CRASH RISK RELATIONSHIPS FOR IMPROVED SAFETY MANAGEMENT OF ROADS

Cenek, P.D.[1] & Davies, R.B.[2]
[1]*Opus International Consultants*
[2]*Statistics Research Associates*

ABSTRACT

This paper presents the results of a first attempt to combine detailed information on road geometry (horizontal curvature, gradient and cross-fall), road surface condition (roughness, rut depth, texture depth and skid resistance), carriageway characteristics (region, urban/rural environment, and traffic flow) and crashes. Such a study was only made possible because of annual surveys of the entire 22,000 lane-km of New Zealand's State Highway network made with SCRIM$^+$ since 1997, which involves simultaneous measurement of road condition and road geometry. Four subsets of road crashes were investigated: 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; and selected injury and fatal crashes occurring in wet conditions. One and two-way tables and Poisson regression modelling were employed to identify critical variables and the form of their relationship with crash risk. The critical variables common to all crash types investigated were horizontal curvature, traffic flow, skid resistance and to a lesser extent lane roughness. The resulting Poisson regression model uses 2nd or 3rd order polynomial functions of these variables to allow for the observed non-linear responses. Therefore, the model can be incorporated in existing road asset management systems. A comparison of observed and predicted crash numbers for different segments of the State Highway network showed that the model can provide estimates of crash numbers that are sufficiently accurate for safety management purposes. For example, the predicted effect of increasing the level of skid resistance was in line with the results from a paired crash site analysis, which considered changes in the number of crashes and road surface skid resistance at two different points in time at specific crash sites.

Key Words: crash rates, crash risk modelling, road surface condition, road geometry, roughness, texture, skid resistance.

1 INTRODUCTION

Each year since 1997, Transit New Zealand surveys the entire sealed length (10,736 km) of New Zealand's State Highway network for road condition (roughness, rutting, texture and skid resistance) and geometry (horizontal curvature, cross-fall and gradient). In addition, to measure the traffic demand imposed on the State Highway network, Transit New Zealand maintains estimates of traffic flows. The flows are typically estimated from individual counts over one-week periods. The traffic monitoring sites are distributed throughout the length of the State Highway network and are counted one, two or three times a year, with some sites counted on a continuous basis. The Land Transport Safety Authority (LTSA), through its crash analysis system (CAS), maintains data on all fatal and injury road crashes attended by the New Zealand Police. The crash data includes details about the location, time, distance, drivers involved, casualties and crash circumstances and cause factors.

These data sources, when combined, enable statistical modelling techniques to match crash rate with road characteristics. Such an analysis allows a broad-brush approach to the entire State Highway network, which is in contrast to studies of individual sites, such as black spot sites. Generally, crash rates in New Zealand are too low to allow consistent conclusions to be drawn about the relationship between road characteristics and road crashes from before and after treatment comparisons at individual crash sites. The kind of analysis presented in this paper, by using data from the whole State Highway network, in effect, combines the data from individual potential crash sites, including those where there were no crashes, and so provides estimates of crash risk that can be used with a degree of confidence to evaluate the cost-effectiveness of road geometry and road condition related safety interventions. However, such an analysis cannot take into account all special features of each section of road, such as specific hazards, and so provides only an average estimate of crash risk.

This paper summarises the results of two analyses of state highway data for the period 1997 to 2002. The first analysis utilised one and two-way tables to provide a preliminary indication as to what road condition and road geometry factors affect crash rate. The second analysis involved Poisson regression modelling to better identify the important predictor variables and how they influence crash rate.

The ability to reliably predict crash rates is very important in the safety management of road networks because it can help in identifying hazardous locations, locations that require treatment and locations where deviations (either higher or lower rates) from expected (predicted) warrant further examination.

2 DATA

2.1 ROAD CHARACTERISTICS

The annual survey of road condition and road geometry over the 6-year period from 1997 to 2002 period was performed by SCRIM$^+$, a truck based multifunctional road monitoring device. Texture (MPD, mm), skid resistance (SCRIM Coefficient), gradient (%), horizontal curvature (radius, m) and cross-fall (%) were recorded over 10m intervals whereas roughness (IRI, m/km) and rut depth (mm) were recorded over 20m intervals.

The 10m data was used as the basis for linking the other datasets used in generating the base file for analysis, which required matching both sides of the road (i.e. increasing and decreasing survey directions). Therefore, multi-lane roads were automatically excluded from the analysis.

Table 1 tabulates the number of 10m segments of state highway over each year of the analysis period that road condition and road geometry data was available for.

Table 1: Number of 10m Road Segments Surveyed with SCRIM+

Nominal Survey Year	Survey Period	No. of 10m Segments	
		Left (Increasing) Lane	Right (Decreasing) Lane
1997	March-May 1997	992649	994692
1998	March-May 1998	1019740	1019371
1999	December 1998 – March 1999	1031110	1025371
2000	December 1999 – May 2000	1046583	1040801
2001	December 2000 – March 2001	1055997	1056202
2002	November 2001 – March 2002	1061474	1062054

With reference to Table 1, the nominal survey year was used for linking road condition and road geometry data with the CAS crash data. Although it might have been better to use the road condition/road geometry data closest to the date of the crash to allow for any intervening maintenance activity, this would have added substantially to the complexity of managing the data and the likelihood of error.

2.2 CRASH DATA

The crash data was extracted from LTSA's CAS database in October 2003 and so is expected to include all reported injury (including fatal) crashes for 1997 to 2002. The statistical analyses were applied to each of the 4 subsets of the crash dataset tabulated in Table 2, with the relevant vehicle movement codes, as used by the LTSA for crash investigation monitoring analysis, summarised in Table 3.

Table 2: Description of Crash Dataset Subsets

Group	Criteria
All	All injury and fatal crashes
Selected	All injury and fatal crashes with LTSA vehicle movement code being one of A, B, C, D, F
Wet*	All injury and fatal crashes with the road wet field being W or the cause code being 801 or 901.
Wet & selected	Satisfying both the wet and selected criteria

* 801/901 = skidding/loss of control crashes

Table 3: Description of Vehicle Movement Codes

Movement Type	Description
A	Overtaking and Lane Change
B	Head On
C	Lost Control or Off Road (Straight Roads)
D	Cornering
F	Rear End

Table 4 shows the number of crashes (including those on multi-lane carriageways) that were able to be located and the percentage they represent of all crashes recorded as occurring on the State Highway network. The reasons why only about 75% could be located are as follows:

- insufficient data about the location
- location data does not correspond to a valid section of state highway
- location not surveyed by SCRIM$^+$

With reference to Table 4, there appears to be a marked improvement between 1997 and 2002 in the recording of crash locations on the State Highway network.

Table 4: Located Road Crash Numbers by Year

Year	Crash Subsets			
	All	**Selected**	**Wet**	**Wet & Selected**
1997	2159 (66%)*	1443 (68%)	550 (66%)	415 (68%)
1998	2112 (70%)	1418 (71%)	444 (66%)	343 (68%)
1999	2222 (72%)	1609 (77%)	551 (73%)	444 (77%)
2000	2115 (74%)	1552 (79%)	418 (77%)	340 (81%)
2001	2452 (76%)	1770 (80%)	494 (73%)	377(76%)
2002	3034 (86%)	2166 (91%)	603 (84%)	460 (89%)
Total	**14094 (74%)**	**9958 (78%)**	**3060 (73%)**	**2379 (76%)**

* ( ) pertains to corresponding percentage of all crashes of that type recorded on the State Highway Network

2.3 LINKING OF CRASH DATA TO ROAD INFORMATION

The original intention was to use crash data held in the accident table contained with Transit New Zealand's RAMM database as this is automatically linked to road condition data also held in RAMM via Transit New Zealand's 'route position' location referencing system (LRS), which is distance based. Route positions provide a unique address for each location on the State Highway network and are measured in increasing direction from the preceding Reference Station (RS). Along state highways, Reference Stations are located at approximately 16 km intervals.

A comparison of all reported injury (including fatal) crash records for the period 1997 to 2002 held in RAMM accident table with those held in the LTSA's Crash Analysis System (CAS), however, showed a significant discrepancy in the number of records. This was attributed to the time delays in entering crash records in the RAMM database and also difficulties with locating crashes in terms of Transit's LRS. Therefore, because of the need to use as large as possible crash database, a decision was made to use the crash data held in CAS. This approach required the development of a procedure to link the spatially referenced crash data to the linearly referenced state highway data.

The following procedure was adopted to systematically link crash data with road condition, road geometry, and traffic data from the road maintenance and management system (RAMM).

Crash map co-ordinates, in terms of XY geocoding, are calculated by CAS from crash reports supplied by the New Zealand Police. These co-ordinates are matched to state highway centreline segments recorded in Critchlow Associates Ltd (www.critchlow.co.nz) database. The centreline segments, in turn, are located to a route position (RP). This operation is also automatically performed within CAS.

Although Critchlow's annually rebuild the links between the centreline data and RAMM, mismatches may occur because of changes to Transit's LRS that occur within a year and between years because of construction and reconstruction. Therefore, CAS derived geocoordinates were additionally matched to the nearest centreline point from Transit's centreline database. Transit's centreline data is automatically linked to the LRS and more accurate than Critchlow's because windy sections are not as simplified. For example, using Critchlow's centreline database, the derived location of RP's can be up to 300m in error (1% error) in the 30km section into the Manawatu gorge.

A comparison of the LRS location of 18172 crashes showed good agreement between both methods for about 70%. Therefore, LTSA/Critchlow derived route positions were used for these crashes. For the remaining 30%, LTSA/Critchlow derived route positions were suitably updated using Transit centreline data to reflect changes such as new RS numbers. Where possible, road features listed in the police crash reports were used in addition to geocoordinates to assist in locating the crashes.

This time consuming exercise highlighted the benefits of using spatial methods, i.e. GPS, for location referencing so as to allow easier integration of crash and state highway data.

3 ONE and TWO-WAY TABLES

3.1 CLASSIFICATION OF DATA

In order to obtain an indication as to what is affecting crash rates, segments of the State Highway network were divided into categories using one or two road characteristics and the average crash rate for each category calculated from the corresponding crash number and road length totals. The road condition and road geometry parameters considered were the average of both (i.e. increasing and decreasing) lanes.

The resulting one and two way tables, although useful for identifying trends, can be misleading for the following two reasons:

- they do not take account of errors in locating crashes;
- the observed variation in crash rate may be due to a variable not included in the table, but correlated with the variable(s) included in the table.

In addition, the calculated crash rates may be subject to substantial statistical error whenever the number of crashes is less than 25.

3.2 ONE-WAY TABLES

The following tables were generated for the 'All Crashes' dataset.

Table 5 shows that as traffic volumes decrease, the crash rate increases. This is as expected because the quality of a road reflects average daily traffic (ADT), with lower ADT suggesting more challenging roads i.e. narrower lanes and more tortuous alignments and it is these road characteristics that would be expected to lead to a higher crash rate.

Table 5: Classification by Average Daily Traffic (ADT) – All Crashes

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 6 shows increasing crash rate as horizontal curvature decreases. However, the smallest curvature grouping is likely to include intersections, so the much higher crash rate observed for curves less than 100m radius may result from other hazards apart from the curve itself.

Table 6: Classification by Horizontal Curvature – All Crashes

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 ≤ R < 100	125	262	518	51
100 ≤ R < 1000	2845	4277	17457	25
1000 ≤ R < 10000	5273	6290	39620	16
10000 ≤ R < 100000	1835	1973	14663	13
R ≥ 100000	20	28	179	16

Transit New Zealand's policy for skid resistance is largely contained within the T/10 specification. This specification was introduced in 1998 and aims to standardise the risk of a wet skid crash across the State Highway network by assigning investigatory skid resistance levels for different site categories, which are related to different friction demands. A description of these site categories and associated investigatory levels are summarised in Table 7.

Tables 8 and 9 show the effect on crash rate of skid resistance level and T/10 site category, respectively. With reference to Table 8, there is an increase in crash rate for lower values of wet road skid resistance as measured in terms of SCRIM Coefficient (SC). A one-way table, such as Table 8, is likely to underestimate the effect of skid resistance on crash rates because surfaces displaying greater skid resistance will be used on more hazardous road sections as a safety measure.

Table 9 indicates that the crash rate is much higher for T/10 site category 1 (roundabouts, railway crossings etc) when compared with site category 4 (normal roads). However, crash rates for site category 1 may be underestimated because, where a road segment has different skid site categories between one (entry) side of the intersection or crossing and the other (exit) side, the lower site category has been used.

Table 7: T/10 Skid Site Categories

Site Category	Description	Notes	Investigatory Level (SC)
5*	Divided carriageway		0.35
4	Normal roads	Undivided carriageways only.	0.4
3	Approaches to road junctions		0.45
2	Curve < 250m radius Gradient > 10%		0.5
1	Highest priority	Railway level crossing, approaches to roundabouts, traffic lights, pedestrian crossings and similar hazards.	0.55

* Not used in analysis

Table 8: Classification by Pavement Skid Resistance – All Crashes

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 ≤ SC < 0.4	294	730	3125	23
0.4 ≤ SC < 0.5	2610	5144	28048	18
0.5 ≤ SC < 0.6	4953	5421	32649	17
0.6 ≤ SC < 0.7	2046	1287	7637	17
SC ≥ 0.7	116	62	372	17

Table 9: Classification by T/10 Skid Site Category – All Crashes

T/10 Skid Site Category	Road Length (km)	Number of Crashes between 1997 & 2002	Total Traffic Exposure (10^6 v-km)	Crash Rate (10^8 vkt)
4	7275	6980	52625	13
3	1264	2935	11165	26
2	1448	2237	6875	33
1	77	493	1004	49

3.3 TWO-WAY TABLES

When the classifying variables are considered two at a time, the crash numbers are much smaller than in the one-way tables and so there is a substantial amount of statistical fluctuation. Therefore crash rates are bolded when the corresponding observed number of crashes is at least 25 since there is not much accuracy in the data when the number of crashes is less.

Table 10 shows the crash rate increases as the radius of curvature decreases and some increase as ADT decreases but less than shown in Table 5. This implies that some of the apparent effect of ADT is reduced when road curvature is allowed for.

Table 10: Crash Rate by Horizontal Curvature and ADT – All Crashes

Horizontal Curvature, R (m)	**Crashes per 10^8 vkt**					
	ADT range (1000 vehicles per day)					
	ADT<1	1≤ADT<2	2≤ADT<5	5≤ADT<10	10≤ADT<20	ADT≥20
10 ≤ R < 100	**53**	**54**	**51**	**48**	**54**	33
100 ≤ R < 1000	**35**	**29**	**26**	**24**	**20**	**13**
1000 ≤ R < 10000	**22**	**19**	**16**	**16**	**15**	**11**
10000 ≤ R < 100000	**16**	**15**	**14**	**13**	**13**	**11**
R ≥ 100000	100	12	12	16	15	0

Table 11 shows that within each range of SCRIM Coefficient values, the crash rate increases as the T/10 skid site category decreases and for each T/10 skid site category the crash rate increases as the level of skid resistance provided by a road surface decreases. The latter effect appears to be strongest for the lowest skid resistance grouping (SC < 0.3), although the accuracy of the calculated crash rate in this case is not high because the number of crashes involved is very small (17 or less).

Table 11: Crash Rate by T/10 Site Category and SCRIM Coefficient – All Crashes

T/10 Skid Site Category	**Crashes per 10^8 vkt**					
	SCRIM Coefficient Range					
	SC< 0.3	0.3≤SC<0.4	0.4≤SC<0.5	0.5≤SC<0.6	0.6≤SC<0.7	SC≥0.7
4	17	**16**	**13**	**13**	**14**	**12**
3	44	**29**	**27**	**26**	**23**	32
2	62	**39**	**33**	**31**	**31**	33
1	0	**44**	**52**	**47**	**47**	40

Table 12 investigates horizontal curvature and skid resistance as classifying variables. Crash rate is shown to increase as the level of skid resistance decreases within each curvature range or as the radius of curvature decreases within each skid resistance range.

Table 12: Crash Rate by Horizontal Curvature and SCRIM Coefficient – All Crashes

Horizontal Curvature, R (m)	**Crashes per 10^8 vkt**					
	SCRIM Coefficient Range					
	SC< 0.3	0.3≤SC<0.4	0.4≤SC<0.5	0.5≤SC<0.6	0.6≤SC<0.7	SC≥0.7
10 ≤ R < 100	55	**48**	**54**	**43**	61	40
100 ≤ R < 1000	55	**30**	**25**	**23**	**23**	24
1000 ≤ R < 10000	13	**19**	**16**	**15**	**15**	14
10000 ≤ R < 100000	32	**21**	**14**	**12**	**13**	14
R ≥ 100000	0	26	15	11	39	0

Table 13 repeats Table 12 for crashes classified as occurring on wet roads. As figures for traffic exposure to wet roads are not available, crash rates are for wet road crashes in terms of total traffic. Therefore the crash rates presented in Table 13 are significantly smaller and display greater statistical error because crash numbers are also smaller. With reference to Table 13, the general form of the results is the same as for Table 12. However, as one might expect, the effect of skid resistance is much stronger.

Table 13: Crash Rate by Horizontal Curvature and SCRIM Coefficient – Wet Crashes

Horizontal Curvature, R (m)	**Crashes per 10^8 vkt**					
	SCRIM Coefficient Range					
	SC< 0.3	0.3≤SC<0.4	0.4≤SC<0.5	0.5≤SC<0.6	0.6≤SC<0.7	SC≥0.7
10 ≤ R < 100	55	17	11	14	5	0
100 ≤ R < 1000	19	**11**	**7**	**5**	**5**	5
1000 ≤ R < 10000	1	**5**	**4**	**3**	**2**	1
10000 ≤ R < 100000	4	**5**	**3**	**2**	**3**	0
R ≥ 100000	0	13	7	4	7	0

4 THE MODEL

A model, which relates a variety of road characteristics exponentially to crash risk, has been developed from a statistical analysis that investigated the dependency of observed crash rates to road condition and road geometry data acquired during annual surveys of the State Highway network. The analysis assumed that the crashes were statistically independent and the number of crashes that occur in each 10m road segment follow a Poisson distribution (of course, for most segments the number of crashes was zero). The fundamental form of the model is given below.

$$\text{Expected number of crashes per year} = ADT.e^{L} \qquad (1)$$

where ADT = is the average daily traffic

L = is the weighted sum of the values of the various road characteristics such as:

- absolute gradient
- horizontal curvature
- cross-fall
- T/10 skid-site category
- skid resistance (SCRIM Coefficient)
- log10(ADT)
- year
- TNZ administration region
- urban/rural classification

The exponent, L, is the sum of a number of variables that are either assigned values depending on the road characteristic (e.g. Urban / Rural road) or are the product of a coefficient multiplied by the value of the road characteristic (e.g. A x Curvature). These values and coefficients were determined by fitting the road data to the variables using the method of maximum likelihood.

The expected number of crashes per year equation given above can be converted to an equation for crash rate (number of crashes per 10^8 vehicle-km) by multiplying by the factor, 10^8/(ADT.365.Road Length). Crash data has been analysed over 10m sections, giving a road length of 10^{-2} km. Therefore, substituting equation 1 gives the crash rate as:

crash rate (crashes per 10^8 vehicle.km)= ADT.e^L x 10^8/(ADT.365.10^{-2})

This simplifies to:

$$\textbf{crash rate} = \frac{10^{10}}{365} e^{L} \qquad (2)$$

A number of analyses were carried out on the different data subsets, comprising 'All Crashes', 'Selected Crashes' (excludes certain types of crashes such as merging and pedestrian crashes), 'Wet Road Crashes', and 'Selected Wet Road Crashes' (only crashes on wet roads, excludes certain types of crashes). While crash rate models have been developed for each of these datasets, only the 'All Crashes' model is discussed in this paper. However, the model coefficients for the 'Selected Crashes', 'Wet Road Crashes', and 'Selected Wet Road Crashes' are given in Table 15.

Within the analysis of the 'All Crashes' data, two models were developed. The first was a complex model, which used spline curves to fit the variables. These curves are illustrated in Figures 1 – 11, and provide a good appreciation of how the various road characteristics considered affect crash rate. However, difficulties associated with applying the spline curves within a spreadsheet precluded the model's widespread use and so a simplified model was developed that used polynomial curves instead to fit the data. This simplified model gave coefficients that are relatively straightforward to apply, and are presented in Section 4.2.

4.1 PREDICTED CRASH RATES FROM COMPLEX MODEL

The complex crash rate model had a number of variables that were used to form the exponent L, and these are listed in Table 14. Figures 1 – 11 illustrate the influence of each of the variables on the crash rate. The other variables in each graph are held constant, taking the default values given in Table 14. The error bounds shown in each plot correspond to a 95% confidence interval.

Table 14: Default Graphing Values for Variables used in the Complex Model (Figures 1 – 11)

variable	value		variable	value
year	2002		log10_iri	0.3
region	R1		rut_depth	3
urban_rural	R		cway_width	12
skid_site	4		texture	1.5
curvature	5000		lanes_category	TwoLane
ADT	1000		irr_width	R
gradient	0		cross_fall	0
SCRIM	0.5			

Figure 6 is difficult to interpret because upward and downward gradients cannot be distinguished. Otherwise the plotted graphs show expected trends though the crash rate relationships shown in Figures 9-11 probably arise from random error and so do not show the true effect of rut depth, carriageway width and texture, respectively.

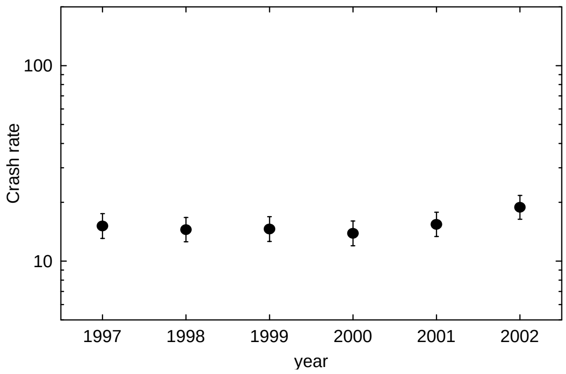

Figure 1 Crash rate versus year

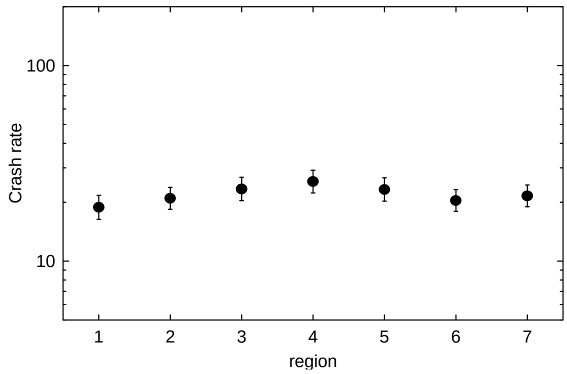

Figure 2 Crash rate versus TNZ region

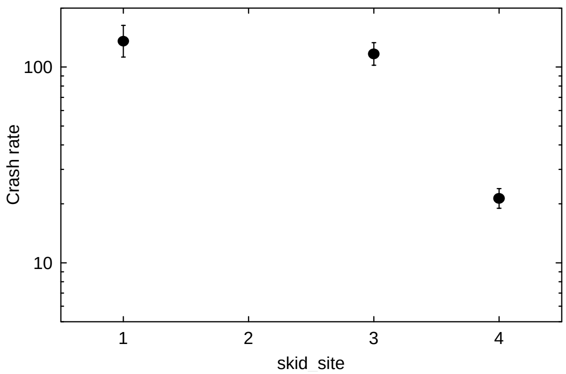

Figure 3 Crash rate versus T/10 skid site

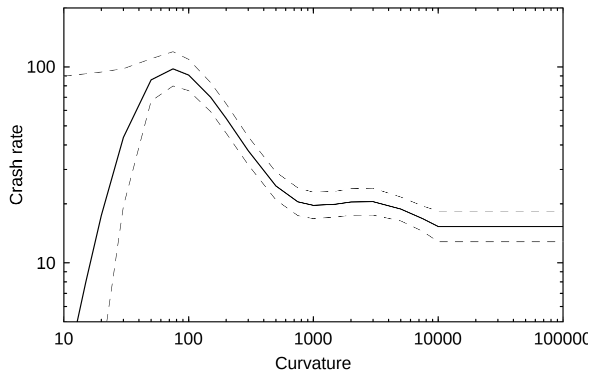

Figure 4 Crash rate versus curvature

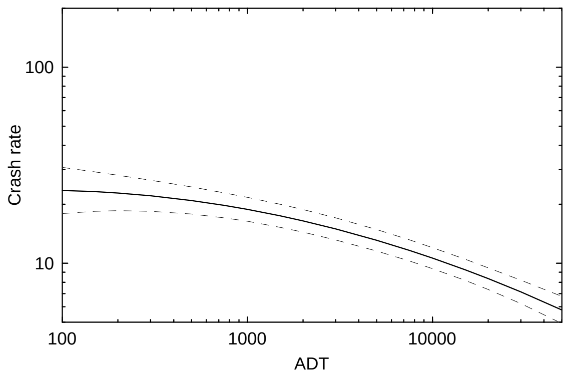

Figure 5 Crash rate versus ADT

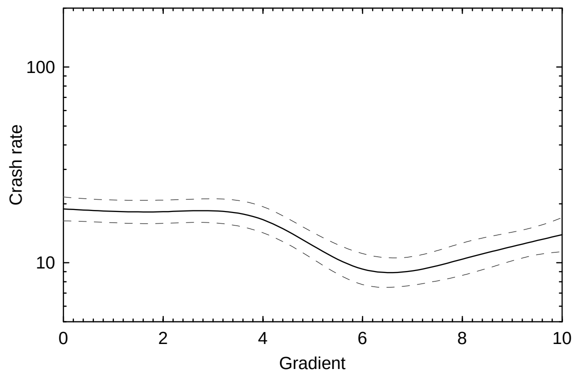

Figure 6 Crash rate versus gradient

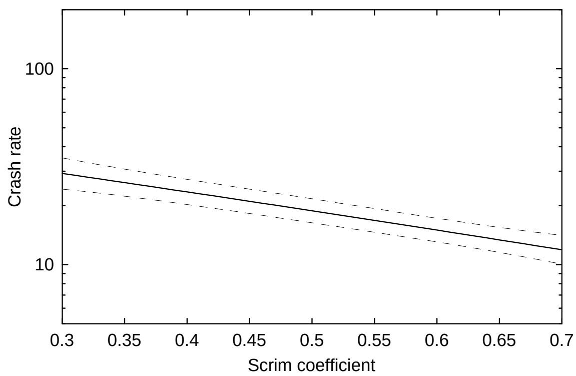

Figure 7 Crash rate versus SCRIM Coefficient

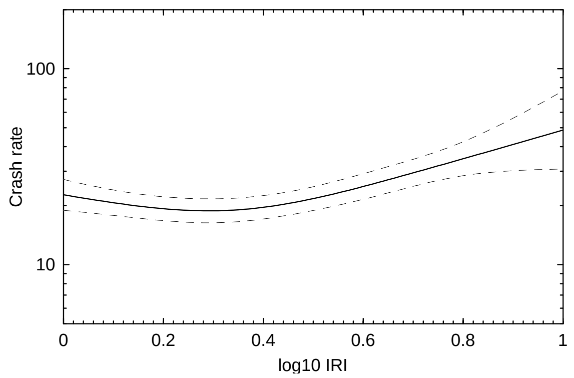

Figure 8 Crash rate versus $\log_{10}$IRI

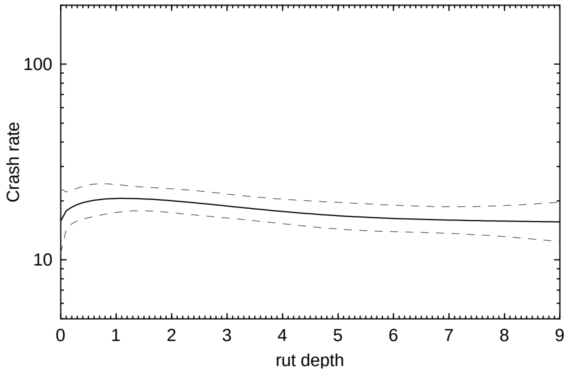

Figure 9 Crash rate versus rut depth

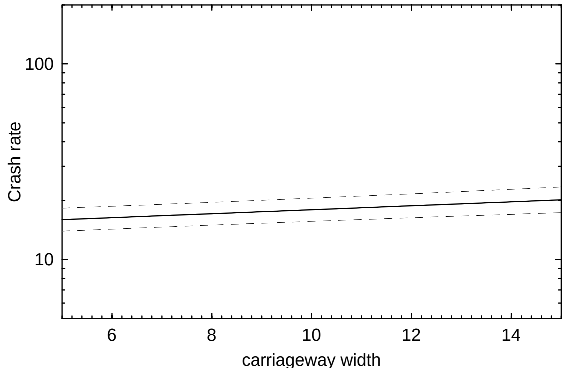

Figure 10 Crash rate versus carriageway width

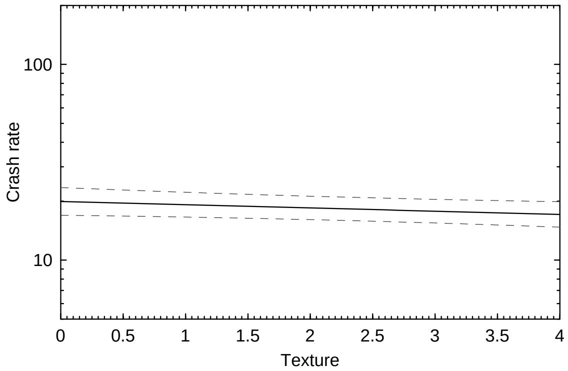

Figure 11 Crash rate versus texture

4.2 SIMPLIFIED CRASH RATE MODEL

A simplified crash rate model was developed from the more complex model to allow ease of use for a wide variety of users. The simplified model employs polynomial equations instead of the spline curves used in the complex model. While slightly less accurate than the spline representations, these polynomial equations are easily represented by a number of coefficients, which are given in Table 15 below.

Limitations in the range of data that was available for the model fitting and the analysis method, means that the model is limited in its applications to the following parameter ranges:

year:	1997 to 2002 (beyond these years requires estimation of the yearly coefficient)
region:	R1 to R7 (= TNZ Administration Regions, where R1=Auckland, R2=Hamilton, R3=Napier, R4=Wanganui, R5=Wellington, R6=Christchurch and R7=Dunedin)
urban_rural:	U (urban) or R (rural)
skid_site:	T/10 site category 1, 3 or 4 (category 2 has been combined into category 4)
curvature:	100 to 10000m radius (absolute value used, i.e. does not differentiate left from right hand curves). For radii outside this range use 100m for values less than 100m and 10000m for values greater than 10000m
ADT:	average daily traffic, unlimited range of values
gradient:	4 to 10 (absolute value is used, and values less than 4 are set equal to 4)
SCRIM:	0.3 to 0.7 SCRIM Coefficient
IRI:	2.0 to 10.0 IRI (m/km) lane roughness

The predicted crash rate is found by applying equation 2, in which L is first evaluated using Table 15. L is the sum of the terms, which are calculated using the coefficients in Table 15. Terms corresponding to categorical variables (i.e. year, region, urban_rural, skid_site) simply take the value of the corresponding coefficient in Table 15, while terms associated with the continuous variables (i.e. curvature, ADT, gradient, SCRIM Coefficient and IRI) are found by multiplying the variable by the corresponding coefficient. An example calculation for determining the crash rate is given in Section 4.3.

Table 15 Coefficients for the Simplified Crash Rate Model

Parameter	All Crashes		Selected Crashes		Wet Road Crashes		Selected Wet Road Crashes	
	coefficient	standard error	coeff.	std. error	coeff.	std. error	coeff.	std. error
constant	2.095	1.76	-0.541	2.01	1.015	3.43	0.008	3.83
year: 1997	0.000		0.000		0.000		0.000	
year: 1998	-0.060	0.03	-0.049	0.04	-0.240	0.07	-0.216	0.08
year: 1999	-0.053	0.03	0.044	0.04	-0.027	0.06	0.059	0.07
year: 2000	-0.118	0.03	-0.014	0.04	-0.331	0.07	-0.240	0.08
year: 2001	0.000	0.03	0.089	0.04	-0.203	0.07	-0.175	0.08
year: 2002	0.198	0.03	0.278	0.04	-0.002	0.07	0.008	0.08
region: R1	0.000		0.000		0.000		0.000	
region: R2	0.108	0.03	0.074	0.04	0.192	0.07	0.188	0.08
region: R3	0.210	0.05	0.206	0.05	0.101	0.10	0.091	0.11
region: R4	0.306	0.04	0.260	0.04	0.565	0.08	0.537	0.09
region: R5	0.224	0.04	0.154	0.05	0.053	0.09	0.041	0.11
region: R6	0.105	0.04	0.090	0.05	0.146	0.09	0.161	0.10
region: R7	0.124	0.04	0.164	0.05	0.045	0.09	0.073	0.10
urban_rural: R	0.000		0.000		0.000		0.000	
urban_rural: U	-0.157	0.03	-0.416	0.04	-0.272	0.06	-0.595	0.09
skid_site: 4	0.000		0.000		0.000		0.000	
skid_site: 3	1.595	0.04	0.569	0.07	1.528	0.08	0.561	0.15
skid_site: 1	1.697	0.08	0.803	0.15	1.175	0.20	0.100	0.47
$\log_{10}$(|curvature|)	-5.360	0.29	-5.036	0.33	-7.426	0.57	-6.329	0.63
$\log_{10}$(|curvature|^2)	0.759	0.05	0.683	0.05	1.048	0.09	0.843	0.10
$\log_{10}$ (ADT)	0.707	0.31	1.129	0.37	2.380	0.71	2.516	0.80
$\log_{10}$ (ADT^2)	-0.173	0.04	-0.247	0.05	-0.401	0.10	-0.424	0.11
|gradient|	-2.598	0.70	-1.411	0.76	-2.913	1.33	-2.802	1.40
|gradient|^2	0.314	0.11	0.202	0.12	0.396	0.21	0.443	0.22
|gradient|^3	-0.012	0.01	-0.009	0.01	-0.017	0.01	-0.022	0.01
SCRIM-0.5	-1.637	0.16	-2.177	0.18	-3.551	0.33	-4.073	0.37
(SCRIM-0.5)^2	-0.090	1.30	1.790	1.47	3.344	2.48	6.220	2.60
$\log_{10}$ (iri)	-10.540	4.48	-18.556	5.96	-7.348	8.48	-17.379	11.50
[$\log_{10}$ (iri)]^2	19.219	8.48	31.537	11.39	10.916	15.65	29.938	21.84
[$\log_{10}$ (iri)]^3	-9.850	4.99	-15.504	6.77	-3.563	8.89	-14.644	12.92

4.3 EXAMPLE CALCULATION USING THE SIMPLIFIED MODEL

The following example shows the procedure for calculating the crash rate using the simplified 'All Crashes' model presented in Section 4.2. First the exponent, L is evaluated, as shown in Table 16.

Table 16 Example Application of Simplified 'All Crashes' Crash Rate Model

parameter	parameter value	calculation value	corresponding coefficient †	product (value x coefficient)
constant		1	2.095	2.095
year	2002	1	0.198	0.198
region	R2	1	0.108	0.108
urban_rural	Rural	1	0.000	0.000
skid_site	4 *	1	0.000	0.000
$\log_{10}$(|curvature|)	300	2.477	-5.360	-13.277
$\log_{10}$(|curvature|^2)	300	4.954	0.759	3.760
$\log_{10}$ (ADT)	10000	4	0.707	2.828
$\log_{10}$ (ADT^2)	10000	8	-0.173	-1.384
|gradient|	0 **	4	-2.598	-10.392
|gradient|^2	0 **	16	0.314	5.024
|gradient|^3	0 **	64	-0.012	-0.768
SCRIM-0.5	0.45	-0.05	-1.637	0.082
(SCRIM-0.5)^2	0.45	0.0025	-0.090	0.000
$\log_{10}$ (iri)	3	0.477	-10.540	-5.029
[$\log_{10}$ (iri)]^2	3	0.228	19.219	4.375
[$\log_{10}$ (iri)]^3	3	0.109	-9.850	-1.070
				sum = **-13.450**

Notes:
† coefficients taken from Table 15
* skid_site category 2 has been combined with skid _site category 4.
**gradients between 0 and 4 default to a value of 4

Next the crash rate is calculated using equation 2: $\frac{10^{10}}{365} e^{L} = \frac{10^{10}}{365} e^{-13.450} = 39.5$

Finally, a correction should be made for the crashes that could not be located on the State Highway network and so were excluded from the analysis and model predictions. Table 4 gives the percentage of 'All Crashes' that could be located in 2002 as 86%. Therefore, multiplying the calculated crash rate by $^{100}/_{86}$ will give the true crash rate in 2002:

crash rate = 39.5 x 100 / 86 **= *45.9 crashes per 10^8 vehicle.km***

It is worth remembering that this figure is derived from reported crashes, and that the actual crash rate, including unreported crashes, will be higher.

4.4 COMPARISON OF FITTED AND OBSERVED CRASHES

The fit of the 'All Crashes' model is tested below by looking at the differences between the measured and predicted crashes. The State Highway network was divided up, using carriageway area and state highway number, to give 136 individual areas. The model was used to predict the number of crashes in each of these areas. The observed numbers are compared with the predicted numbers in Figure12 below. Figure 13 shows a plot of the residuals (i.e. the differences between the observed and predicted values), which have been normalised as shown in equation 3.

$$\text{residual} = \frac{\text{observed} - \text{predicted}}{\sqrt{\text{predicted}}} \qquad (3)$$

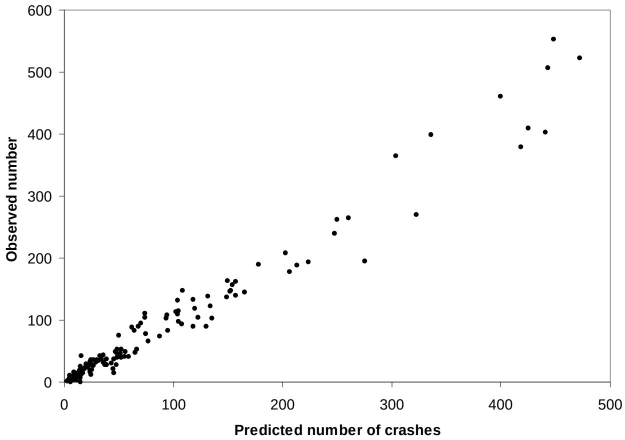

Figure 12 Predicted versus Observed Crashes

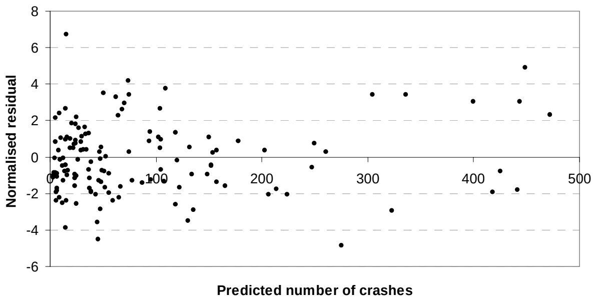

Figure 13 Normalised Residual Plot

Ideally there should be few normalised residuals outside the range –2 to 2, indicating that the model fits perfectly. The actual range of residuals is more like –4 to 4, with a few outside this range, showing that the model fits the data well, but not perfectly.

4.5 EFFECT OF SKID RESISTANCE

A previous study (Cenek et al, 2002) using paired crash site analysis, which considered changes in the number of crashes and road surface skid resistance at two different points in time at specific crash sites, found the 95% confidence interval for the crash rate reduction factor per 0.1 increase in SCRIM Coefficient to be:

- (1.2, 1.7) for a comparison of 1995 and 1998 data
- (1.1, 1.8) for a comparison of 1995 and 1999 data

Applying the simplified model for the 'Wet Selected Crashes' data subset with only a linear function of SCRIM Coefficient gave the 95% confidence interval for the crash rate reduction factor per 0.1 increase in SCRIM Coefficient as (1.4, 1.7). Since this is in general agreement with the previous estimates, there can be a degree of confidence that the simplified model can provide estimates of crash rates that are sufficiently accurate for safety management purposes.

5 CONCLUDING REMARKS

The Poisson regression model presented appears to work in a reasonably satisfactory way and produces results that, for the most part, make sense. For example, curvature has a strong effect on crash rate as expected. There is also a strong effect for skid resistance and a weaker effect for lane roughness. However, as this has been a retrospective analysis (as opposed to a designed experiment), it is not possible to be sure that the predictor variables used in the regression analysis are really the ones affecting the crash rates. In particular, it is likely that ADT is a general indicator of road quality and this is leading to the observed drop in crash rate as ADT increases.

The simplified model in its current form is sufficiently robust for the following four applications:

1. To improve the understanding of the factors affecting crash risk and the relative importance of different factors.
2. To improve the management of the highway network by estimating the effect on crash numbers of changes in standards for curvature, skid resistance and roughness.
3. To identify black spot regions where, because of factors not included in the model, crash rates are much higher than predicted by the model. It may also be possible to detect white spots where crash rates are lower, although this is less likely to be successful.
4. To use the model to help evaluate the effect of an actual change in road construction or management policy in a Transit New Zealand administration region by comparing the observed and predicted number of crashes.

6 REFERENCES

Cenek, P.D, Loader, M.D.J. and Davies, R.B. (2002): *Statistical Analyses of State Highway Crash and Skid Resistance Data: 1995-2000*, Central Laboratories Report No. 02-529295.00 prepared for Transit New Zealand.

TNZ (2002): TNZ T/10, Specification for Skid Resistance Investigation and Treatment Selection, Transit New Zealand.

7 ACKNOWLEDGEMENT

Preparation of this paper was funded by Transfund New Zealand as part of research contract PR3-0709.