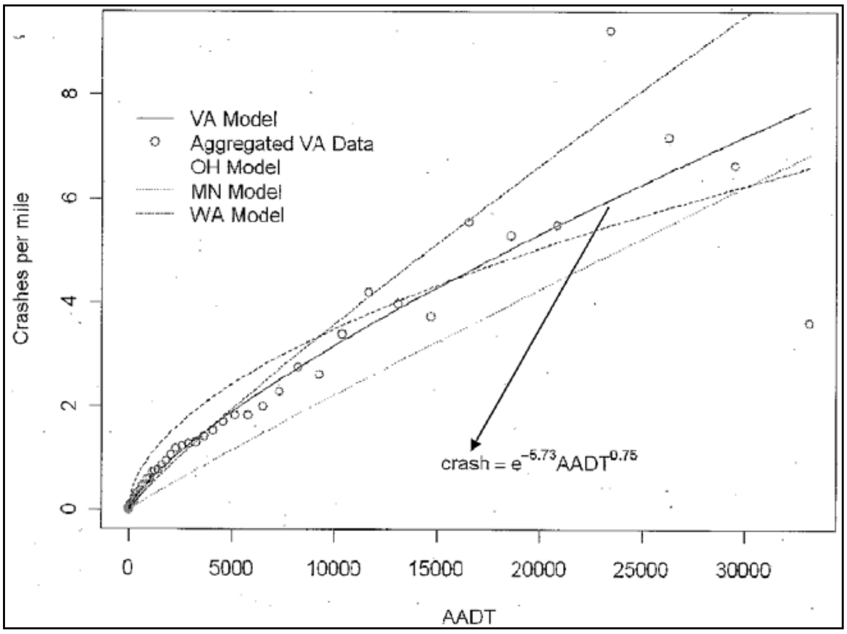
$$Crashes = \exp[a] \times [AADT]^{b} \times [SL] \quad \text{(Equation 2.9)}$$

where:

Crashes	predicted crashes per mile per year
AADT	average annual daily traffic
SL	segment length
a and b	regression parameters

The authors collected road and crash data from 139,635 sites (41,796 miles) along Virginia two-lane highways. The fitting of the model is shown in figure 2.13. The improved fit of the Virginia model confirms the need for area-specific models.

Figure 2.13 Crashes versus AADT



2.3.5 Montella, Colantuoni and Lamberti (2008)

Montella et al (2008) developed crash prediction models for rural roads in Italy. A 127.5km section of motorway A16 between Naples and Canosa was selected. A generalised linear model was used in the form of equation 2.10.

$$E(\Lambda) = L \times \exp[a_0 + a_1 \times \ln[AADT]] \times \exp[\Sigma_i [b_i \times x_i]] \quad \text{(Equation 2.10)}$$

where:

E(Λ)	predicted annual crash frequency
L	segment length
AADT	average annual daily traffic

a_i, b_i model parameters

x_i explanatory variables additional to L and AADT

The authors investigated 24 explanatory variables and developed two models. The first examined all crashes and found the following significant variables:

- inverse of the radius of the horizontal curve
- operating speed (85th percentile)
- length of the tangent preceding the curve
- difference between friction demand and friction supply on horizontal curve
- crashes in the year 2002
- total deflection angle
- upgrade (positive grade).

The second model only included severe crashes (injuries and fatalities). It was found that all the significant variables from the previous model (with the exception of the length of the tangent preceding the curve) were significant in this model too. Additional significant variables included:

- if a concrete barrier is present in the median
- if a curve is preceded by another curve or by a tangent shorter than 250m
- percentage of heavy traffic
- if a bridge is present on the section of road.

The Akaike Information Criterion and R^2 for the all crash models were 4792.6 and 54.2% respectively, and for severe crash models 5501.9 and 65.8% respectively.

2.3.6 Mattar-Habib, Polus and Farah (2008)

Mattar-Habib et al (2008) investigated the relationship between road consistency and probability of road crashes. A consistency model from a previous study (Polus et al 2004) related two-lane highway consistency with the bounded area between the profile and the average speed; and the standard deviation of speeds, as outlined in equation 2.11:

$$C = 2.808 \times \exp[-0.278 \times Ra \times \sigma] \qquad \text{(Equation 2.11)}$$

where:

C consistency of a highway segment

Ra normalised area bounded by average speed profile of cars and average operating speed

σ standard deviation of car speeds

An enhanced, integrated consistency model was developed to include variability of speed profiles with relation to hilly and mountainous roads, in particular for trucks and other heavy vehicles.

The Poisson distribution was used to model the road crash probability, with the Poisson parameter equating to the expected number of crashes. The expected number of crashes was modelled by a generalised linear model, as outlined in equation 2.12.

$$\lambda_i = \alpha AADT_i^{\beta} exp(\gamma_1[Length_i] + \gamma_2[RC_i]) \qquad \text{(Equation 2.12)}$$

where:

λ_i	expected number of crashes on segment i
α, β, γ_1, γ_2	model parameters
$AADT_i$	average annual daily traffic on segment i
$Length_i$	length of segment i
RC_i	road consistency of segment i

Two models were tested. The first was developed from 26 road sites in Northern Israel, and the second from 83 in Saxony, Germany. Figures 2.14 and 2.15 graphically illustrate the relationship between the average number of crashes and road consistency, and the crash probability distribution respectively. Figure 2.14 clearly shows that improvements to road consistency cause a reduction in the average number of crashes.

Figure 2.14 Average crash number per year vs road consistency

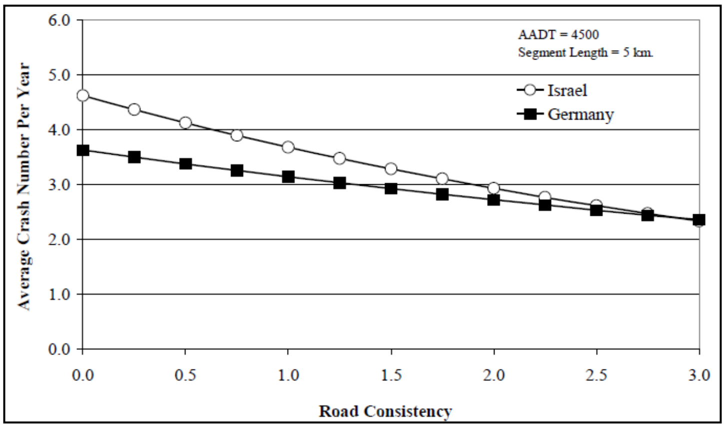
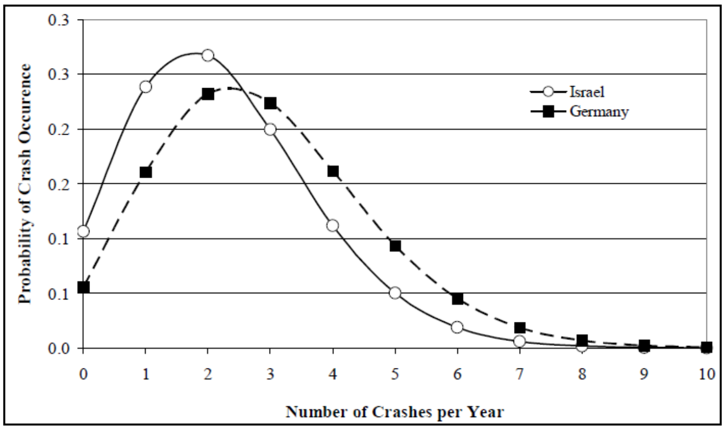


Figure 2.15 Probability of crash occurrence



2.4 Summary

The types of crash relationships and crash prediction models that have been developed in New Zealand and internationally and the key variables identified for each of the models are summarised in table 2.8. The research contributed to the selection of variables and development of models.

Table 2.8 Prediction model types and key variables

Study and model type	Country	Key variables and other findings
Jackett (1992) – models for isolated rural curves (figure produced)	New Zealand	Related crashes to % speed drop entering isolated curves. Found crashes increased as speed drop increased. Found consistency of alignment was a key factor
Chadfield (1992) – seal width (lane and shoulders) crash rates	New Zealand	Calibrated USA/Australian research on crash rates by volume and seal width to New Zealand conditions
Koorey and Tate (1997) – mean speed and speed drop of crashes and crash costs (includes severity)	New Zealand	Found crash costs (number and severity) increased as mean speed increased. Curve crash rates increased as mean reduction in speed between road elements (200m sections) increased.
Turner (2000) – flow-only crash prediction models for rural highway by crash type and terrain type	New Zealand	Negative binomial flow-only models developed for each crash type. Exponents for rear-end (other than one turning right) and overtaking crashes greater than one (increasing crash rate as volume increased). Less than one for all other crash types. Overtaking became more of a problem as volumes increased.
Cenek and Davies (2004) – crash prediction model including volume, road geometry and surface condition	New Zealand	Found that crash rates generally decreased with increasing AADT, horizontal curvature (radius) and SCRIM coefficient (skid resistance). Found crash rates increased with seal width.
Turner et al (2004) – crash prediction models with roadside hazard variable	New Zealand	Found that crash rates increased as the length of severe hazards increase in the first 3m from the edge-line or edge of seal. Also found crashes increased with severity of curves and seal width.
Tate and Turner (2007) – shoulder width and crash rates	New Zealand	Crash rates reduced as shoulder width increased.
Cenek and Henderson (2009) – crash prediction models for curves	New Zealand	Curve crashes reduced as traffic volumes increased and skid resistance improved. Crash rates increased as the curve got longer and speed drop increased. Also did work around improving accuracy of geometric and crash data.
Zegeer et al (1988) – multi-variable crash prediction models for rural mid-block	USA	Developed a five-point roadside hazard rating. Up to 65% crash saving if the poorest roadside rating could be improved to the best. Up to a 44% saving if roadside recovery distance was increased to 20 feet. Also crash savings from increasing the lane and shoulder widths.
Zegeer et al (1992) – multi-variable crash prediction models for rural curves	USA	Model showed that curve flattening could reduce crashes by up to 80%, while lane widening and addition of sealed or gravel shoulders could reduce crashes by between 20% and 33%. Spiral transition had a modest effect of a 5% reduction.
Fitzpatrick et al (2008) – examined driveway density	USA	Developed an accident modification factor for driveway density. Increase in the number of driveways increased the risk of crashes occurring.

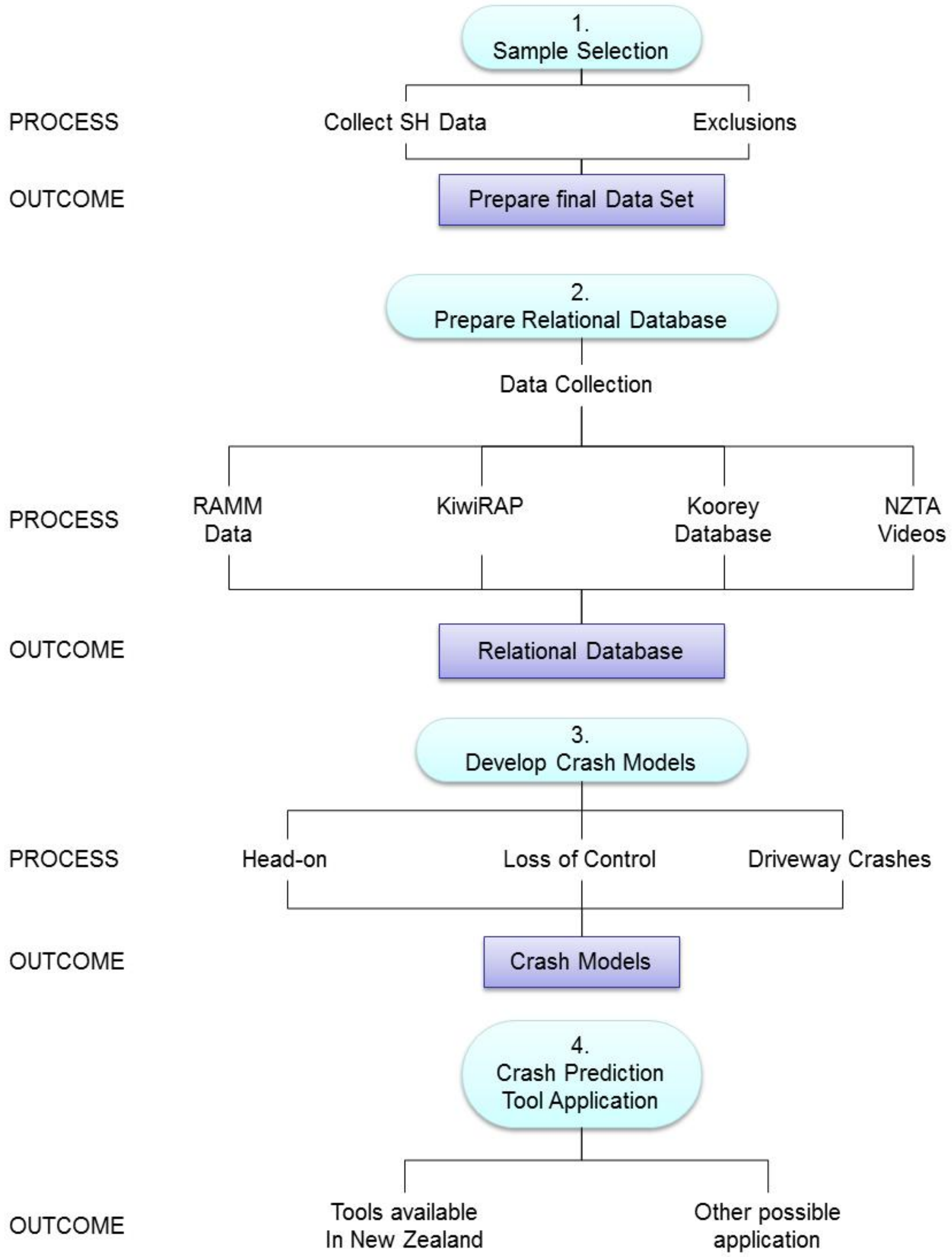
Study and model type	Country	Key variables and other findings
Rengarasu et al (2009) – compared models developed using homogenous road segments (eg curve) with those for 1km fixed lengths	Japan	Found that the model developed with homogenous road segments had a greater accuracy than the model based on 1km road segments. The following variables increased the number of crashes – AADT, lane and shoulder width, tunnels and bridges, increased bendiness and density of adjoining land-use. Hilliness and number of lanes decreased the crash rate.
Mattar-Habib et al (2008) – crash rate versus road consistency	Israel and Germany	Found that the average number of crashes reduced as road consistency improved.

3 Study methodology

3.1 Overview of the study methodology

This chapter provides an introduction to the overall study methodology. As previously outlined the aim was to develop crash models to evaluate the crash rate for individual sections of the rural road network. To achieve this aim, the research built on previous studies by using the methodology outlined in figure 3.1.

Figure 3.1 Overall study methodology



The following section provides a summary of the methodology for each of the four key stages.

3.2 Sample selection

The research aimed to develop crash prediction models for rural two-lane highways. Consequently, sections of the state highway lying in urban areas, on motorways, or having features such as narrow bridges and railway crossings were excluded from the sample set for this study. Further exclusions were required due to incomplete or erroneous data.

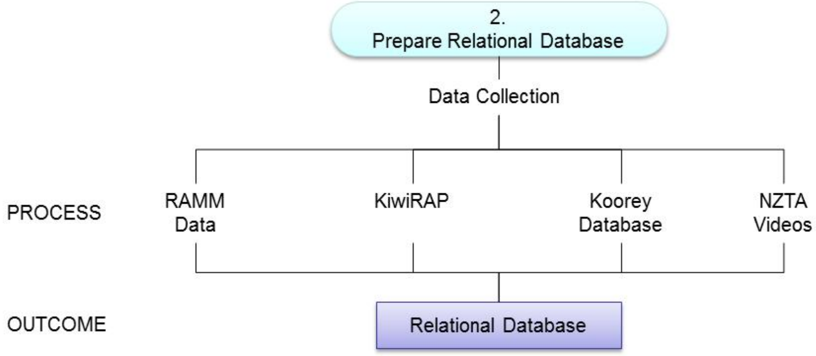
The sample selection process began with a total of 11,282km of highway data being available. Where data for variables were incomplete or missing, the corresponding elements were excluded (see section 4.5).

A total of 6829km of New Zealand state highways were selected for the modelling process following the sample selection process. Chapter 4 describes the procedures that were undertaken for selecting and excluding state highway sections.

3.3 Development of the relational database

The development of the relational database is outlined in figure 3.2. This process included collecting data from several sources.

Figure 3.2 Methodology for developing a relational database



The procedure used for data collection was an extension of the previous pilot study. Variables that were important in the pilot study were used in this study, with a few modifications. With the assistance of a panel of experts (as part of the steering group), many other variables were considered and were used where literature verified the inclusion of data.

The 6879km of state highway network were split into 30,577 straight and curve elements and predictor variables were found for each element. Data was collected from four electronic sources and processed using Visual Basic macros. The final processed data was stored in the relational database.

Crash data for the selected sections of the state highway was collected and stored in the relational database. The more detailed methodology for the data collection is outlined in chapter 5 and is followed by a breakdown of the crash data collection and analysis in chapter 6.

3.4 Development of crash prediction models

Ten crash prediction models were developed for loss-of-control, head-on and driveway-related crashes on straight and curved rural road sections. Chapter 7 describes the crash modelling methodology adopted and details the various models that were developed.

4 Sample selection

4.1 Sample selection procedure

The entire New Zealand state highway network was analysed to identify only those two-lane highways that lay in rural areas, had a speed limit of 100km/h and did not have narrow bridges or railway crossings located on them. Table 4.1 describes the various categories of exclusions that were implemented.

Table 4.1 Exclusion categories and descriptions

Exclusion type	Description of excluded sections
Carriageway	• Sections having more than or less than two lanes • Sections having a carriageway width greater than 12m • Sections lying in urban areas • Motorways and ramps
Speed limit	• Sections having a speed limit of less than 100km/h
Narrow bridges	• Sections with single lane bridges or narrow bridges
Railway crossings	• Sections with a railway crossing
Incomplete data	• Sections with missing or erroneous data

Sections 4.2 to 4.5 provide detailed descriptions of the methodology for exclusion of sections for each of the criteria listed in table 4.1.

4.2 Carriageway exclusions

The carriageway table in RAMM divides the entire New Zealand state highway network into sections of varying length and provides information on carriageway characteristics such as the number of lanes, carriageway width, surrounding environment (whether rural or urban), pavement type and traffic volume (AADT), if available, for each of the sections.

Data stored in the carriageway table was used to exclude sections of the state highway which were not considered relevant to the scope of this research.

4.2.1 Exclusion criteria

The criteria used for the exclusion of carriageway sections are described in table 4.2.

Table 4.1 Criteria for exclusion of state highway sections

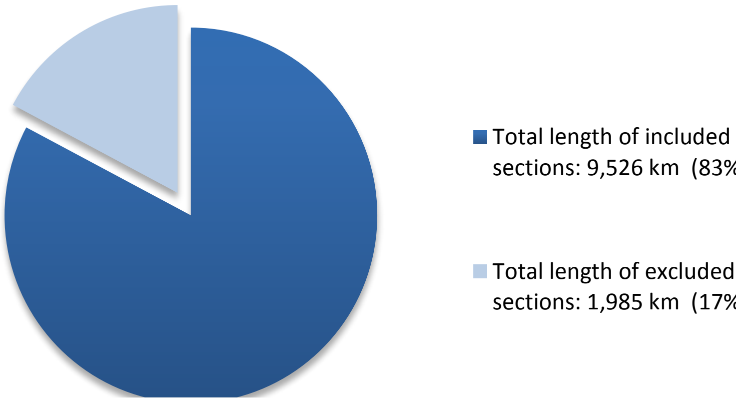
Parameter	Criteria	Description
Surrounding environment	CJNEX_urban_rural_CODE = urban	Sections of state highway lying in urban environments were excluded from the sample set. This was achieved by filtering the urban/rural classification fields in the Carr_way database.
Motorway sections and on/off-ramps	Cway_hierarchy = NSHS MOTORWAY	The carriageway hierarchy field in RAMM was used to identify SH sections that were part of a motorway or consisted of on and off-ramps.

Parameter	Criteria	Description
Sections with multiple lanes	Lanes >2 and lanes <2	Single-laned and three/four-laned sections and sections with passing lanes were excluded. This was achieved by excluding all carriageway elements with a number of lanes greater than or less than two.
Excessive carriageway width	Cway_width >12m	Sections associated with a carriageway width of more than 12m were also excluded from the sample set, since these are likely to have more than two lanes. Although most of these exclusions had already been made while excluding according to the number of lanes, some cases of two-laned sections with a carriageway width of more than 12m were also observed. Exclusion according to this additional criteria would thus be successful in removing these sections from the sample set.

4.2.2 Length of excluded sections

The RAMM database contains data for 11,511km of state highways. Out of this, 1985km of sections were excluded according to the carriageway criteria mentioned above. This resulted in a sample set consisting of 9526km of state highways. This is shown in figure 4.1.

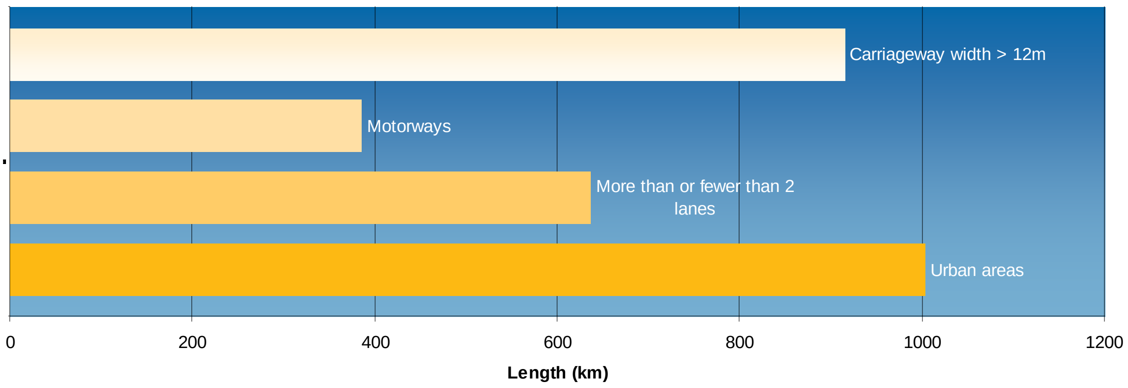
Figure 4.1 Length and proportion of excluded sections due to carriageway criteria



4.2.3 Carriageway exclusions by type

Of the 1985km of sections excluded, just over 1000km lay in urban areas; 637km of excluded sections had more than or fewer than two lanes, while 386km of sections were part of a motorway. A total of 914km of sections had a carriageway width of more than 12m. However, it is to be noted that a significant proportion of the 914km was already excluded on the basis of the number of lanes. Figure 4.2 shows a summary of carriageway exclusions by type.

Figure 4.2 Exclusions by type



4.3 Exclusions by speed limit

Although sections lying in urban areas were part of the carriageway exclusions detailed above, this only resulted in the removal of sections with speed limits of 70km/h or less. Some sections with speed limits of between 70km/h and 100km/h were still present in the sample set and had to be excluded.

Locations of state highway sections having a speed limit of less than 100km/h were extracted from an internal unpublished speed limit database maintained by the NZTA. The start and end route station (RS)/route position (RP) locations of state highway sections with a speed limit of less than 100km/h were extracted from this database and used for further exclusions in the sample set.

4.4 Excluding railway crossings and narrow bridges

4.4.1 Railway crossings

Locations of railway crossings on the state highway network were identified through data collected during pavement surveys conducted by the NZTA. During these surveys, the presence of a railway crossing on the state highway network was recorded in the form of an 'event'. An event code ('F') was assigned to the RS/RP of the location of the railway level crossing, and this was also subsequently recorded in the RAMM database.

At least one (and frequently more than one) entry was recorded for each lane (increasing/decreasing) on which a railway crossing was encountered during the survey. This resulted in multiple records in the database corresponding to the same railway crossing. All RS/RP locations in the database were thus manually analysed to combine rows which referred to the same railway crossing. A distance of 50m was excluded before and after each 'combined' location included in the database. This resulted in a list of unique railway crossing RS/RP location ranges for exclusion from the sample set.

4.4.2 Bridges

The locations of narrow bridges and single-lane bridges on the state highway network were extracted from the signs database in RAMM. This database contains a record of the location of various types of signs on

New Zealand state highways. The sign types in table 4.3 were used to determine the locations of narrow bridges.

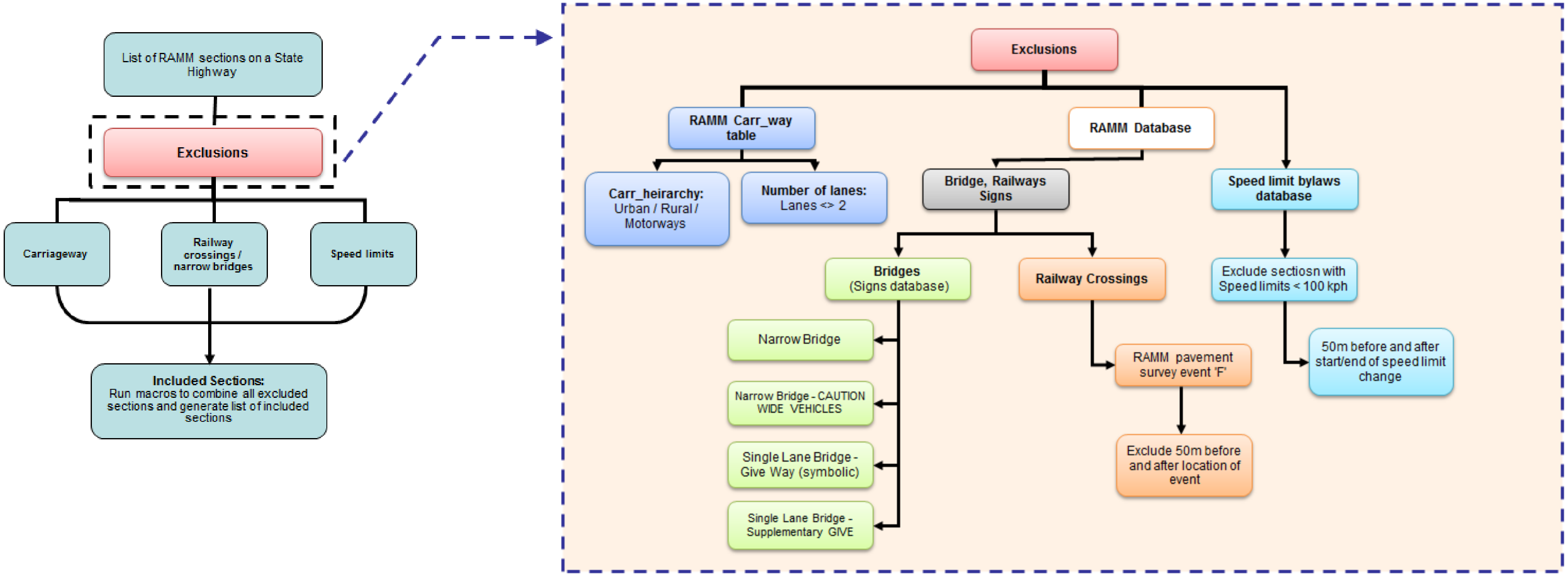
Table 4.3 RAMM narrow bridge/single-lane bridge sign types

Sign code	Sign description
PW44	Narrow bridge
PW44.1	Narrow bridge - Caution Wide Vehicles
RG19	Single lane bridge - Give Way (symbolic)
RG19.1	Single lane bridge - Supplementary Give Way

Similar to the case of railway crossings, a sign was present on both approaching directions (increasing/decreasing) for each narrow bridge or single-lane bridge. Multiple rows of the signs database thus referred to the location of the same bridge, and this was corrected by manual analysis to identify unique bridge locations. A distance of 50m was again taken before and after the locations of signs adjacent to the bridge, resulting in a range of RS/RP locations for exclusion from the sample set.

Figure 4.3 summarises the various categories and mechanisms of the exclusions.

Figure 4.3 Section exclusion process



4.5 Further exclusions

4.5.1 Data incompleteness

The state highway network was grouped into one of the following element types: curves or straights. Where a section of an element contained missing data or errors, the whole element was excluded. This is further explained in section 5.2.

4.5.2 Short section lengths

Elements with lengths less than 40m were excluded. Short straights could be misinterpreted as inflections in reverse curves or part of a compound curve while small curves were removed because of the very small number of crashes on them (relating to crash location error). This is further explained in section 5.2.

4.6 Final data set

The methodology detailed above resulted in a list of excluded state highway sections by category.

Individual state highways were separately analysed to identify sections for exclusion. The high-speed geometry table from the RAMM database (which divides each state highway into 10m sections) was used to identify those 10m sections that were to be included in the final sample set for this study. This process of exclusion by 10m sections also allowed detailed road curvature and alignment data to be directly extracted from the RAMM database at a later stage. Macros were developed for automating the exclusion process.

A total of 4435km of state highway sections were excluded, resulting in a sample set of 6829km (61%) of state highway for which further data was collected and analysed as part of the model development stage of this study. Appendix A provides a summary of the total length and excluded length of each state highway.

The spreadsheet that resulted from the sample selection provided the baseline data which was enhanced during the data collection phase, as outlined in chapters 5 and 6.

5 Data collection – traffic and geometry

The included sections of the New Zealand state highway network (6829km) were split into 17,087 curved elements (2195km) and 13,490 straight elements (4634km). A relational database was developed in Excel which outlined all the elements of the state highway network with their corresponding variables.

This section describes the procedures used for data collection and the data sources. Variable development is outlined within the remainder of this chapter. The final data set contributed to the completion of the relational database, which would then be used for the crash prediction modelling process.

5.1.1 Data sources

This study drew upon data from four sources:

1. RAMM database contains roading data for road assessment and maintenance management for New Zealand state highways. Most of the data used in this study was derived from this database.
2. KiwiRAP is a road assessment programme which is associated with the International Road Assessment Programme (iRAP). The data from KiwiRAP was used in assessing the level of roadside hazards along the state highways.
3. Koorey database was compiled as part of Koorey's thesis (2009). 'Cleaned' crash data for years 2002 to 2006 and approach speed data was used from this database which originally sourced its data from CAS and the RAMM database.
4. NZ Transport Agency videos were used to identify the location of accessways along certain sections of state highways.

5.1.2 Selected variables

The pilot study used 28 variables to develop crash prediction models from 200 separate 400m rural road sections. It found the following 10 variables to be significant:

- traffic flow (AADT)
- unsealed shoulder width
- seal width
- combined point hazards
- combined accesses
- distance to non-traversable slope/perpendicular deep drain
- average absolute gradient
- average curvature
- SCRIM coefficient
- horizontal consistency (percentage change in speed).

The main study, as reported in this document, used 12 variables from the pilot study. The relationships between the variables in both studies and the reason for any differences are shown in table 5.1.

Table 5.1 Differences between main study and pilot study variables

Variables used in main study	Variables used in pilot study	Reason for discrepancy
Traffic flow (AADT)	Traffic flow (AADT)	Same variable used.
Seal width (m)	Seal width (m)	Same variable used.
Average absolute gradient	Average absolute gradient	Same variable used.
SCRIM coefficient	SCRIM coefficient	Same variable used.
Minimum curve radius (m)	Average curvature (inverse of radius) (m^{-1})	Minimum radius identified the tightest component of the curve which is believed to better correspond to the level of crashes, as opposed to the average radius.
KiwiRAP – roadside hazard ratings	Number of point hazards Distance to non-traversable slope/perpendicular deep drain	KiwiRAP was recommended in the Pilot Study since the two variables used were found to be highly correlated and an integrated variable was needed. KiwiRAP data also requires far less time to collect than the other variables and allows for a much larger sample size.
Approach speed (km/h) (and curve radius, as above)	Horizontal consistency (percentage change in speed)	Horizontal consistency was found to give illogical results, with vehicles speeding up as they enter curve elements or slowing down into straight elements. Approach speed was chosen as it is a surrogate to horizontal consistency.
Accessway trips (per day)	Number of accesses along each 200m section	The number of accesses was found to under represent the vehicle activity manoeuvring around driveways. Instead, the total number of trips originating from driveways was used.
Straight and curve elements Element length	Not used in pilot study	One of the objectives of the main study was to use homogenous rural road sections as opposed to fixed length sections, which are represented by these variables.
Regions	Not used in pilot study	This variable was included to distinguish the differences in the quality of state highways within New Zealand regions. Refer to section 5.6.
Mean texture depth	Not used in pilot study	A macro surface texture variable was identified to complement the micro texture variable (SCRIM)
Not used in this study	Unsealed shoulder width	This variable was not included in this study as it is not readily available in electronic form.

The following sections present the data collection and variable distribution for the exhaustive set of model variables.

5.2 Element type and element length

Here state highway sections were grouped into homogenous straights and curves (referred to as element types in this report). The pilot study used fixed-length homogenous elements. Recent research by Rengarasu et al (2009) looked at both homogenous sections and fixed-length sections when developing crash prediction models and found homogenous sections performed better. Hence homogenous sections were used.

A method developed by Cenek et al (2004) grouped 10m adjacent sections into curve and straight elements. The 10m segment data was derived from the RAMM database. Straight and curve elements were defined using a 30m rolling average of the radius and the sign (or direction) of the radius. Curve segments had an average radius less than 800m and all three 10m segments had the same directions. A straight segment occurred when either the rolling average was greater than 800m or there were differences in the sign of the radius.

Where consecutive segments were classified as the same type, these sections were grouped together and considered a single element.

Two special cases of curves occurred: when there was a sudden change in direction of the curve (reverse curve) or a small straight segment separated two curves (compound curves). These special cases would both produce very small straight elements as a result of using the above methodology. These straight elements were not considered true straights but rather a product of the special cases. For this reason, straights less than 40m were omitted. Another reason for removing these short straight elements was that crashes occurring on these elements ran the risk of being recorded on neighbouring elements.

The proportion of each element type is shown in figure 5.1. Straight and curve elements correspond to 4634km and 2195km of the state highway network sample set respectively. Figure 5.2 shows the distribution of the lengths of each element type.

Figure 5.1 Proportion of straight and curve elements

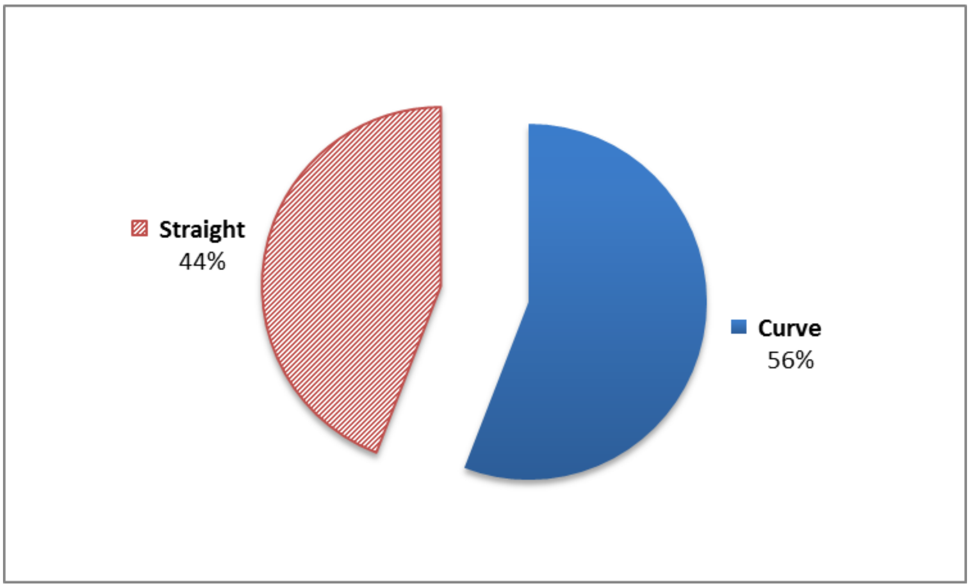
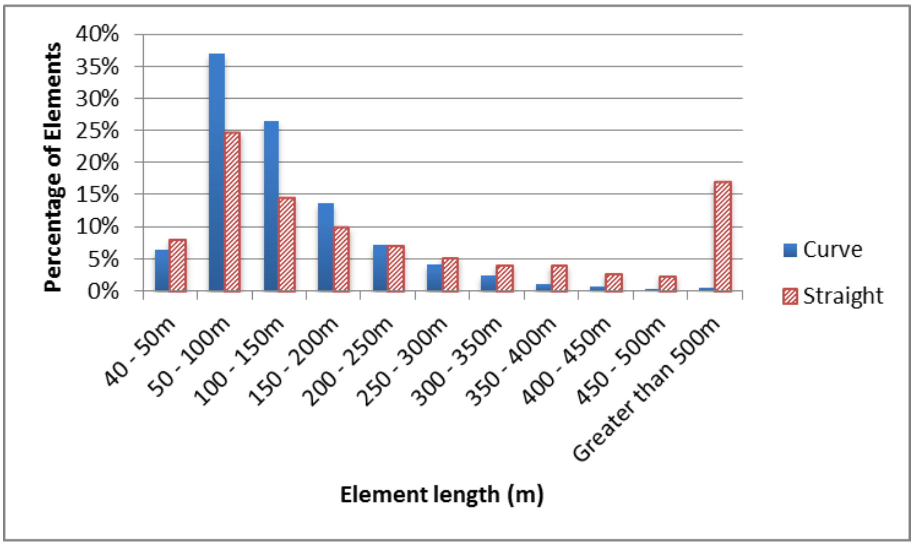


Figure 5.2 Distribution of element lengths

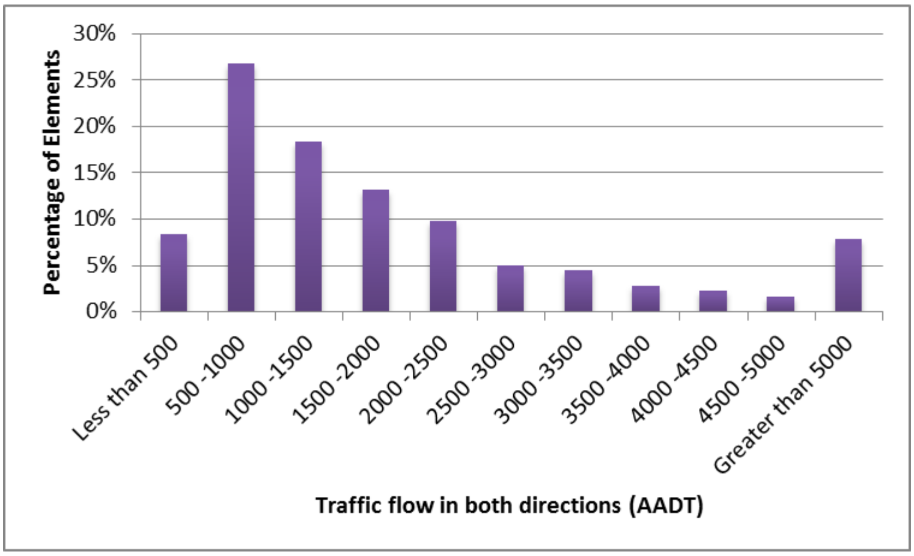


The median element length is 124m for curves and 169m for straights. Long straight elements between 500m and 1km comprise 7% of the straight elements, with 10% of these having a length greater than 1km.

5.3 Traffic flow

Traffic flows as AADT were collected from the RAMM database for both directions of travel. Figure 5.3 illustrates the distribution of traffic flows over all the elements in both directions.

Figure 5.3 Distribution of average annual daily traffic



The median traffic flow is 1410 vehicles/day and two thirds of the elements have flows with fewer than 2000 vehicles/day.

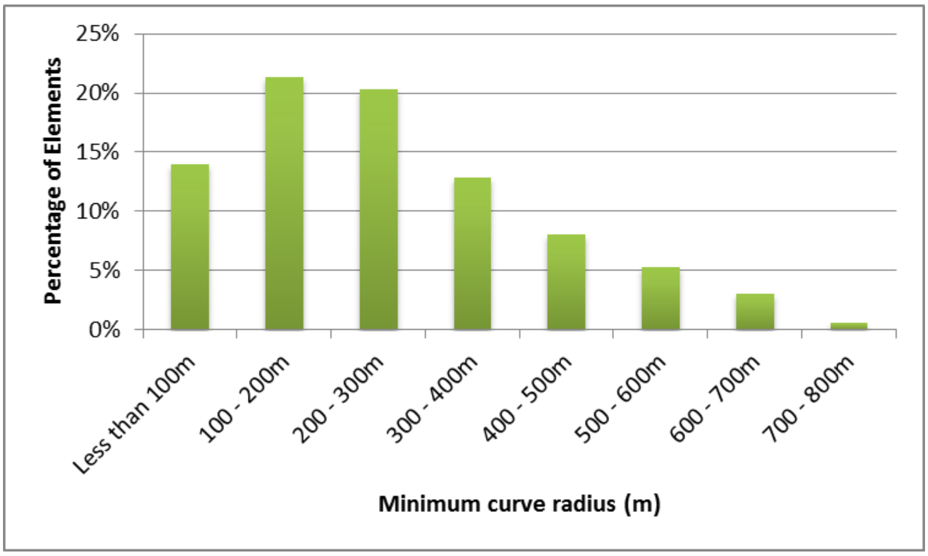
5.4 Road geometry

5.4.1 Radius

The radius of a given element was derived from the RAMM database and is represented by the minimum radius within the element. The pilot study used an average radius which was a result of using fixed length sections (where sections included both straight and curve elements). Since curve elements were isolated in the main study, the tightest point of the curve could be identified and used.

Figure 5.4 illustrates the distribution of minimum radii associated with the curve elements. Straight elements are not displayed because they are much greater than curve elements (800m). Radii are also not considered important for crashes on straight elements, therefore they are not represented in figure 5.4.

Figure 5.4 Distribution of curve radii



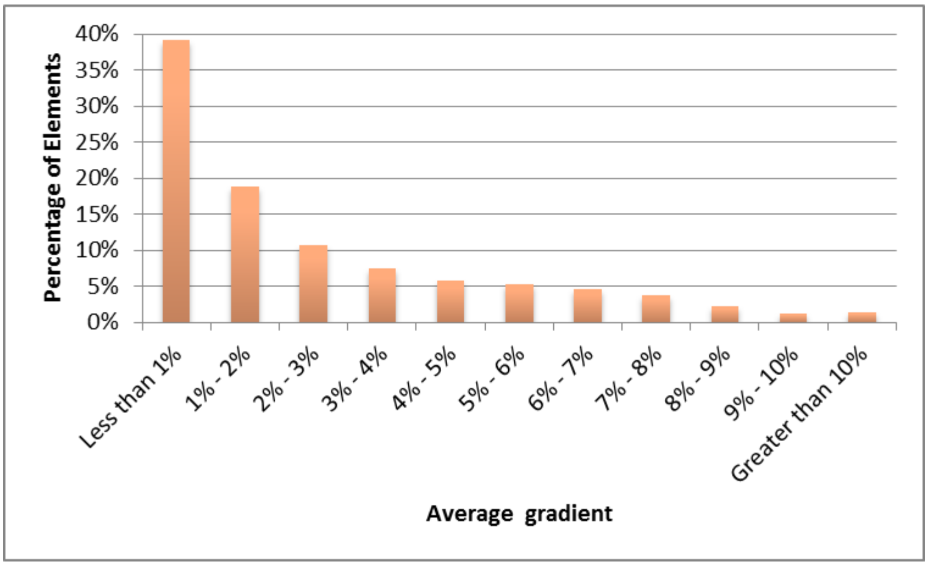
The median minimum curve radius is 220m.

5.4.2 Gradient

The gradient of the elements was derived from the RAMM database and is represented as a percentage. Figure 5.5 illustrates the distribution of the average gradient of an element as an absolute value.

One of the issues with this variable is how to analyse the case when the element crosses undulating hills. Where there are combined ascents and descents, the average gradient will be small. An alternative to this is to add all the gradients in an element together and treat descents as ascents. This would represent the 'steepness' of the element but would be large for very long elements. To produce a more balanced representation, the average gradient was chosen and represented as a scalar value.

Figure 5.5 Distribution of gradients



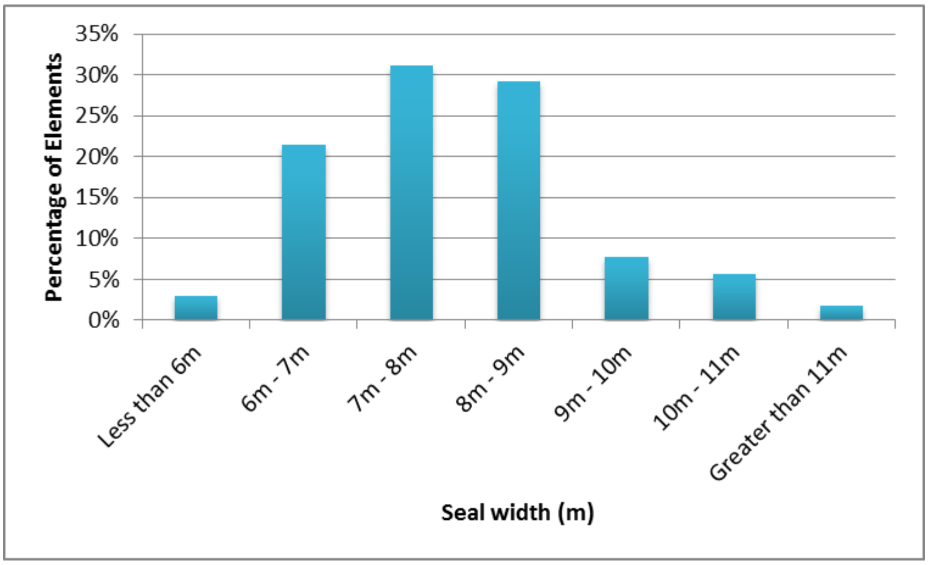
Up to 58% of elements have an average gradient of less than 2%, which is considered quite flat.

5.4.3 Seal width

The seal width of an element was collected from the RAMM database. The measurement taken for the seal width is illustrated in figure 5.6. Figure 5.7 illustrates the distribution of average seal widths.

Figure 5.6 Seal width measurement

Figure 5.7 Distributions of seal widths



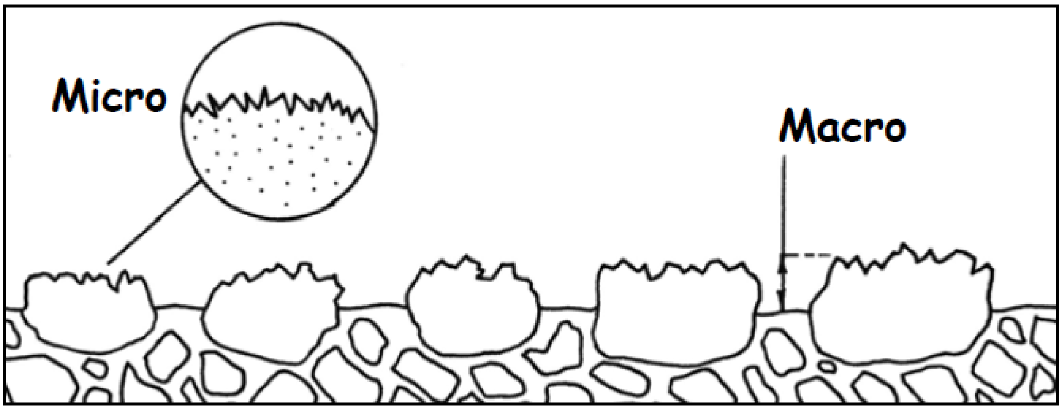
It was found that 82% of elements had an average seal width between 6m and 9m.

Note that wide seals (greater than 12m) were excluded.

5.5 Road surface characteristics

Road surfacing can be characterised by microtexture and macrotexture. Microtexture relates to the dry road surface friction and macrotexture to wet surface friction. A visual representation of these characteristics is illustrated in figure 5.8.

Figure 5.8 Cross-section of road surface illustrating micro and macrotexture (Naidu, Klaruw Systems)



A surrogate for microtexture is SCRIM coefficient and a surrogate for macrotexture is mean texture depth (MTD).

SCRIM coefficients were used in the pilot study and were identified as a key variable in rural road crash prediction modelling. This main study included MTD to further describe the macrotexture component of road surface characteristics.

5.5.1 SCRIM coefficient

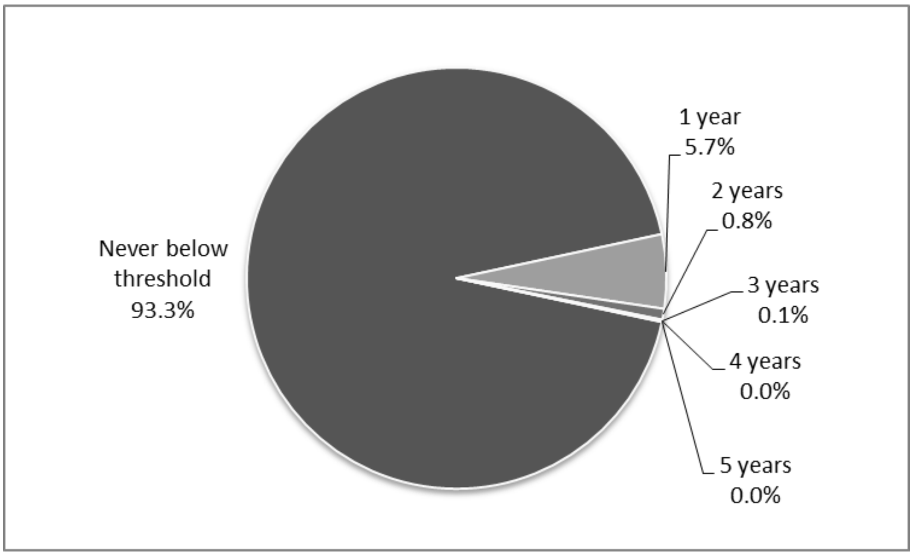
SCRIM coefficients were extracted from the RAMM database for the five-year period 2002 to 2006, corresponding to the five years of crash data collected. A stretch of road with a SCRIM coefficient below 0.4 may result in increased risk of skidding (NZ Transport Agency 2010b). This threshold was used in the development of the variable.

The variable used in this study was found by identifying the number of years the element had a SCRIM coefficient below the threshold. This number was then divided by the number of years assessed (five years for this study) and stored as a decimal (between 0 and 1). This value can be interpreted as the proportion of time the element had SCRIM coefficients below the threshold.

For this study, if an element falls below the threshold for one year then it will have a variable value of 0.2, while an element with five years below the threshold will have a value of 1.

Figure 5.9 illustrates the distribution of years when an element fell below the SCRIM coefficient threshold.

Figure 5.9 Number of years that SCRIM was below the threshold (2002–06)



Of the elements, 6.7% had at least one year where the SCRIM coefficient fell below the threshold. This corresponds to 395km of the state highway network.

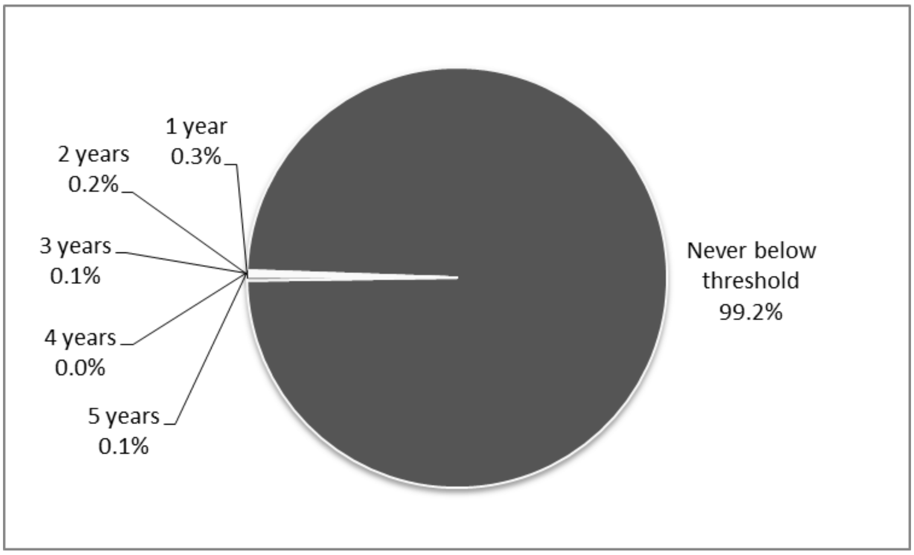
5.5.2 Mean texture depth

Mean texture depth (MTD) was extracted from the RAMM database for the five-year period 2002–06. A stretch of road with MTD below 0.7mm may increase the risk of wet road crashes (NZ Transport Agency 2010a).

This variable is represented in the same way as the SCRIM coefficient variable. It represents the proportion of time that the element had a MTD below the threshold.

Figure 5.10 illustrates the distribution of years when an element fell below the MTD threshold.

Figure 5.10 Number of years that mean texture depth was below the threshold (over five-year period)



Of the elements, 0.8% had at least one year when the MTD fell below the threshold. This corresponds to 25km of the state highway network. Given the low occurrence of poor texture depth, it is unlikely that the modelling will fully explain the importance of texture depth for crash occurrence.

5.6 Regional groupings

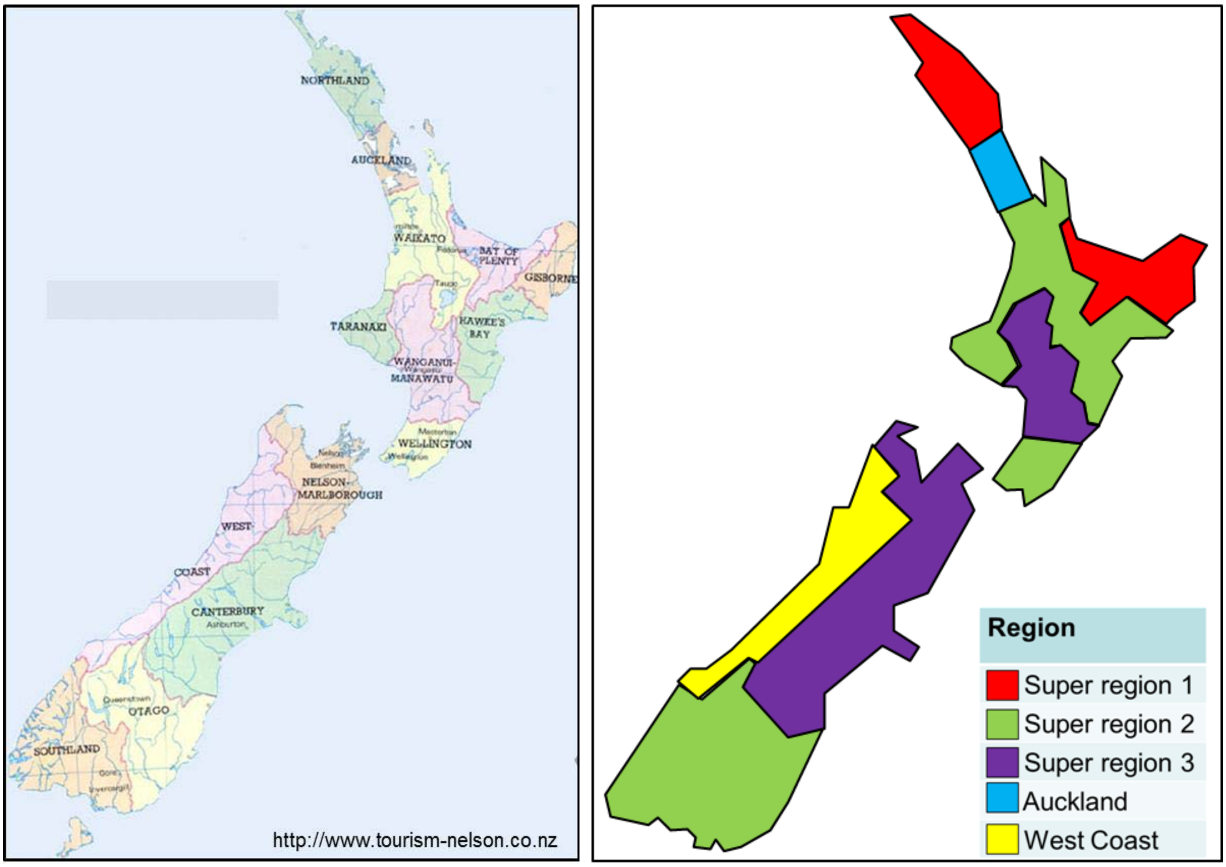
The New Zealand state highway network is divided into 14 regions, shown in figure 5.11. The scoping study identified socio-economic and weather effects that would not be included in the final model. In order to represent these effects, the regions were clustered together to form super regions.

The scoping study grouped the regions based on the following criteria:

- open road 85th percentile speed
- regional under reporting of serious crashes
- percentage of state highway crashes in wet weather
- percentage of state highway midblock 100km/h alcohol-related crashes
- percentage of state highway midblock 100km/h crashes in dark conditions
- percentage of state highway midblock 100km/h crashes relating to cornering.

Five regional groupings were identified and shown in figure 5.11 and table 5.2. The distribution of elements in each grouping is shown in figure 5.12.

Figure 5.11 New Zealand regional groupings

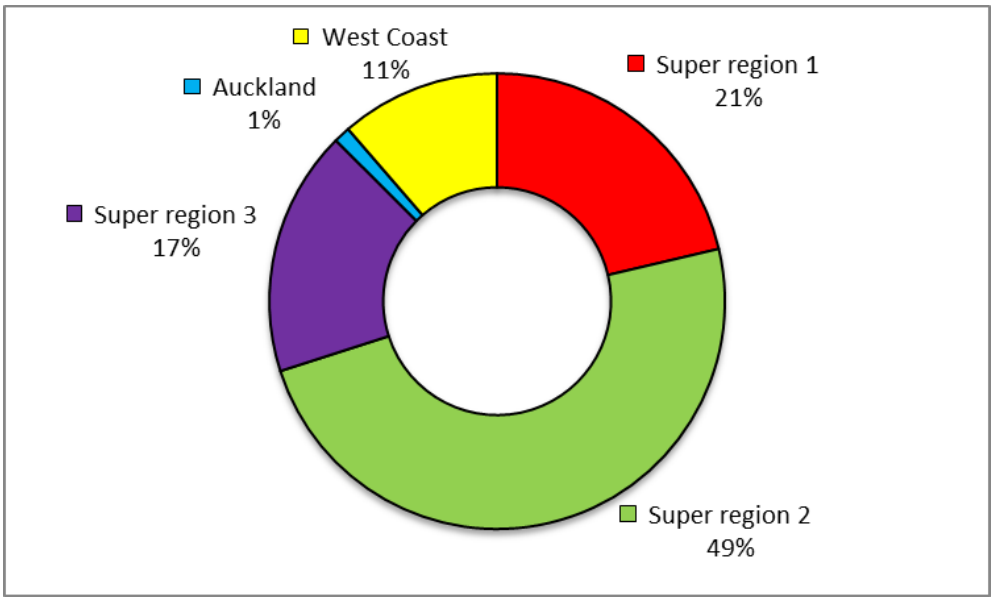


http://www.tourism-nelson.co.nz

Table 5.2 Grouping of regions

Super region 1	Northland Gisborne	Bay of Plenty
Super region 2	Waikato Hawkes Bay Taranaki	Wellington Otago Southland
Super region 3	Manawatu-Wanganui Canterbury	Nelson Marlborough
Auckland	Auckland	
West Coast	West Coast	

Figure 5.12 Proportion of elements within each region



The small proportion of elements in Auckland is due to the large amount of the network located within urban areas and comprising multiple lanes and motorways.

5.7 Roadside hazard assessment – KiwiRAP

The pilot study identified the KiwiRAP as a way of assessing the severity of roadside hazards and their offset from the carriageway. The two variables used in the pilot study (number of point hazards and the distance to a non-traversable slope or perpendicular deep drain) were found to be highly correlated, meaning the variables contributed to models being produced with conflicting results. KiwiRAP required far less labour to process data and is available to the future model practitioners.

Roadside hazard data was collected from the KiwiRAP dataset. Risk is assessed as being a value from one (low risk) to four (high risk) and is the combination of the severity of the hazard (outlined in table 5.3) and the offset of the hazard from the carriageway (Tate et al 2009). This is known as the risk code and is shown in table 5.4.

Table 5.3 KiwiRAP hazard severity description

Severity	Description
Negligible	Resulting in minor property damage.
Rigid barriers	Collision with steel beam or guard fence.
Moderate	Hazard likely to result in minor injury crash or moderate property damage.
Severe	Likely to result in serious or fatal crash

Table 5.4 KiwiRAP severity risk code

Hazard severity	Offset from carriageway		
	Less than 4m	4m to 9m	Greater than 9m
Negligible	1	1	1
Rigid barriers	2	1	1
Moderate	3	2	1
Severe	4	3	2

Research by the Land Transport Safety Authority (now part of the NZTA) found that severe hazards, especially those within a 4m edge of the edge line, were more likely to cause crashes than moderate and low severity hazards (Beca 2004 and Turner et al 2004) For this reason, this study focused on modifying the risk code to focus on severe hazards only. All other hazards were set to one, and severe hazards remained unchanged. This is shown in table 5.5.

Table 5.5 Modified KiwiRAP severity risk code

Hazard severity	Offset from carriageway		
	Less than 4m	4m to 9m	Greater than 9m
Negligible	1	1	1
Rigid barriers	1	1	1
Moderate	1	1	1
Severe	4	3	2

The risk to road users does not increase linearly as the hazard severity increases. The increase in risk from a moderate hazard to a severe hazard is far greater than the increase from a rigid barrier to a moderate risk. KiwiRAP outlines this relationship (Tate et al 2009) and is illustrated in figure 5.13 where risk codes are converted to a risk weighting. The risk weighting is the roadside hazard variable used in this study. The equations for converting from risk code to risk weighting are shown in table 5.6.

Figure 5.13 Relationship between risk code and weighting

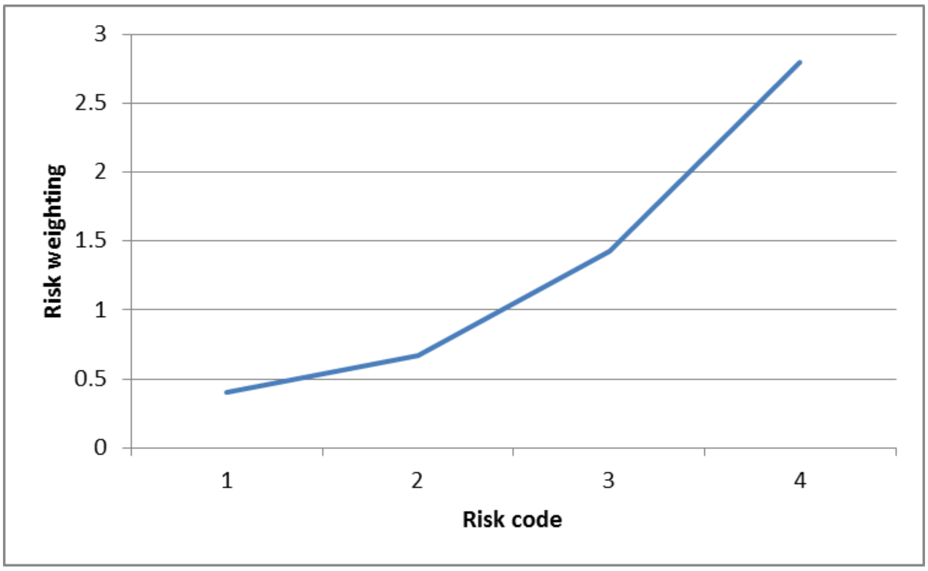
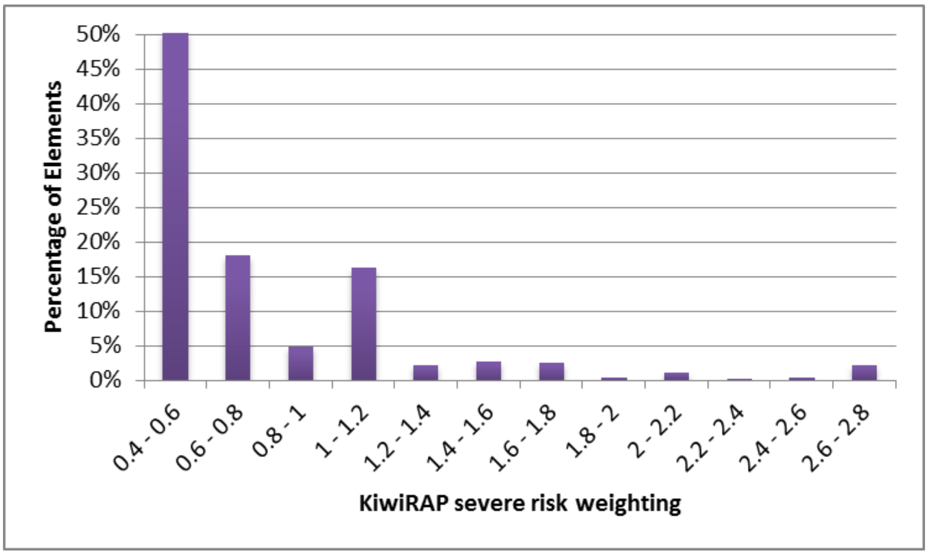


Table 5.6 Conversion from risk code to risk weighting

Risk code between		Equation to convert to risk weighting
lower value	upper value	
1	1.9	0.27 × [Risk Code] + 0.13
2	2.9	0.76 × [Risk Code] – 0.85
3	4	1.37 × [Risk Code] – 2.68

The distribution of severe risk weighting over all the elements is illustrated in figure 5.14. Of the elements, 36% do not contain any severe hazards (having a weighting of exactly 0.4), 3% contain only severe hazards and the remaining 61% contain some levels of severe hazards within the element.

Figure 5.14 Distribution of KiwiRAP severe risk weighting



5.8 Approach speed/horizontal consistency

Horizontal consistency was used as a measure in the pilot study to identify changes in driving conditions/geometry, eg travelling from a straight section onto a curved section. This was derived by using the RAMM data for the radius and the speed environment variable to calculate the percentage change in speed from before entering the element to within the element.

When this model was applied to the straight and curve elements in the main study, it generated incorrect results (significant speed increases from some straight to curve elements and some significant speed decreases from curve to straight elements). These issues were believed to be associated with the development of the model using fixed length sections as opposed to homogenous straight and curve sections. For this reason, the horizontal consistency model was not used in the main study.

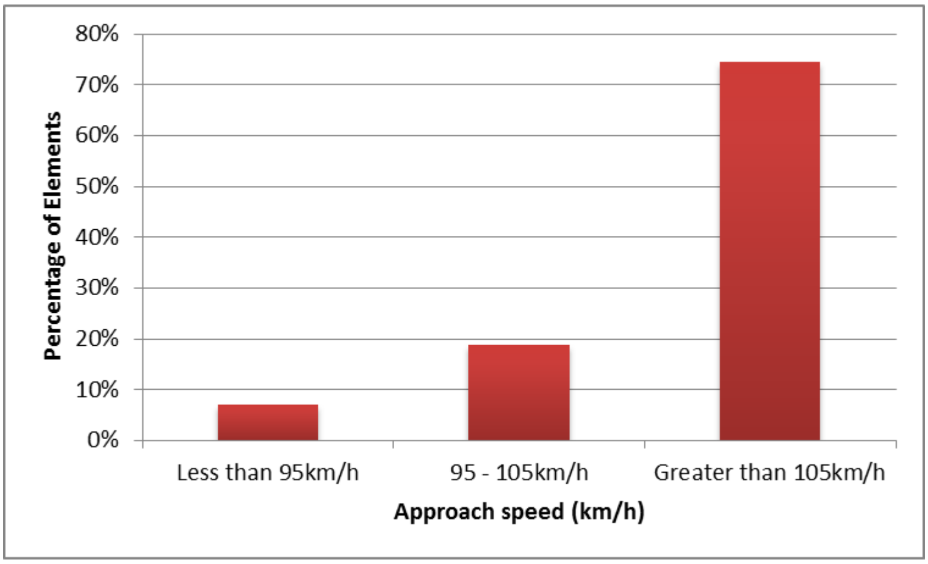
As an alternative to using a horizontal consistency model, the two data sets of element radius and the approach speed were used as inputs into the relational database.

Koorey used equation 5.1 to determine the road geometry advisory speed *(AS)*, where *R* is the horizontal radius and *e* is the superelevation or ‘cross-fall’. The speed environment used from the Koorey database is a 500m rolling average for every 10m element.

$$\boldsymbol{AS = -(0.10795R) + \sqrt{(0.10795R)^2 + 127R(0.+e)}} \qquad \text{(Equation 5.1)}$$

The approach speed data was sourced from the speed environment in the Koorey database and captured the approach speed for the 500m preceding the element. The data is illustrated in figure 5.15. The speed approaches were capped at 106km/h to better represent driver behaviour.

Figure 5.15 Distribution of approach speeds for the preceding 500m



5.9 Driveway/accessway trip density

Data was collected to develop a small-scale model to understand the impact of accessways on the safety of the rural state highway. To identify the sample set, crash data from CAS was analysed to identify the sections of state highway where driveway crashes had occurred. The state highways with the highest proportion of this crash type were selected, which resulted in a sample set of 1474km. Those sections are highlighted in white in figure 5.16. No data was collected from the West Coast super region.

Figure 5.16 State highways used in driveway study

Driveway data was collected by undertaking a visual inspection of the NZ Transport Agency recorded video to identify the number of accessways and associated land use present on the state highway.

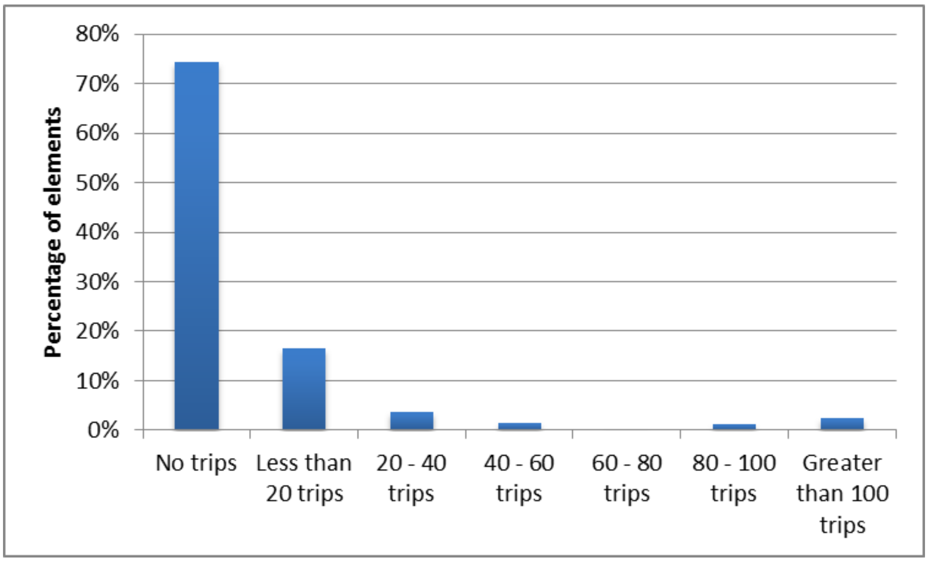
The land use associated with driveways was converted into the expected number of trips per day. A trip is considered any vehicle movement which enters or exits the driveway. Table 5.7 outlines the land uses, the expected number of trips and any assumptions made in determining these values.

Table 5.7 Trip generation by land use

Land use		Assumed trips generated (per day)	Assumptions
Rural residential (per letterbox)		8	Based single-family detached housing from ITE 7th edition
Commercial/industrial	Low/farm	16	Household (8 trips) and four employees (8 trips)
	Medium	80	Six employees (18 trips) and 31 visitors (62 trips)
	High	150	Twelve employees (36 trips) and 57 visitors (114 trips)

The distribution of trip generation per day is illustrated in figure 5.17. Of the surveyed elements, 74% of road sections generate no trips (have no driveways or accessways), 22% generate less than 60 trips per day and the remaining 4% generate at least 60 trips per day.

Figure 5.17 Distribution of trips generated



5.10 Summary of variables

A summary of all the variables used in the modelling is outlined in table 5.8.

Table 5.8 Summary of variables

Variable name	Units	Comments
Element type	–	Either a curve or straight element.
Traffic flow	AADT	
Length	m	Element length.
Region	–	The super-region in which the element is located.
Width	m	Average seal width of the element.
Grade	(decimal)	Absolute average gradient of the element.
KiwiRAP	–	Average severe risk weighted KiwiRAP value of the element.
Approach speed	kph	Average speed environment over the preceding 500m to the element.
SCRIM proportion	(decimal)	Proportion of time the element has a SCRIM coefficient below the threshold of 0.4
Mean textured depth (MTD) proportion	(decimal)	Proportion of time the element has a MTD value below the threshold of 0.7mm
Radius	m	Minimum curve radius
Trips	veh/day	Expected number of trips generated per day along the element (both entry and exit).

6 Data collection – crash data

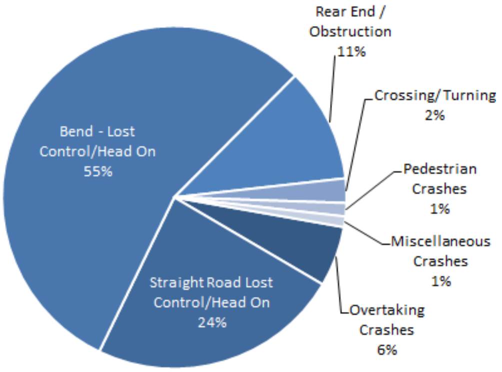
This chapter introduces the crash data used in developing the crash prediction models. Crashes were assigned to elements based on location and stored in the relational database.

6.1 Understanding crashes on rural roads in New Zealand

Loss-of-control and head-on crashes account for the majority (79%) of crashes occurring on two-lane rural state highways in New Zealand, as shown in figure 6.1. Speed was identified as a factor in 22% of all crashes, while road factors contributed to 20%.

The crash prediction modelling undertaken during this study focused on these crash types, with the further consideration of crashes occurring at accessways.

Figure 6.1 Crashes on New Zealand rural state highways (2002–06)



Source: NZTA CAS database.

6.2 Crash data

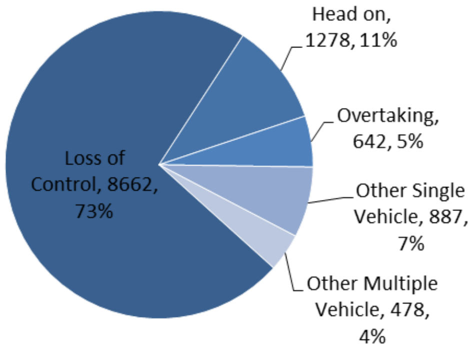
Koorey (2009) developed a validated crash database covering the entire New Zealand state highway network for the 2002–06 period by removing errors in crash coding and location referencing. This database was used to obtain crash data for the sections selected for modelling.

Although this data is not recent, the study team considered its use beneficial, as an analysis of a more recent crash period would necessitate labour intensive re-examination of the crash data to check for errors and inconsistencies in coding.

6.3 Key crash movements

Figure 6.2 shows the proportion of crashes by type for the selected rural road sections. Loss-of-control crashes account for the overwhelming majority (73%) of the 12,818 crashes in the sample set, followed by head-on (11%) and overtaking (5%) crashes. Of the remaining crash types, single vehicle crashes (such as collision with obstacle crash) are found to be more prevalent than crashes involving multiple vehicles (such as rear-end crashes).

Figure 6.2 Crashes by type on the study sample (excludes driveway crashes)



The differences between figures 6.1 and 6.2 relate to the differences in the road types included. For example the significant reduction in rear-end crashes in figure 6.2 (under other multiple vehicle), could be related to the removal of motorway sections.

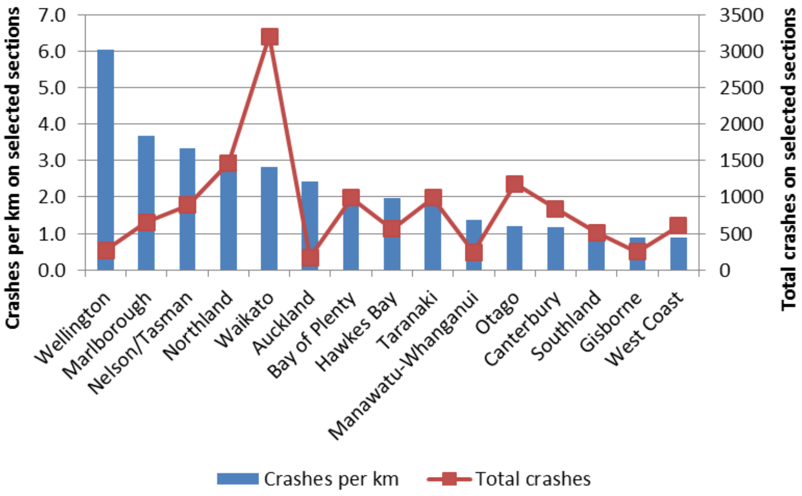
In addition to the main crash types illustrated above, 504 driveway crashes were also observed on the selected rural road sections. These crashes were used to develop driveway specific crash models, as described in section 7.5.

6.4 Crashes by region

Figure 6.3 illustrates the total number of crashes (red line series) and the number of crashes per kilometre (blue bar series) observed on selected rural roads in 15 New Zealand regions.

The two aspects depicted tend to vary because of differences in the lengths of selected state highways between regions. For instance, Wellington is observed to have one of the lowest total crash numbers but the highest crash rate per kilometre on the selected sections. On the other hand, Waikato has the highest observed crash numbers, with a higher crash rate per kilometre compared with most other regions.

Figure 6.3 Crashes by region

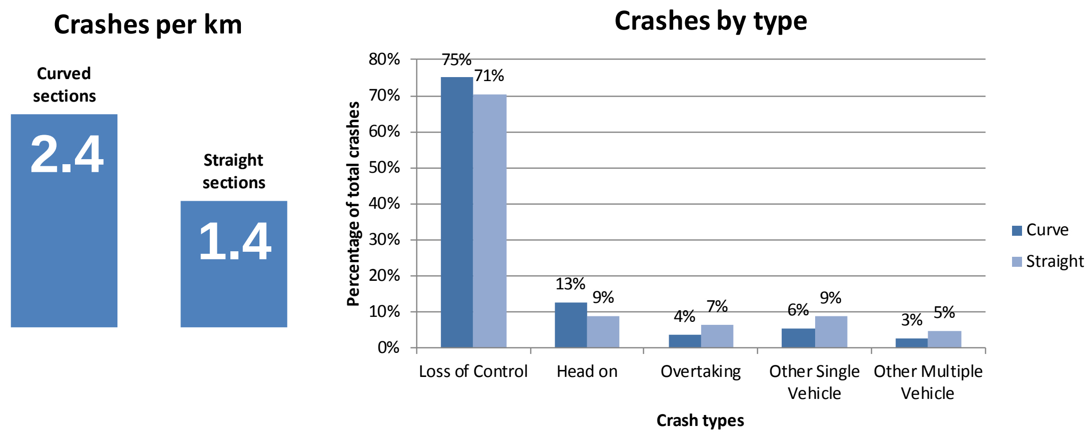


6.5 Crashes by element

Analysis of crashes by section type shows that curved sections have roughly 70% more crashes per kilometre than straight sections.

As can be seen in figure 6.4, the proportion of crashes by type at straight and curved sections is similar to that of the overall sample, with the majority being loss-of-control and head-on crashes. However, subtle differences are seen in the proportions of crashes between the two geometric element types, with overtaking and other single- and multi-vehicles crashes being more frequent on straight sections.

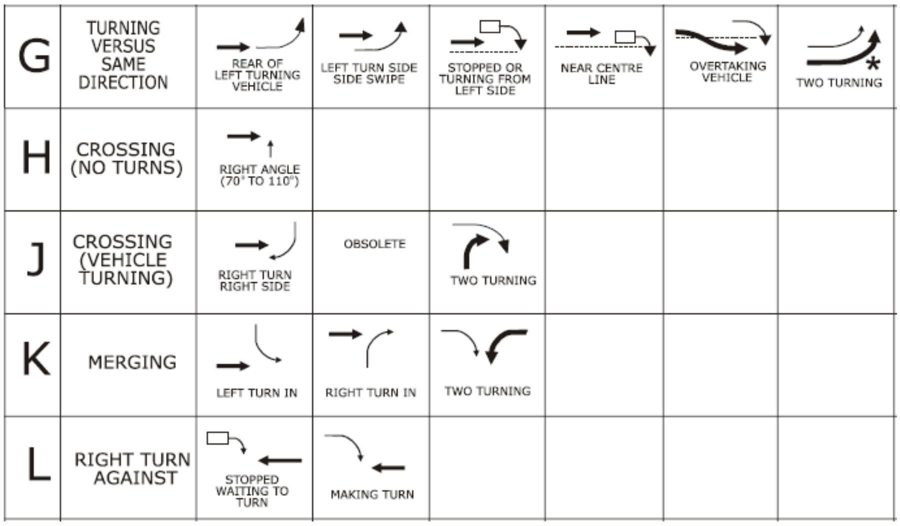
Figure 6.4 Crashes by type



6.6 Driveway-related crashes

For the accessway modelling, data was collected for crashes occurring at driveways identified from CAS for the 1474km of rural state highways included in the study. These crashes were further restricted to movements G, H, J, K, L (see figure 6.5) to take into account inaccuracies in the coding of driveway-related crashes in CAS.

Figure 6.5 Driveway-related crash movements



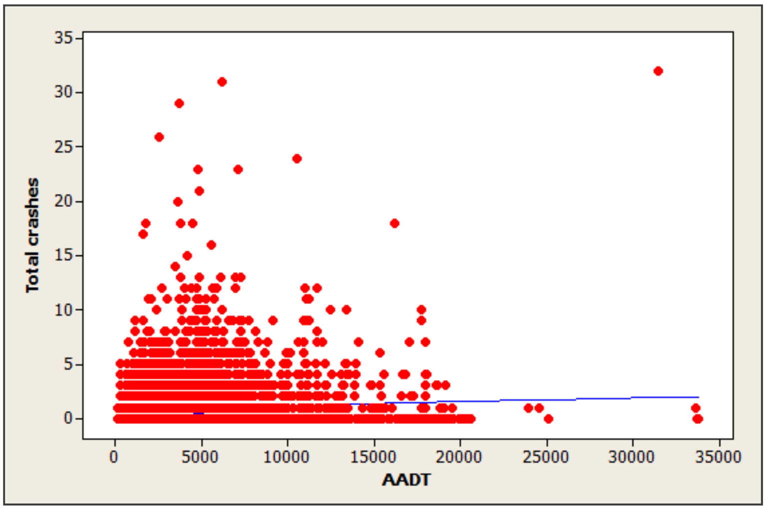
A total of 277 driveway-related crashes were observed on the selected sections, and were used to develop the model for driveway crashes.

6.7 Crash relationship between predictor variables

6.7.1 Traffic volume

Figure 6.6 plots the total number of injury crashes against the daily traffic volume (AADT) for the selected straight and curved rural roads sections. The figure shows a gradual increasing trend in crashes with increasing traffic volume, as shown by the blue linear trend line.

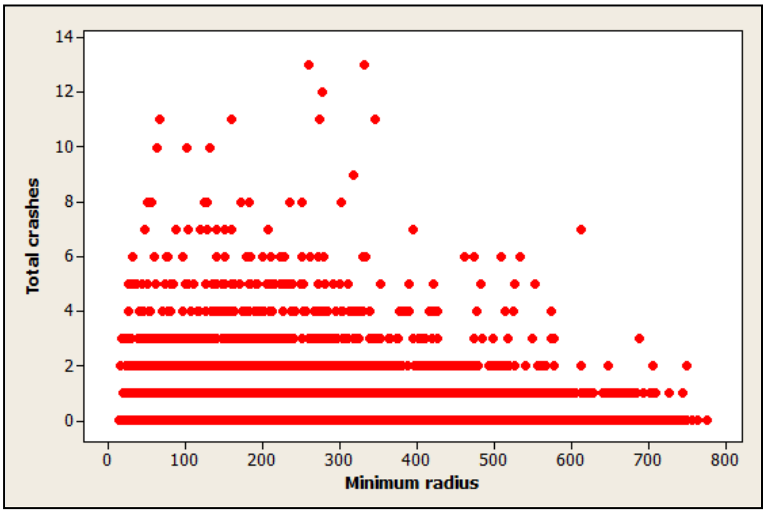
Figure 6.6 Injury crashes vs AADT



6.7.2 Minimum curve radius

The scatter plot of injury crashes and the minimum radius of the curved sections in figure 6.7 shows that tight curves are associated with higher crashes than more gradual curves on the rural state highway network.

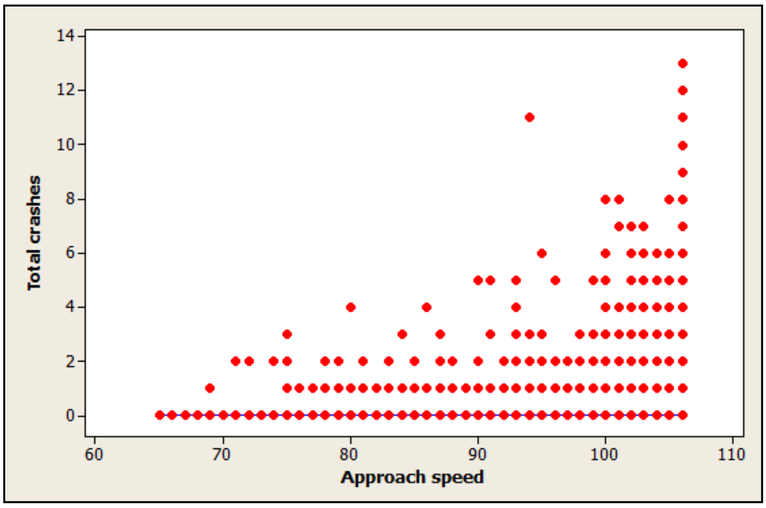
Figure 6.7 Injury crashes vs minimum curve radius



6.7.3 Approaching vehicle speed on curves

Figure 6.8 shows a gradual trend towards more crashes with an increase in the speed of vehicles as they start navigating the curved section.

Figure 6.8 Injury crashes vs approaching vehicle speed



Plots of the other predictor variables against the total number of injury crashes did not indicate a clear trend. While this is likely to be a result of the inherent variation within the large sample set, this lack of a clear trend supports the approach of building crash models for key crash movements, which are expected to show a more significant relationship with key geometric and operational variables.

Scatter plots of key predictor variables against injury crashes are provided in appendix E.

6.8 Variable correlations

Identification of variable correlations is required to avoid having two or more significantly correlated variables in the same prediction model. In such cases the variability in one variable does, to a certain extent, predict the variability in the correlated variable.

Only three significant correlations were observed in the dataset. These are highlighted in table 6.1. The full correlation matrix for all variables is attached as appendix C of this report.

Table 6.1 Variable correlations

Correlation	Description
Traffic flow (AADT) and seal width (0.67)	Both traffic volume and seal width are important predictor variables, and have been included in the models. Where high traffic volumes are present, wider seal widths are expected.
Curvature and gradient (0.29)	Mountainous roads are expected to be hilly (large gradients) and windy (sharp curves) while roads on plateaus or plains are usually flat and relatively straight.
Curvature and approach speed (-0.24)	Most drivers speed up as they enter straight elements (with small curvature) and slow down as they enter curve elements (large curvatures)

These correlations were considered in the models developed. Correlation between curvature and gradient and correlation between curvature and approach speed were found to be small enough to be negligible. The correlation between traffic flow and seal width was large and was taken into consideration when developing models.

Correlations between element types, regions, traffic flows, crash types and other variables do not need to be considered in the modelling methodology as they are not considered surrogates for other variables.

6.9 Rural roads relational database

The relational database provided each of the data sets as outlined in chapters 5 and 6 and was used to develop the crash prediction models. The data is also available for use by university students and other New Zealand researchers for further analysis. Instructions for using this database are given in appendix D.

7 Crash prediction models

Ten crash prediction models were developed for loss-of-control, head-on and driveway-related crashes on straight and curved rural road sections. The following section describes the crash modelling methodology adopted and details the various models that were developed.

7.1 Modelling methodology

Crash prediction models are mathematical models that relate crashes to road user volumes and other road layout and operational features. Crash prediction models are cross-sectional regression models. With crashes being discrete events, typically following a Poisson or negative binomial distribution, traditional regression analysis methods such as normal linear regression are not suitable. The models used in crash prediction modelling are developed using generalised linear modelling methods. Generalised linear models were first introduced to road crash studies by Maycock and Hall (1984), and extensively developed in Hauer and Hakkert (1989). These models were further developed and fitted using crash data and traffic counts in the New Zealand context for motor-vehicle-only crashes by Turner (1995).

The aim of this modelling exercise was to develop relationships between the mean number of crashes (as the response variable), and traffic flows, as well as non-flow predictor variables. Typically the models take the multiplicative form

$$A = b_0 x_1^{b_1} \ldots x_i^{bi} e^{b_{i+1} x_{i+1}} \ldots e^{b_n x_n} \qquad \text{(Equation 7.1)}$$

where A is the fitted annual mean number of crashes, the x_1 to x_i are measurement variables, such as average daily flows of vehicles, and the x_{i+1} to x_n are categorical variables, recording the presence, for example, of a cycle installation, and the b_1 ,...,b_n are the model coefficients.

7.1.1 Model development process

Once a functional model form has been selected, in this case the power model, generalised linear models are then developed for each crash type using either a negative binomial or Poisson distribution error structure.

Software has been developed in order to fit such models (ie to estimate the model coefficients). The popular Bayesian Information Criterion (BIC) was used in this study as the preferred criterion to decide when the addition of a new variable was worthwhile. Goodness of fit testing of all models was also undertaken by using software written in the form of macros.

7.1.2 Model interpretation

Once models have been developed, in some simple cases the relationship between crashes and predictor variables can be interpreted. Caution should always be exercised when interpreting relationships as two or more variables can be highly correlated. However, the modelling process described in the previous sections usually means that variables in the ‘preferred’ models are not highly correlated because the method acknowledges that adding a variable correlated to those already in an existing model does not improve the fit of the model compared with the addition of important non-correlated variables. Likewise, functional forms that deviate from a power function are also difficult to interpret. In these situations it is always best to plot the relationship.

In models with a power function form, where the variables are not correlated, an assessment of the relationship can be carried out. For a typical model with a power-function form and two continuous variables (such as flows or speeds), this takes the following form:

$$A = b_0 x_1^{b_1} x_2^{b_2} \qquad \text{(Equation 7.2)}$$

Where:

A	is the annual mean number of crashes
x_1 , x_2	are continuous flow or non-flow variables
b_0, b_1 and b_2	are model parameters.

In this model form, the parameter b_0 acts as a constant multiplicative value. If the number of reported injury crashes is not dependent on the values of the two-predictor variables (x_1 and x_2), then the model parameters b_1 and b_2 are zero. In this situation the value of b_0 is equal to the mean number of crashes. The value of the parameters b_1 and b_2 indicate the relationship that a particular predictor variable has (over its flow range) with a crash occurrence. There are five types of relationship for this model form, as presented in figure 7.1 and discussed in table 7.1.

Figure 7.1 Relationship between crashes and predictor variable x for different model exponents (b_1)

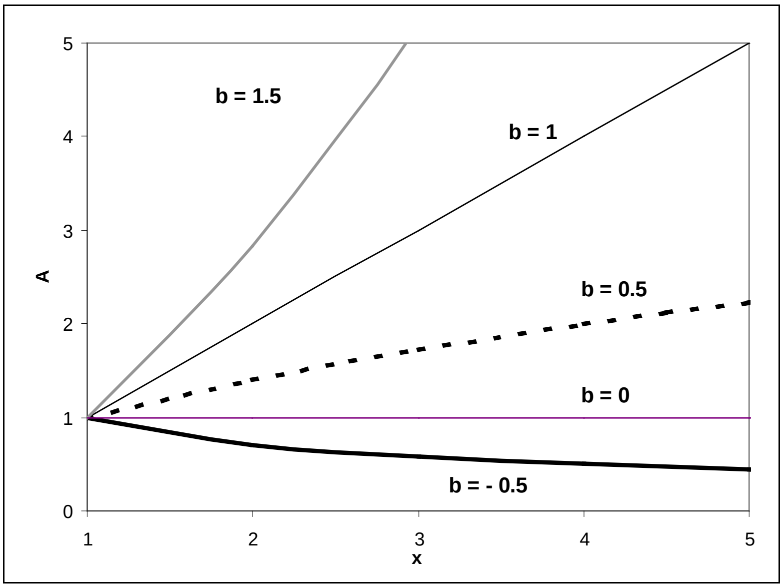


Table 7.1 Relationship between predictor variable and crash rate

Value of exponent	Relationship with crash rate
$b_i > 1$	For increasing values of the variable, the number of crashes will increase, at an increasing rate
$b_i = 1$	For increasing values of the variable, the number of crashes will increase, at a constant (or linear) rate
$0 < b_i < 1$	For increasing values of the variable, the number of crashes will increase, at a decreasing rate
$b_i = 0$	There will be no change in the number of crashes with increasing values of the variable
$b_i < 0$	For increasing values of the variable, the number of crashes will decrease

Generally, models of this form have exponents between $b_i = 0$ and $b_i = 1$, with most flow variables having an exponent close to 0.5, ie the square root of flow. In some situations, however, parameters have a value outside this range.

7.2 Models developed

Crash models were developed for the two key crash types on midblock rural road sections: head-on (11% of crashes) and loss-of-control (73% of crashes). In addition, models were developed for crashes occurring at driveways and accesses onto the rural road network.

Recognising the distinct characteristics of straight and curved rural road sections, separate models were also developed for the key crash types by geometric element type. The exception was driveway crashes, for which only models for combined straight and curved sections were built.

Furthermore, during the model development stage it became clear that the statistically significant model often did not contain the parameters that were of most interest to practitioners. Therefore, two distinct models were developed in these situations:

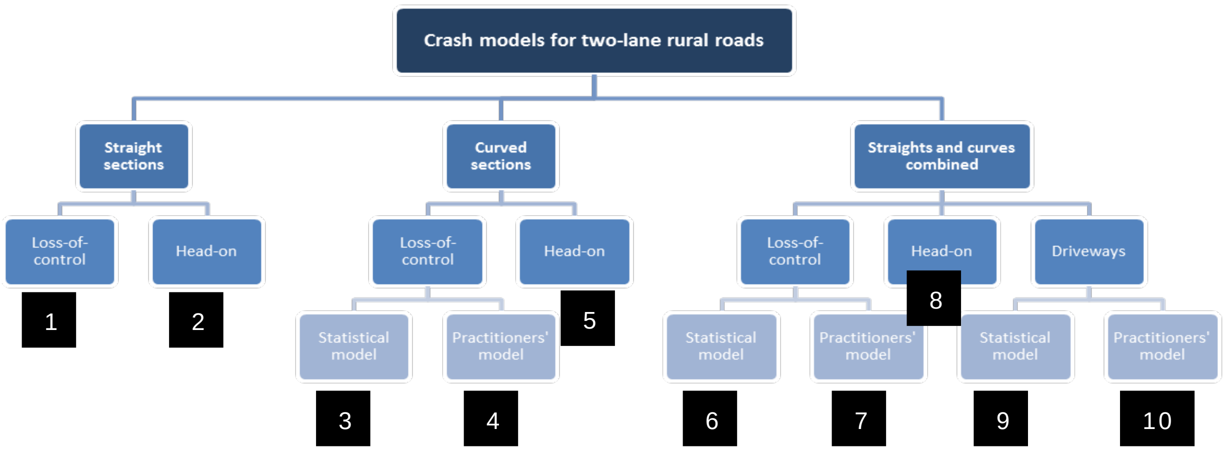
1 Statistical model: this was the best performing model as shown by the model Akaike Information Criterion (AIC) and BIC values. The 95%ile confidence level was selected for these models (ie we were 95% confident that the variable had an effect on crashes – there was only a 5% likelihood the variable had no effect or an opposite effect to that presented).

2 Practitioners' model: while not necessarily achieving 95% confidence, this model contained additional variables (compared with the statistical model) that were of interest to practitioners. Criteria for variables to be included were related to industry expert opinion, but were typically where the confidence level for the variable was over 70%, and often over 80%. These models may be used in crash prediction toolkits or industrial processes.

Separate statistical and practitioners' models were developed for loss-of-control crashes (on curves, and for combined straight and curved sections) and for driveway crashes.

Figure 7.2 illustrates the various models developed during this study.

Figure 7.2 Crash models for two-lane rural roads



7.3 Head-on crash models

Separate head-on crash models were developed for straight and curved rural road sections. In addition, a combined model considering both element types was also built.

7.3.1 Straight sections

Equation 7.3 presents the preferred model.

$$A_{HO,\ straight} = 7.971E\text{-}09\ (AADT)^{0.9177}\ (length)\ (regionid)\ exp^{(0.1196\ (width) + 13.97\ (grade) + 1.711(SCRIMPROP))} \qquad \text{(Equation 7.3)}$$

Where $A_{HO,\ straight}$ is the number of predicted head-on injury crashes per year on straight rural road sections. A full list of variables can be found in table 5.8. The values of 'regionid' are given in table 7.2.

Table 7.2 'Regionid' values for straight head-on model

Region	Region ID
Super region 1	**1**
Super region 2	**0.6954**
Super region 3	**0.7424**
Auckland	**0.3732**
West Coast	**0.9169**

The error structure for the model was found to be negative binomial. Equation 7.3 was found to have a p-value of zero. While the low p-value suggests that the model is not well-fitting, this is likely to be influenced by unmeasured characteristics between state highways in different regions.

The relationship between AADT and head-on crashes is observed to be near-linear, which suggests that the higher opposing traffic flow on high-volume rural roads results in more head-on crashes. Further, the model also indicates that wider and steeper sections are associated with more head-on crashes. SCRIM also affects the number of head-on crashes, and the model shows adequately surfaced sections having fewer head-on crashes.

Among the five regions, super region 1 (Northland, Gisborne, Bay of Plenty) has the highest prevalence of head-on crashes on straight sections, while Auckland has the lowest.

7.3.2 Curved sections

Equation 7.4 presents the preferred model.

$$A_{HO,\ curve} = 1.7216E\text{-}08\ (AADT)^{0.921}\ (length)^{1.051}\ (regionid)\ exp^{(0.043\ (width) + 6.77\ (grade) + 1.568\ (SCRIMPROP) + 58.98(1/rad\_min)} \qquad \text{(Equation 7.4)}$$

Where $A_{HO,\ curve}$ is the number of predicted head-on injury crashes per year on curved rural road sections. The values of 'regionid' are given in table 7.3.

Table 7.3 ‘Regionid’ values for curve head-on model

Region	Region ID
Super region 1	1
Super region 2	0.9546
Super region 3	0.7242
Auckland	0.4216
West Coast	0.9618

Equation 7.2 was found to have a p-value of zero.

Similar to straight sections, the model for curves shows a near-linear relationship between AADT and head-on crashes. Seal width and gradient appear to have a less pronounced effect for curves compared with straights. However, the ‘tightness’ of the curve, as measured by the inverse of the minimum curve radius, is shown to be a key factor affecting the risk of colliding with an on-coming vehicle while navigating the curve in a high-speed environment.

Super region 1 (Northland, Gisborne, Bay of Plenty) again has the highest prevalence of head-on crashes on curved sections, while Auckland has the lowest.

7.3.3 Straight and curved sections

Equation 7.5 presents the preferred model.

$$A_{HO,\ combined} = \quad 1.070\text{E-}08\ (AADT)^{0.92}\ (length)\ (regionid)\ \exp^{(0.077\ (width) + 9.167(grade) + 1.593\ (SCRIMPROP) + 55.09\ (1/rad\_min) + 0.478(curve))}$$

(Equation 7.5)

Where $A_{HO,\ combined}$ is the total number of predicted head-on injury crashes per year on rural road sections, irrespective of geometric element type. The values of ‘regionid’ are given in table 7.4.

Table 7.4 ‘Regionid’ values for combined head-on model

Region	Region ID
Super region 1	1
Super region 2	0.8243
Super region 3	0.7272
Auckland	0.4030
West Coast	0.9318

Equation 7.5 was found to have a p-value of zero.

The combined model for straight and curved rural roads supports the results of the individual models for straight and curved sections. The ‘curve’ variable in this model also shows that the risk of a head-on crash occurring is significantly greater on a curved section than on a straight section of similar gradient, seal width and surface friction. For instance, a curved section having a minimum radius of 200m is predicted to have twice the occurrence risk of a head-on crash than a straight section with similar cross section, traffic flow and length.

7.4 Loss-of-control crash models

Similar to head-on crashes, separate models were built for loss-of-control crashes on straight and curved sections, with an additional model considering both straights and curves.

7.4.1 Straight sections

Equation 7.6 presents the preferred model.

$$A_{LOC,\ straight} = 2.062E\text{-}06\ (AADT)^{0.74}\ (length)^{0.77}\ (regionid)\ exp^{\ (0.052(width)\ +\ 2.573\ (grade)\ +\ 0.067(kiwirap\_sev)\ +\ 0.625(SCRIMPROP)\ +\ 1.202(MDTPROP))}$$

(Equation 7.6)

Where $A_{LOC,\ straight}$ is the total number of predicted loss-of-control injury crashes per year on straight rural road sections. The values of ‘regionid’ are given in table 7.5.

Table 7.5 ‘Regionid’ values for straight loss-of-control model

Region	Region ID
Super region 1	1
Super region 2	0.8919
Super region 3	0.7230
Auckland	0.4082
West Coast	0.5952

Equation 7.6 has a p-value of 0.066 indicating that the model satisfies standard goodness-of-fit criteria.

The results of the model highlight the strong relationship between loss-of-control crashes and traffic volume (AADT). While this relationship is not linear, as with head-on crashes, the coefficient suggests that loss-of-control crashes on straight rural road sections increase at a decreasing rate with an increase in traffic volume.

The presence of roadside hazards affects the occurrence of injury loss-of-control crashes. For instance, the presence of a severe hazard within 4m of the edge of the carriageway results in a 20% increase in the predicted risk of loss-of-control crashes.

Adequate road surfacing (SCRIM and MTD) is also shown to be important in enabling vehicles to maintain a good ‘grip’ on the pavement and reducing the risk of losing control in high-speed environments. The effect of seal width and gradient, while still important, is observed to be less pronounced than that predicted by the head-on crash models. The effect of seal width may be explained by the high correlation with traffic volume.

7.4.2 Curved sections

7.4.2.1 Statistical model

Equation 7.7 presents the preferred model.

$$A_{LOC,\ curve,\ stat} = 4.403E\text{-}08\ (AADT)^{0.753}\ (length)^{1.106}\ (regionid)\ exp^{(2.69(grade)+0.024(approach\ speed)\ +\ 1.42(SCRIMPROP)\ +\ 42.62(1/rad\_min))}$$

(Equation 7.7)

Where $A_{LOC,\ curve,\ stat}$ is the total number of predicted loss-of-control injury crashes per year on curved rural road sections. The values of 'regionid' are given in table 7.6.

Table 7.6 'Regionid' values for curve loss-of-control statistical model

Region	Region ID
Super region 1	1
Super region 2	0.9873
Super region 3	0.9343
Auckland	0.4839
West Coast	0.8061

Equation 7.7 was found to have a p-value of zero.

The best performing model statistically for loss-of-control crashes on curved sections highlights the importance of traffic volume, gradient, approach speed into the curve, SCRIM and minimum curve radius on the risk of occurrence of loss-of-control crashes. The coefficients for AADT and gradient are observed to be similar to those for straight sections.

7.4.2.2 Practitioners' model

Seal width is considered to be an important design criterion for curves; however, this was not significant in the statistical model. This variable was therefore included in the practitioners' model, where there is a lower level of confidence that the variable is important. Given this lower level of confidence that a variable is important (statistically significant) readers should be cautious in applying the practitioners' model. More confidence should be placed on the statistical model. Equation 7.8 presents the preferred model.

$$A_{LOC,\ curve,\ prac} = 4.486E\text{-}08\ (AADT)^{0.724}\ (length)^{1.104}\ (regionid)\ exp^{(0.026(width) + 2.685(grade) + 0.024(approach\ speed) + 1.421(SCRIMPROP) + 42.75(1/rad\_min))} \quad \text{(Equation 7.8)}$$

Where $A_{LOC,\ curve,\ prac}$ is the total number of predicted loss-of-control injury crashes per year on curved rural road sections. The values of 'regionid' are given in table 7.7.

Table 7.7 'Regionid' values for curve loss-of-control practitioners' model

Region	Region ID
Super region 1	1
Super region 2	0.9930
Super region 3	0.9370
Auckland	0.4887
West Coast	0.8224

Equation 7.8 was found to have a p-value of zero.

The variable coefficients for the practitioners' model are broadly similar to those of the statistical model, with the exception of the additional seal width variable.

7.4.3 Straight and curved sections

7.4.3.1 Statistical model

Equation 7.9 presents the preferred model.

$$A_{LOC,\ combined,\ stat} = 2.214E\text{-}07\ (AADT)^{0.735}\ (length)^{0.83}\ (regionid)\ exp^{(0.04(width)\ +\ 2.892(grade)\ +\ 0.019(approach\ speed)\ +\ 1.193(SCRIMPROP)\ +\ 38.556(1/rad\_min)\ +\ 0.175(curve))} \quad \text{(Equation 7.9)}$$

Where $A_{LOC,\ combined,\ stat}$ is the total number of predicted loss-of-control injury crashes per year on all rural road sections irrespective of geometric element type. The values of 'regionid' are given in table 7.8.

Table 7.8 'Regionid' values for combined loss-of-control statistical model

Region	Region ID
Super region 1	1
Super region 2	0.9330
Super region 3	0.8162
Auckland	0.4438
West Coast	0.7068

Equation 7.9 was found to have a p-value of zero.

The model for straight and curved sections supports the results of the previously reported loss-of-control models. The 'curve' variable in this model also shows that the risk of occurrence of a loss-of-control crash is higher on curved sections than on a straight section of similar gradient, seal width and SCRIM.

7.4.3.2 Practitioners' model

The mean texture depth of the road surfacing, which was shown to be an important factor affecting loss-of-control crashes on straight rural road sections, was not present in the statistical model. This variable was therefore artificially introduced in the practitioners' model. Equation 7.10 presents the preferred model.

$$A_{LOC,\ combined,\ prac} = 2.256E\text{-}07\ (AADT)^{0.735}\ (length)^{0.83}\ (regionid)\ exp^{(0.04(width)\ +\ 2.888(grade)\ +\ 0.018(approach\ speed)\ +\ 1.195(SCRIMPROP)\ +\ 0.204(MDTPROP)\ +\ 38.18(1/rad\_min)\ +\ 0.177(curve))} \quad \text{(Equation 7.10)}$$

Where $A_{LOC,\ combined,\ prac}$ is the total number of predicted loss-of-control injury crashes per year on all rural road sections irrespective of geometric element type. The values of 'regionid' are given in table 7.9.

Table 7.9 'Regionid' values for combined loss-of-control practitioners' model

Region	Region ID
Super region 1	1
Super region 2	0.9346
Super region 3	0.8176
Auckland	0.4429
West Coast	0.7081

Equation 7.10 was found to have a p-value of zero.

The practitioners' model broadly resembles the statistical model, with the addition of the MTD assessment of the section providing additional insight into the safety benefits of adequate surfacing.

7.5 Driveway crashes

Due to the smaller sample set of driveway-related crashes, models were only developed for combined straight and curved rural road sections.

7.5.1 Statistical model

Equation 7.11 presents the preferred model.

$$A_{Dwy,\ stat} = 3.107E\text{-}13\ (AADT)^{0.528}\ (length)\ (regionid)\ exp^{(0.46(kiwirap\_sev) + 0.133(approach\ speed) + 0.003(trips)} \qquad \text{(Equation 7.11)}$$

Where $A_{Dwy,\ stat}$ is the total number of predicted driveway-related injury crashes per year occurring on the given rural road section. The values of 'regionid' are given in table 7.10.

Table 7.10 'Regionid' values for driveway statistical crash model

Region	Region ID
Super region 1	1
Super region 2	0.6205
Super region 3	0.3911
Auckland	1.331
West Coast	N/A#

The sample set excluded driveway-related crashes from region 5.

Equation 7.11 was found to have a p-value of zero.

The statistical model for driveway-related crashes quantifies the increased risk of occurrence of driveway-related crashes on sections with a higher traffic volume (although the effect is less than that seen in the other crash models), severe roadside hazards, higher vehicle speeds approaching the rural road section, and higher driveway-induced trips.

Super region 4 (Auckland) was found to have the highest prevalence of crashes out of all the super regions. This is believed to be related to the small sample set taken from this region in developing the driveway crash models. This is the only model where super region 1(Northland, Gisborne, Bay of Plenty) does not have the highest prevalence of crashes.

7.5.2 Practitioners' model

Seal width and mean texture depth were artificially introduced into the statistical model, in view of their importance for designers and highway operations and maintenance personnel. Equation 7.12 presents the preferred practitioners' model.

$$A_{Dwy,\ prac} = 5.122E\text{-}13\ (AADT)^{0.406}\ (length)\ (regionid)\ exp^{(0.098(width)+0.482(kiwirap\_sev)+0.13(approach\ speed) + 1.084(MDTPROP) + 0.003(trips))}$$

(Equation 7.12)

Where $A_{Dwy, prac}$ is the total number of predicted driveway-related injury crashes per year occurring on the given rural road section. The values of 'regionid' are given in table 7.11.

Table 7.11 'Regionid' values for straight head-on model

Region	Region ID
Super region 1	1
Super region 2	0.6144
Super region 3	0.4331
Auckland	0.7653
West Coast	N/A#

The sample set excluded driveway-related crashes from region 5.

Equation 7.12 was found to have a p-value of zero.

7.6 Model validation

Three 100km sections located in different parts of New Zealand were selected to assess the accuracy of model predictions over a substantial section of the state highway. The predicted number of crashes and the actual recorded number of crashes were compared over the five-year period from 2002 to 2006. Geometric data for the selected 100km sections was extracted from the database developed for this study.

The three 100km sample sections selected were:

1. SH2 in the Bay of Plenty, from Athenree Gorge to Pukehina. This section has moderate hills, with longer stretches of straight elements and an average AADT of 10,172 making it the section with greatest traffic flow. There are 71 curve elements averaging 170m in length and 54 straight elements averaging 415m.
2. SH10 in Northland, from Awanui to Pakaraka. This section is relatively flat and has an average AADT of 2857. There are 229 curve elements averaging 138m in length and 203 straight elements averaging 264m.
3. SH1S in Canterbury, from Clarence to Spotswood. This section is located along the coast and winds around the cliff face. It has an average AADT of 2390 and has the tightest curves of the three sections with an average curve radius of 253m. There are 292 curve elements averaging 121m in length and 182 straight elements averaging 260m.

These sections originate from different macro regions and vary in their alignments. All models could be tested on all sections except for driveway crashes in the third section.

The results of applying the models are shown in table 7.12. The percentage difference specifies the percentage increase of the recorded crashes vs the predicted. A result of 10% means the model is over predicting crashes by 10%, while -5% means the model is under predicting crashes by 5%.

Table 7.12 Comparison between recorded and predicted crashes

Crash type	Element type	Model type	Percentage difference (over predicting)		
			Section 1 SH2	Section 2 SH10	Section 3 SH1S
Loss of control	Straight	All	57%	14%	-2%
	Curve	Statistical	82%	-2%	-19%
		Practitioners'	87%	-3%	-17%
	Combined	Statistical	60%	4%	-14%
		Practitioners'	60%	4%	-14%
Head on	Straight	All	72%	17%	-39%
	Curve	All	5%	108%	-50%
	Combined	All	32%	51%	-45%
Driveway	Combined	Statistical	-7%	11%	N/A
		Practitioners'	2%	5%	N/A

Section 1 was found to over predict many models. This is believed to be due to high traffic volumes and long element lengths. Section 2 displays typical results, except for head-on curve crashes where it significantly over predicts. Section 3 under predicts, which could be due to the relatively flat environment.

These three sections show varied levels of over and under predicting by the models. None of the models consistently over or under predict for all the sections. The models do seem to generally over predict at the high volume site (eg section1). This may be as a result of additional/enhanced safety measures on this section of road. Many of the higher volume roads have had such treatments, eg better delineation and shoulder rumble strips. .

7.7 Application of the models

The following section presents examples of the practical application of the models to scenarios often encountered by highway safety engineers.

Hypothetical straight and curved state highway sections were assumed, and common state highway safety improvement scenarios identified for each. Next, the crash prediction models developed during this study were applied to calculate the number of injury crashes on the hypothetical sections, thereby estimating the safety benefits (through lower crash numbers) expected from the improvements.

7.7.1 Straight sections

A hypothetical 500m straight state highway section carrying 4000 vehicles per day (AADT) was selected for this example. The section was assumed to lie on relatively flat terrain (gradient of 2%), have a seal width of 7m, severe roadside hazards (weighted KiwiRAP severity rating of 2.8) and inadequate surface characteristics (SCRIM and mean texture depth were below the threshold for 60% of the previous five-year period).

The example used the crash prediction models developed for loss-of-control and head-on crashes on straight sections to assess the cumulative safety benefits of improving the road surface and subsequently mitigating the presence and severity of roadside hazards. Results are provided in figure 7.3.

Figure 7.3 Application of crash models for straight sections

	Base scenario		Improve road surfacing		Mitigate roadside hazards	
Model Parameters	Straights – Loss-Of-Control	Straights – Head-On	Straights – Loss-Of-Control	Straights – Head-On	Straights – Loss-Of-Control	Straights – Head-On
AADT	4000	4000	4000	4000	4000	4000
Length	500	500	500	500	500	500
Seal width	7	7	7	7	7	7
Gradient	0.02	0.02	0.02	0.02	0.02	0.02
KiwiRAP weighted severity rating	2.8		2.8		0.7	
% time SCRIM below threshold	0.6	0.6	0	0	0	0
% time MTD below threshold	0.6		0		0	
Injury crashes per year	**0.618**	**0.069**	**0.206**	**0.025**	**0.180**	**0.025**
			Crash reduction compared to Base		Further reduction in crashes	
			-67%	-64%	-13%	0%

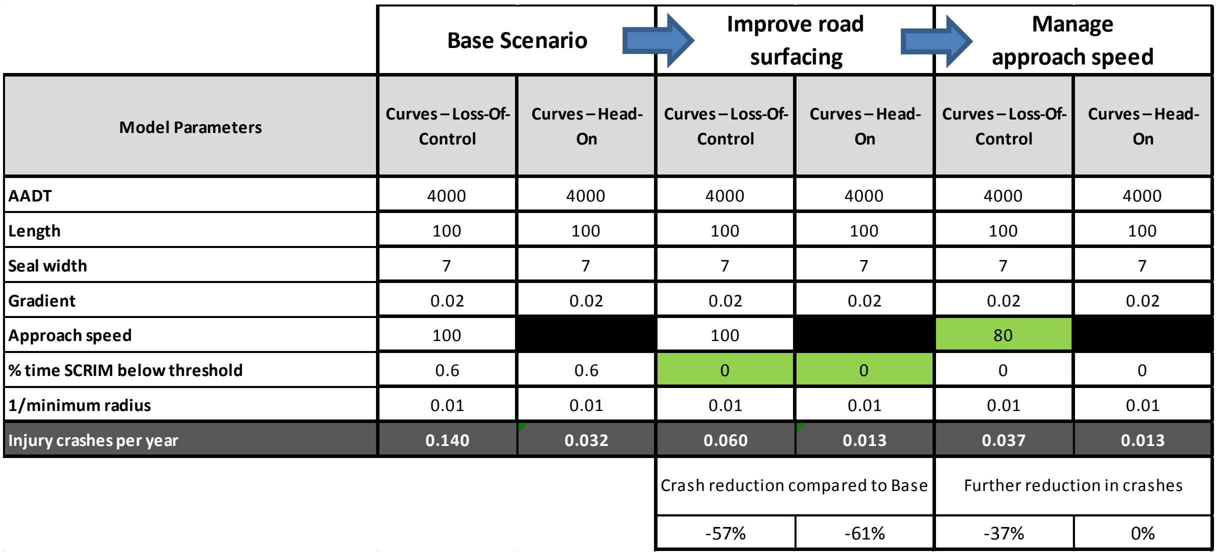
The model results indicate substantial reductions in loss-of-control and head-on crashes with improvements to the macro- and micro-texture of the road surface. Managing the severity of the roadside environment results in a further reduction in the number of loss-of-control crashes.

7.7.2 Curved sections

A 100m long curved section carrying 4000 vehicles per day (AADT) and with a minimum radius of curvature of 100m was considered for this analysis. The section was located on flat terrain (gradient of 2%) and had a seal width of 7m. Vehicles were assumed to enter the curve at a speed of 100km/h. The section was also assumed to have inadequate surface characteristics, with SCRIM and MTD being below the 'safe' threshold for 60% of the previous five-year period.

Figure 7.4 shows the number of crashes predicted by the head-on and loss-of-control (practitioners') models for the base scenario and subsequent improvements to micro-texture (SCRIM) and managing the speed environment (reducing entering vehicles' speed to 80km/h).

Figure 7.4 Application of crash models for curved sections



	Base Scenario		Improve road surfacing		Manage approach speed	
Model Parameters	**Curves – Loss-Of-Control**	**Curves – Head-On**	**Curves – Loss-Of-Control**	**Curves – Head-On**	**Curves – Loss-Of-Control**	**Curves – Head-On**
AADT	4000	4000	4000	4000	4000	4000
Length	100	100	100	100	100	100
Seal width	7	7	7	7	7	7
Gradient	0.02	0.02	0.02	0.02	0.02	0.02
Approach speed	100	■	100	■	80	■
% time SCRIM below threshold	0.6	0.6	0	0	0	0
1/minimum radius	0.01	0.01	0.01	0.01	0.01	0.01
Injury crashes per year	**0.140**	**0.032**	**0.060**	**0.013**	**0.037**	**0.013**
			Crash reduction compared to Base		Further reduction in crashes	
			-57%	-61%	-37%	0%

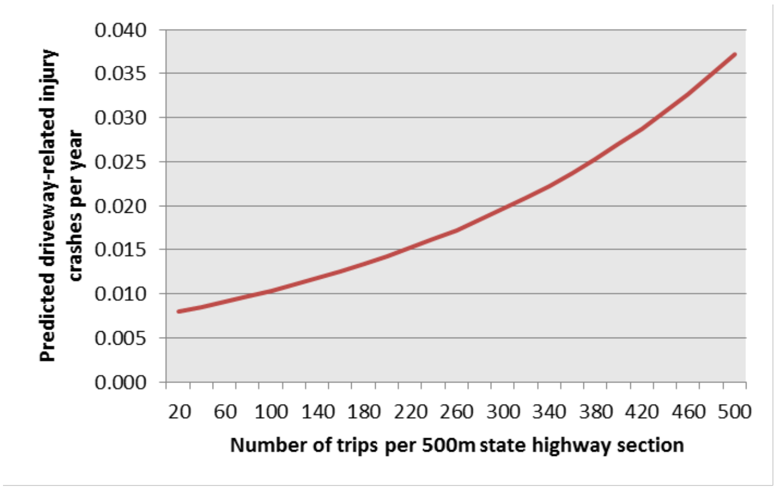
The predicted crash numbers again highlight the significant safety benefits of improving the micro-texture of the road surface. Reducing the speed environment prior to entering a fairly tight curve also results in a large decrease in the number of loss-of-control crashes.

7.7.3 Driveway-related crashes

The practitioner's model for driveway-related crashes enables engineers to assess the crash risk of the amount of vehicles in and out of a series of accessways. It is the amount of activity (vehicle trips) that is used rather than the number of driveways. The assumed state highway section for this assessment was 500m long, carried 4000 vehicles per day (AADT), had a seal width of 7m, little or no severe roadside hazards, and 100km/h vehicle speeds.

Figure 7.5 plots the predicted driveway-related crashes for various trip numbers on the assumed section.

Figure 7.5 Relationship between driveway-related crashes and trips



7.8 Scaling prediction models to total crashes

The crash prediction models only predict reported injury crashes for their particular crash type. In order to estimate the total number of reported injury crashes on a midblock element, the results of the crash prediction model need to be scaled up, in order to include other crash types.

Table 7.13 provides the factors required to scale up the crash predictions from various models to predict all reported injury crashes. It is recommended that both the head-on and loss of control models be used to produce a more accurate base prediction for scaling.

Table 7.13 Scaling factors of models

Models used on element	Scaling factor	Equation
Head-on and loss of control	16%	$A_{Total} = (A_{HO} + A_{LOC}) \times 1.16$
Loss of control only	27%	$A_{Total} = A_{LOC} \times 1.27$

Where A_{Total} is the total number of predicted injury crashes per year, A_{HO} is the result of the head-on crash prediction model and A_{LOC} is the result of the loss-of-control crash prediction model.

Scaling of 'head-on only' has not been included as the scaling factor is very large and scaled results are not expected to be accurate.

The NZ Transport Agency (2010b) *Economic evaluation manual* can be used to estimate total injury crashes (taking reported crashes into account) and to estimate likely numbers of non-injury crashes.

7.9 Tools for predicting crashes in high-speed environments

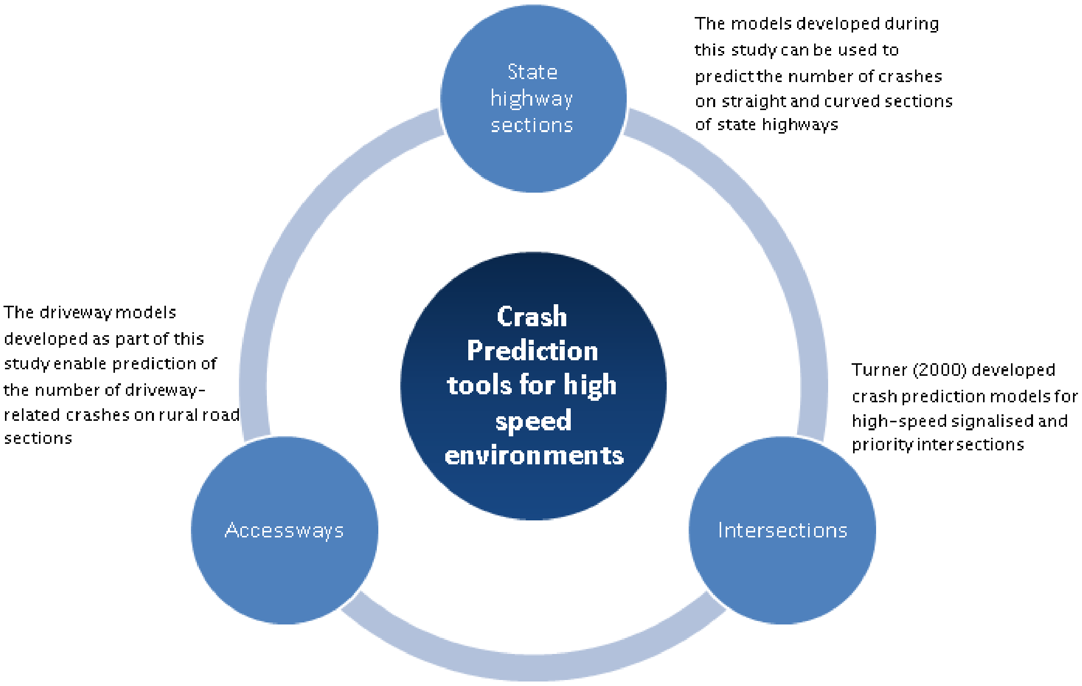
Data from state highways across New Zealand was utilised in this study to build crash prediction models for two-lane rural roads. These models represent a significant upgrade of existing prediction tools and supplement other safety assessment tools for high-speed environments. However, their applicability is limited to the specific environment they were developed for.

The high-speed rural road network can broadly be split into three key categories:

1. Two-lane rural roads
2. 'Enhanced' rural roads (eg passing lanes, three- and four-laned sections)
3. Motorway interchanges.

Crash prediction tools are already available for two-lane rural roads (see figure 7.6) and high-speed interchanges (through motorway crash prediction software such as ISAT). However, there is currently a dearth of internationally recognised tools for assessing the safety of three- or four-laned rural roads, and further research is needed to develop accurate models for predicting crashes in these environments.

Figure 7.6 Crash prediction tools for high-speed environments



Note: Minor intersections and low-volume side roads in high-speed environments do not experience the same turning volumes as major intersections, and can be modelled using the driveway crash models.

8 Summary and conclusions

8.1 Crash models

Ten crash prediction models were developed during this study for predicting loss-of-control, head-on and driveway-related crashes on rural roads. Loss-of control and head-on crashes collectively account for almost 80% of crashes on rural state highways in New Zealand. In addition, crashes occurring at driveways and accessways along the state highway network were considered in the crash modelling to understand the safety impacts of increased usage of accesses in high-speed environments.

By separately considering homogeneous straight and curved sections on the state highway network, the developed models represent an update of the crash prediction tools currently available for rural roads in New Zealand. In addition, the models quantify the safety impact of key road features, many of which can be influenced or changed by highway safety engineers, including:

- traffic flow (AADT)
- segment length
- minimum radius of curvature
- average gradient
- seal width
- SCRIM coefficient
- mean texture depth
- region
- KiwiRAP roadside hazard rating
- approaching vehicle speed
- number of accessway trips.

The models support previous research findings which show that wider and steeper roads, and those carrying more traffic, have a higher number of crashes. Traffic volume is particularly important for head-on crashes, as evidenced by the near-linear relationship between crashes and daily traffic volume. However loss-of-control and driveway-related crashes also show a significant relationships with traffic volume.

The models also indicate the significant benefits that can be achieved by improving the condition of the road surface, particularly the micro-texture. Models for both straight and curved sections show large reductions in the number of loss-of-control and head-on crashes when the condition of the road surface is improved. While the micro-texture, which has been measured through the SCRIM variable in this study, is shown to be the most important measure of the road surface because of its significant in both the straight and curved section models, the macro texture (measured by MTD) was shown to be important for loss-of-control crashes, particularly on straight segments.

Specific results for straight and curved sections are provided in the following sections.

8.1.1 Straight sections

Roadside hazards (measured by the KiwiRAP hazard rating) are a significant variable in the model for loss-of-control crashes occurring on straight segments. Removal of these hazards has the potential to substantially reduce the injury crash risk; for instance, removal of a severe hazard that lies within 4m of the edge of the carriageway is predicted to reduce loss-of-control injury crashes by up to 20%.

8.1.2 Curved sections

The radius of curvature affects the ease of navigation, and therefore safety, of curved rural road segments. This study has identified the minimum curve radius as the critical factor affecting the safety of a curved rural road segment.

The models also show that seal width and gradient, while still important from a safety perspective, have a less pronounced effect on safety on curved segments than on straight segments. For example, the risk of a head-on crash on a curved segment with a minimum radius of 200m is predicted to be twice that of a straight segment of similar length and other characteristics.

Curves which have a lower speed environment on the approach also have a lower risk of loss-of-control crashes, especially in mountainous terrain where such crashes can have severe consequences.

It was surprising that roadside hazards did not feature as a key variable for curves. It may be that the KiwiRAP roadside scores are not a suitable variable for curves and another factor needs to be developed and tested. The more detailed roadside variables considered in the pilot study (stage 2) were found to be important, so that may be a starting point for the development of a new variable.

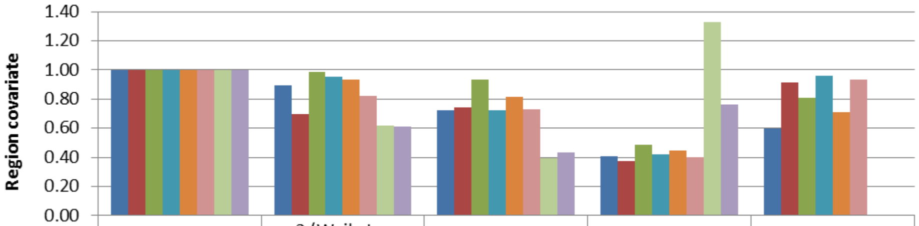
8.2 Regions

New Zealand state highways have been grouped into five super regions based on speed, under-reporting of serious crashes, and the percentage of dry weather, alcohol-related and night-time crashes. The models include regional covariates to enable assessment of crash risk by the location of the rural road section.

Overall, region 1 (Northland, Gisborne and Bay of Plenty) was found to have the highest incidence of loss-of-control and head-on crashes. Region 4 (Auckland) was shown to have the lowest incidence of head-on and loss-of-control crashes, but the highest number of driveway-related crashes. While, this to an extent is likely to be a result of the extremely limited length of state highways selected from this region, it does indicate that the prevalence of accesses on state highways in this region is a significant safety concern.

Among the other regions, region 2 (Waikato, Hawke’s Bay, Taranaki, Wellington, Otago and Southland) had more loss-of-control crashes than region 3 (Manawatu-Whanganui, Canterbury and Nelson-Marlborough) and the West Coast. On the other hand, the West Coast had more head-on crashes, followed by region 2 and then region 3 (refer to figure 8.1).

Figure 8.1 Comparison of model region coefficients



	1(Northland, Gisborne, Bay of Plenty)	2 (Waikato, Hawkes Bay, Taranaki, Wellington, Otago, Southland)	3 (Manawatu-Wanganui, Canterbury, Nelson-Marlborough)	4 (Auckland)	5 (West Coast)
Straights Loss-Of-Control	1.00	0.89	0.72	0.41	0.60
Straights Head-On	1.00	0.70	0.74	0.37	0.92
Curves Loss-Of- Control	1.00	0.99	0.93	0.48	0.81
Curves Head-On	1.00	0.95	0.72	0.42	0.96
Straights and curves combined Loss-Of-Control	1.00	0.93	0.82	0.44	0.71
Straights and curves combined Head-On	1.00	0.82	0.73	0.40	0.93
Driveway-related Driveways - Statistical	1.00	0.62	0.39	1.33	
Driveway-related Driveways - Practitioners	1.00	0.61	0.43	0.77	

■ Straights Loss-Of-Control ■ Straights Head-On
■ Curves Loss-Of- Control ■ Curves Head-On
■ Straights and curves combined Loss-Of-Control ■ Straights and curves combined Head-On
■ Driveway-related Driveways - Statistical ■ Driveway-related Driveways - Practitioners

8.3 Relational database

Data on road geometry, cross sections, surfacing, roadside environment and traffic volume for 6874 km of two-lane rural state highways in New Zealand was collected as part of this study. In addition, driveway density and trip information was collected for a smaller sample of 1474km.

A relational database was developed to contain this information in a format that will enable easy analysis by university students and New Zealand researchers in the future.

8.4 Accessway crashes

Traffic volume, roadside environment, approaching vehicle speed and the number of access vehicle trips were highlighted as important factors affecting crashes at driveways and accessways along the rural state highway network.

The NZTA has developed policies to limit the number of accessways on rural state highways. Indeed, model results show that an increase in the number of trips originating from side-accesses raises the number of crashes occurring at driveways.

Turning vehicles at accessways located in winding mountainous regions often suffer from reduced visibility around bends, which increases the crash risk. Roadside hazards such as trees can also result in visibility deficiencies for turning vehicles even on relatively straight sections. The combination of reduced visibility and high speed in such situations can result in severe safety consequences.

8.5 Developing a safety analysis tool for New Zealand

There is potential for the inclusion of the developed models in a safety analysis software tool that can assist in the use of the crash models in New Zealand. Investigations by the study team have led to the conclusion that this is best accomplished through spreadsheet-based software similar in structure to ISAT, a crash prediction tool for motorway interchanges.

The IHSDM suite of software was initially identified as a suitable medium for developing a New Zealand version for local use. However, discussions between the study team and the US Federal Highways Administration (FHWA) identified barriers to updating the in-built models currently used by IHSDM, thereby limiting its widespread applicability in New Zealand.

In contrast, a spreadsheet–based tool would be easier to develop and maintain, and would also allow ease of widespread application in evaluating the safety impacts of various state highway designs and improvements.

8.6 Further research

The next step in this work could include:

- Undertake predictions, using the models, across the entire network (at overlapping 100km sections), moving some sections that are outliers into other regions, or adding new regional categories and then rerunning the models with roads reclassified into these new regions. This should ideally improve the fit of the models.
- Look at developing shorter approach speed profiles to better understand horizontal consistency and test these in the models.
- Investigate the relationship between seal width and traffic volume and consider developing separate models for different flow bands.
- Look at special cases of curves, such as reverse and compound curves and develop crash prediction models for these cases.
- Look at the effect of delineation and how it differs within each of the regions. This may explain some of the inner-region variability and improve the model fits.

9 References

American Association of State Highway Transportation Officials (AASHTO) (2010) *Highway safety manual*. Washington: AASHTO.

Beca (2004) Assessing the crash risk implications of roadside hazards. Internal LTSA (now NZTA) project/research report.

Cenek, PD and RB Davies (2004) Crash risk relationships for improved safety management of roads. In *Proceedings Towards Sustainable Land Transport Conference,* Wellington, New Zealand, 22–24 November 2004.

Cenek, PD and RJ Henderson (2009) Crash risk model for curves. Interim unpublished report (17 June 2009). Wellington: NZ Transport Agency.

Cenek, PD, RB Davies and RJ Henderson (2004) The effect of skid resistance and texture on crash risk. In *Proceedings Towards Sustainable Transport Conference,* Wellington, New Zealand, 22–24 November 2004.

Chadfield, E (1992) Review of cross section guidelines for two lane rural roads. Transit NZ internal file note.

Fitzpatrick, K, EL Park and WH Schneider (2008) Potential accident modification factors for driveway density on rural highways. *Transportation Research Record: Journal of the Transportation Research Board 2083*: 49–61.

Garber, NJ, PR Haas and A Gosse (2010) Safety performance function for two-lane roads in Virginia. *Transportation Research Board 89th Annual Meeting,* Washington DC.

Hauer, E and S Hakkert (1989) The extent and some implications of incomplete crash reporting. *Transportation Research Record: Journal of the Transportation Research Board 1185*: 1–10.

Jackett, MJ (1992) On which curves do accidents occur? A policy advisory speed signs. In *Volume 1 Proceedings IPENZ Annual Conference.*

Koorey, GF (2009a) *Incorporating safety into rural highway design*. PhD thesis, University of Canterbury, New Zealand

Koorey, GF (2009b) Road data aggregation and sectioning considerations for crash analysis. *Transportation Research Record: Journal of the Transportation Research Board 2103*: 61–68.

Koorey, GF and FN Tate (1997) Review of accident analysis procedures for project evaluation manual. *Transfund NZ research report 85.* 54pp.

Landon-Lane, R (2007) *Driver behaviour at low-visibility, low-radius curves*. Master of Engineering in Transportation Degree, University of Canterbury, New Zealand.

Mattar-Habib, C, A Polus and H Farah (2008) Further evaluation of the relationship between enhanced consistency model and safety of two-lane rural roads in Israel and Germany. *European Journal of Transport and Infrastructure Research 8,* no.4: 320–332.

Maycock, G and RD Hall (1984) Crashes at four-arm roundabouts. *TRRL laboratory report LR1120.*

Montella, A, L Colantuoni and R Lamberti (2008) Crash prediction models for rural motorways. *Transportation Research Board, Transportation Research Record 2083*: 180–189

Naidu, T (nd) *Klaruw systems: skid resistance corrections – the sustainable solution*. Accessed 14 September 2011. www.soci.org/News/~/media/Files/Conference Downloads/Road Maintenance Treatments Mar 10/1500TimNaidu.ashx

New Zealand Transport Agency (NZTA) (2010a) *T10: 2010 Specification for state highway skid resistance*. pp5–9.

New Zealand Transport Agency (NZTA) (2010b) *Economic evaluation manual, volume 1*. Accessed 27 November 2012. www.nzta.govt.nz/resources/economic-evaluation-manual/volume-1/index.html

Polus, A and C Mattar-Habib (2004) A new consistency model for rural highways and its relationship to safety. *Journal of Transportation Engineering 130*, no.3: 286–293.

Rengarasu, TM, T Hagiwara and M Hirasawa (2009) Effects of road geometry and cross-section variables on traffic accidents. *Transportation Research Board, Transportation Research Record 2102*: 34–42.

Tate, FN and S Turner (2007) Road geometry and drivers speed choice. Paper presented at *IPENZ Transportation Group Conference,* Tauranga, New Zealand, 2–10 October 2007. 15pp.

Tate, FN, G Waibl and C Brodie (2009) *KiwiRAP summary report*. Wellington: NZ Transport Agency.

Turner, S (1995) Estimating accidents in a road network. PhD, University of Canterbury

Turner, S (2000) Accident prediction models. *Transfund NZ research report 192*. Wellington: Transfund NZ.

Turner, S, AD Dixon and G Wood (2004) *Assessing the crash risk implications of roadside hazards.* Traffic Management Workshop: Technical Conference, Wellington, New Zealand, September 2004.

Turner, S and F Tate (2009) Relationship between road geometry, observed travel speed and rural accidents. *NZ Transport Agency research report 371.* 72pp.

Turner, S, F Tate and R Singh (2011) Next generation of rural roads crash prediction models – pilot study, *NZ Transport Agency research report 437.* 86pp.

Zegeer, C, J Hummer, D Reinfurt, L Herf and W Hunter (1987) Safety effects of cross-section design for two-lane roads. *Report no. FHWA-RD-87/008*. Washington DC: (US) Federal Highway Administration (FHWA).

Zegeer, C, D Reinfurt, W Hunter, J Hummer, R Stewart and L Herf (1988) Accident effects of sideslope and other roadside features on two-lane roads. *Transportation Research Board: Transportation research record 1995.*

Zegeer, C, JR Stewart, F Council, D Reinfurt and E Hamilton (1992) Safety effects of geometric improvements on horizontal curves. *Transportation Research Board: Transportation research record 1356*, pp11–19.

Appendix A: Selected state highways

SH Number	Total length (km)	Excluded length (km)	Incomplete data (km)	% Excluded	Remaining length (km)
1N	1115	572	202	69%	341
1B	45	6	11	38%	28
1S	968	267	271	56%	430
1X	4	1	3	100%	0
2	959	194	177	39%	588
2A	6	6	0	100%	0
2B	4	0	4	100%	0
3	490	162	96	53%	232
3A	16	1	2	17%	13
4	237	19	16	15%	201
5	246	41	31	29%	174
6	1163	133	110	21%	919
6A	7	7	0	100%	0
7	272	26	35	22%	212
7A	9	3	2	53%	4
8	457	29	40	15%	387
8A	21	0	1	5%	20
8B	3	0	3	100%	0
10	104	12	7	18%	85
11	30	6	13	66%	10
12	218	27	26	24%	165
14	109	21	47	62%	41
15A	8	1	8	100%	0
16	119	50	21	60%	48
17	32	15	17	100%	0
18	283	11	272	100%	0
18A	4	4	0	100%	0
20	38	38	0	100%	0
20A	8	8	0	100%	0
20B	2	1	2	100%	0
21	7	0	3	56%	3
22	13	6	2	61%	5
23	43	4	4	18%	35
24	13	3	1	24%	10
25	231	53	38	39%	141
25A	28	0	1	4%	27
26	96	28	17	46%	52
27	92	5	17	24%	70
28	21	0	3	16%	18
29	68	22	24	68%	22
30	220	44	45	40%	132
30A	4	4	0	100%	0
31	56	1	5	11%	50
32	96	3	2	5%	91
33	36	5	3	24%	27
34	25	4	2	25%	19
35	334	36	20	17%	278
36	48	7	41	100%	0
37	7	0	0	5%	7
38	79	8	40	61%	31
39	57	7	5	21%	45
41	59	1	5	11%	52
43	137	10	30	29%	97
44	5	5	0	100%	0

SH Number	Total length (km)	Excluded length (km)	Incomplete data (km)	% Excluded	Remaining length (km)
45	105	26	21	45%	58
46	19	0	0	1%	19
47	46	0	5	12%	41
48	7	1	1	27%	5
49	36	5	5	25%	27
50	97	12	15	28%	70
50A	17	2	2	25%	13
53	18	3	7	55%	8
54	57	7	13	35%	37
56	23	2	8	43%	13
57	64	11	12	36%	41
58	15	8	1	60%	6
60	116	23	22	38%	72
62	13	0	13	100%	0
63	117	15	23	33%	79
65	71	2	5	10%	64
67	51	12	11	45%	28
67A	9	1	3	44%	5
69	33	2	4	19%	27
71	6	1	0	6%	6
73	235	44	132	74%	60
73A	4	4	0	100%	0
74	29	28	0	100%	0
74A	2	2	0	100%	0
75	77	25	12	48%	40
77	94	8	11	20%	75
78	1	1	0	100%	0
79	61	4	16	33%	41
80	55	2	4	10%	49
82	71	4	4	11%	63
83	109	6	19	23%	84
84	3	3	0	100%	0
85	165	11	10	13%	144
86	5	1	4	100%	0
87	114	9	14	20%	91
88	13	13	0	100%	0
90	59	2	3	9%	54
93	43	2	1	7%	40
94	254	16	28	17%	211
95	20	4	1	22%	16
96	90	4	6	11%	80
97	20	0	19	100%	0
98	22	0	2	12%	19
99	92	10	5	16%	78
Total	**11309**	**2255**	**2181**	**39%**	**6874**

Appendix B: Modelling result tables

Model	a (s.e.)	b (s.e.)	c (s.e.)	regionid (s.e.)	d (s.e.)	e (s.e.)	f (s.e.)	g (s.e.)	h (s.e.)	i (s.e.)	j (s.e.)	k (s.e.)	l (s.e.)	Overdispersion parameter (s.e.)	LL	AIC	BIC
Straights – loss-of-control	−13.0917 (0.2254)	0.7395 (0.0352)	0.7695 (0.0174)	1 – 0.0000 2 – −0.1144 (0.0472) 3 – −0.3243 (0.0595) 4 – −0.8959 (0.1600) 5 – −0.5189 (0.1000)	0.0515 (0.0206)	2.5728 (0.9953)	0.0666 (0.0420)		0.6246 (0.3291)	1.2015 (0.4357)				0.6414 (0.0469)	−8138	16,301	16,395
Straights – head-on	−18.6474 (0.5065)	0.9177 (0.0870)	1.0	1 – 0.0000 2 – −0.3633 (0.1110) 3 – −0.2979 (0.1368) 4 – −0.9856 (0.3462) 5 – −0.0868 (0.2439)	0.1196 (0.0475)	13.9734 (2.1317)			1.7110 (0.6946)					0.7587 (0.2205)	−2075	4170	4245
Curves – loss-of-control – statistical	−16.9384 (0.5176)	0.7532 (0.0263)	1.1056 (0.0345)	1 – 0.0000 2 – −0.0128 (0.0499) 3 – −0.0680 (0.0613) 4 – −0.7258 (0.1818) 5 ––0.2156 (0.0901)		2.6895 (0.8067)		0.0236 (0.0045)	1.4200 (0.2110)		42.6223 (2.8794)			1.2143 (0.0727)	−8769	17,562	17,655

Model	a (s.e.)	b (s.e.)	c (s.e.)	regionid (s.e.)	d (s.e.)	e (s.e.)	f (s.e.)	g (s.e.)	h (s.e.)	i (s.e.)	j (s.e.)	k (s.e.)	l (s.e.)	Overdispersion parameter (s.e.)	LL	AIC	BIC
Curves – loss-of-control – engineering	−16.9198 (0.5178)	0.7242 (0.0368)	1.1040 (0.0345)	1 – 0.0000 2 – −0.0070 (0.0501) 3 – −0.0651 (0.0614) 4 – −0.7161 (0.1821) 5 – −0.1955 (0.0916)	0.0260 (0.0225)	2.6849 (0.8070)		0.0235 (0.0045)	1.4213 (0.2111)		42.7518 (2.8800)			1.2145 (0.0727)	−8769	17,563	17,660
Curves – head-on	−17.8774 (0.5782)	0.9211 (0.0773)	1.0507 (0.0735)	1 – 0.0000 2 – −0.0465 (0.1037) 3 – −0.3227 (0.1334) 4 – −0.8636 (0.3686) 5 – −0.0389 (0.2035)	0.0430 (0.0457)	6.7677 (1.6295)			1.5684 (0.3857)		58.9765 (4.9036)			1.4881 (0.2967)	−2491	5005	5094
Straights and curves combined – loss-of-control – statistical	−15.3231 (0.3908)	0.7354 (0.0254)	0.8295 (0.0158)	1 – 0.0000 2 – −0.0693 (0.0344) 3 – −0.2031 (0.0427) 4 – −0.8124 (0.1213) 5 – −0.3470 (0.0672)	0.0401 (0.0153)	2.8915 (0.6177)		0.0185 (0.0036)	1.1927 (0.1719)		38.5559 (2.5793)	0.1753 (0.0336)		0.9033 (0.0424)	−16,988	34,004	34,121
Straights and curves combined – loss-of-control – engineering	−15.3046 (0.3914)	0.7351 (0.0254)	0.8301 (0.0158)	1 – 0.0000 2 – −0.0676 (0.0345) 3 – −0.2014 (0.0427) 4 – −0.8145 (0.1213) 5 – −0.3452 (0.0672)	0.0399 (0.0153)	2.8881 (0.6179)		0.0184 (0.0036)	1.1951 (0.1719)	0.2036 (0.2328)	38.1826 (2.6206)	0.1768 (0.0336)		0.9036 (0.0425)	−16,988	34,006	34,131

Model	a (s.e.)	b (s.e.)	c (s.e.)	regionid (s.e.)	d (s.e.)	e (s.e.)	f (s.e.)	g (s.e.)	h (s.e.)	i (s.e.)	j (s.e.)	k (s.e.)	l (s.e.)	Overdispersion parameter (s.e.)	LL	AIC	BIC
Straights and curves combined – head-on	−18.3529 (0.3380)	0.9202 (0.0577)	1.0	1 – 0.0000 2 – −0.1932 (0.0759) 3 – −0.3185 (0.0956) 4 – −0.9088 (0.2539) 5 – −0.0706 (0.1558)	0.0771 (0.0329)	9.1672 (1.3002)			1.5927 (0.3311)		55.0926 (4.5372)	0.4783 (0.0691)		1.1211 (0.1860)	−4,577	9,177	9,273
Straights and curves combined – driveway – statistical	−28.8000 (4.9290)	0.5282 (0.1018)	1.0	1 – 0.0000 2 – −0.4773 (0.1545) 3 – −0.9388 (0.2976) 4 – 0.2862 (0.3048) 5 – n/a			0.4601 (0.1519)	0.1334 (0.0467)					0.0031 (0.0007)	1.6474 (0.4451)	−896	1,809	1,866
Straights and curves combined – driveway – engineering	−28.3000 (4.9010)	0.4058 (0.1377)	1.0	1 – 0.0000 2 – −0.4871 (0.1562) 3 – −0.8369 (0.3063) 4 – −0.2675 (0.3064) 5 – n/a	0.0978 (0.0746)		0.4817 (0.1525)	0.1295 (0.0464)		1.084 (0.7796)			0.0032 (0.0007)	1.6420 (0.4449)	−894	1810	1884

Model form: $\exp^{(a)}\ (AADT)^{b}(length)^{c}\exp^{(regionid+d(width)+e(grade)+f(kiwirap\_sev)+g(approach\ speed)+h(SCRIMPROP)+i(MDTPROP)+j(1/rad\_min)+k(curve)+l(trips))}$

AADT = the average AADT over 5 years

Regionid = 1(Northland, Gisborne, Bay of Plenty); 2 (Waikato, Hawkes Bay, Taranaki, Wellington, Otago, Southland); 3 (Manawatu-Wanganui, Canterbury, Nelson-Marlborough); 4 (Auckland); 5 (West Coast)

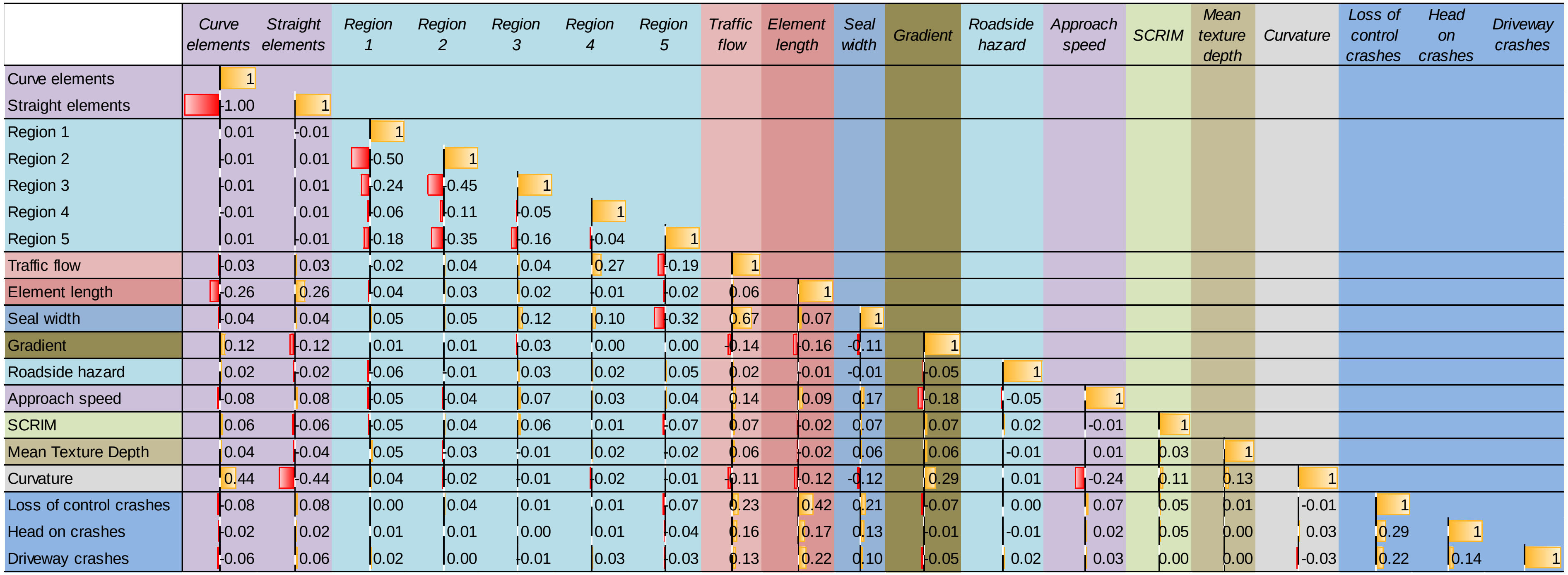
1/rad_min = the inverse of the minimum radius measurement in the segment

curve = 0 if a straight segment, 1 if a curve segment

SCRIMPROP = fraction of time the SCRIM reading was under 0.4

MDTPROP = fraction of time the mean texture depth reading was under 0.7

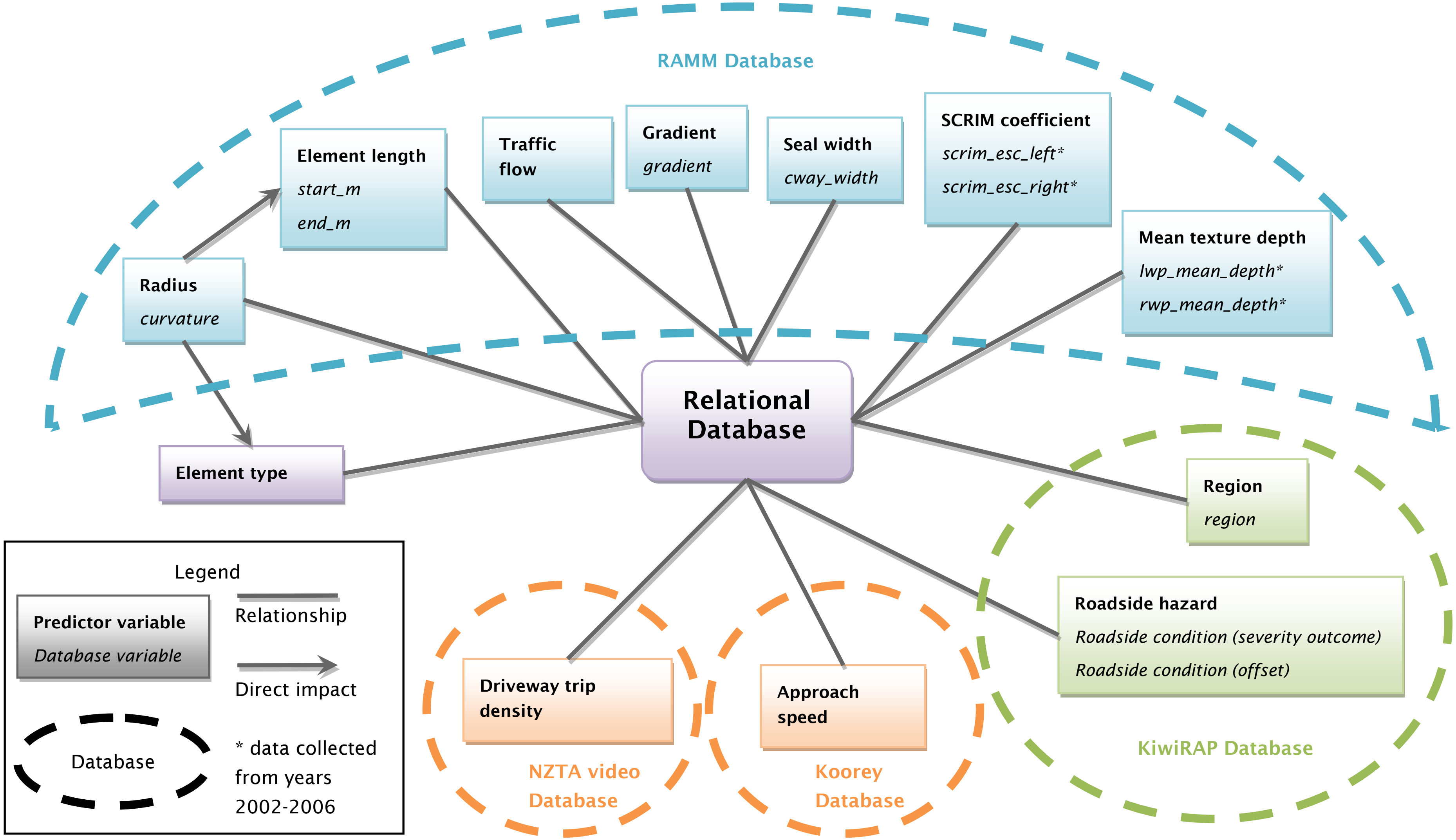
Appendix C: Correlation matrix



	Curve elements	*Straight elements*	*Region 1*	*Region 2*	*Region 3*	*Region 4*	*Region 5*	*Traffic flow*	*Element length*	*Seal width*	*Gradient*	*Roadside hazard*	*Approach speed*	*SCRIM*	*Mean texture depth*	*Curvature*	*Loss of control crashes*	*Head on crashes*	*Driveway crashes*
Curve elements	1																		
Straight elements	-1.00	1																	
Region 1	0.01	-0.01	1																
Region 2	-0.01	0.01	-0.50	1															
Region 3	-0.01	0.01	-0.24	-0.45	1														
Region 4	-0.01	0.01	-0.06	-0.11	-0.05	1													
Region 5	0.01	-0.01	-0.18	-0.35	-0.16	-0.04	1												
Traffic flow	-0.03	0.03	-0.02	0.04	0.04	0.27	-0.19	1											
Element length	-0.26	0.26	-0.04	0.03	0.02	-0.01	-0.02	0.06	1										
Seal width	-0.04	0.04	0.05	0.05	0.12	0.10	-0.32	0.67	0.07	1									
Gradient	0.12	-0.12	0.01	0.01	-0.03	0.00	0.00	-0.14	-0.16	-0.11	1								
Roadside hazard	0.02	-0.02	-0.06	-0.01	0.03	0.02	0.05	0.02	-0.01	-0.01	-0.05	1							
Approach speed	-0.08	0.08	-0.05	-0.04	0.07	0.03	0.04	0.14	0.09	0.17	-0.18	-0.05	1						
SCRIM	0.06	-0.06	-0.05	0.04	0.06	0.01	-0.07	0.07	-0.02	0.07	0.07	0.02	-0.01	1					
Mean Texture Depth	0.04	-0.04	0.05	-0.03	-0.01	0.02	-0.02	0.06	-0.02	0.06	0.06	-0.01	0.01	0.03	1				
Curvature	0.44	-0.44	0.04	-0.02	-0.01	-0.02	-0.01	-0.11	-0.12	-0.12	0.29	0.01	-0.24	0.11	0.13	1			
Loss of control crashes	-0.08	0.08	0.00	0.04	0.01	0.01	-0.07	0.23	0.42	0.21	-0.07	0.00	0.07	0.05	0.01	-0.01	1		
Head on crashes	-0.02	0.02	0.01	0.01	0.00	0.01	-0.04	0.16	0.17	0.13	-0.01	-0.01	0.02	0.05	0.00	0.03	0.29	1	
Driveway crashes	-0.06	0.06	0.02	0.00	-0.01	0.03	-0.03	0.13	0.22	0.10	-0.05	0.02	0.03	0.00	0.00	-0.03	0.22	0.14	1

Appendix D: Rural roads database usage guide

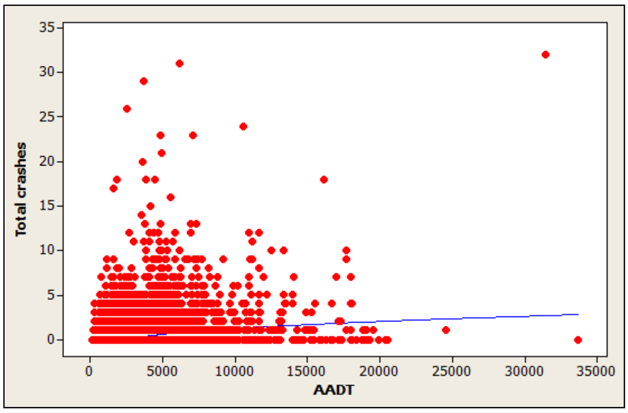
The following diagram outlines how the databases interact with the predictor variables. It also identifies where certain variables are used to find the value of other variables. Database variable names have also been included.


RAMM Database
Element length
start_m
end_m
Traffic flow
Gradient
gradient
Seal width
cway_width
SCRIM coefficient
scrim_esc_left*
scrim_esc_right*
Mean texture depth
lwp_mean_depth*
rwp_mean_depth*
Radius
curvature
Element type
Relational Database
Region
region
Roadside hazard
Roadside condition (severity outcome)
Roadside condition (offset)
KiwiRAP Database
Driveway trip density
NZTA video Database
Approach speed
Koorey Database
Legend
Predictor variable
Database variable
Relationship
Direct impact
Database
* data collected from years 2002-2006

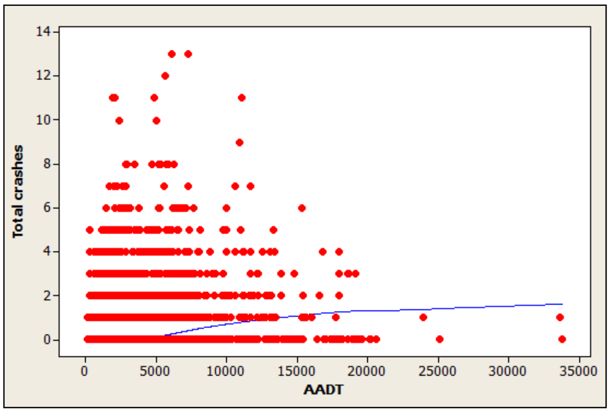
Appendix E: Predictor variable relationships

AADT

Straight sections

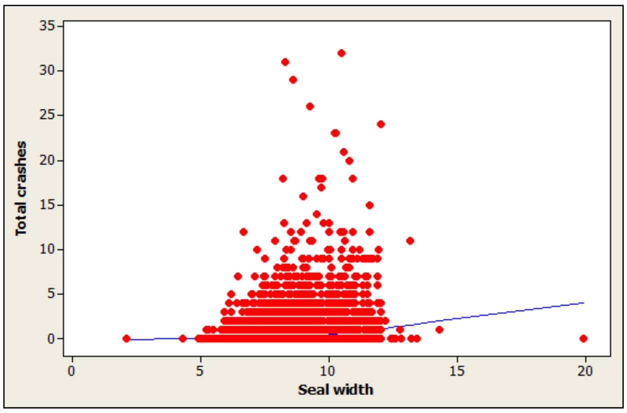


Curved sections

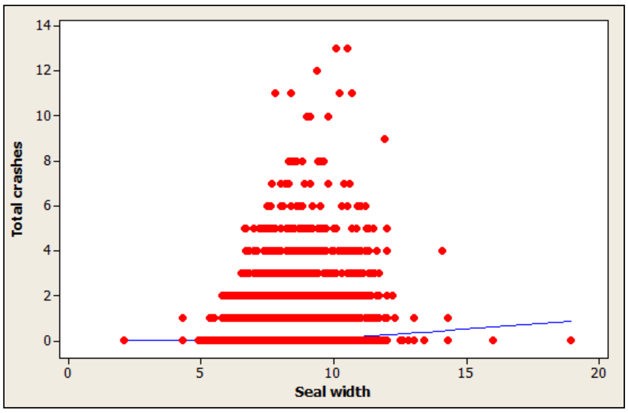


Seal width

Straight sections

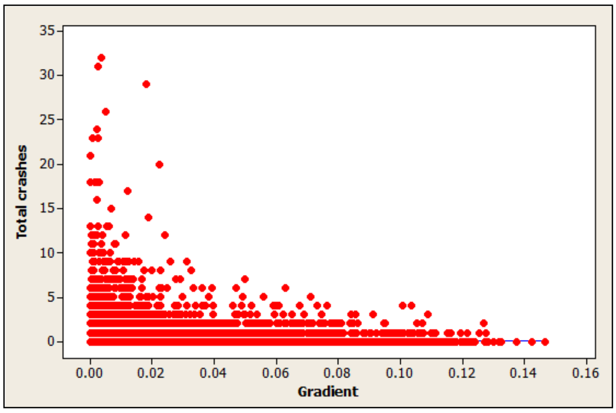


Curved sections

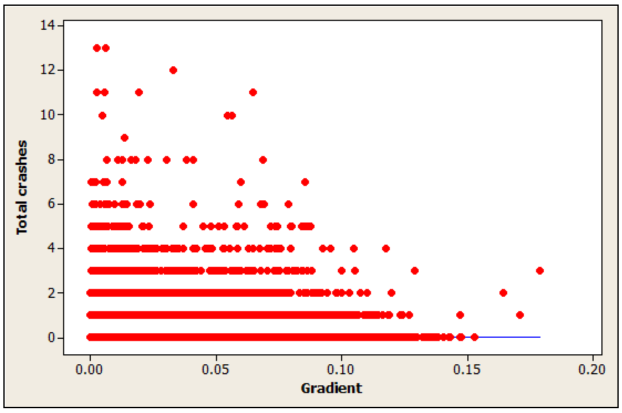


Gradient

Straight sections

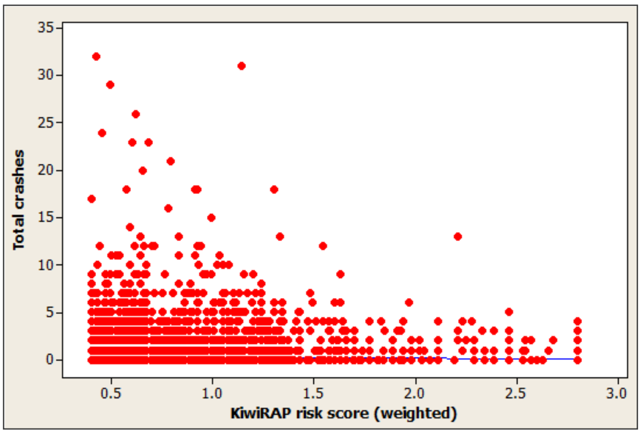


Curved sections

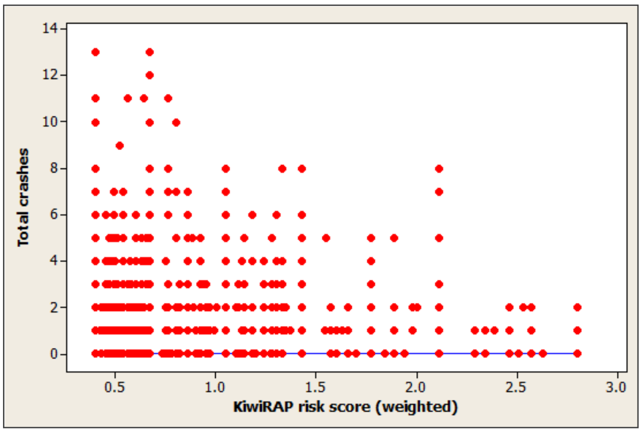


Roadside hazards

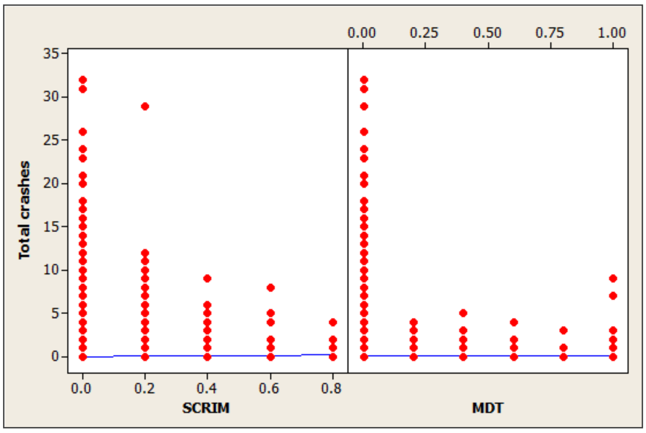
Straight sections



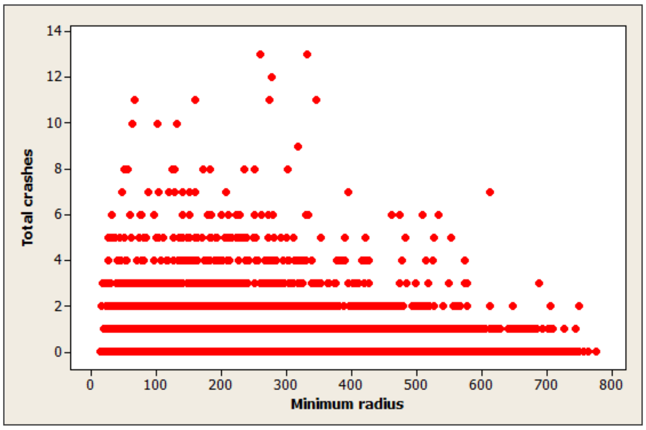
Curved sections



SCRIM



Minimum radius (curves only)



Approach speed (curves only)