Crash risk relationships

1 Introduction

Each year since 1997 the New Zealand State Highway system has been surveyed for road characteristics such as curvature, skid resistance, roughness, texture. The LTSA collects data on reported injury road crashes. Transit New Zealand has estimates of the traffic levels on the State Highways.

This data enables us to use statistical modelling techniques to match crash rates with the road characteristics.

Our analysis gives a broad brush approach of the whole network. It is to be contrasted with studies of individual sites such as *black spot* sites. In general, crash rates are too low to give much information on individual sites analysed by themselves. The kind of analysis given here, by using the data from the whole network, in effect, combines the data from the individual potential crash sites including those where there were no crashes, and so provides usable estimates of risk. It cannot, of course, take account of all special features of each section (such as specific hazards) of road so gives some kind of average figure. In particular, if one was using the model derived here for estimating the risk associated with a particular piece of road one might also want to adjust for any unusual features of that piece of road.

This paper describes two methods of analysis. The first uses one and two way tables to provide a simple preliminary view of the data. The second uses a Poisson regression model.

A description of the data and its preliminary processing is given in section 2. The analysis using the one and two way tables is in section 3.

The Poisson regression model is described in section 4 and the model is fitted in section 5. Two versions of the model are tried. The first is a larger version including all of the predictor variables. It uses spline functions of the more important variables to allow for a non-linear response. The second version includes only the more important predictor variables and avoids the use of the spline functions since these would be hard to use separately from the fitting program. Section 5 also includes two examples showing how to calculate the crash risks predicted by the model.

Section 6 shows the results of additional testing of the model.

Section 7 shows a possible application of the model: the effect on crash numbers of upgrading the skid resistance.

Section 8 includes a summary of the main results, some comments on their credibility and some suggestions for future work.

Some notes on the data are given in section 9.

2 The data

2.1 Linking the data

We have four main sets of data for each of the years 1997 to 2002

- Data collected by high speed data collection machine at 10 metre intervals
  - texture
  - scrim coefficient
  - scrim mssc
  - gradient
  - curvature
  - crossfall
- Data collected by high speed data collection machine at 20 metre intervals
  - reading date
  - roughness
  - rut depth mean
  - rut depth standard deviation
- Carriage way data
  - region
  - urban or rural
  - adt estimate
  - other road information
- Crash data
  - crash date
  - crash cause
  - crash severity
  - movement type
  - road dry, wet, icy
  - other crash information

I have used the 10 metre data as the base for our model and for linking the other files to. The 20-metre data has some overlapping segments, which would cause difficulties for the linking. Also, the more important variables for this study are in the 10-metre data set.

The actual base file is obtained by matching the left and right hand sides of the road so for each 10 metre segment we the data from the left and right hand sides of the road.

The analyses carried out so far require data to be present from both sides of the road so dual carriageway roads are automatically excluded.

I matched the carriageway data to the 10 metre data by requiring the road-ids to match and the midpoint of the 10 metre data to fall inside the starting and end points of each segment defined in the carriageway data.

For the crash data I matched the crash data year, road_id, rp with the road_id and 10-metre segment.

2.2 Length of road surveyed

Table 1 shows the number of 10 metre blocks surveyed each year.

Table 1: number of 10 metre segments surveyed

Year	Left lane	Right lane
1997	992649	994692
1998	1019740	1019371
1999	1031110	1025371
2000	1046583	1040801
2001	1055997	1056202
2002	1061474	1062054

There has been a slight increase each year.

2.3 When was high-speed data collected?

Table 2 shows the months where the majority of the 20 metre data was collected. (Presumably the 10 metre data was collected at the same time).

Table 2: months when 20 metre data was collected

Nominal year	Principal months of measurement
1997	March-May 1997
1998	March-May 1998
1999	December 1998 – March 1999
2000	December 1999 – May 2000
2001	December 2000 – March 2001
2002	November 2001 – March 2002

I have used the nominal year for matching the crash data with the high-speed data.

It might have been better to use the nearest high speed data measurement, but also allow for roads being resealed. However, this would have added substantially to the complexity of managing the data and the likelihood of error.

2.4 Which years do we have high-speed data for?

High-speed data was not collected every year. Table 3 shows the years in which data was collected.

Table 3: years when was data collected

Year	Geometry	Scrim	Roughness	Texture	Rut depth
1997	no	no	yes	yes	yes
1998	yes	yes	yes	yes	yes
1999	no	one side	yes	yes	yes
2000	yes	yes	yes	yes	yes
2001	yes	yes	yes	yes	yes
2002	yes	yes	yes	yes	yes

I used 1998 and 2000 year data for 1997 and 1999 respectively for *geometry* and *scrim*.

2.5 What years do we have crash data for?

The data was extracted from the CAS database in October 2003 data so should include all reported injury (including fatal) crashes for 1997 to 2002

2.6 Crash data subsets

The analyses were applied to each of 4 subsets of the crash dataset described in Table 4.

Table 4: subsets of crash dataset

Group	Criteria
All	All injury and fatal crashes
Selected	All injury and fatal crashes with MVMT_IDA 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 was 801 or 901.
Wet & selected	Satisfying both the wet and selected criteria

The MVMT_IDA codes are described in Table 5.

Table 5: movement code A descriptions

A	Overtaking and Lane Change
B	Head On
C	Lost Control or Off Road (Straight Roads)
D	Cornering
E	Collision with obstruction
F	Rear End
G	Turning Versus Same Direction
H	Crossing (No Turns)
J	Crossing (Vehicle Turning)
K	Merging
L	Right Turn Against
M	Manoeuvring
N	Pedestrians Crossing Road
P	Pedestrians Other
Q	Miscellaneous

2.7 Fraction of crashes identified

Not all crashes could be located on the road network. This can arise for three reasons

- insufficient data about location
- location data does not correspond to a valid road
- road not in high speed survey.

Table 6 shows the total number of crashes per year (on the state highway system) and Table 7 shows the number located.

Table 6: total crashes by year

Year	All	Selected	Wet	Selected - wet
1997	3254	2129	832	614
1998	3024	1993	673	505
1999	3081	2088	750	574
2000	2839	1966	541	422
2001	3230	2215	677	493
2002	3548	2393	715	515
Total	**18976**	**12784**	**4188**	**3123**

Table 7: crashes located

Category	All	Selected	Wet	Selected - wet
1997	2159	1443	550	415
1998	2112	1418	444	343
1999	2222	1609	551	444
2000	2115	1552	418	340
2001	2452	1770	494	377
2002	3034	2166	603	460
Total	**14094**	**9958**	**3060**	**2379**

Table 8 shows the percent of crashes located.

Table 8: percent of crashes located

Category	All	Selected	Wet	Selected - wet
1997	66%	68%	66%	68%
1998	70%	71%	66%	68%
1999	72%	77%	73%	77%
2000	74%	79%	77%	81%
2001	76%	80%	73%	76%
2002	86%	91%	84%	89%
Total	**74%**	**78%**	**73%**	**76%**

Note the lower rates for 1997 and 1998 and the higher rate for 2002.

The analyses in this paper also exclude crashes on dual lane carriageways. However, they are included in Table 7.

3 One and two-way tables

I look at crash rates when the road segments are divided into categories by one or two characteristics of the road. This gives a simple illustration of how crash rates vary with each of these characteristics.

They also show the number of crashes and the road lengths corresponding to each category.

While these tables give a useful preliminary look at the data, they can be misleading for 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 variables included in the table.

As an example of the second point: suppose we plot the crash rate against the skid resistance as in Table 11 below. Locations of high hazard, such as intersections, curves and *black spots* are supposed to have had their skid resistance raised. Hence it would be quite possible for Table 11 to show an *increase* of crash rate with skid resistance. In fact, Table 11 does show a *decrease* in crash rate with skid resistance. However it is likely that the effect we see is biased towards zero.

So these tables should be regarded only as providing indications as to what is affecting crash rates.

I am considering the road rather than the sides of the road. The surface characteristics and geometry are the averages of the values on each side. Skid-site is the *minimum* of the skid-sites on each side.

In each case the total traffic is in millions of kilometres. Crash rate is in crashes per 100 million vehicle kilometres.

The crash rate may be subject to substantial statistical error when the number of crashes is less than 50.

3.1 All crashes – one-way tables

Table 9: average daily traffic

ADT range	road length	number of crashes	total traffic	crash rate
<200	68	14	26	55
>=200,<500	650	111	517	21
>=500,<1000	2103	862	3268	26
>=1000,<2000	2646	1817	8323	22
>=2000,<5000	2538	3374	18144	19
>=5000,<10000	1485	3672	21548	17
>=10000,<20000	503	2329	14941	16
>=20000,<50000	109	660	5579	12
>=50000	0	0	158	0

This table shows the increasing crash rate as ADT decreases. The *quality* of a road reflects the ADT. So lower ADT suggests more bends, narrower roads, generally more difficult roads and it is these road characteristics that would be expected to lead to a higher crash rate.

Table 10: curvature

Curvature range	road length	number of crashes	total traffic	crash rate
>=10,<100	125	262	518	51
>=100,<1000	2845	4277	17457	25
>=1000,<10000	5273	6290	39620	16
>=10000,<100000	1835	1973	14663	13
>=100000	20	28	179	16

This table shows increasing crash rate as the radius of curvature decreases. The very small radii curvatures may indicate intersections, so the much higher rate may result from other hazards apart from the curve itself.

Table 11: scrim

Scrim range	road length	number of crashes	total traffic	crash rate
<0.3000	18	40	150	27
>=0.3000,<0.4000	294	730	3125	23
>=0.4000,<0.5000	2610	5144	28048	18
>=0.5000,<0.6000	4953	5421	32649	17
>=0.6000,<0.7000	2046	1287	7637	17
>=0.7000	116	62	372	17

This shows an increase in crash rate for the lower values of skid resistance. A table one-way like this is likely to understate the effect of skid resistance since the skid resistance is likely to be increased on more hazardous sections of road as a safety measure.

Table 12: skid site

Skid site	road length	number of crashes	total traffic	crash rate
4	7275	6980	52625	13
3	1264	2935	11165	26
2	1448	2237	6875	33
1	77	493	1004	49

Descriptions of the skid sites are given in Table 58. Table 12 shows a much higher crash rate level for skid site 1 (round-abouts, railway crossings etc) compared with skid site 4 (normal roads). Note that in this part of the analysis where a road has different skid site levels on each side of the road I have used the minimum (most dangerous) level, so crash rates for level 1 are likely to be under-estimated. Also note that the road length assigned to what is essentially a point on the road is somewhat arbitrary so crash rates for skid-site 1 may also be somewhat arbitrary.

3.2 All crashes – two-way tables

This section considers the factors considered in section 3.1 two at a time. The road length, numbers of crashes, total travel and crash rates are given in separate two-way tables. Crash numbers in the individual cells in the two way tables are much smaller than in the one-way tables so there is a substantial amount of statistical fluctuation.

The first set of two tables is for curvature and ADT as the classifying variables.

Table 13: curvature and ADT – road length

	ADT range (1000 vehicles per day))					
Curvature	<1	>=1,<2	>=2,<5	>=5,<10	>=10,<20	>=20
>=10,<100	52	44	19	6	2	0
>=100,<1000	935	839	627	305	109	28
>=1000,<10000	1399	1320	1372	835	284	61
>=10000,<100000	429	436	512	331	105	19
>=100000	3	5	5	4	1	0

Table 14: curvature and ADT - number of crashes

	ADT range (1000 vehicles per day))					
Curvature	<1	>=1,<2	>=2,<5	>=5,<10	>=10,<20	>=20
>=10,<100	31	74	72	42	33	10
>=100,<1000	429	773	1170	1063	655	187
>=1000,<10000	425	768	1605	1910	1236	346
>=10000,<100000	97	200	521	641	397	117
>=100000	5	2	5	11	5	0

Table 15: curvature and ADT – total traffic

	ADT range (1000 vehicles per day)					
Curvature	<1	>=1,<2	>=2,<5	>=5,<10	>=10,<20	>=20
>=10,<100	59	138	142	88	61	31
>=100,<1000	1230	2660	4416	4421	3275	1454
>=1000,<10000	1918	4139	9833	12118	8417	3195
>=10000,<100000	597	1369	3710	4826	3117	1045
>=100000	5	17	43	68	34	11

Table 16: curvature and ADT - crash rate

	ADT range (1000 vehicles per day)					
Curvature	<1	>=1,<2	>=2,<5	>=5,<10	>=10,<20	>=20
>=10,<100	53	54	51	48	54	33
>=100,<1000	35	29	26	24	20	13
>=1000,<10000	22	19	16	16	15	11
>=10000,<100000	16	15	14	13	13	11
>=100000	100	12	12	16	15	0

This shows the crash rate increasing as the radius of curvature decreases and some increase as ADT decreases but less than we saw in Table 9 (so some of the apparent effect of the ADT variable is reduced when we allow for curvature).

The second set of two way tables is for skid-site and skid resistance as classifying variables.

Table 17: skid site and scrim – road length

	Scrim range					
Skid site	<0.3	>=0.3,<0.4	>=0.4,<0.5	>=0.5,<0.6	>=0.6,<0.7	>=0.7
4	11	170	1781	3649	1546	85
3	2	43	367	613	219	12
2	3	74	427	647	265	17
1	0	4	24	33	13	0

Table 18: skid site and scrim - number of crashes

	Scrim range					
Skid site	<0.3	>=0.3,<0.4	>=0.4,<0.5	>=0.5,<0.6	>=0.6,<0.7	>=0.7
4	17	322	2650	3094	811	36
3	10	163	1249	1259	220	13
2	12	200	942	832	211	12
1	0	35	216	191	36	1

Table 19: skid site and scrim – total traffic

	Scrim range					
Skid site	<0.3	>=0.3,<0.4	>=0.4,<0.5	>=0.5,<0.6	>=0.6,<0.7	>=0.7
4	102	1954	19865	24327	5846	291
3	23	553	4599	4929	972	41
2	19	510	2871	2696	680	36
1	5	79	415	403	76	3

The interesting thing to note from Table 17 and Table 19 is that there is very little relationship between the skid site and the scrim range. See Table 58. Skid-site 4 road segments tend to be above the investigatory level; skid-site 1 tend to be below (remembering, of course, that we are classifying the road by the lowest skid-site of its two sides).

Table 20: skid site and scrim - crash rate

	Scrim range					
Skid site	<0.3	>=0.3,<0.4	>=0.4,<0.5	>=0.5,<0.6	>=0.6,<0.7	>=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

This shows within each scrim range the crash rate increasing as the skid-site decreases and for each skid site the crash rate increasing as the skid-resistance gets smaller. The latter effect is strongest for the very low skid-resistances, although note the small number of crashes involved so accuracy is not high.

The third set of two-way tables is for curvature and scrim as classifying variables.

Table 21: curvature and scrim – road length

	Scrim range					
Curvature	<0.3	>=0.3, <0.4	>=0.4, <0.5	>=0.5, <0.6	>=0.6, <0.7	>=0.7
>=10,<100	0	11	44	49	16	1
>=100,<1000	4	104	773	1336	564	36
>=1000,<10000	9	131	1300	2638	1102	60
>=10000,<100000	3	45	482	918	359	17
>=100000	0	0	6	10	3	0

Table 22: curvature and scrim - number of crashes

	Scrim range					
Curvature	<0.3	>=0.3, <0.4	>=0.4, <0.5	>=0.5, <0.6	>=0.6, <0.7	>=0.7
>=10,<100	2	31	122	74	23	1
>=100,<1000	17	287	1767	1684	430	25
>=1000,<10000	11	290	2429	2819	644	27
>=10000,<100000	9	118	810	835	184	9
>=100000	0	2	11	9	6	0

Table 23: curvature and scrim – total traffic

	Scrim range					
Curvature	<0.3	>=0.3, <0.4	>=0.4, <0.5	>=0.5, <0.6	>=0.6, <0.7	>=0.7
>=10,<100	4	65	228	170	38	3
>=100,<1000	31	960	7034	7284	1871	105
>=1000,<10000	82	1526	14994	18306	4252	199
>=10000,<100000	29	550	5689	6805	1459	64
>=100000	0	8	73	81	15	1

Table 24: curvature and scrim - crash rate

	Scrim range					
Curvature	<0.3	>=0.3, <0.4	>=0.4, <0.5	>=0.5, <0.6	>=0.6, <0.7	>=0.7
>=10,<100	55	48	54	43	61	40
>=100,<1000	55	30	25	23	23	24
>=1000,<10000	13	19	16	15	15	14
>=10000,<100000	32	21	14	12	13	14
>=100000	0	26	15	11	39	0

Again note the increasing crash rate as the skid resistance decreases within each curvature range or as radius of curvature decreases within each scrim range.

3.3 Wet crashes – one-way tables

This section repeats the previous tables where we count only crashes classified as occurring on wet roads. We don't have traffic figures for wet roads. Hence crash rates are for wet road crashes in terms of the total traffic. So the rates are much smaller than in the previous section. Crash numbers are also smaller so there is more statistical error.

The general form of the results is the same as before. However, as one might expect, the effect of skid-resistance is much stronger.

Table 25: average daily traffic

ADT range	road length	number of crashes	total traffic	crash rate
<200	68	3	26	11.7
>=200,<500	650	19	517	3.7
>=500,<1000	2103	180	3268	5.5
>=1000,<2000	2646	397	8323	4.8
>=2000,<5000	2538	779	18144	4.3
>=5000,<10000	1485	784	21548	3.6
>=10000,<20000	503	483	14941	3.2
>=20000,<50000	109	112	5579	2.0
>=50000	0	0	158	0.0

Table 26: curvature

Curvature range	road length	number of crashes	total traffic	crash rate
>=10,<100	125	67	518	12.9
>=100,<1000	2845	1058	17457	6.1
>=1000,<10000	5273	1272	39620	3.2
>=10000,<100000	1835	349	14663	2.4
>=100000	20	10	179	5.6

Table 27: scrim

Scrim range	road length	number of crashes	total traffic	crash rate
<0.3000	18	10	150	6.7
>=0.3000,<0.4000	294	219	3125	7.0
>=0.4000,<0.5000	2610	1190	28048	4.2
>=0.5000,<0.6000	4953	1064	32649	3.3
>=0.6000,<0.7000	2046	232	7637	3.0
>=0.7000	116	7	372	1.9

Table 28: skid site

Skid site	road length	number of crashes	total traffic	crash rate
4	7275	1384	52625	2.6
3	1264	603	11165	5.4
2	1448	640	6875	9.3
1	77	96	1004	9.6

3.4 Wet crashes – two-way tables

The road lengths and total travel are the same as for all crashes so this section gives just the number of crashes and the crash rates. The general appearance of the data is as with the all-crash data. However, the numbers of crashes, particularly around the borders of the tables tend to be quite small so it is difficult to make precise interpretations.

Table 29: curvature and ADT - number of crashes

	ADT range (1000 vehicles per day)					
Curvature	<1	>=1,<2	>=2,<5	>=5,<10	>=10,<20	>=20
>=10,<100	4	22	19	8	8	6
>=100,<1000	92	189	327	270	151	29
>=1000,<10000	80	154	333	395	244	66
>=10000,<100000	24	32	98	105	79	11
>=100000	2	0	1	6	1	0

Table 30: curvature and ADT - crash rate

	ADT range (1000 vehicles per day)					
Curvature	<1	>=1,<2	>=2,<5	>=5,<10	>=10,<20	>=20
>=10,<100	6.8	16.0	13.4	9.1	13.1	19.6
>=100,<1000	7.5	7.1	7.4	6.1	4.6	2.0
>=1000,<10000	4.2	3.7	3.4	3.3	2.9	2.1
>=10000,<100000	4.0	2.3	2.6	2.2	2.5	1.1
>=100000	40.1	0.0	2.3	8.8	2.9	0.0

Table 31: skid site and scrim - number of crashes

	Scrim range					
Skid site	<0.3	>=0.3,<0.4	>=0.4,<0.5	>=0.5,<0.6	>=0.6,<0.7	>=0.7
4	2	90	577	559	141	3
3	1	39	286	251	23	1
2	7	78	276	211	57	3
1	0	10	37	38	8	0

Table 32: skid site and scrim - crash rate

	Scrim range					
Skid site	<0.3	>=0.3,<0.4	>=0.4,<0.5	>=0.5,<0.6	>=0.6,<0.7	>=0.7
4	2.0	4.6	2.9	2.3	2.4	1.0
3	4.4	7.1	6.2	5.1	2.4	2.5
2	36.4	15.3	9.6	7.8	8.4	8.3
1	0.0	12.6	8.9	9.4	10.5	0.0

Table 33: curvature and scrim - number of crashes

	Scrim range					
Curvature	<0.3	>=0.3, <0.4	>=0.4, <0.5	>=0.5, <0.6	>=0.6, <0.7	>=0.7
>=10,<100	2	11	24	24	2	0
>=100,<1000	6	108	462	366	89	5
>=1000,<10000	1	71	545	543	102	2
>=10000,<100000	1	28	153	128	38	0
>=100000	0	1	5	3	1	0

Table 34: curvature and scrim - crash rate

	Scrim range					
Curvature	<0.3	>=0.3, <0.4	>=0.4, <0.5	>=0.5, <0.6	>=0.6, <0.7	>=0.7
>=10,<100	54.9	16.9	10.5	14.1	5.3	0.0
>=100,<1000	19.4	11.3	6.6	5.0	4.8	4.8
>=1000,<10000	1.2	4.7	3.6	3.0	2.4	1.0
>=10000,<100000	3.5	5.1	2.7	1.9	2.6	0.0
>=100000	0.0	13.1	6.9	3.7	6.5	0.0

4 The model

I suppose each side of each 10-metre length of road can *generate* crashes at the rate (per year)

$$a \exp(L) \tag{1}$$

where a is the ADT (per side) and L is a linear combination of the road characteristics including

- a constant,
- gradient,
- curvature,
- crossfall,
- skid-site classification,
- skid resistance,
- log(ADT),
- year,
- region,
- urban/rural classification.

The coefficients in the linear combination are the unknown parameters to be estimated.

See section 5.4 for a numerical example of the calculation of the crash risk from the fitted model.

Since we are taking the exponential of L, a linear combination of the road characteristics, the actual model is multiplicative.

Note that the average daily traffic (ADT) appears in the model in two places, a in equation (1) and as a component of L. These could have been combined into a single term in L. However, by using the formulation in (1) the component in L is present only if the crash risk (expected number of crashes per 100 million vehicle kilometres) depends on ADT. When there is dependence, this dependence is modelled by the size of the coefficient of log(ADT) in L. The crash risk is given by

$$\frac{10^{10}}{365} \exp(L)\,. \tag{2}$$

The *actual* rate that crashes are reported in a 10 metre length of road is the average of *generating* rates over the 10 metre lengths in its immediate neighbourhood (on the same road) and summed over the two sides of the road. In the results reported here the average is over 10 metre lengths within 100 metres of the length being considered. Typically, this gives an average over 210 metres on each side of the road. There is no weighting down of the more distant 10 metre lengths. This averaging allows for error in reporting the location and the possibility that a crash ends at a location some distance from the piece of road involved in generating the crash.

Because we are combining the sides of the road we don't have to know the directions of vehicles involved in the crash.

The model assumes that the crashes are statistically independent and the number in each 10-metre segment follows a Poisson distribution. (Of course, for most segments the number will be zero).

Fitting is by maximum likelihood and uses my C++ libraries for matrix manipulation and automatic differentiation from http://www.robertnz.net and a new array and statistical modelling package currently under development.

5 Poisson regression model fitting

I carried out the analyses for each of the four categories of crash data described in Table 4.

I carried out two versions of the analyses; the first when all available predictor variables of interest were included and a second that included only the most important variables.

5.1 Notes on the variables

Three variables required special processing.

5.1.1 Curvature

See section 9.2 for the handling of values outside the range 3 to 100,000 and of missing values.

5.1.2 IRI roughness

The IRI roughness measurement appears to be strongly affected by curvature and gradient. This appears to be either a problem with the measurement equipment or the actual definition of roughness and is unlikely to be a real effect. In the analyses in this section the values of IRI have been adjusted to attempt to compensate for the measurement bias. See section 9.3 for more details.

5.1.3 Gradient

See section 9.4. In all years, except 1998, there appears to be a bias in the gradient measurements. I have, therefore, used 1998 gradient data for all gradients. Where 1998 gradient data is not available, I have put the gradient to 0.

5.2 The large model

Table 35 lists the predictor variables included in the large model.

Table 35: variables included in the large model

Variable	Description
year	Categorical variable showing the calendar year (1997 to 2002)
region	Categorical variable showing the region (R1 to R7)
urban_rural	Categorical variable taking the values R (rural) and U (urban)
skid_site	Categorical variable showing the skid-site category (1, 3 or 4 – category 2 is combined into 4)
spline6(log10_curvature)	Six point spline function of log(absolute curvature); range of curvature 10 to 10,000
poly2_log10_ADT	2^{nd} degree polynomial function of the log(ADT)
spline6(gradient)	Six point spline function of absolute gradient; range 0 to 10
poly2_scrim-0.5000	2^{nd} degree polynomial function of the (scrim – 0.5)
spline4(log10_iri)	Four point spline function of adjusted $\log_{10}$(IRI); range of $\log_{10}$(IRI) is 0 to 1
skid_site*(scrim-0.5000)	The interaction between the skid_site category and the scrim coefficient
spline5(sqrt_rut_depth)	Five point spline function of the square root of the rut depth; range of rut depth is 0 to 9
cway_width	The carriageway width
texture	Texture
lanes_category	A categorical variable taking the values TwoLane or MultiLane (usually corresponding to a passing lane).
irr_width	Irregular width indicator
crossfall	Crossfall
abs_crossfall	The absolute value of the crossfall

Where a spline function transformation is involved the range of the variable is curtailed to the range shown in the table. Values above or below this range are changed to the upper or lower end of the range respectively.

Model fitting is by the method of maximum likelihood. Table 36 shows a reference number for each analysis, the number of crashes included in the analysis and the value of the log-likelihood function at the maximum. This value is useful when comparing the large and the simplified model.

Table 36: large model fit details

	All	Selected	Wet	Wet selected
Analysis reference	200	210	220	230
Number of crashes	12083	8423	2619	2013
Maximum of the log-likelihood function	-81061.10	-60581.60	-21270.53	-16937.62

The following section shows a analysis of variance tables derived from the model fitting program and graphs of the predicted value of crash rates as function of the predictor variables. The analysis of variance tables are used for deciding which variables should be included in the simplified model.

I have not shown the estimates of the coefficients in the large model as these would be very difficult to use independently of the fit program.

5.2.1 *Analysis of variance*

Table 37 and Table 38 show two versions of the analysis of variance table for the four categories of crash data.

Table 37: large model analysis of variance - terms added last

Predictor variable	**df**	**1% pt.**	**Chi-squared values**			
			All	Selected	Wet	Wet selected
year	5	15.09	119.43	98.98	36.65	28.54
region	6	16.81	92.36	56.15	92.82	60.53
urban_rural	1	6.63	58.44	90.18	27.23	41.51
skid_site	2	9.21	1857.40	82.05	346.90	19.15
spline6(log10_curvature)	5	15.09	901.06	853.04	495.91	463.86
poly2_log10_ADT	2	9.21	376.15	266.50	86.35	53.69
spline6(gradient)	5	15.09	214.80	15.29	18.88	10.93
poly2_scrim-0.5000	2	9.21	129.89	146.95	127.45	137.90
spline4(log10_iri)	3	11.34	65.77	52.39	20.51	16.36
skid_site*(scrim-0.5000)	2	9.21	27.16	12.50	13.26	2.56
spline5(sqrt_rut_depth)	4	13.28	24.43	3.21	2.15	0.69
cway_width	1	6.63	26.41	0.09	16.91	0.40
texture	1	6.63	4.37	0.64	0.92	0.15
lanes_category	1	6.63	3.47	0.05	5.32	0.04
irr_width	1	6.63	0.34	1.88	4.47	1.33
crossfall	1	6.63	0.05	0.00	0.41	0.42
abs_crossfall	1	6.63	0.35	0.01	1.11	0.35

Table 38: large model analysis of variance - terms added sequentially

Predictor variable	df	1% pt.	Chi-squared values			
			All	Selected	Wet	Wet selected
year	5	15.09	102.10	56.56	46.76	39.46
region	6	16.81	122.23	93.91	81.25	73.10
urban_rural	1	6.63	203.20	148.60	3.89	64.69
skid_site	2	9.21	2015.70	206.72	315.64	45.64
spline6(log10_curvature)	5	15.09	2036.80	2874.40	1365.80	1620.80
poly2_log10_ADT	2	9.21	281.44	300.17	55.99	48.30
spline6(gradient)	5	15.09	258.68	15.26	24.10	12.42
poly2_scrim-0.5000	2	9.21	125.05	148.28	147.93	172.11
spline4(log10_iri)	3	11.34	56.12	51.36	18.46	15.23
skid_site*(scrim-0.5000)	2	9.21	28.89	12.59	13.86	2.61
spline5(sqrt_rut_depth)	4	13.28	27.39	3.06	2.70	0.80
cway_width	1	6.63	23.42	0.04	13.29	0.42
texture	1	6.63	4.27	0.64	0.82	0.16
lanes_category	1	6.63	3.48	0.03	5.50	0.07
irr_width	1	6.63	0.31	1.89	4.24	1.24
crossfall	1	6.63	0.02	0.00	0.43	0.46
abs_crossfall	1	6.63	0.35	0.01	1.11	0.35

The *chi-squared values* show the result of a statistical test for the statistical significance of each variable. If the chi-squared value is greater than the value in the column labelled 1% point the test is showing statistical significance at the 1% level.

In Table 37 the test of a variable is in the presence of all the other variables. The test is based on what are essentially the type III sums of squares commonly favoured by users of the SAS statistical program. This can be misleading if some of the variables are highly correlated with each other. In this case it is possible for the correlated variables to appear as not significant when one carries out the test of each variable in the presence of the other variables as in Table 37 when, in fact, at least some of them are important. In Table 38 the test of each variable is in the presence only of the variables above it in the table. The test is based on what are essentially the type I sums of squares commonly favoured by users of the S-plus and R statistical programs. This table can also be misleading. But if both sets of tests are giving similar results then one can be fairly sure one is not missing significant effects because of correlated variables. So it is worth looking at both tables.

The values in Table 38 are dependent on the order the values are added into the model. The order selected starts with 3 variables that we are going to include, whether or not they are significant, and then the remainder in approximate order of significance.

The chi-squared values for the variables below the line for *spline4(log10_iri)* all seem moderately small although somewhat more than the 1% point. I am inclined to regard these variables as non-significant. Our model does not consider all factors that could influence crashes and it is entirely possible that these are increasing the chi-squared values. There maybe also some deviation from the Poisson assumption. For the variables *spline4(log10_iri)* and above the chi-squared values tend to be substantially larger than the 1% point and these are the ones included in the simplified model of section 5.3.

Generally the size of the chi-squared values decreases as we go from left to right corresponding to the small numbers of accidents.

The gradient effect is large only in the column corresponding to all crashes suggesting that its effect is mainly on crashes that are excluded from the selected crashes and not important for the wet crash risk.

The scrim effect tends to increase as we go from left to right despite the decreasing numbers of crashes, especially in Table 38, showing that skid resistance is more important for wet crashes.

It isn't surprising that crossfall is not significant as the way it is introduced into the model is too simplistic.

5.2.2 Predicted crash rates

In order to see how each variable in the model affects the crash rate I have graphed the crash rate predicted by the model as each variable, in turn is varied. For the terms not being varied I have used the following values.

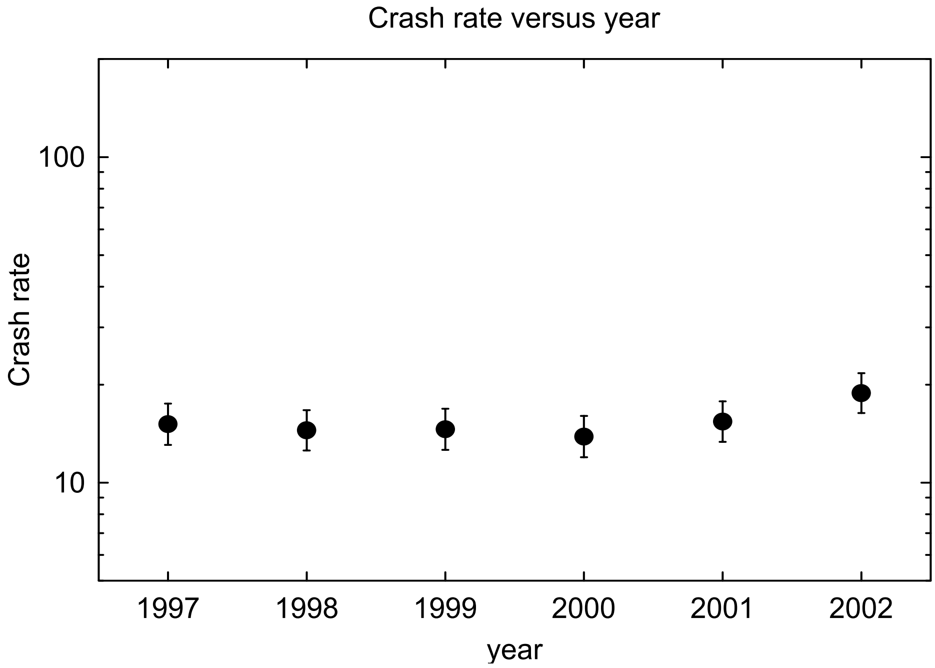
Table 39: default values of variables

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

I have not included the graphs when the effects of the variables are very small.

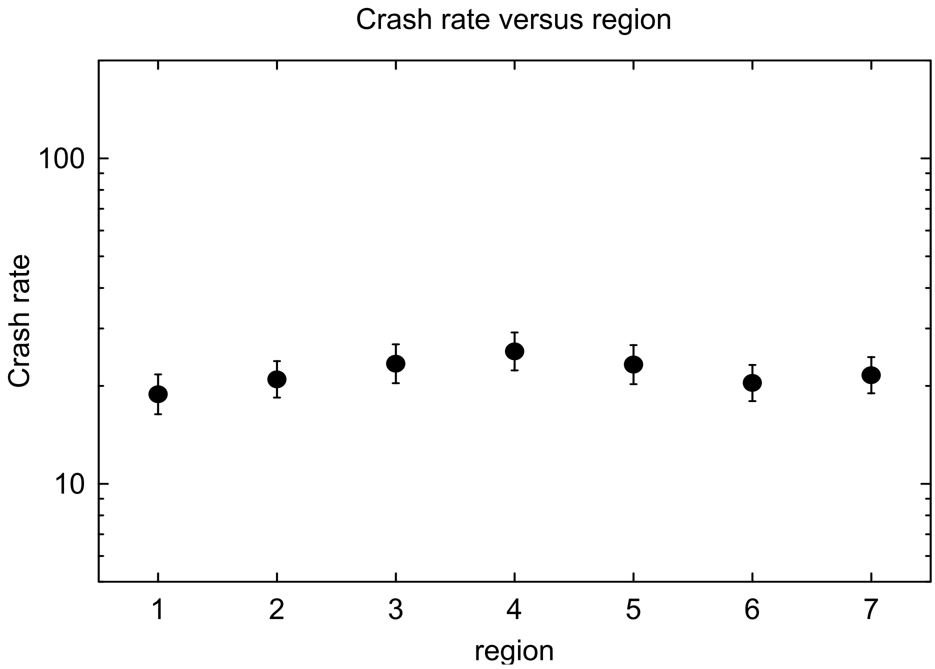
Data category: all crashes

Figure 1



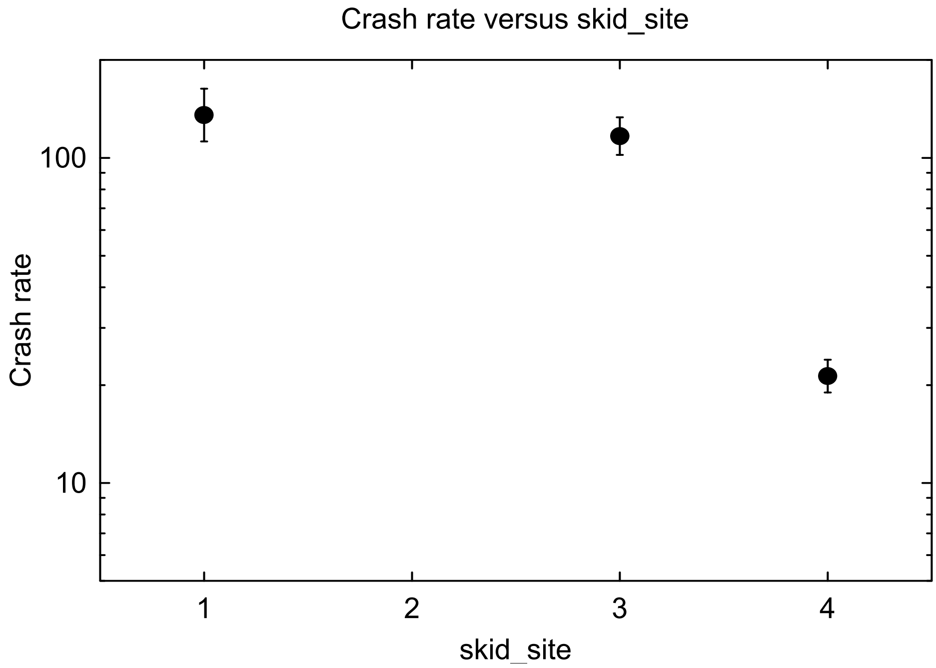
The error bounds show 2 standard deviations (roughly 95% confidence). The year-to-year differences are, at least partially, due to the changes in the fraction of crashes that could be located.

Figure 2



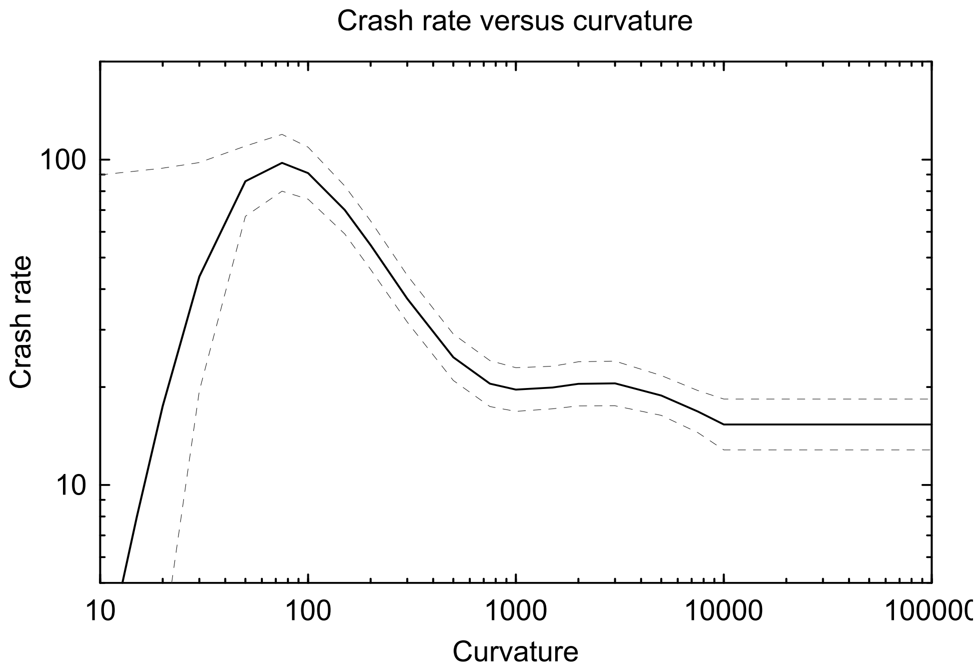
There is relatively little difference in the crash rates in the different regions.

Figure 3



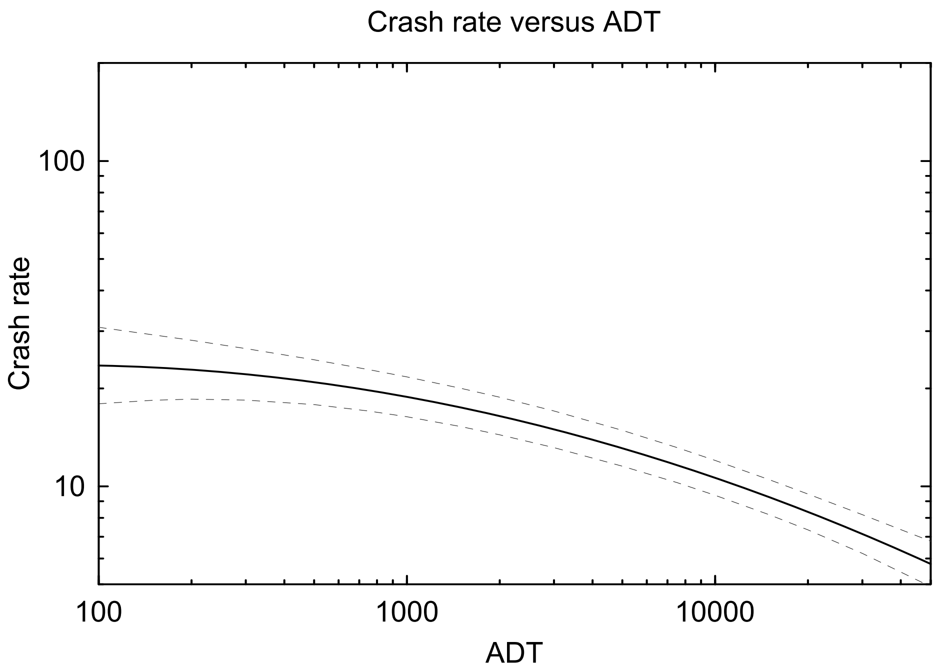
Skid-site 2 data has been amalgamated into skid-site 4 since skid-site 2 is determined by road geometry and we would prefer to have the effect explained by the geometry variables. The diagram shows skid-sites 1 and 3 having roughly the same crash-rate and this is very much higher than for skid-site 4. See section 9.1 for the definitions of skid-site.

Figure 4



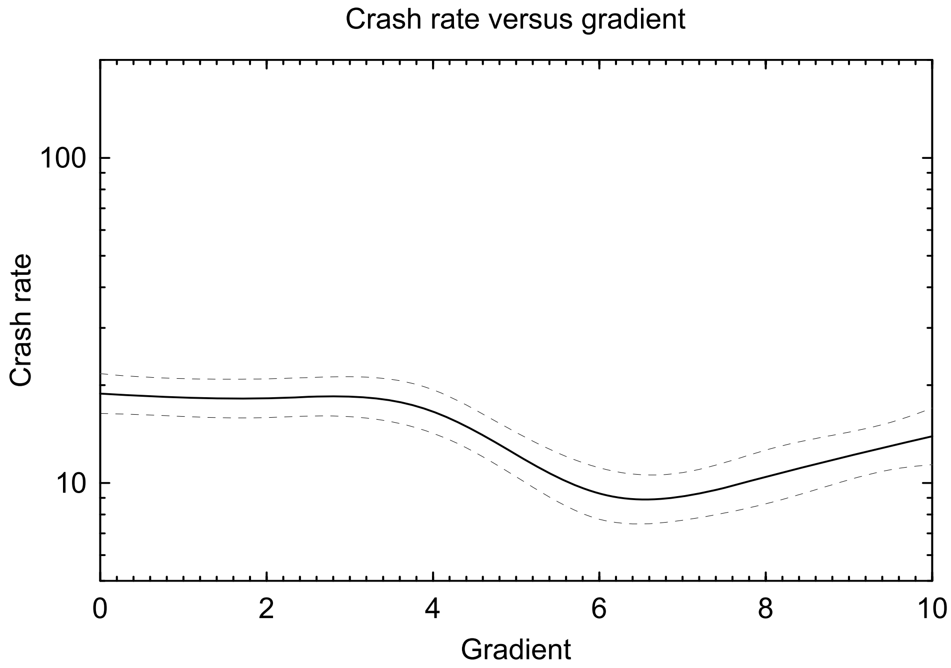
Curvature radii less than 10 have been amalgamated as have curvatures greater than 10,000. The graph shows the constant level for curvatures 10,000 to 100,000. The spline fit may have exaggerated wriggles in this curve. However, the general appearance is a rapid increase in the crash-rate as the radius of curvature goes from 1000 to 100 and then a levelling off or possibly a decrease.

Figure 5



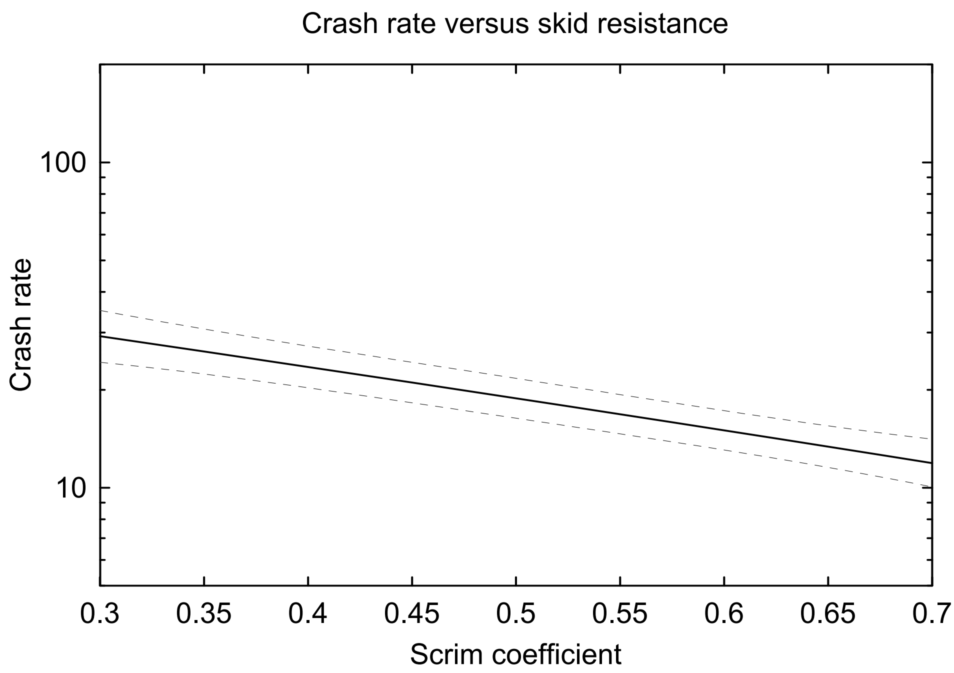
This shows a drop in crash-rate as average daily traffic increases.

Figure 6



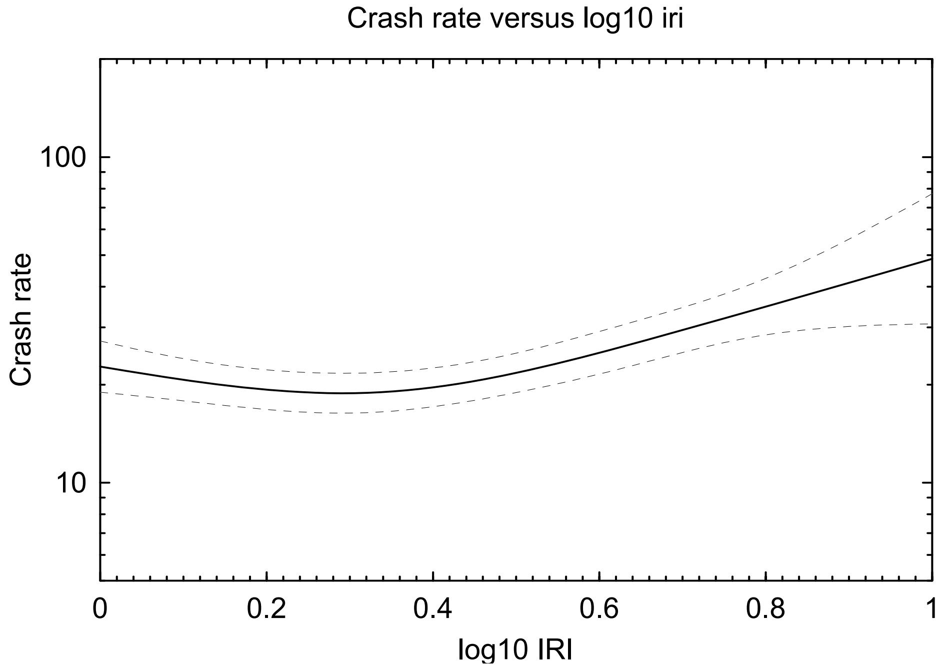
This shows a strong effect of gradient. Because we can't tell the direction of travel of a vehicle involved in a crash we can't distinguish between uphill and downhill. The graph shows the net effect on crash rate when uphill and downhill effects are combined. The curve shows no effect until the gradient reaches 4% and then there is a drop in crash rate. At slopes greater than around 7% there seems to be an increase again.

Figure 7



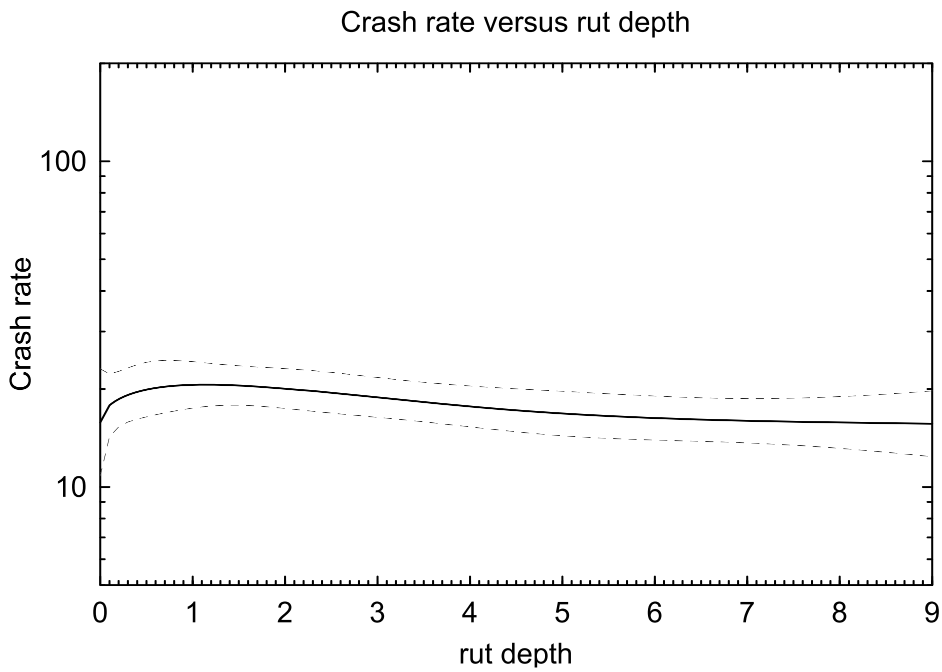
This shows a reduction in crash-rate as the skid resistance increases.

Figure 8



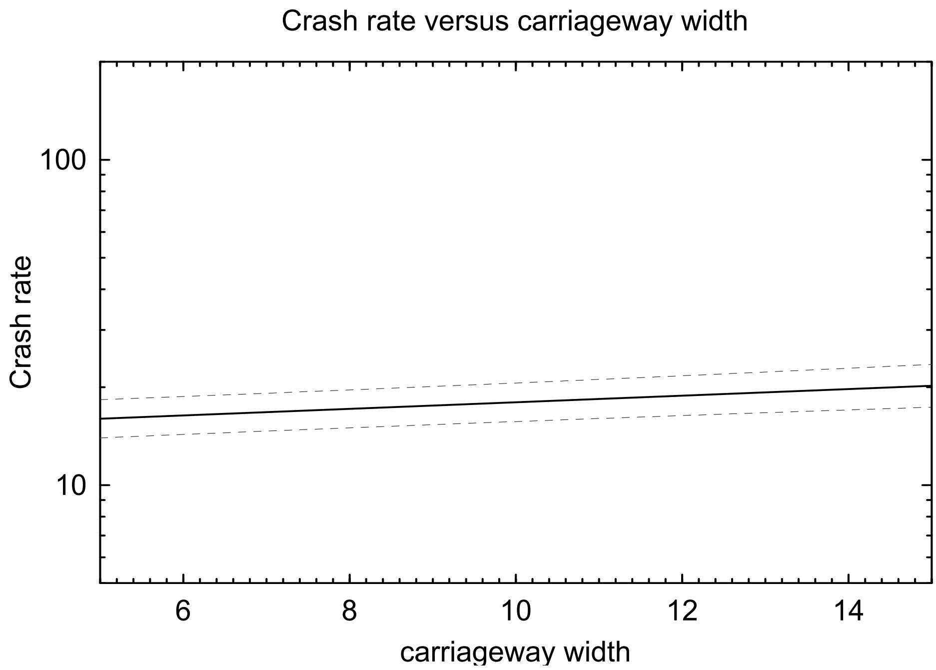
This shows an increase in crash rate as IRI increases, particularly for $\log_{10}(\text{IRI}) > 0.4$; that is IRI > 2.5. There is a suggestion of an increase in crash-rate for very low IRI. Possibly, this is a consequence of the adjustment process.

Figure 9



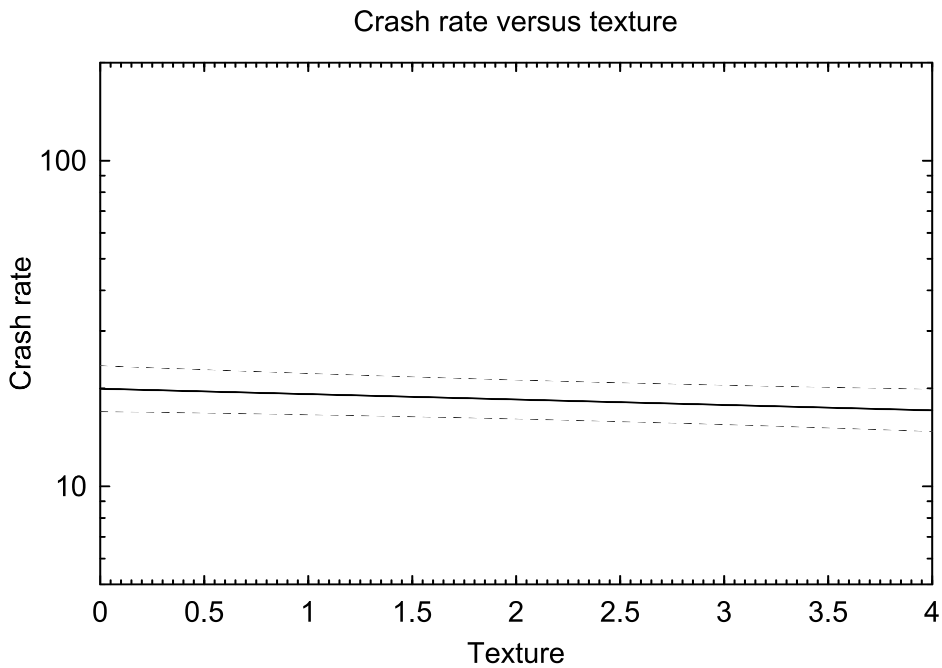
There seems to be a slight decrease in crash rate as rut depth increases. This is not the direction one would expect.

Figure 10



There seems to be an increase is crash rate as carriageway with increases.

Figure 11



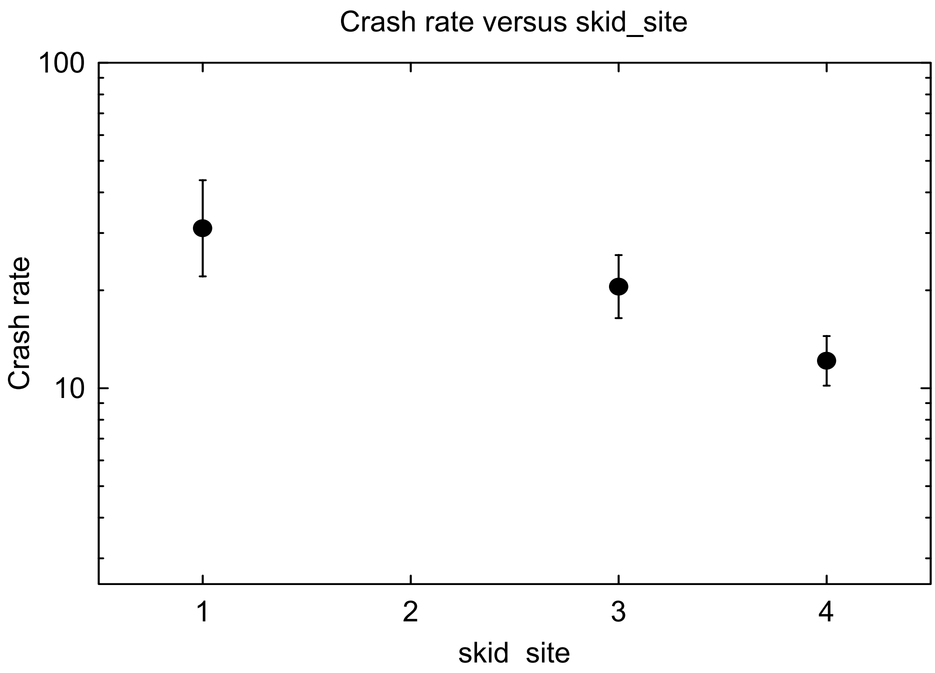
There seems to be a slight decrease in crash rate as texture increases.

Data category: selected crashes

Because of the smaller number of crashes the scale of the crash-rate axis is changed. The ratio of crash-rates covered by the axis is as before.

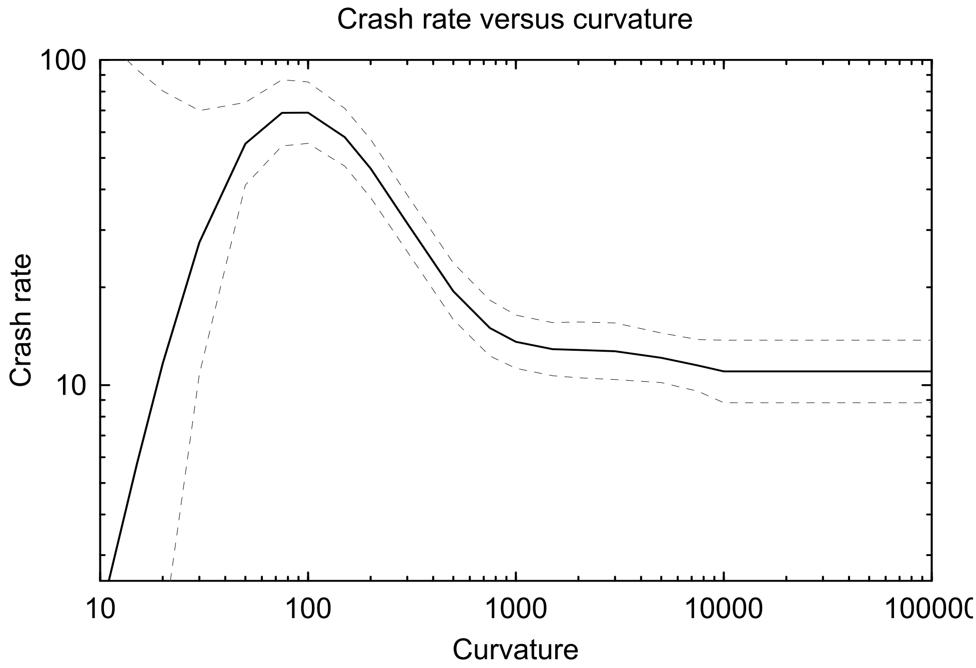
Year and region graphs are similar to those for all crashes and are omitted.

Figure 12



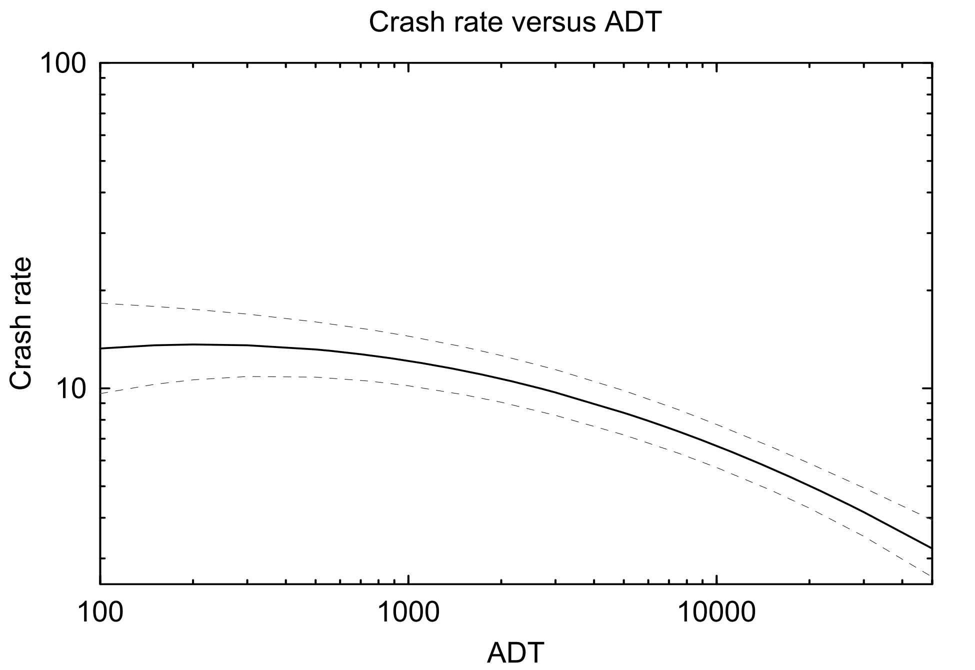
Skid site shows smaller differences than for all crashes – the selected crashes do not include a number of intersection and crossing crashes.

Figure 13



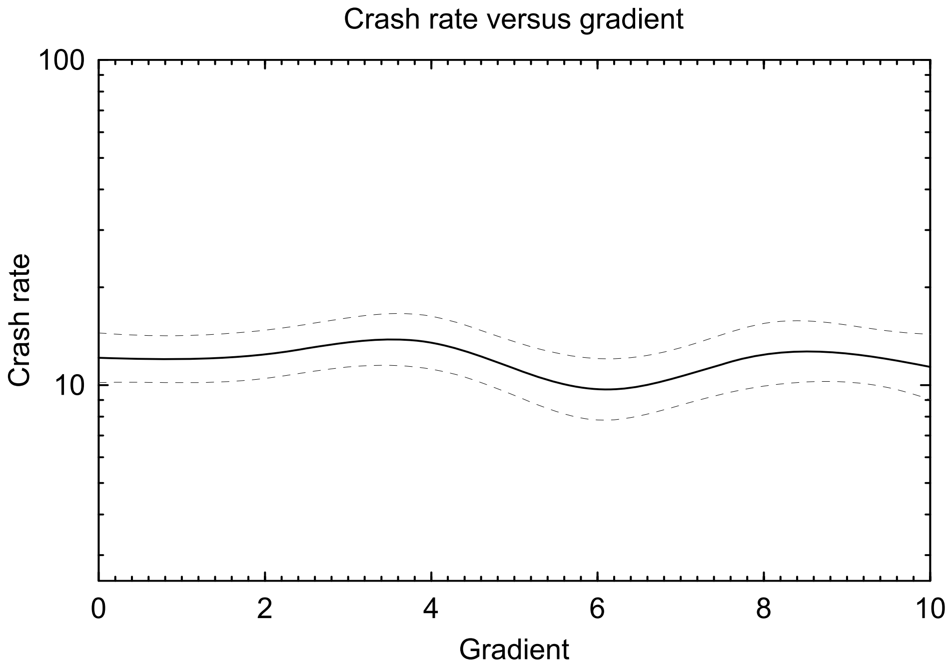
The curvature graph is basically similar to that for all crashes.

Figure 14



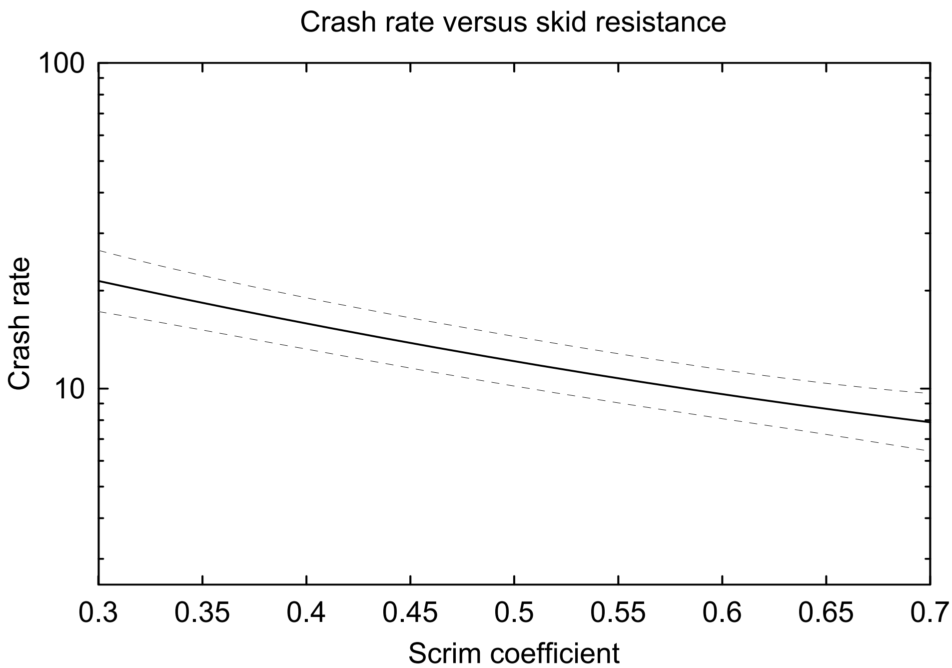
ADT is similar to that for all crashes.

Figure 15



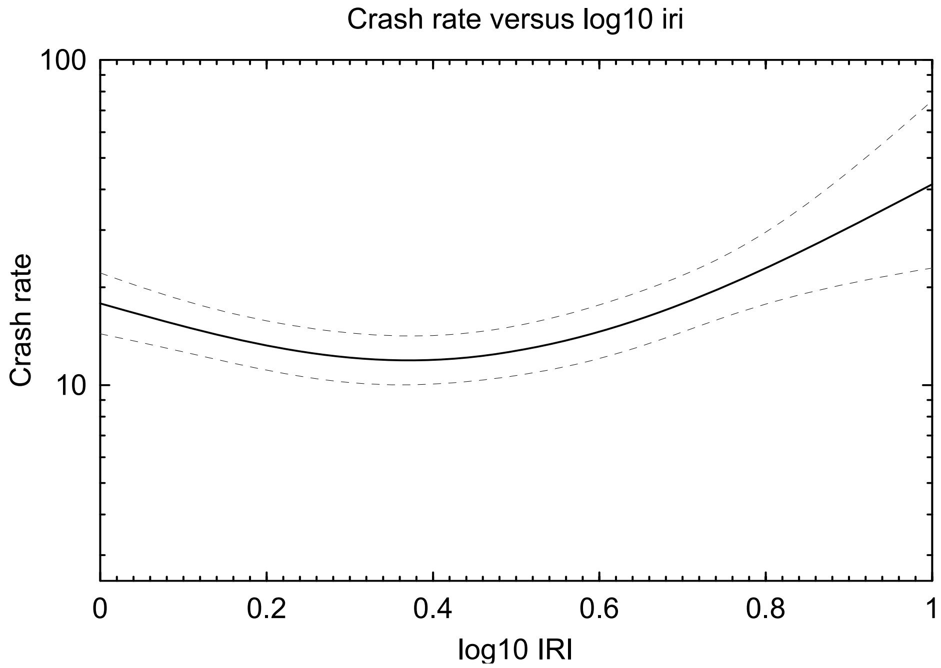
The gradient effect apparent for all crashes has largely disappeared.

Figure 16



The skid resistance effect is similar to that for all crashes.

Figure 17



The IRI effect is similar to that for all crashes.

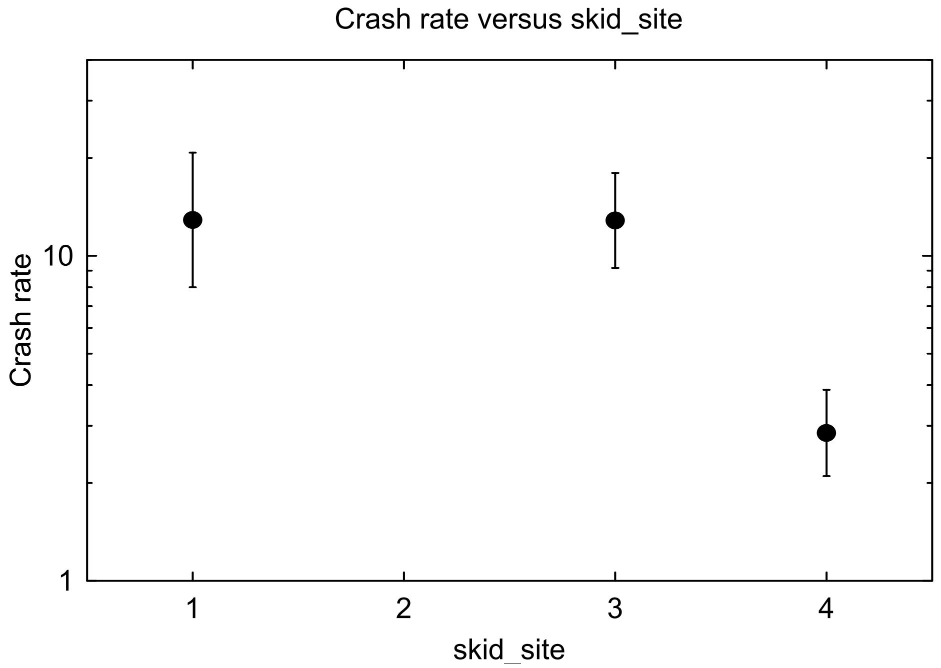
Texture, carriageway width and rut-depth show little effect and the graphs are omitted.

Data category: wet crashes

Because of the smaller number of crashes the scale of the crash-rate axis is changed. The ratio of crash-rates covered by the axis is as before.

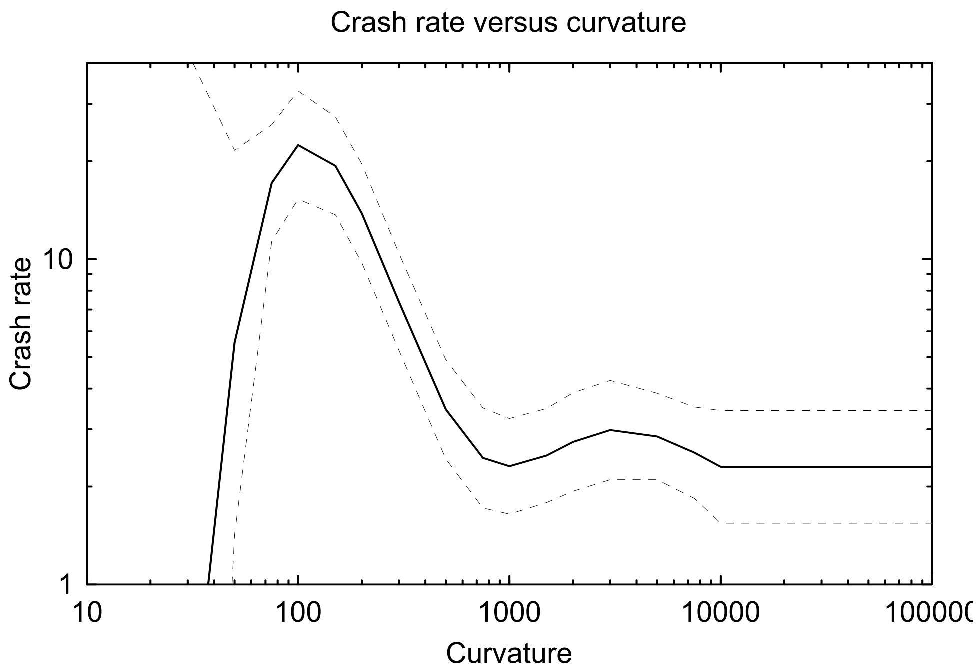
Year and region graphs are basically similar to the preceding year and region graphs and are omitted.

Figure 18



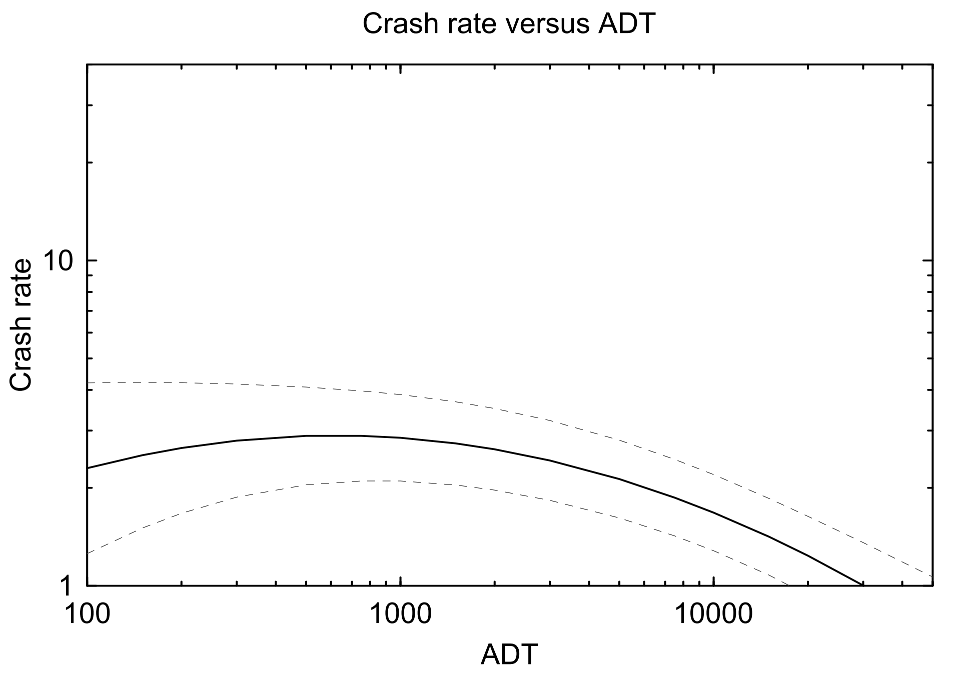
This is similar to the graph for all crashes.

Figure 19



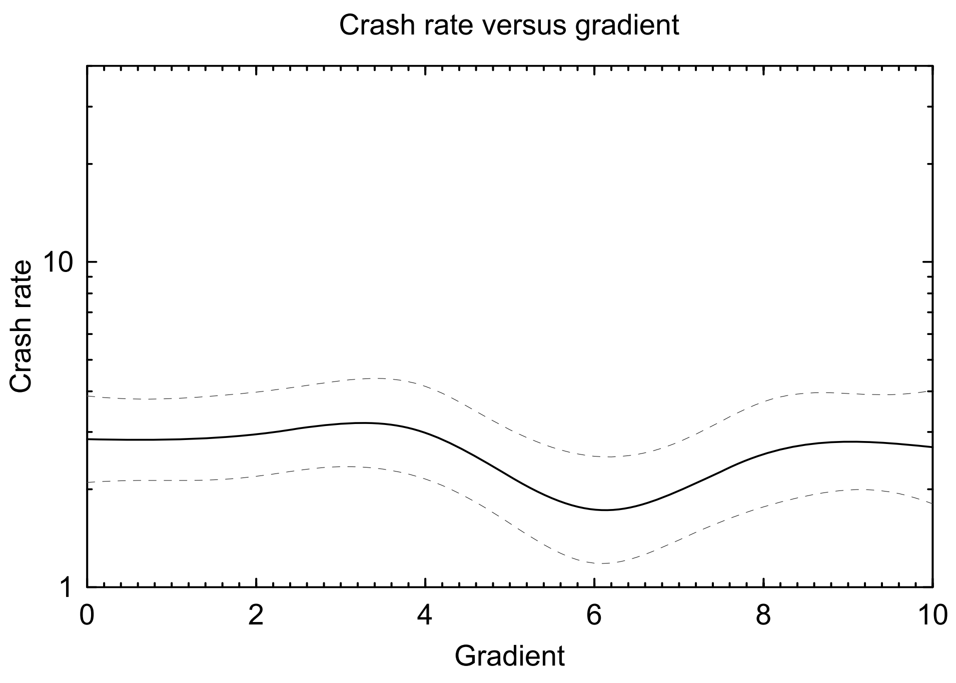
The effect of curvature is similar to the previous graphs except that the effect is much stronger.

Figure 20



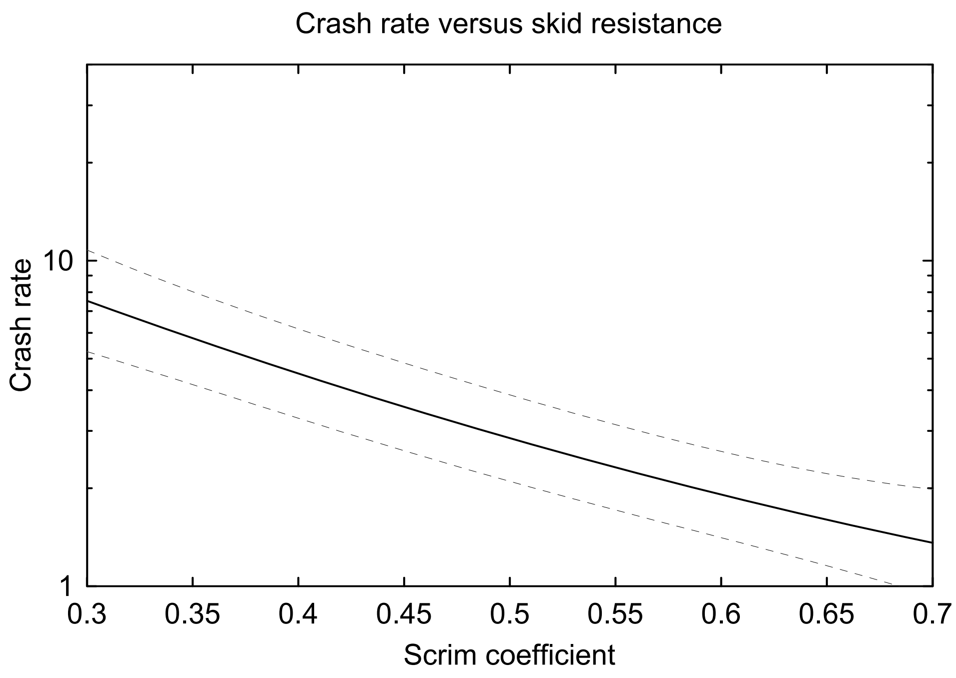
The effect of ADT is similar to that in the preceding graphs.

Figure 21



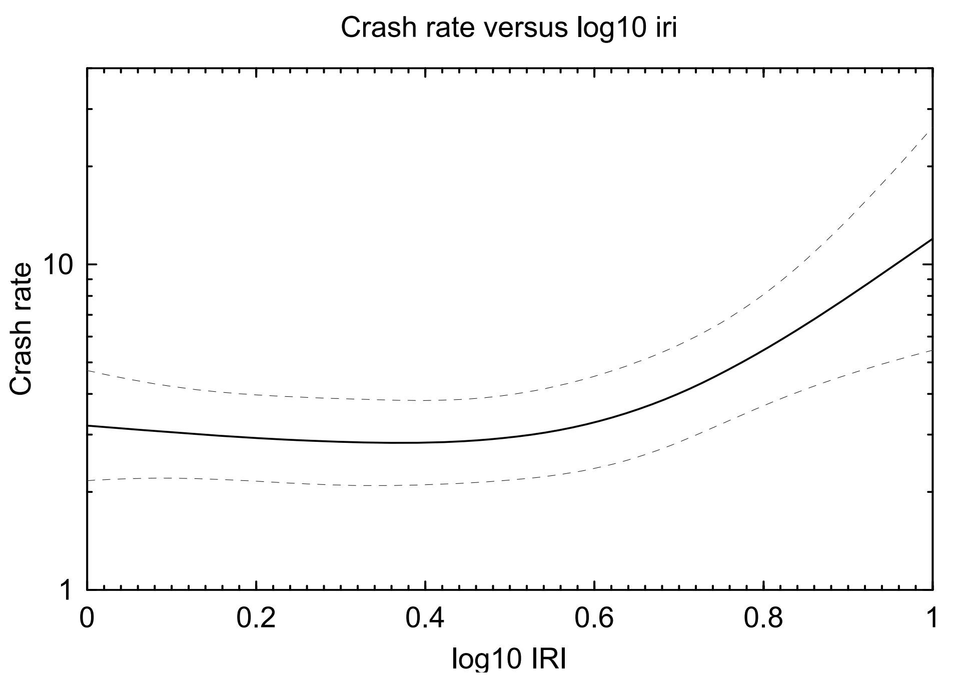
The effect of gradient is small although somewhat similar to that for all crashes.

Figure 22



The effect of scrim is similar to that in preceding graphs except that the effect is much stronger.

Figure 23



The effect of IRI is similar to that for all crashes.

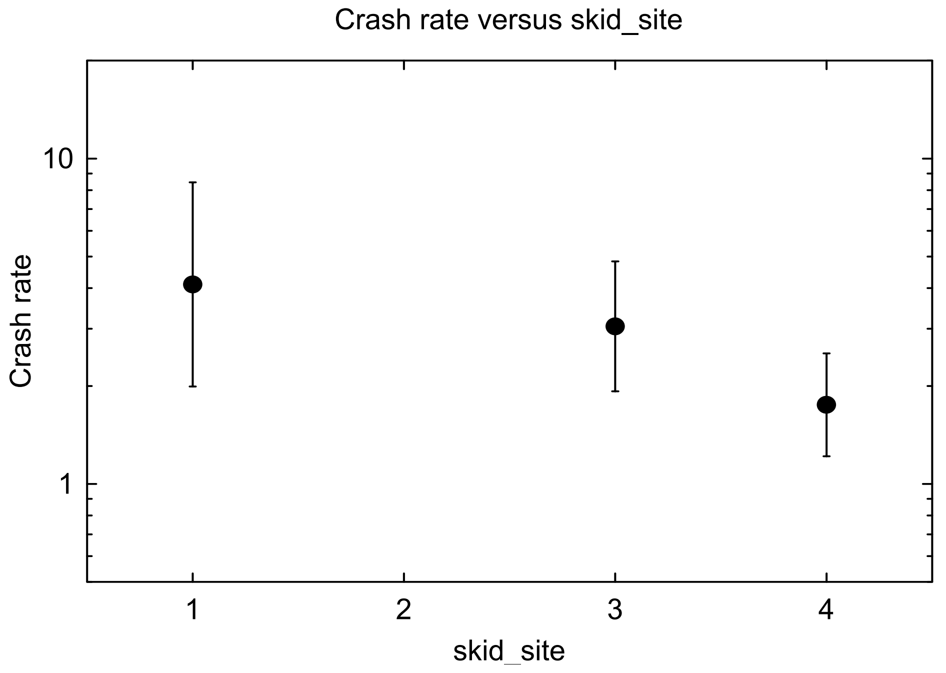
Texture, carriageway width and rut-depth show little effect and the graphs are omitted.

Data category: wet selected crashes

Because of the smaller number of crashes the scale of the crash-rate axis is changed. The ratio of crash-rates covered by the axis is as before.

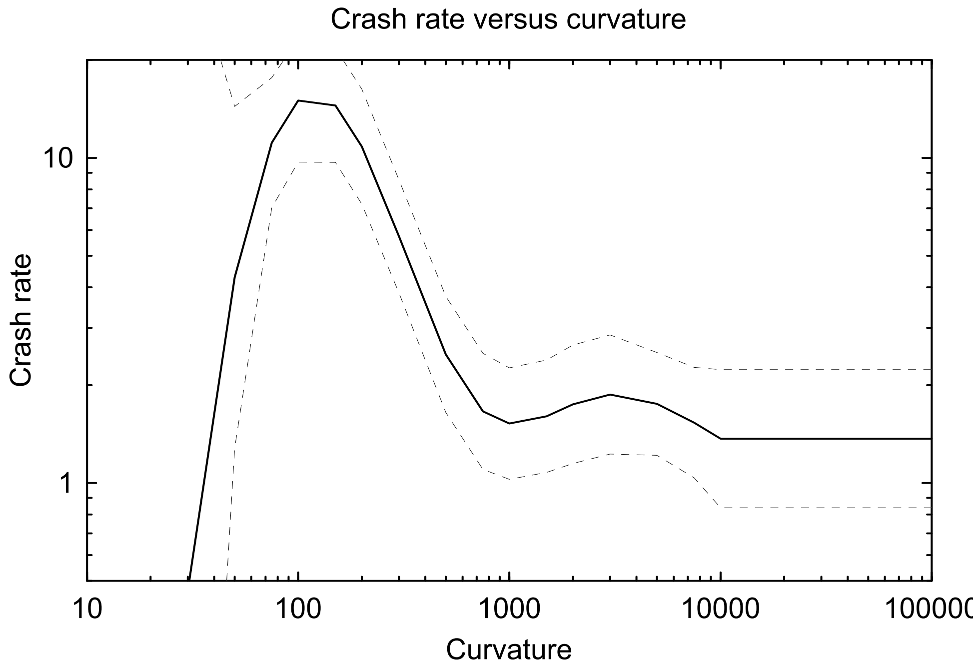
The graphs for year and region are basically similar to the ones for all crashes and are omitted.

Figure 24



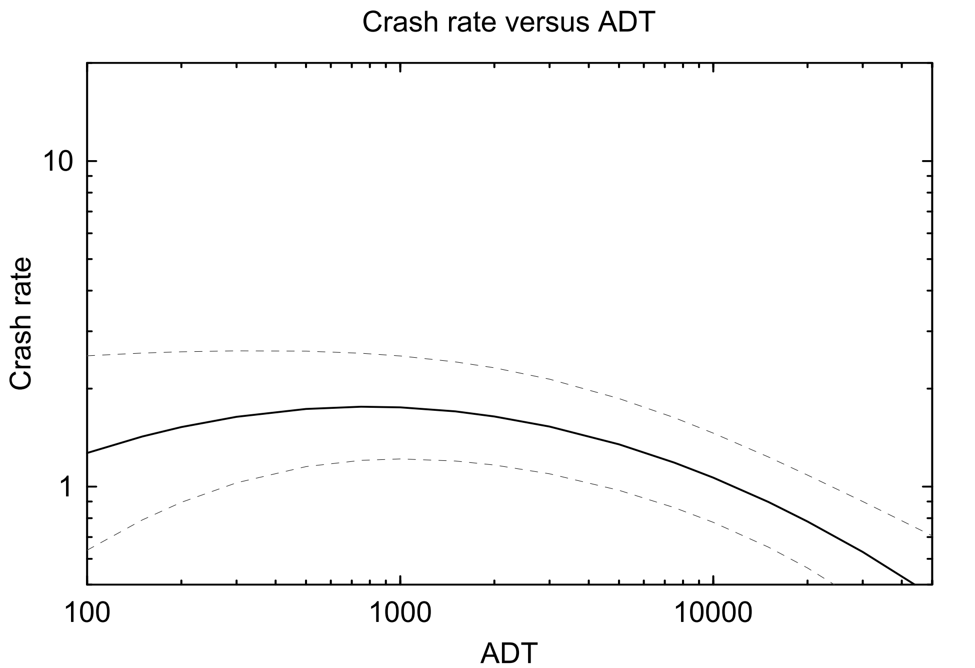
The graph for skid-site is similar to that for selected crashes.

Figure 25



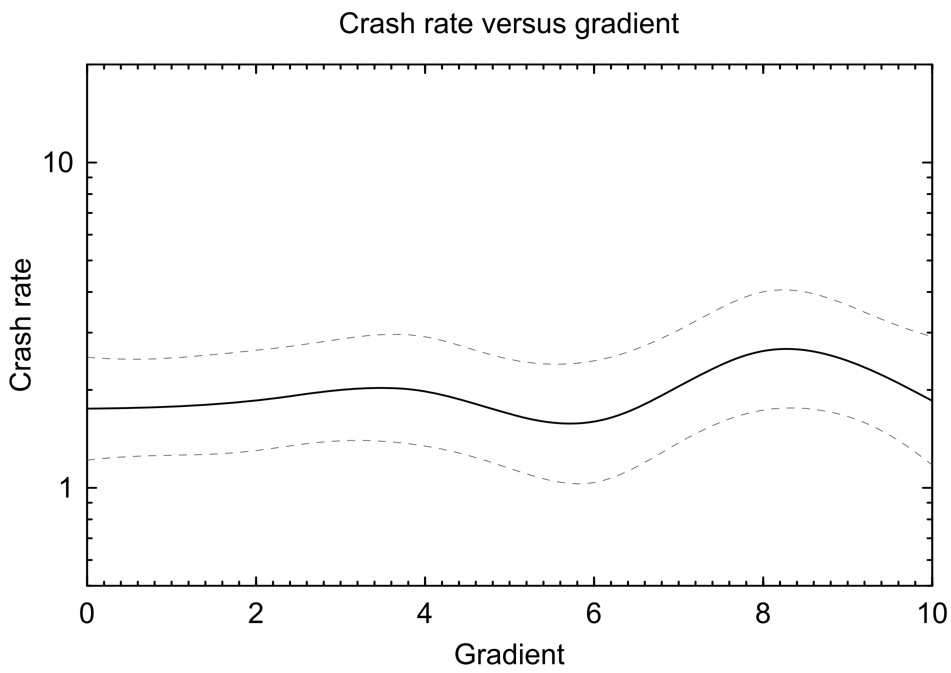
The graphs for curvature are similar to that for wet crashes.

Figure 26



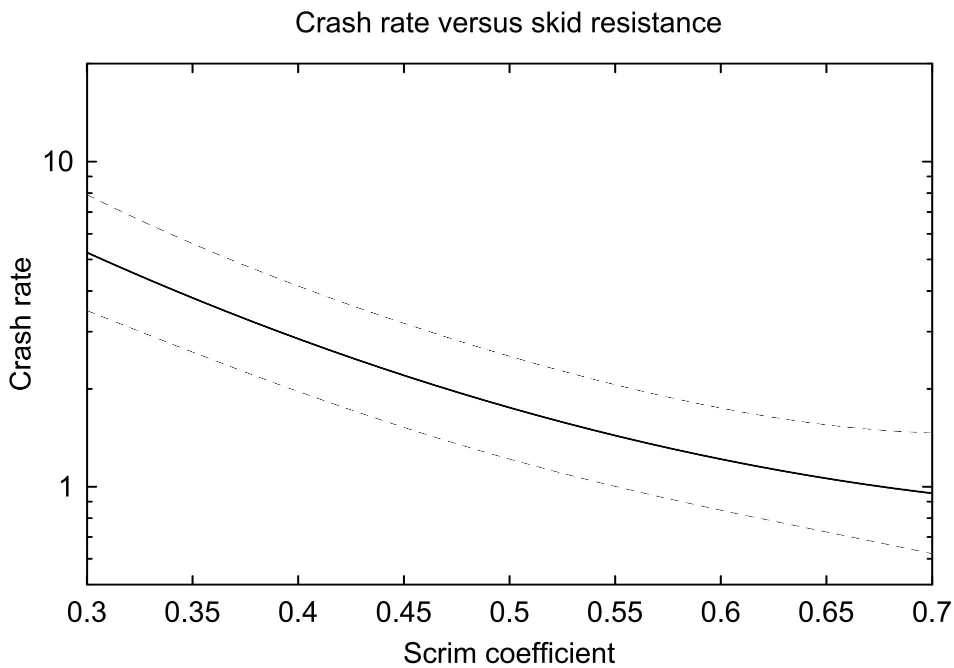
The graph for ADT is basically similar to the preceding graphs for ADT.

Figure 27



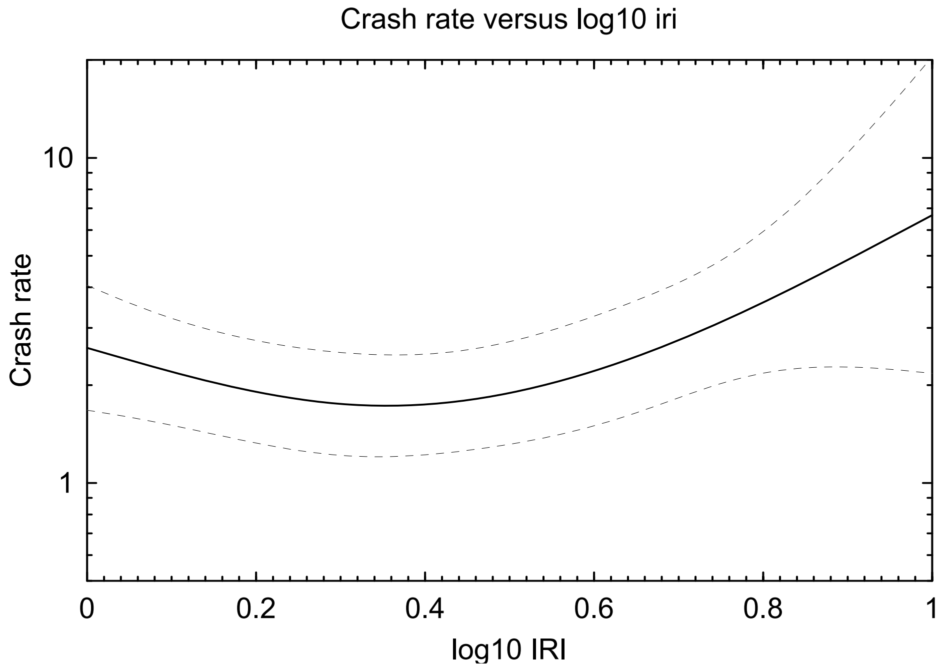
The effect of gradient is small although somewhat similar to that for all crashes.

Figure 28



The graph for skid resistance is similar to that for all wet crashes.

Figure 29



The graph for IRI is similar to that for all crashes, but much less accuracy because of the smaller number of crashes in the wet selected category.

Texture, carriageway width and rut-depth show little effect and the graphs are omitted.

5.3 The simplified model

In this version all variables below IRI in Table 37 are omitted as it is likely that they would not be statistically significant if we were able to fully model that random aspects of the model.

I have replaced the spline functions in the large model 2nd or 3rd degree polynomials. In addition I have amalgamated values outside the ranges shown in Table 40. While these don't fit as well as the spline functions it is much simpler for someone to reproduce the calculation, say on a spreadsheet.

Table 40 lists the variables included in the simplified model.

Table 40: variables included in the simplified model

Variable	Description
year	Categorical variable showing the calendar year (1997 to 2002)
region	Categorical variable showing the region (R1 to R7)
urban_rural	Categorical variable taking the values R (rural) and U (urban)
skid_site	Categorical variable showing the skid-site category (1, 3 or 4 – category 2 is combined into 4)
poly2_log10_ curvature	2nd degree polynomial function of the $\log_{10}$(absolute curvature); range of curvature is 100 to 10000.
poly2_log10_ADT	2nd degree polynomial function of the $\log_{10}$(ADT)
poly3_gradient	3rd degree polynomial function of the absolute gradient; range of gradient is 4 to 10.
poly2_scrim-0.5000	2nd degree polynomial function of the (scrim – 0.5)
poly3_log10_iri	3rd degree polynomial function of the adjusted $\log_{10}$(IRI); range of $\log_{10}$(IRI) is 0.3 to 1.

Table 41 shows the model fit details for the simplified model.

Table 41: simplified model fit details

	All	Selected	Wet	Wet selected
Analysis reference	300	310	320	330
Number of crashes	12144	8468	2630	2023
Maximum of the log-likelihood function	-81487.45	-60901.73	-21399.62	-17051.71

Compare Table 36 and Table 41 – there are slightly more crashes in Table 41. This is because we have fewer variables and hence we have fewer road segments rejected through missing values. The maxima of the log-likelihood functions are quite a bit smaller in the simplified model. This suggests we have thrown out statistically significant variables. The most important source of the difference will be the replacement of the spline function of curvature by the polynomial function.

5.3.1 *Analysis of variance*

As before, there are two versions of the analysis of variance table.

Table 42: simplified model analysis of variance - terms added last

Predictor variable	df	1% pt.	Chi-squared values			
			All	Selected	Wet	Wet selected
year	5	15.09	129.46	104.26	42.57	28.46
region	6	16.81	94.18	50.37	90.92	59.55
urban_rural	1	6.63	29.58	103.07	17.99	46.91
skid_site	2	9.21	2165.60	86.93	363.82	14.37
poly2_log10_abs_curvature	2	9.21	1518.50	1584.00	822.72	765.94
poly2_log10_ADT	2	9.21	385.84	371.41	79.36	63.34
poly3_abs_gradient	3	11.34	257.35	10.73	27.23	5.49
poly2_scrim-0.5000	2	9.21	110.97	144.03	135.17	161.77
poly3_log10_iri	3	11.34	48.10	36.14	20.31	12.90

Table 43: simplified model analysis of variance - terms added sequentially

Predictor variable	df	1% pt.	Chi-squared values			
			All	Selected	Wet	Wet selected
year	5	15.09	92.51	55.76	49.80	38.67
region	6	16.81	121.43	90.88	89.90	72.85
urban_rural	1	6.63	217.14	143.32	6.79	57.79
skid_site	2	9.21	2157.40	202.28	341.68	33.58
poly2_log10_abs_curvature	2	9.21	1960.70	2806.00	1247.40	1501.70
poly2_log10_ADT	2	9.21	282.67	298.14	53.88	47.15
poly3_abs_gradient	3	11.34	263.60	12.37	29.72	6.28
poly2_scrim-0.5000	2	9.21	125.67	152.17	144.93	174.34
poly3_log10_iri	3	11.34	48.10	36.14	20.31	12.90

With the exception of gradient and IRI most of the chi-squared values are well above the 1% points. They are similar to the values obtained for the large model in Table 37 and Table 38.

5.3.2 Regression coefficients

The following tables show the fitted coefficients for the Poisson regression model described in section 4. See section 5.4 for examples in the use of these values.

Table 44: regression coefficients for *all crashes*

parameter	estimate	st. error	ratio	sig. prob.
constant	2.095	1.76	1.192	0.233
year:1997	0.000			
year:1998	-0.060	0.03	-1.879	0.060
year:1999	-0.053	0.03	-1.641	0.101
year:2000	-0.118	0.03	-3.625	0.000
year:2001	0.000	0.03	0.013	0.989
year:2002	0.198	0.03	6.437	0.000

parameter	estimate	st. error	ratio	sig. prob.
region:R1	0.000			
region:R2	0.108	0.03	3.196	0.001
region:R3	0.210	0.05	4.595	0.000
region:R4	0.306	0.04	8.153	0.000
region:R5	0.224	0.04	5.441	0.000
region:R6	0.105	0.04	2.610	0.009
region:R7	0.124	0.04	3.017	0.003
urban_rural:R	0.000			
urban_rural:U	-0.157	0.03	-5.438	0.000
skid_site:4	0.000			
skid_site:3	1.595	0.04	43.818	0.000
skid_site:1	1.697	0.08	21.753	0.000
log10_abs_curvature**1	-5.360	0.29	-18.383	0.000
log10_abs_curvature**2	0.759	0.05	16.054	0.000
log10_ADT**1	0.707	0.31	2.249	0.025
log10_ADT**2	-0.173	0.04	-3.951	0.000
abs_gradient**1	-2.598	0.70	-3.697	0.000
abs_gradient**2	0.314	0.11	2.880	0.004
abs_gradient**3	-0.012	0.01	-2.217	0.027
scrim-0.5000**1	-1.637	0.16	-10.509	0.000
scrim-0.5000**2	-0.090	1.30	-0.069	0.945
log10_iri**1	-10.540	4.48	-2.353	0.019
log10_iri**2	19.219	8.48	2.266	0.023
log10_iri**3	-9.850	4.99	-1.973	0.048

Table 45: regression coefficients for *selected crashes*

parameter	estimate	st. error	ratio	sig. prob.
constant	-0.541	2.01	-0.270	0.788
year:1997	0.000			
year:1998	-0.049	0.04	-1.264	0.206
year:1999	0.044	0.04	1.136	0.256
year:2000	-0.014	0.04	-0.358	0.720
year:2001	0.089	0.04	2.303	0.021
year:2002	0.278	0.04	7.456	0.000
region:R1	0.000			
region:R2	0.074	0.04	1.836	0.066
region:R3	0.206	0.05	3.873	0.000
region:R4	0.260	0.04	5.792	0.000
region:R5	0.154	0.05	3.086	0.002
region:R6	0.090	0.05	1.874	0.061
region:R7	0.164	0.05	3.404	0.001
urban_rural:R	0.000			

parameter	estimate	st. error	ratio	sig. prob.
urban_rural:U	-0.416	0.04	-10.152	0.000
skid_site:4	0.000			
skid_site:3	0.569	0.07	7.928	0.000
skid_site:1	0.803	0.15	5.218	0.000
log10_abs_curvature**1	-5.036	0.33	-15.358	0.000
log10_abs_curvature**2	0.683	0.05	12.725	0.000
log10_ADT**1	1.129	0.37	3.034	0.002
log10_ADT**2	-0.247	0.05	-4.705	0.000
abs_gradient**1	-1.411	0.76	-1.852	0.064
abs_gradient**2	0.202	0.12	1.715	0.086
abs_gradient**3	-0.009	0.01	-1.603	0.109
scrim-0.5000**1	-2.177	0.18	-11.957	0.000
scrim-0.5000**2	1.790	1.47	1.221	0.222
log10_iri**1	-18.556	5.96	-3.114	0.002
log10_iri**2	31.537	11.39	2.770	0.006
log10_iri**3	-15.504	6.77	-2.289	0.022

Table 46: regression coefficients for *wet crashes*

parameter	estimate	st. error	ratio	sig. prob.
constant	1.015	3.43	0.296	0.768
year:1997	0.000			
year:1998	-0.240	0.07	-3.654	0.000
year:1999	-0.027	0.06	-0.425	0.671
year:2000	-0.331	0.07	-4.774	0.000
year:2001	-0.203	0.07	-2.979	0.003
year:2002	-0.002	0.07	-0.032	0.974
region:R1	0.000			
region:R2	0.192	0.07	2.717	0.007
region:R3	0.101	0.10	1.010	0.312
region:R4	0.565	0.08	7.314	0.000
region:R5	0.053	0.09	0.565	0.572
region:R6	0.146	0.09	1.660	0.097
region:R7	0.045	0.09	0.497	0.619
urban_rural:R	0.000			
urban_rural:U	-0.272	0.06	-4.242	0.000
skid_site:4	0.000			
skid_site:3	1.528	0.08	18.625	0.000
skid_site:1	1.175	0.20	5.749	0.000
log10_abs_curvature**1	-7.426	0.57	-13.100	0.000
log10_abs_curvature**2	1.048	0.09	11.272	0.000
log10_ADT**1	2.380	0.71	3.344	0.001
log10_ADT**2	-0.401	0.10	-4.040	0.000

parameter	estimate	st. error	ratio	sig. prob.
abs_gradient**1	-2.913	1.33	-2.184	0.029
abs_gradient**2	0.396	0.21	1.919	0.055
abs_gradient**3	-0.017	0.01	-1.692	0.091
scrim-0.5000**1	-3.551	0.33	-10.869	0.000
scrim-0.5000**2	3.344	2.48	1.349	0.177
log10_iri**1	-7.348	8.48	-0.866	0.386
log10_iri**2	10.916	15.65	0.698	0.485
log10_iri**3	-3.563	8.89	-0.401	0.689

Table 47: regression coefficients for *wet selected crashes*

parameter	estimate	st. error	ratio	sig. prob.
constant	0.008	3.83	0.002	0.998
year:1997	0.000			
year:1998	-0.216	0.08	-2.877	0.004
year:1999	0.059	0.07	0.809	0.418
year:2000	-0.240	0.08	-3.060	0.002
year:2001	-0.175	0.08	-2.231	0.026
year:2002	0.008	0.08	0.111	0.912
region:R1	0.000			
region:R2	0.188	0.08	2.360	0.018
region:R3	0.091	0.11	0.820	0.412
region:R4	0.537	0.09	6.120	0.000
region:R5	0.041	0.11	0.387	0.699
region:R6	0.161	0.10	1.598	0.110
region:R7	0.073	0.10	0.704	0.481
urban_rural:R	0.000			
urban_rural:U	-0.595	0.09	-6.849	0.000
skid_site:4	0.000			
skid_site:3	0.561	0.15	3.787	0.000
skid_site:1	0.100	0.47	0.212	0.832
log10_abs_curvature**1	-6.329	0.63	-10.042	0.000
log10_abs_curvature**2	0.843	0.10	8.076	0.000
log10_ADT**1	2.516	0.80	3.126	0.002
log10_ADT**2	-0.424	0.11	-3.746	0.000
abs_gradient**1	-2.802	1.40	-2.005	0.045
abs_gradient**2	0.443	0.22	2.059	0.039
abs_gradient**3	-0.022	0.01	-2.092	0.036
scrim-0.5000**1	-4.073	0.37	-11.096	0.000
scrim-0.5000**2	6.220	2.60	2.392	0.017
log10_iri**1	-17.379	11.50	-1.511	0.131
log10_iri**2	29.938	21.84	1.371	0.170
log10_iri**3	-14.644	12.92	-1.134	0.257

5.3.3 Predicted crash rates

I show only the graphs for curvature, gradient, scrim and IRI. The general form of the graphs is similar to those in the previous section, subject to the limitations imposed by the use of the polynomial transformation.

All crashes – simplified model

Figure 30

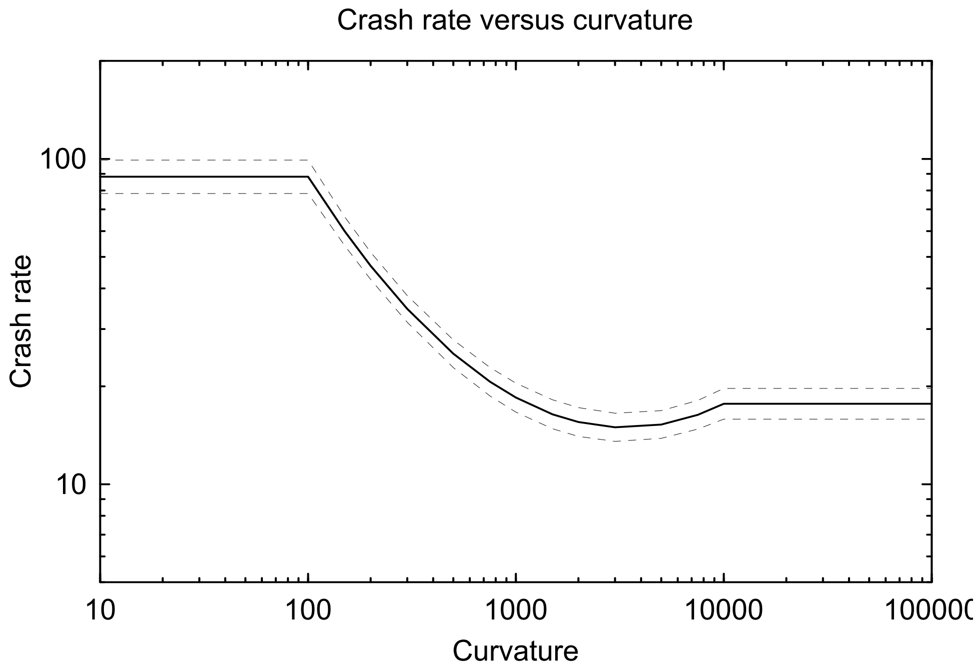


Figure 31

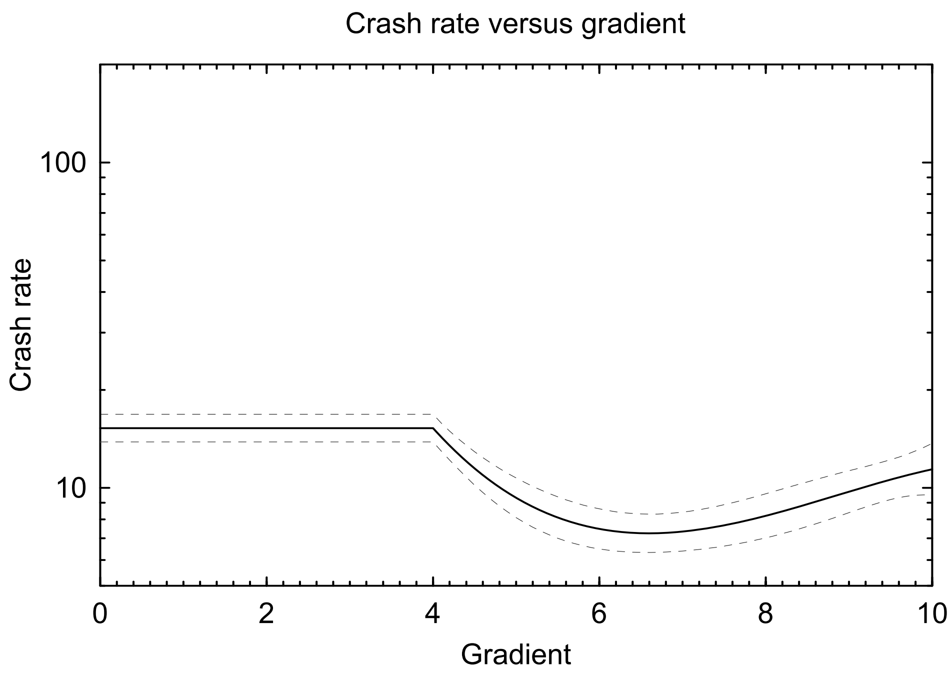


Figure 32

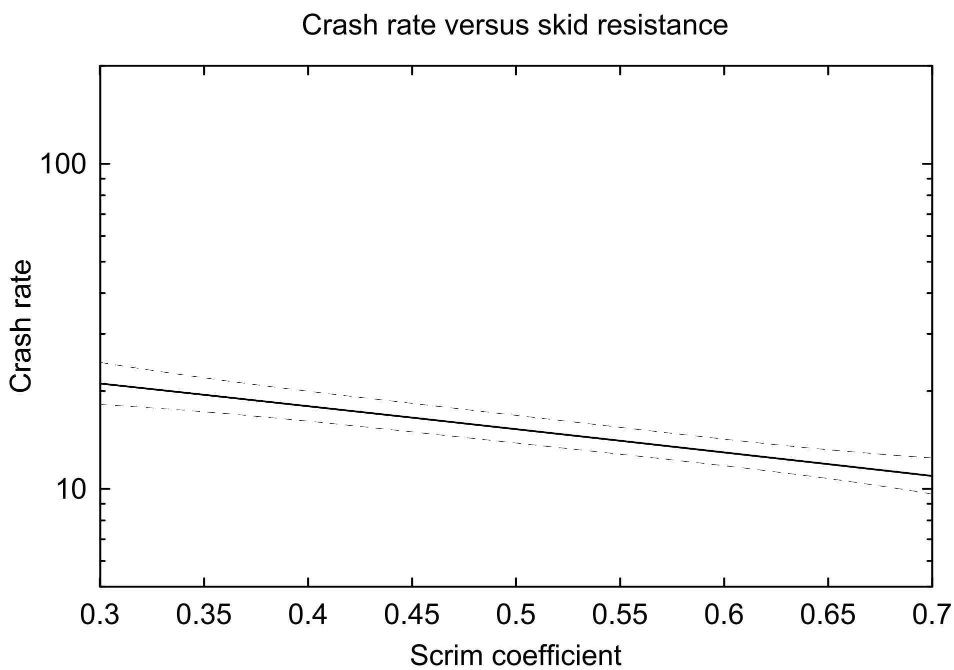
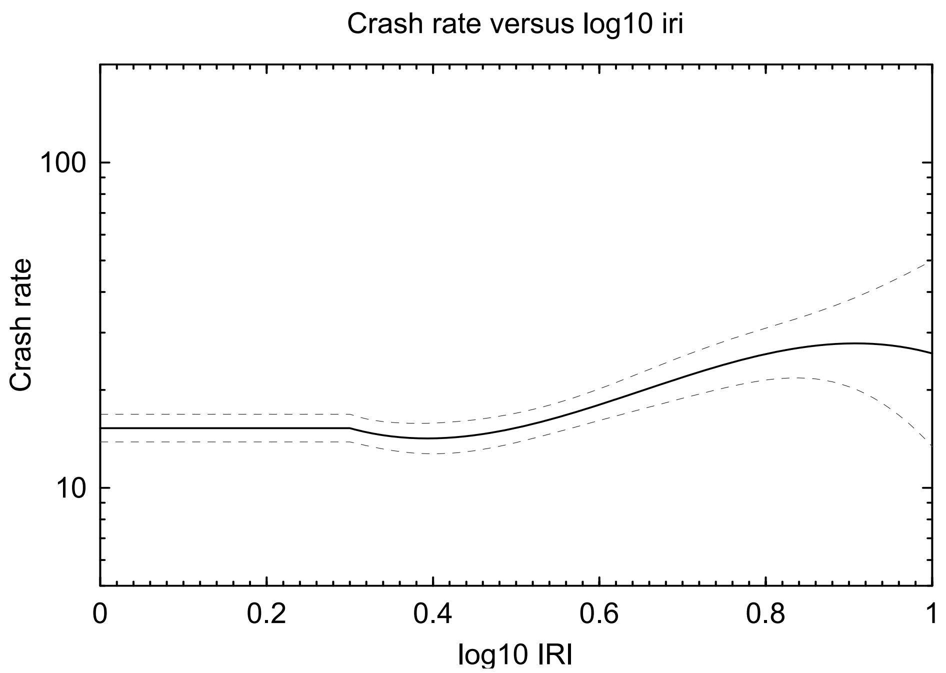


Figure 33



Selected crashes

Figure 34

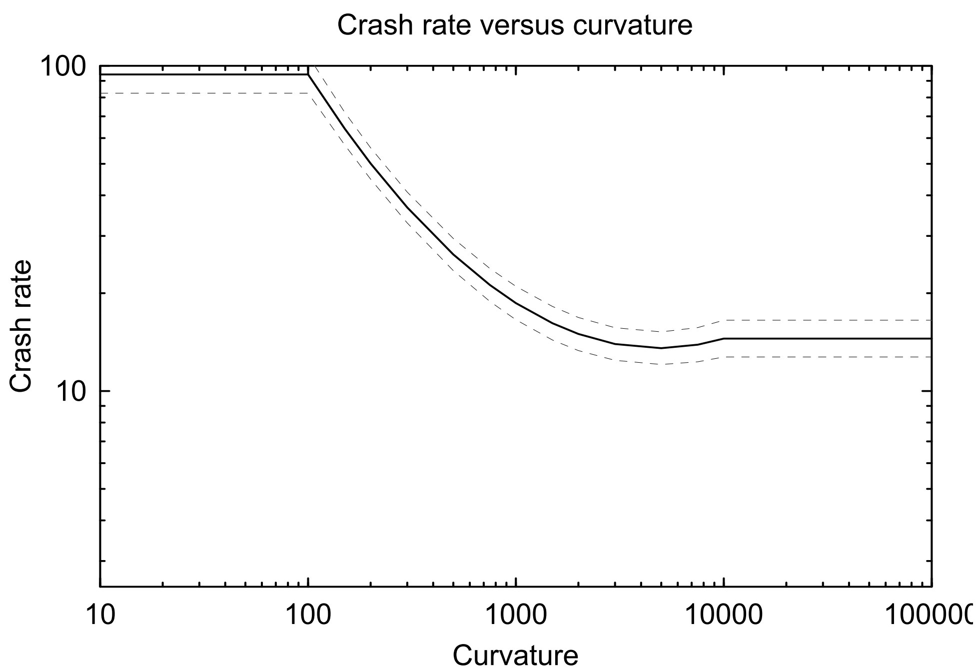


Figure 35

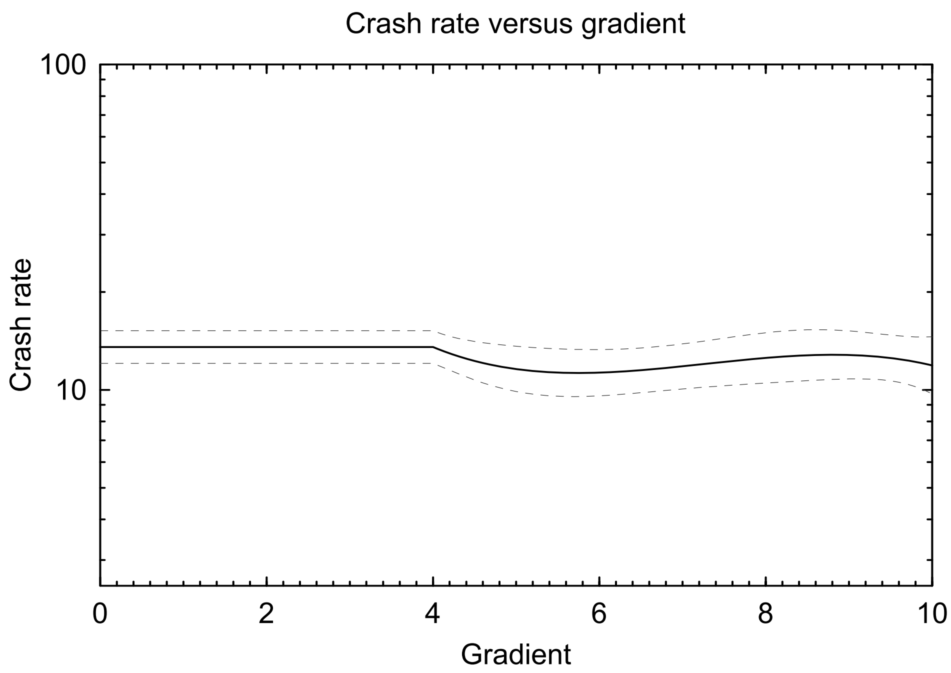


Figure 36

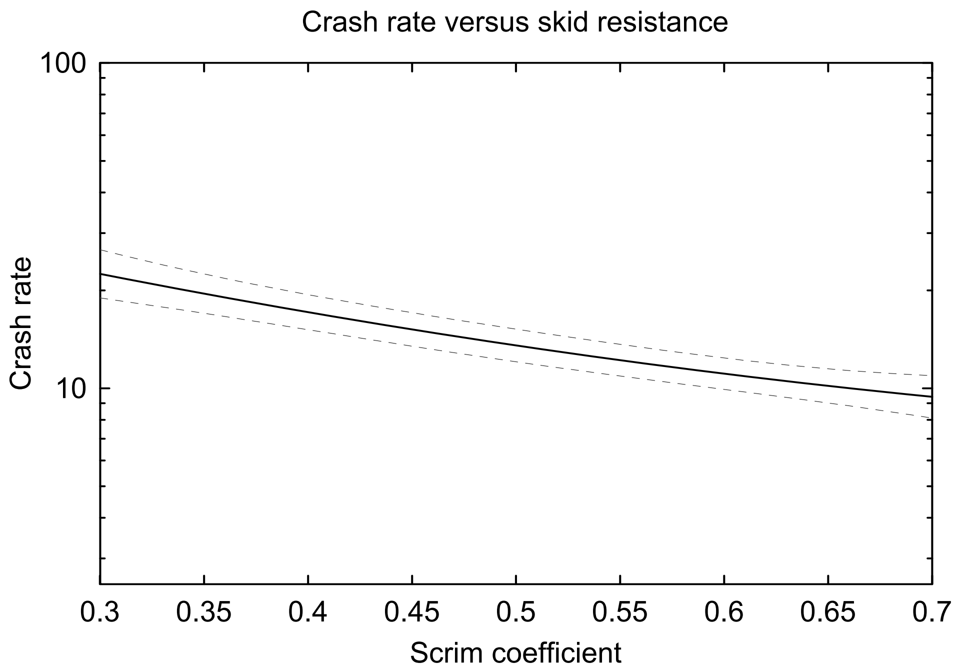
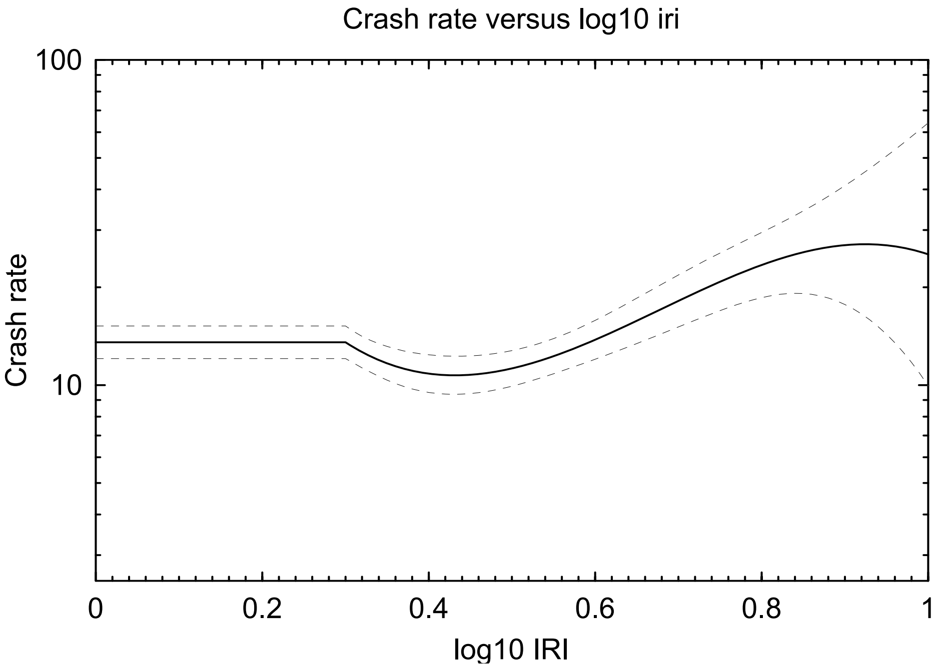


Figure 37



Wet crashes

Figure 38

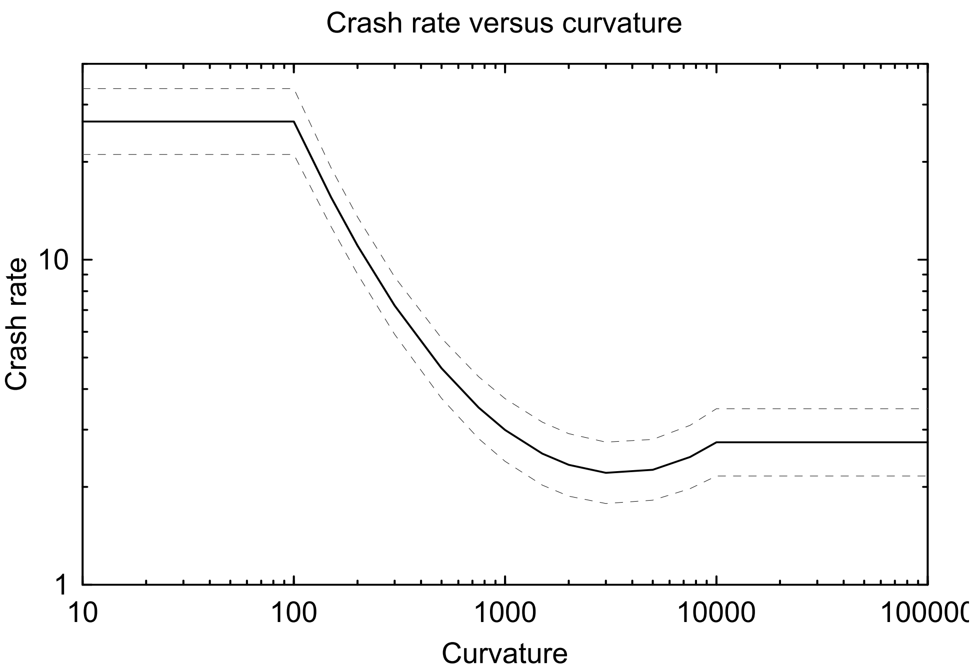


Figure 39

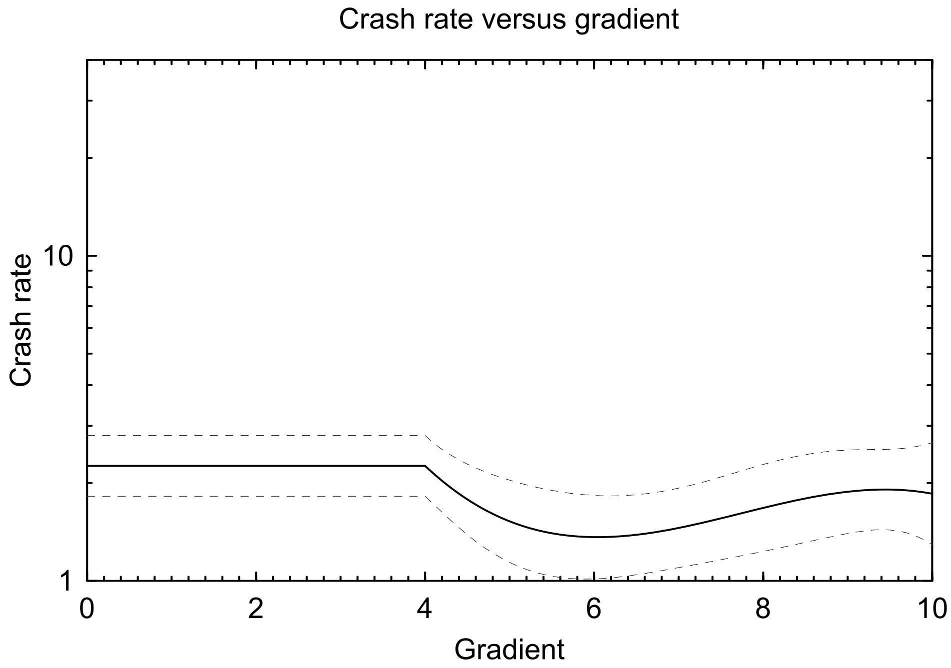


Figure 40

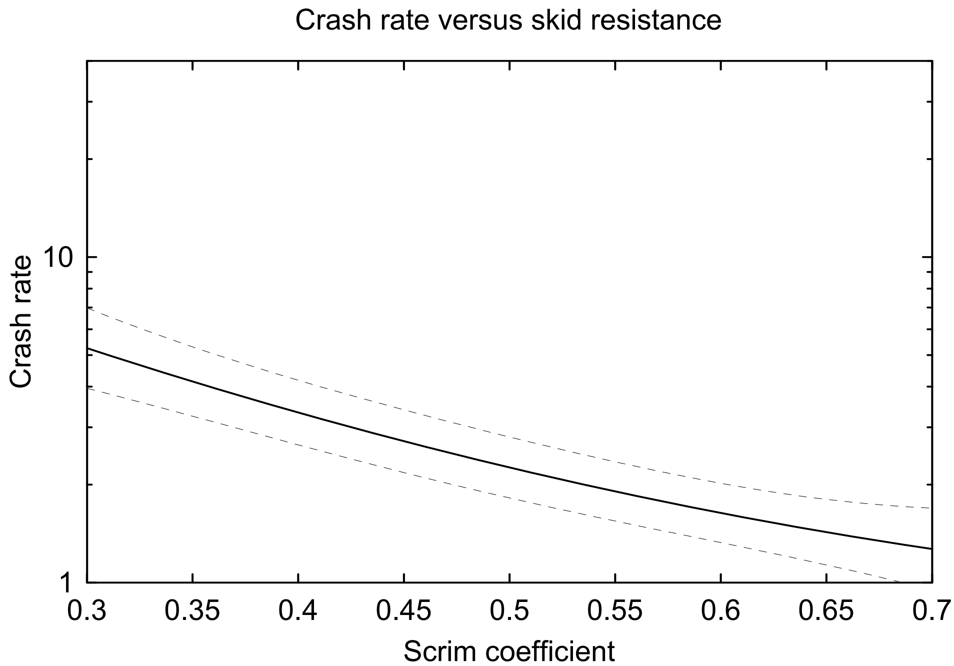
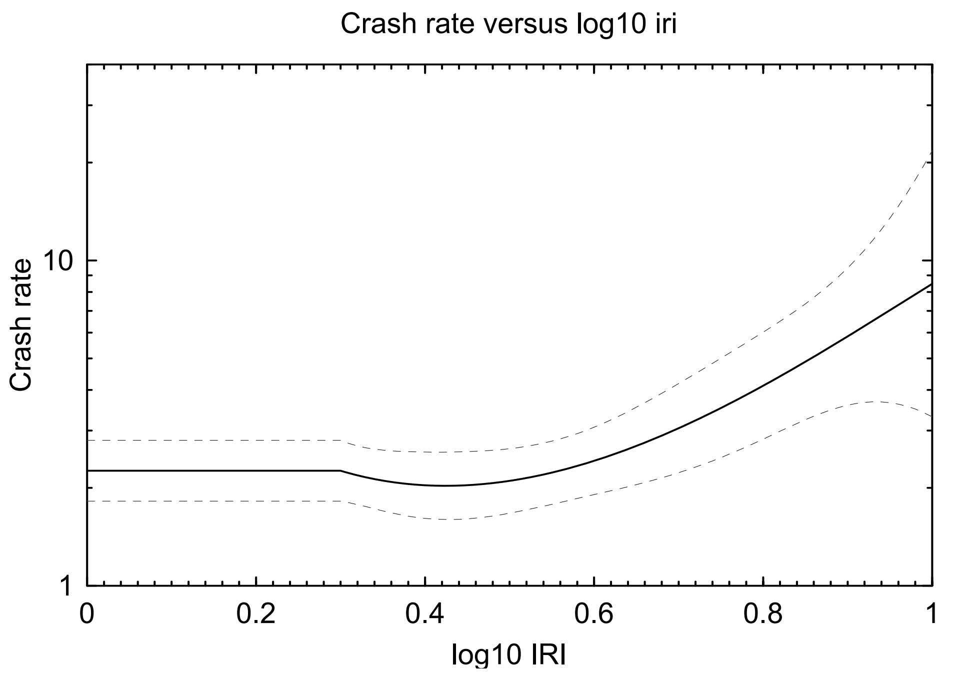


Figure 41



Wet selected crashes

Figure 42

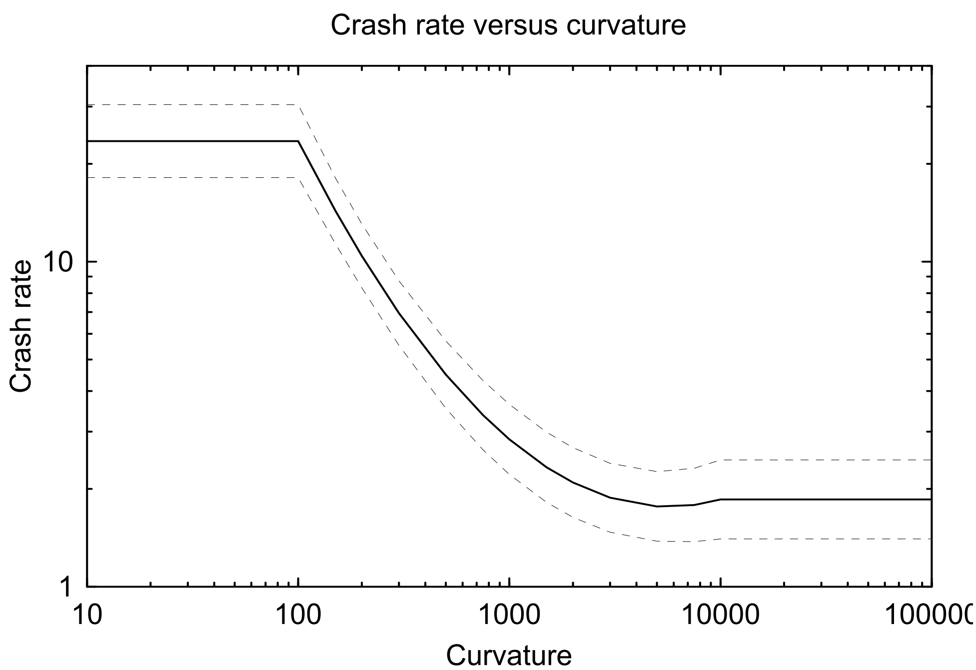


Figure 43

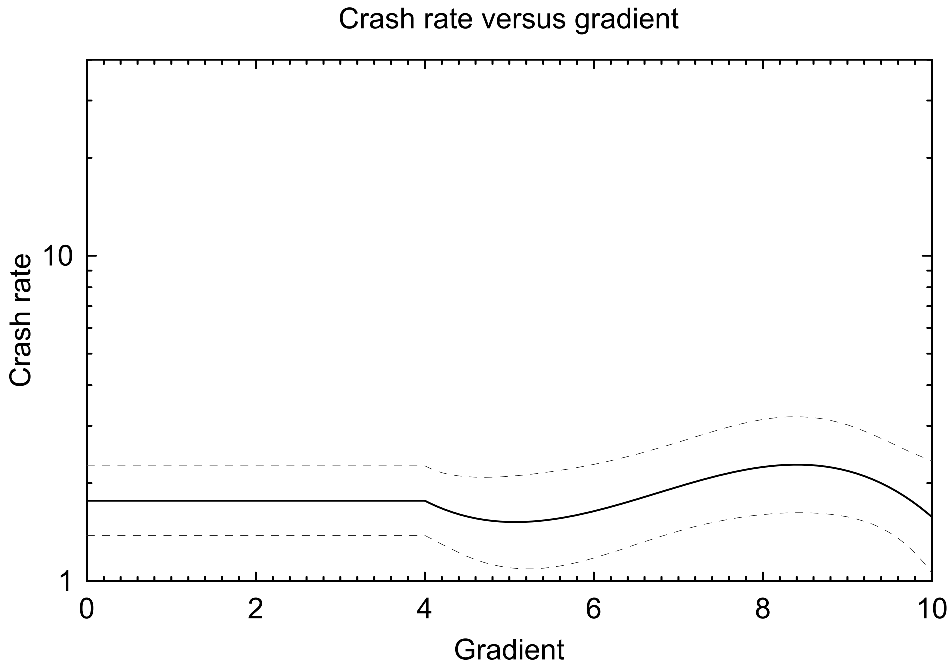


Figure 44

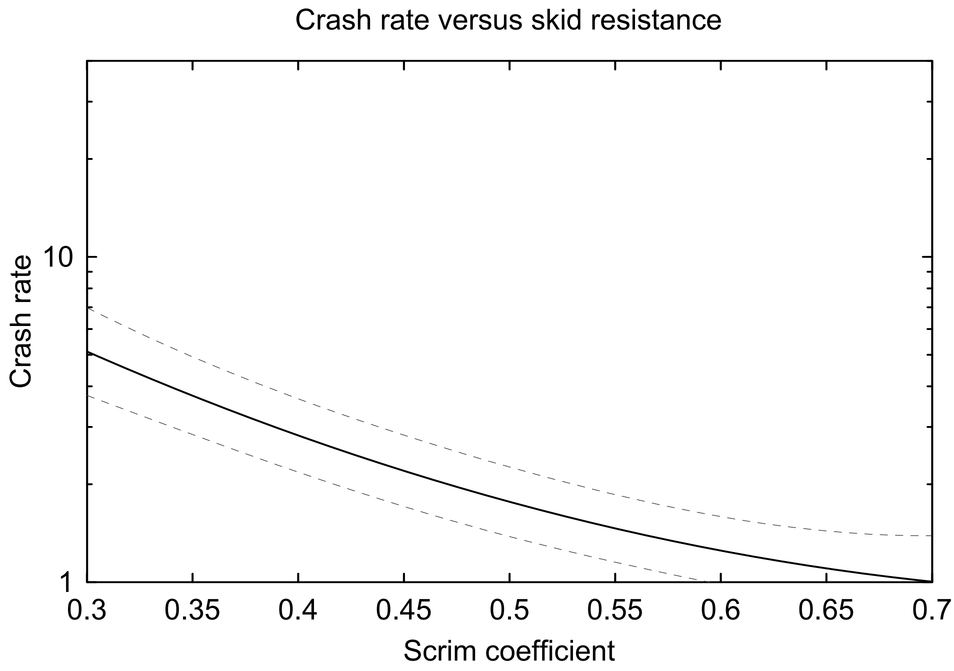
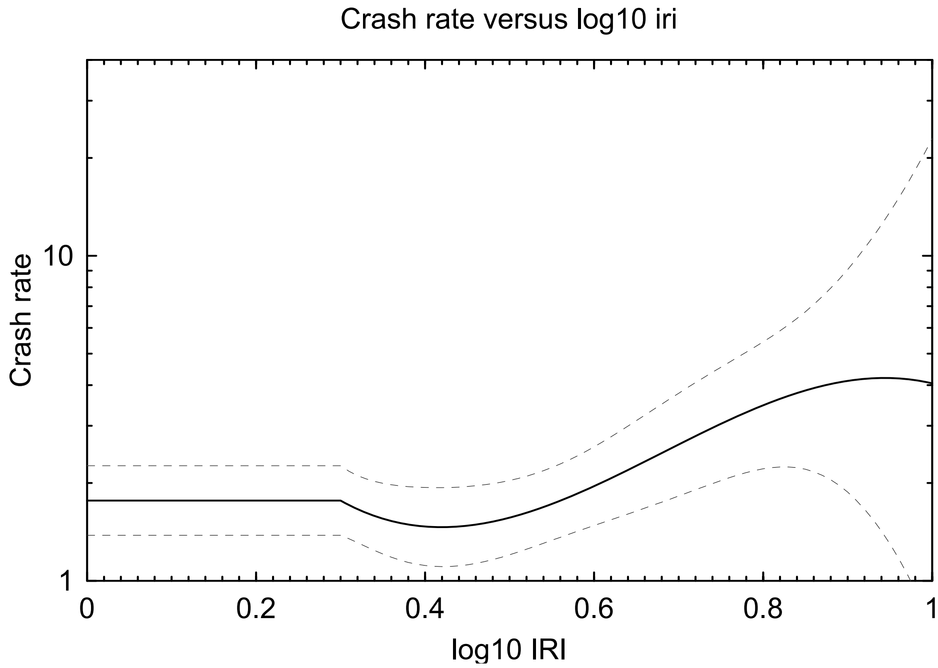


Figure 45



5.4 Example of calculation of predicted crash rate

Here are some examples of the calculation of the predicted crash rate using the simplified model.

5.4.1 Example 1

In this example we are fitting *wet selected crashes*.

Table 48: predicted crash rate example 1

variable	value	transformed variable	coefficient	transformed value	product
		constant	0.008	1	0.008
year	2002	year:1997	0.000		0.000
		year:1998	-0.216		0.000
		year:1999	0.059		0.000
		year:2000	-0.240		0.000
		year:2001	-0.175		0.000
		year:2002	0.008	1	0.008
region	R1	region:R1	0.000	1	0.000
		region:R2	0.188		0.000
		region:R3	0.091		0.000
		region:R4	0.537		0.000
		region:R5	0.041		0.000
		region:R6	0.161		0.000
		region:R7	0.073		0.000

variable	value	transformed variable	coefficient	transformed value	product
urban_rural	R	urban_rural:R	0.000	1	0.000
		urban_rural:U	-0.595		0.000
skid_site	4	skid_site:4	0.000	1	0.000
		skid_site:3	0.561		0.000
		skid_site:1	0.100		0.000
curvature	5000	$\log_{10}$(curvature)	-6.329	3.699	-23.412
		$\{\log_{10}$(curvature)$\}^2$	0.843	13.682	11.540
ADT	1000	$\log_{10}$(ADT)	2.516	3	7.549
		$\{\log_{10}$(ADT)$\}^2$	-0.424	9	-3.814
gradient	0	gradient	-2.802	4	-11.210
		(gradient)^2	0.443	16	7.092
		(gradient)^3	-0.022	64	-1.404
scrim	0.5	scrim-0.5	-4.073	0	0.000
		(scrim-0.5)^2	6.220	0	0.000
IRI	1.995	$\log_{10}$(IRI)	-17.379	0.3	-5.214
		$\{\log_{10}$(IRI)$\}^2$	29.938	0.09	2.694
		$\{\log_{10}$(IRI)$\}^3$	-14.644	0.027	-0.395
Sum					**-16.557**

The entries under *value* are the values of the variables for the segment for which we want the crash rate.

The values under *transformed value* are the values after the transformation in the column *transformed variable*. Note that curvatures and gradient are the absolute values. IRI is the value after adjusting for curvature – think of it as the value measured by a device which is not affected by curvature. Values of curvature, gradient and IRI which fall outside the ranges shown in Table 49 should be replaced by the nearest bound.

Table 49: ranges of variables

Variable	lower bound	upper bound
curvature	100	10,000
gradient	4	10
IRI	1.99526	10

So gradient has been replaced by 4 in Table 48.

For the *factors* (year, region, skid_site, urban-rural) the transformed value is 1 in the line corresponding to the factor level we want; otherwise it is zero (or leave blank).

The values under *coefficient* are the *estimates* from the regression model (Table 47 for *wet selected crashes*).

The values under *product* are the products of the previous two columns. The sum of the products is the value of *L* in equation (1). To get the crash rate (in crashes per 100 million vehicle km) calculate

$$\exp(L) \times 10^{10} / 365 . \quad (3)$$

In this example we get 1.8. This is the value predicted by the model.

There should be a correction to allow for the fraction of crashes located. Say divide by 0.89 from Table 8. Finally, of course, only a relatively small fraction of crashes are reported and our figure is concerned only with *reported* crashes.

The figure of 1.8/0.89 = 1.9 crashes per 100 million vehicle km appears low because we are counting only *wet* selected crashes whereas 100 million vehicle km it is compared with is for all driving conditions – both wet and dry. We don't have the fraction of time the road is wet (preferably weighted by usage) so we can't give crash rates exclusively for wet roads.

5.4.2 Example 2

In this example we are fitting *all crashes*.

Table 50: predicted crash rate example 2

variable	value	transformed variable	coefficient	transformed value	product
		constant	2.095	1	2.095
year	2000	year:1997	0.000		0.000
		year:1998	-0.060		0.000
		year:1999	-0.053		0.000
		year:2000	-0.118	1	-0.118
		year:2001	0.000		0.000
		year:2002	0.198		0.000
region	R3	region:R1	0.000		0.000
		region:R2	0.108		0.000
		region:R3	0.210	1	0.210
		region:R4	0.306		0.000
		region:R5	0.224		0.000
		region:R6	0.105		0.000
		region:R7	0.124		0.000
urban_rural	U	urban_rural:R	0.000		0.000
		urban_rural:U	-0.157	1	-0.157
skid_site	3	skid_site:4	0.000		0.000
		skid_site:3	1.595	1	1.595
		skid_site:1	1.697		0.000
curvature	100000	$\log_{10}$(curvature)	-5.360	4	-21.441
		$\{\log_{10}(\text{curvature})\}^2$	0.759	16	12.152
ADT	10000	$\log_{10}$(ADT)	0.707	4	2.827
		$\{\log_{10}(\text{ADT})\}^2$	-0.173	16	-2.769
gradient	0	gradient	-2.598	4	-10.393
		$(\text{gradient})^2$	0.314	16	5.029
		$(\text{gradient})^3$	-0.012	64	-0.759

variable	value	transformed variable	coefficient	transformed value	product
scrim	0.4	scrim-0.5	-1.637	-0.1	0.164
		$(\text{scrim-0.5})^2$	-0.090	0.010	-0.001
iri	1.995	$\log_{10}(\text{IRI})$	-10.540	0.3	-3.162
		$\{\log_{10}(\text{IRI})\}^2$	19.219	0.09	1.730
		$\{\log_{10}(\text{IRI})\}^3$	-9.850	0.027	-0.266
sum					**-13.265**

The calculations are as before. The coefficients are from Table 44.

Applying formula (3) we get a crash rate of 48 crashes per 100 million vehicle km. The correction for the fraction of data located is to divide by 0.74 from Table 8.

5.5 Comparison with previous studies of the effect of skid resistance

Previous studies using the paired crashed sites which found 95% confidence intervals for the reduction in crash rate per 0.1 increase in scrim as

- (1.8, 2.9)
- (1.1,1.8)
- (1.2,1.7).

We can't do direct comparison with the analyses in the previous section because I have allowed for curvature in the graph of log(crashrisk) versus scrim coefficient. Therefore I repeated the fit of the simplified model for the wet selected crashes with only a linear function of the scrim coefficient.

Table 51: simplified model with linear scrim term

	Wet selected
Analysis reference	340
Number of crashes	2023
Maximum of the log-likelihood function	-17054.15

This gave a 95% confidence interval for the reduction in crash rate per 0.1 increase in scrim coefficient of (1.44,1.66) which is in general agreement with the previous estimates. Probably, we should regard the confidence interval as claiming more accuracy than is actually the case since there are effects not taken into account in the model.

6 Additional testing

This section looks at tests of the fit of the model and some modifications of the model.

6.1 Comparison of fitted and observed counts

This is an attempt to see how well the model fits.

The highway network was divided into segments by partitioning by carriageway area and state highway number. That is two sections of road are in the same partition if they are in the same carriageway area and on the same state highway. This gave 136 partitions. Then the model was used to predict the number of crashes in each of these. The observed numbers were compared with the predicted number. Figure 46 shows the comparison for all crashes and Figure 47 shows the normalised residual

$$\frac{\text{Observed} - \text{Predicted}}{\sqrt{\text{Predicted}}}$$

in terms of *predicted*.

Figure 46

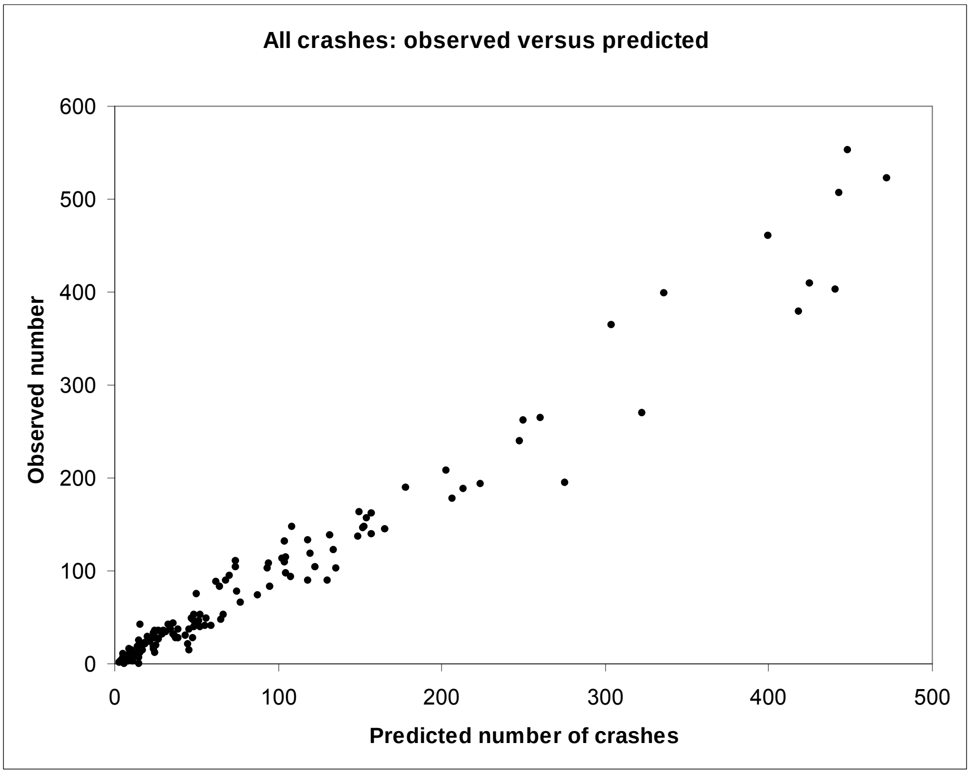
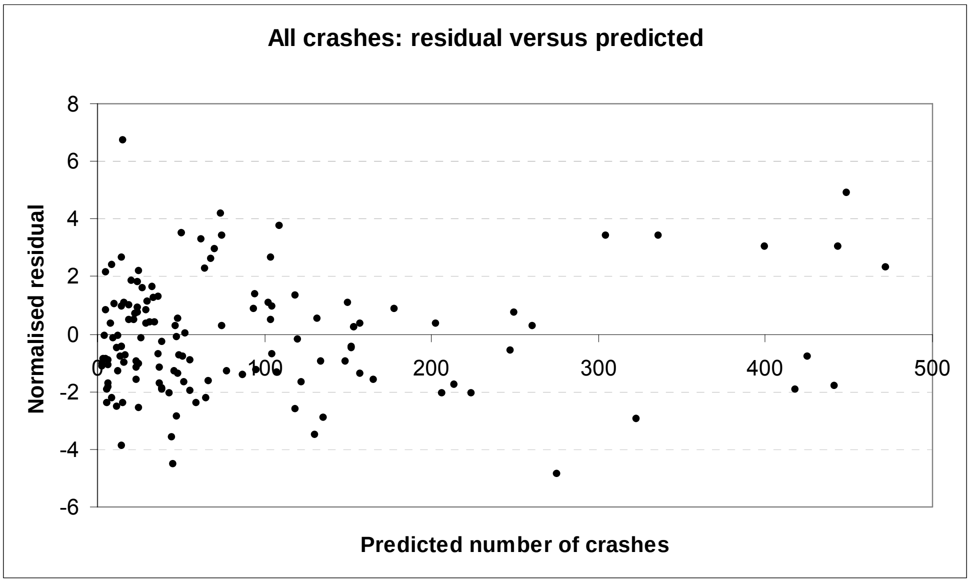


Figure 47



If the model was fitting perfectly there would be few points outside the range –2 to 2. The actual range of points is more like –4 to 4 with a few outside this range (particularly to the left of the graph). So the model doesn't fit perfectly.

Here are the corresponding graphs for the wet road crashes.

Figure 48

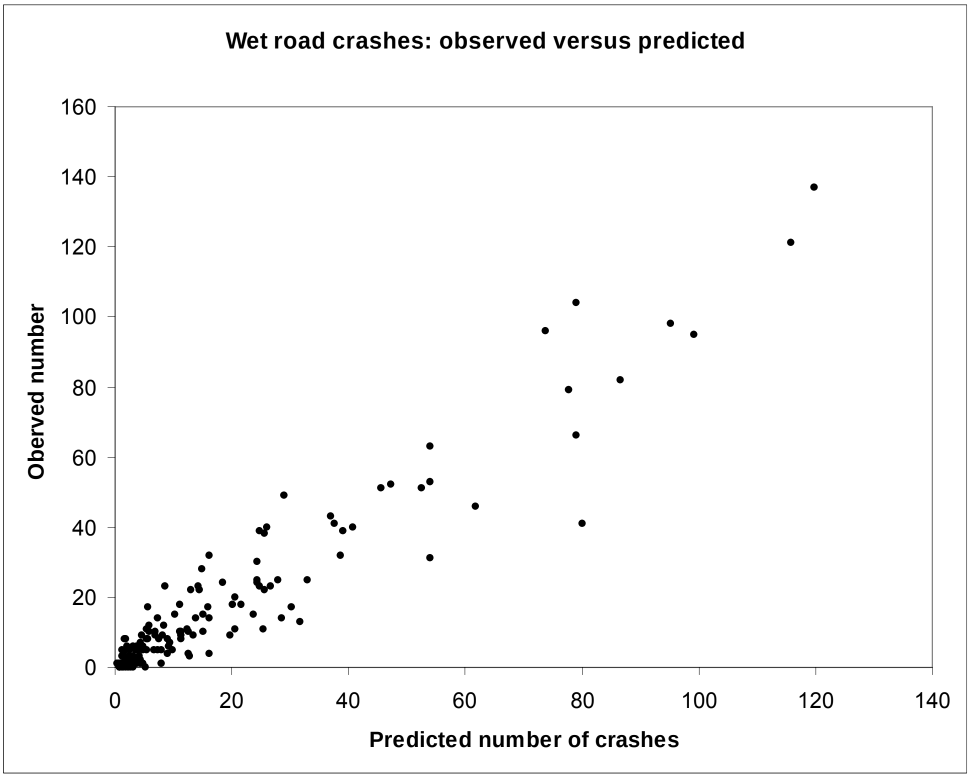
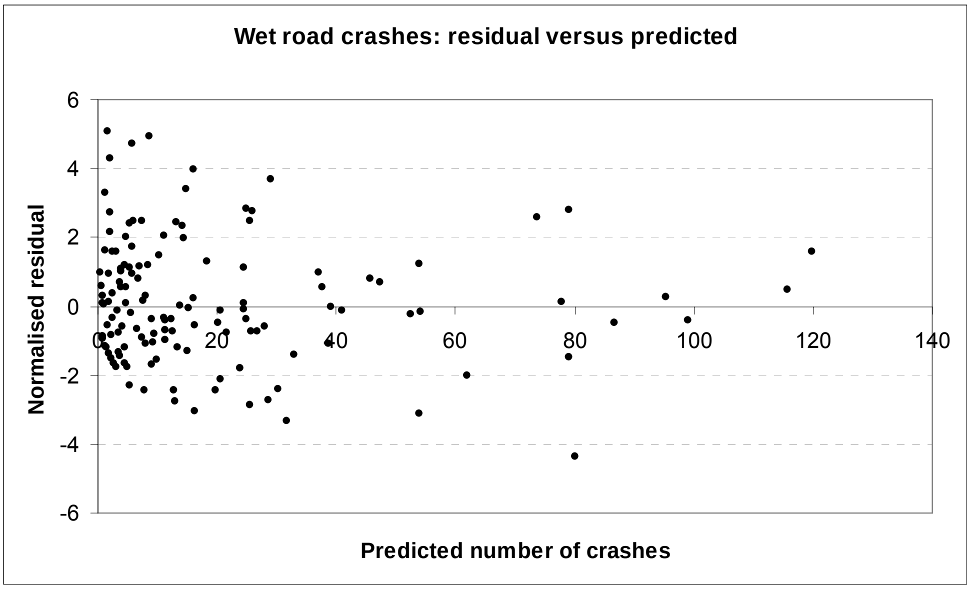


Figure 49



The quality of the fit is about the same as for all crashes.

One could identify the particular points where the fit is bad and see if there was a data problem or where there were special risk factors. There are no points that are dramatically failing to fit so this is not done in the present study.

6.2 Interaction between curvature and scrim

The simplified analysis was run for the wet-selected-crashes with an additional term to allow for an interaction between curvature and scrim. This is to see if we can detect a difference in the effect of scrim on roads with different curvature. In fact, the effect was not statistically significant – the analysis could not detect such an effect.

Table 52: effect of interaction term

	Without interaction term	With interaction term
Analysis reference	330	350
Number of crashes	2023	2023
Maximum of the log-likelihood function	-17051.71	-17051.69

Table 52 shows the maximum of the log-likelihood function when we don't or do include the interaction term. The interaction involves 2 degrees of freedom so the difference in not statistically significant (the difference would need to be at least 3 for 5% significance).

6.3 Replication

In order to check the stability of the model the simplified version for selected-crashes was fitted using only the 1997-1999 data and again for the 2000-2002 data. Figure 50 shows the predicted crash risk as in section 5.3.3. For the 1997-1999 data the estimate is for 1997 and for the 2000-2002 data the estimate is for 2000. These two years were chosen because they have similar year effects in Table 45. In this case, I am not showing the amalgamated parts for curvature radii less than 100 or greater than 10000. I have shaded the area between the confidence intervals dark gray for the 1997-1999 data and light gray for the 2000-2002 data. Where the two confidence intervals overlap the colour is black.

There is a lot of overlap showing that the estimates of the crash rates as a function of curvature are very similar.

Figure 50

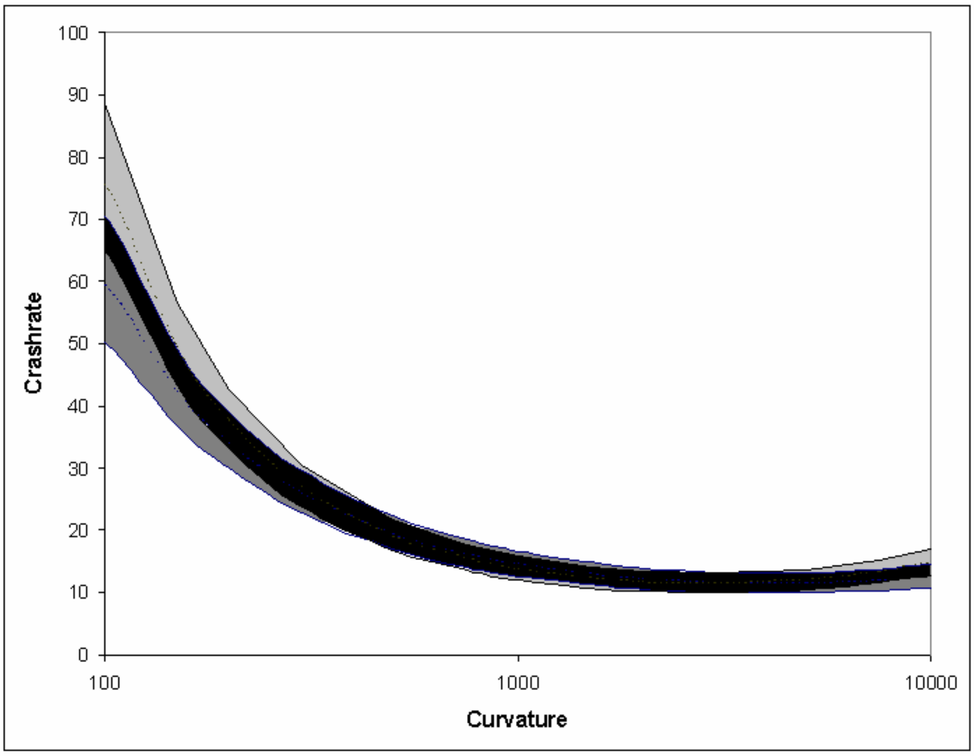
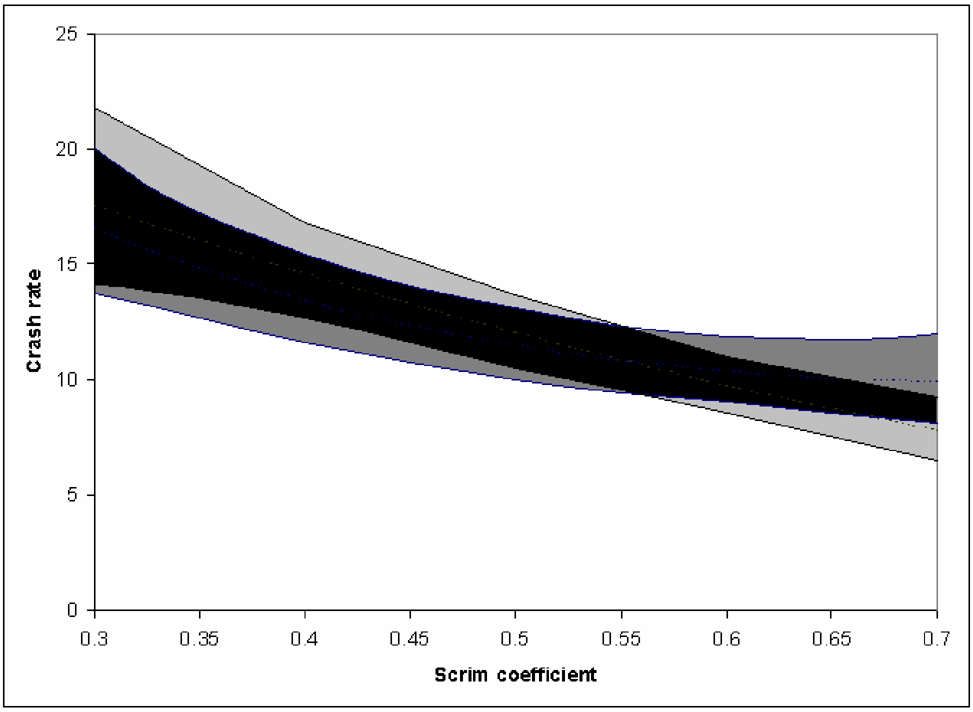


Figure 51 is the corresponding graph for skid resistance.

Figure 51



Again there is a lot of overlap but there is not enough data for one to be sure that the graphs for the two periods are very similar.

6.4 Effect of the averaging

These calculations were made with an earlier version of the model that did not include gradient or IRI. However, I do not expect the results to be substantially different from those that would have been obtained using the model of section 5.3.

The model fitted in section 5 averages the prediction from the log Poisson model (1) over 21 adjacent 10 metre segments, that is the segments within a range of 10 segments of the one in which we are actually making the estimate. This section investigates the effect of altering the number of segments being averaged. I consider only the simplified model and either the all-crash or wet-crash data.

Table 53: effect of averaging length on the all-crashes model

Averaging length	Analysis reference	log likelihood	skid-site chi-squared	curvature chi-squared	scrim chi-squared
30	368	-83944.20	1769.1	1115.2	131.48
90	367	-83838.54	2098.9	1117.0	132.96
110	366	-83828.64	2148.3	1190.8	134.45
130	365	-83838.81	2153.1	1216.7	135.38
210	300	-83987.63	1832.2	1265.1	144.15

Table 53 shows the effect on the model for all-crashes when we change the averaging effect. The chi-squared values are the ones analogous to those in

Table 42. The last line when we are averaging over 210 metres corresponds to the model used in section 5. The best likelihood is obtained in the third line, when we are averaging over 110 metres. A change in the log likelihood of 2 or more can be considered as statistically significant and the changes we are observing are rather more than this. The chi-squared value for skid-site is maximum in the fourth line and the other chi-squared values increase as we proceed down the table.

Table 54 shows the effect on the model for wet-crashes when we change the averaging effect.

Table 54: effect of averaging length on the wet-crashes model

Averaging length	Analysis reference	log likelihood	skid-site chi-squared	curvature chi-squared	scrim chi-squared
30	362	-17583.73	33.3	677.4	126.2
210	330	-17466.10	31.3	989.4	183.6
310	360	-17458.85	28.2	1019.3	194.0
410	361	-17477.55	32.7	999.8	186.3

In this case the best likelihood is obtained when we are averaging over 310 metres. This is also the best value for the curvature and scrim chi-squared values. In this case the effect on chi-squared is quite strong.

Here is a possible explanation for the different averaging lengths that seem to be optimum. For all-crashes there are a lot that have their locations well determined either because they occur in urban areas or are associated with intersections or crossings. Hence, not too much averaging is required. The wet-crashes may be more dominated by the open-road crashes with less well determined locations, so more averaging is required. It is also possible that for the high curvature or the low skid-resistance longer stretches are proportionally more hazardous than shorter stretches and the averaging process is able to model this.

6.5 Advisory speed

Advisory speed is calculated from the formula

$$AS = -\left(\frac{107.95}{H}\right) + \sqrt{\left(\frac{107.95}{H}\right)^2 + \left(\frac{127{,}000}{H}\right)\left(0.3 + \frac{\lambda X}{100}\right)} \qquad (4)$$

where AS = RGDAS advisory speed (km/h)
X = % crossfall (sign relative to curvature)
H = absolute curvature (radians/km) = (1000m / R)
λ = 1 if crossfall is to be included.

If we don't include crossfall ($\lambda = 0$) then advisory speed can be used in place of curvature in the regression models and will give similar results. If advisory speed is taken as an indication of the *difficulty* of segment of road then including crossfall should improve the fit.

Several fits with different values of λ were tried using the small model but with the curvature term replaced by a 6 point spline function of the advisory speed over the range 15km/hr to 115km/hr. Crossfalls were limited to the range –30% to 30%. Curvatures less than 10 in absolute value were replaced by 10 or –10 depending on their sign. Table 55 shows the log-likelihoods calculated by the fit program.

Table 55: Effect of crossfall

Crossfall multiplier	Log likelihood
1	-81487.10
0.25	-81474.50
0	-81469.22
-0.25	-81465.32
-0.5	-81466.33
-1	-81487.50

The best fit (largest log likelihood) has λ between –0.25 and –0.5. This is the wrong sign, but is not significantly different $\lambda = 0$ (not including crossfall) if we use the same intuitive criteria that were used in section 5.

Thee are graphs of crossfall versus curvature in section 9.6 which show we are interpreting the signs of the crossfall and curvature correctly.

7 Application of the model

7.1 What-if study

This is an illustration of the use of the model.

We consider increasing the skid resistance of the sections of road classified as skid-site 2 (curve with less than 250m radius or gradient greater than 10%; investigatory level 0.5). We choose one of the following minimum levels for scrim: 0.4, 0.5, 0.6. If the actual scrim value is below this the road surface is upgraded to raise the scrim value to this amount; otherwise it is left the same. We can choose to do this for all the roads that fall short of the minimum scrim level or only those with ADT above some prescribed level.

Our what-if study is for 2001. Table 56 shows the actual road and crash data for 2001 for the road segments classified as skid-site 2.

Table 56: actual road and crash data (for 2001)

Length of road sides	1992
Number of crashes	381
Number of wet crashes	80

Table 57 shows the reduction in the predicted number of crashes 2001 if the upgrade had been done before 2001.

Table 57: what-if study on skid resistance

			All crashes		Wet crashes	
minimum scrim	fix for ADT≥	fix length	predicted crashes	saved crashes	predicted crashes	saved crashes
0	0	0	370	0	101	0
0.4	0	120	369	2	99	2
0.5	0	719	357	13	90	11
0.6	0	1574	330	41	76	25
0.4	1000	93	369	2	99	2
0.5	1000	545	358	12	91	10
0.6	1000	1055	333	37	78	23
0.4	5000	18	370	1	100	1
0.5	5000	98	365	5	97	4
0.6	5000	169	356	14	92	9

The first line is for no upgrade, the rest are for the values of minimum scrim value and ADT shown in the first two columns. It is supposed that the two sides of the road are handled independently and the column *fix length* shows the length of side that needs to be upgraded. The analysis is carried using the model for all crashes and the model for all wet crashes. The table shows the predicted number of crashes for each model and then the reduction in the number of crashes compared with the first line.

The numbers of saved crashes is less for the wet crashes than for total crashes. This is probably because

- not all crashes that actually took place on wet roads were classified as wet or
- scrim does have some effect on dry roads or
- higher values of scrim are associated with other road characteristics which improve safety.

The table suggests one should consider raising minimum scrim to 0.6 on skid-site 2 roads with ADT ≥ 1000.

This example should be regarded as illustrative only at this time. In particular, I have not tried to estimate accuracy.

8 Discussion

8.1 Main results

The Poisson regression model appears to work in a reasonably satisfactory way and produces results that, for the most part, make sense.

There is a strong effect of curvature, as expected. There is also a strong effect from the skid-resistance measurement and a weaker effect due to roughness. The effect of skid resistance was in line with the values found from the paired crash site analyses.

There also seems to be an effect due to roughness.

There is an effect from gradient that seems unexpected. Possibly this is due to problems with the gradient measurement.

The other surface measurements did not lead to relationships that we could use.

8.2 Credibility of results

This is a retrospective analysis (as opposed to a designed experiment). So 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 predictor variables are subject to measurement error. In particular the SCRIM+ technology is quite new and one must expect some problems with it.

Where predictor variables are subject to random error, one will find a decrease in the size of the effect detected by the model. Thus regression coefficients are likely to be under-estimated.

There appears to be error that is not accounted for by the Poisson model. This is partly due to errors in the predictor variables, including ADT, and road properties that are not accounted for in the model. For the data involving the wet crashes, variation in the percent of time the road is wet is possible source of additional error. There has been no attempt to model the special characteristics of skid-site 1 events. In particular, there is no information on the traffic crossing the highway at intersections.

All this means that confidence intervals will be wider than those calculated by the model – possibly twice as wide.

Nevertheless the analysis does seem to be giving sensible results – but it is important that as far as possible they are cross-checked against other studies.

8.3 Applications

There are four main applications of the analyses:

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. An example of the type of calculation is given is section 7.

3. To identify *blackspot* 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 *whitespots* where crash rates are lower, although this is less likely to be successful. An example of the type of analysis but for rather longer sections of road than would normally be used in a *blackspot* study is given in section 6.1.
4. It *may* be possible to use the model to help evaluate the effect of an actual change in road construction or management policy in a region by comparing the observed and predicted number of crashes.

8.4 Further work

It may be possible to model the error structure of the model more accurately than the simple Poisson model of section 4 and so get more objective criteria for including variables and for calculating confidence intervals.

There needs to be more investigation of interaction between the predictor variables. This applies particularly to crossfall which probably has different effects at different curvatures. The same might also apply to carriageway width. Both of these variables show up as non-significant but there might be some effect which will show up only when we look at interactions. Some further checking in interaction between skid resistance and the other variables would be appropriate. There was an attempt at this in section 6.2, but there needs to be further analysis.

The black spot detection described in section 8.3 needs investigation.

There needs to be an investigation of sudden changes in the road characteristics – particularly curves following a period of straight. Preliminary investigations did not pick up any effect, but this needs to be looked at again.

It may be possible to look at other criteria for selecting crashes, particularly for deciding which types of crashes are affected by the different road characteristics.

It may be possible to do a more accurate analysis of skid-site one locations taking account of the different types of sites.

Dual lane roads including motorways were not considered in this analysis and it may be possible to include these.

The method of locating the crashes needs to be reviewed, especially as centre-line data is updated. In may be possible to include the estimated accuracy of the location in the modelling process.

9 Notes

9.1 Skid-site definitions

The descriptions of skid-site definitions are shown in Table 58. Divided carriageways are not included in this study so skid-site 5 is not relevant. We want the effect of curvature to be modelled by the curvature term rather than the skid-site term so I have absorbed skid-site 2 into skid-site 4 in the Poisson regression model fits.

Table 58: Skid-site descriptions

skid site	Description	Notes	scrim site investigatory level
4	Normal roads	All normal roads. (Undivided carriageways only)	0.4
3	Approaches to road junctions	Approaches to road junctions.	0.45
2	Curve <250m rad. Gradient>10%	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
5	Divided carriageway	Divided carriageways	0.35

9.2 Curvature

The curvature is measured as radius of curvature in metres. A value of 100,000 appears to mean straight so I replaced all values greater than or equal to 100,000 by 100,000. The values of 0 and 1 seemed to be special codes which I interpreted as missing. Where values were present on one side of the road and missing on the other I replaced the missing values by the values on the other side of the road.

9.3 Roughness

As noted in a previous study, the roughness measurement seems to be severely affected by curvature. The following graphs show log IRI fitted as a function of curvature and gradient. Curvature is a spline fit and gradient is just a quadratic. Left and right hand sides are handled separately. These graphs use all the data but one gets similar graphs if the age of the pavements since reconstruction are limited to two years. So the effect is most likely to be a measurement or definition effect rather than a real effect due to wear.

Figure 52

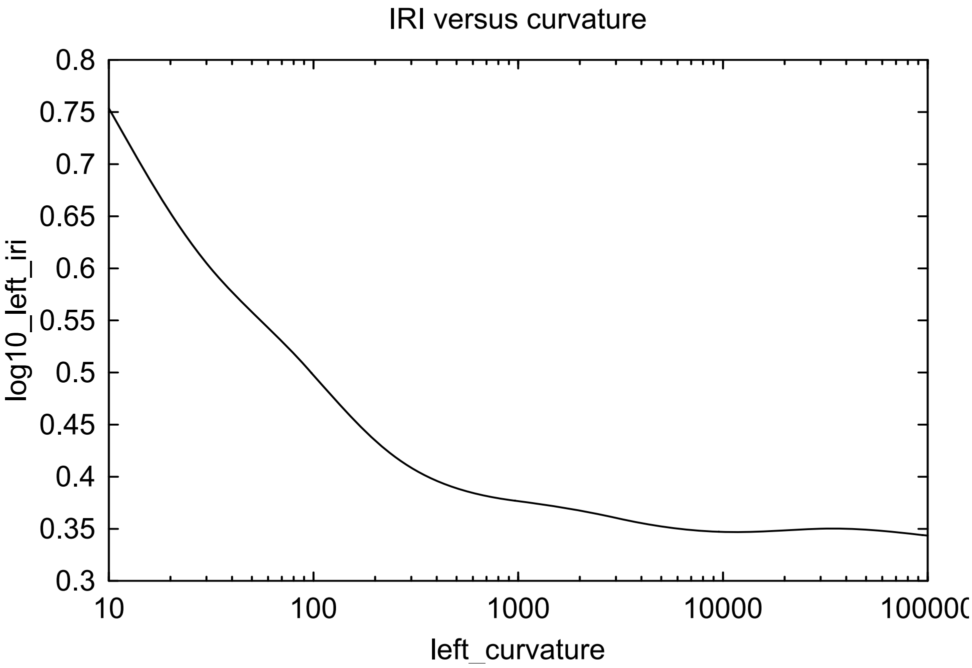


Figure 53

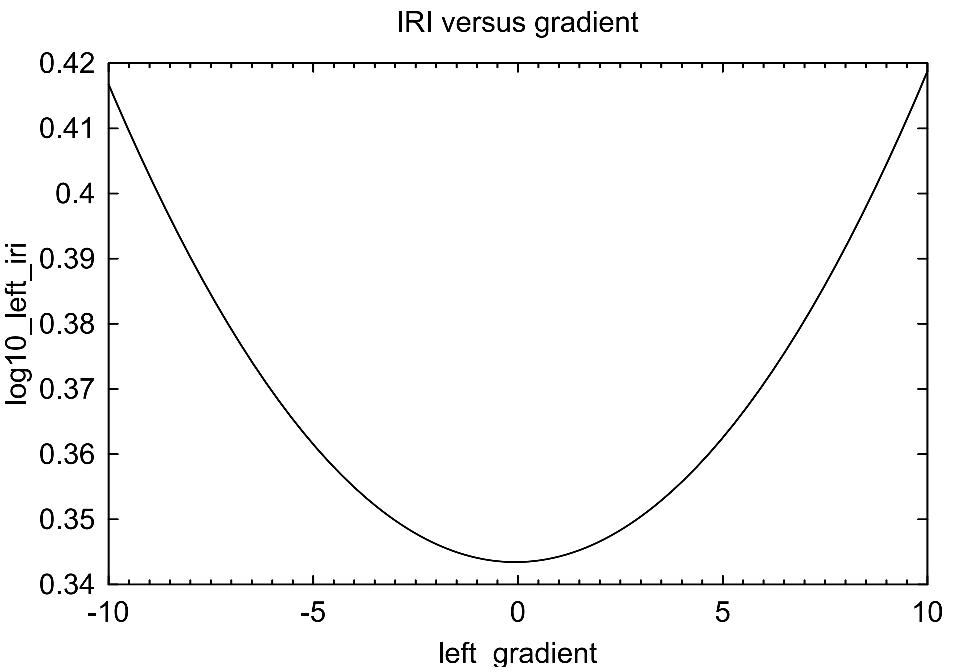


Figure 54

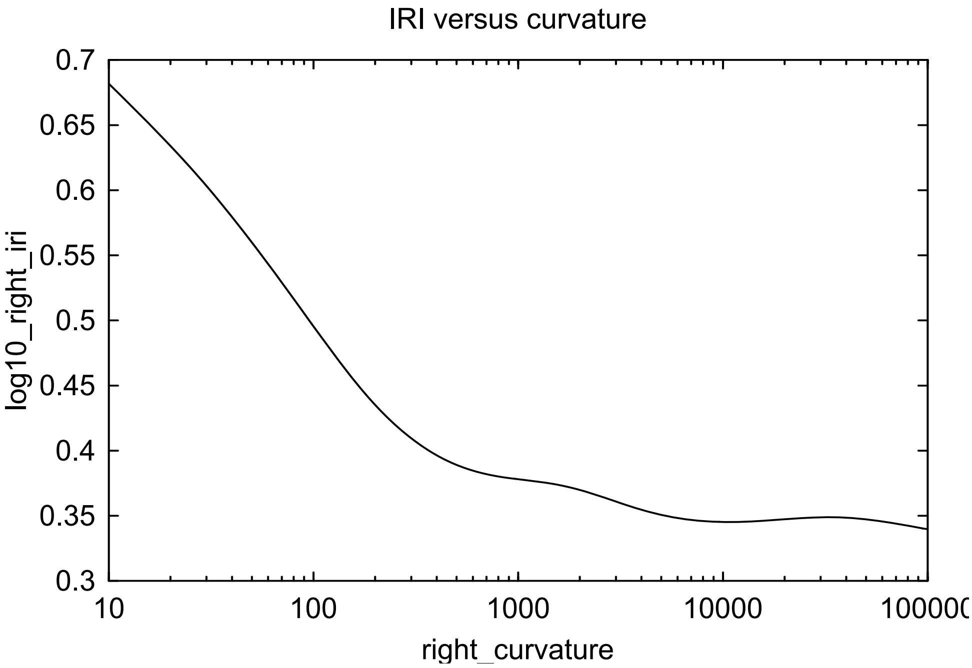
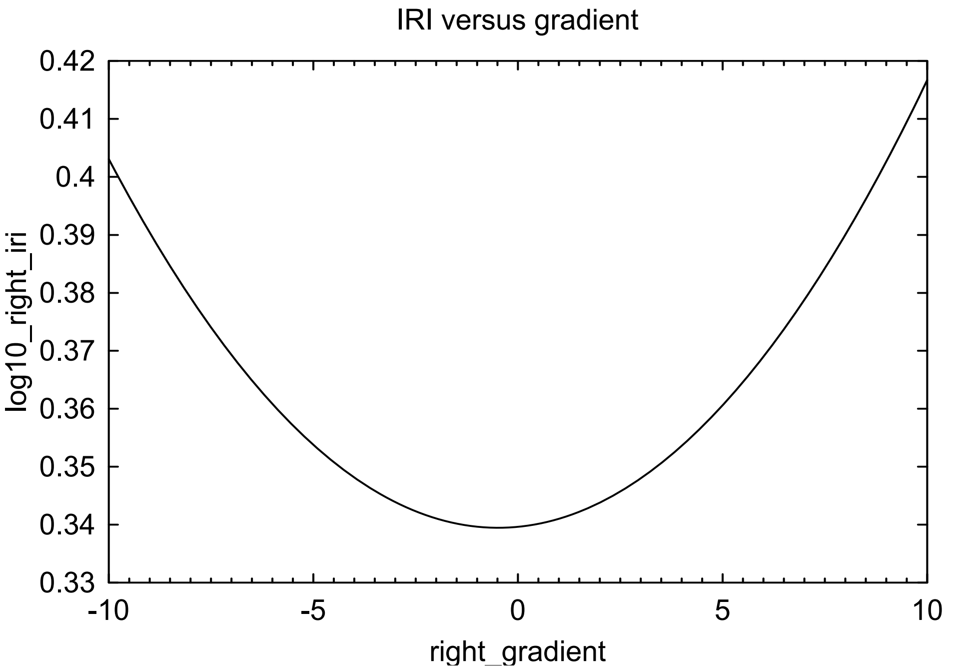


Figure 55



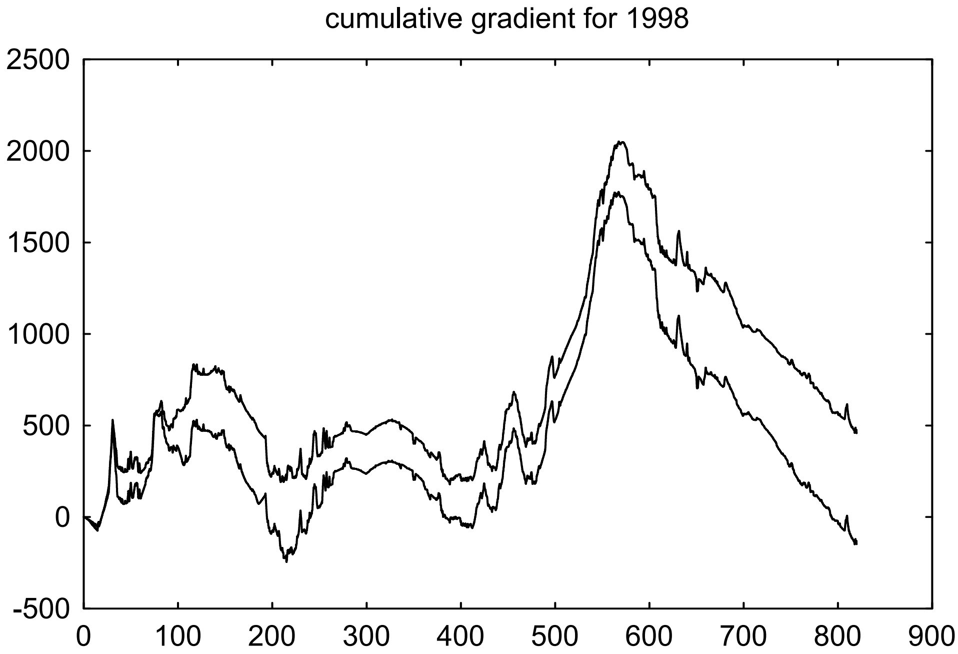
It seems best to try to adjust the IRI measurements to remove the effect of curvature and gradient. This is to reduce the possibility of the IRI effect being confused with the curvature and gradient effects. In addition, since a non-linear transformation is carried out on the IRI value, the bias due to curvature and gradient would be likely to distort the estimation of any effect of IRI.

The correction was carried out by subtracting the fitted effects identified in these graphs from the log IRI.

9.4 Gradient

The following four graphs show the cumulative sum of the gradient as we proceed down state highway 1 in the North Island for the years 1998, 2000, 2001, 2002. The two lines show the forward and reverse direction (with the sign reversed).

Figure 56



It is too much to expect these graphs to show an accurate altitude profile of state highway 1 because of the accumulation of measurement errors. However, apart from the 1998 graph they seem to be dominated by something other than the gradient. Even the 1998 graph seems to be showing the Dessert Road summit as too high. It should be around 1000 metre.

Figure 57

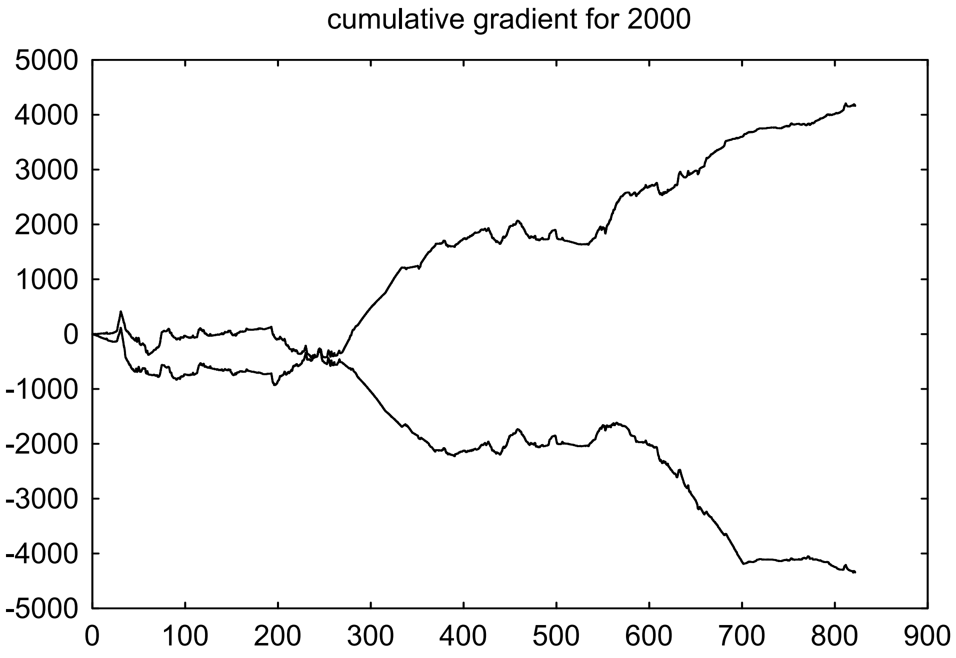


Figure 58

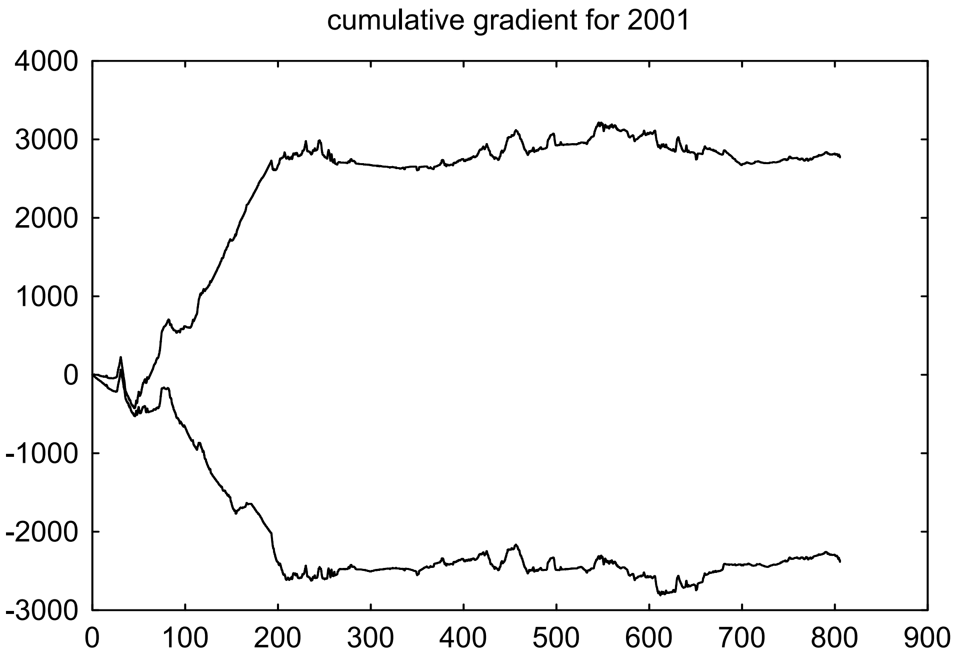
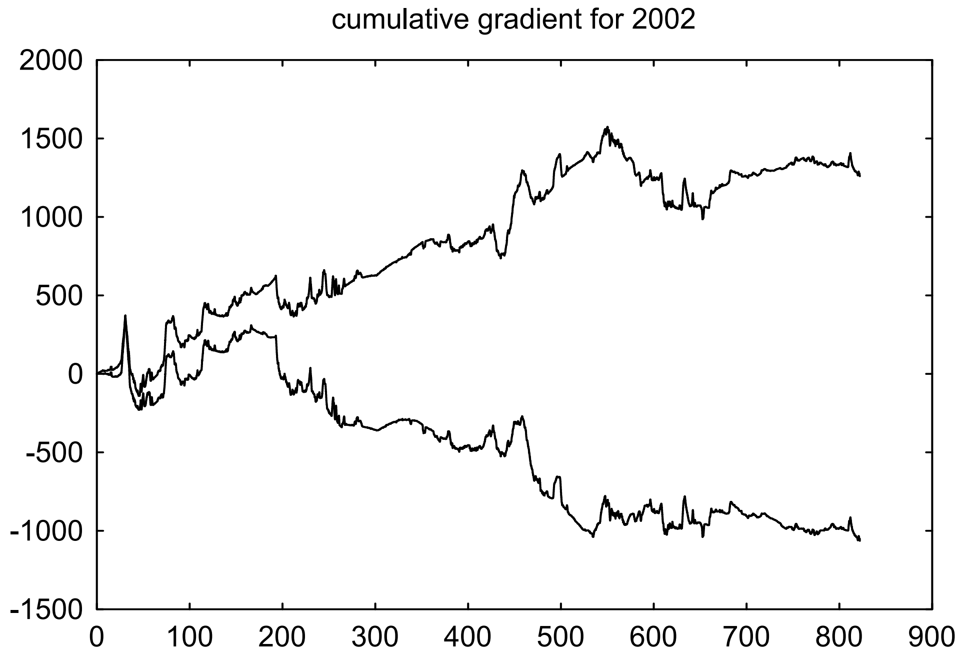


Figure 59



9.5 Carriageway width versus ADT

The analysis in section 5.2.1 did not find an effect of carriageway width large enough to be included in the simplified model. Actual statistical significance was not clear because the analysis was not able to fully model the error structure of the data. The purpose of this section is to note that the effect of carriageway width could be masked by the ADT effect. Figure 60 shows a boxplot relating carriageway width to ADT. I am using the carriage way widths recorded in the year 2000 survey, restricting attention to two lane roads and have deleted any observation where the width is greater than 20 metres. The ten metre segments were grouped according to their ADT values as shown in Table 59 and then a boxplot drawn for the carriageway widths for each of the widths. The whiskers show the maximum and minimum and the box shows the upper and lower quartiles so half of the observations lie within the box. The bar through the box shows the median so half the observations lie above the bar and half below.

Figure 60

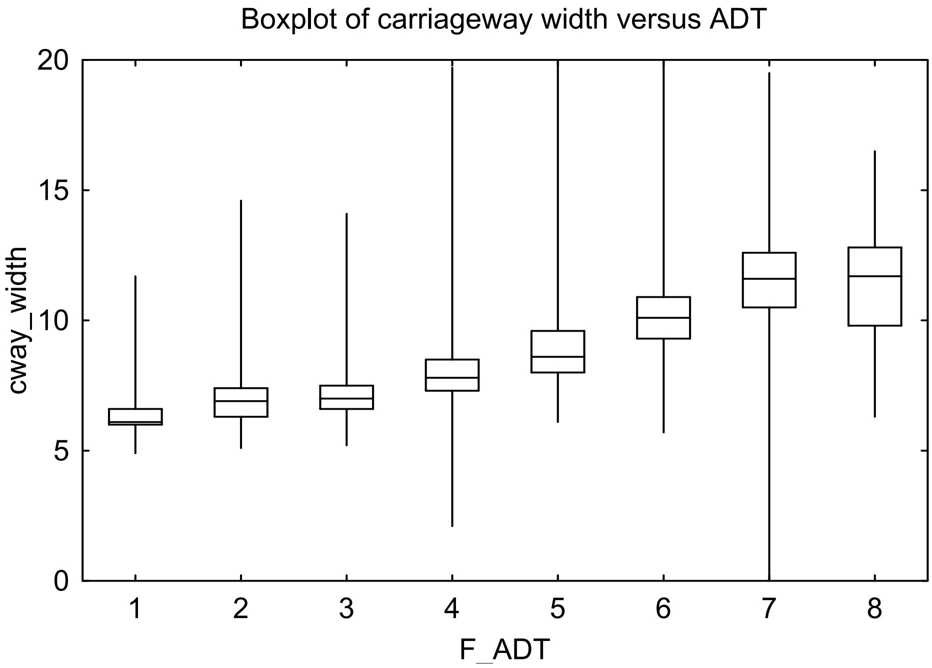


Table 59: Grouping of ADT values

ADT group (F_ADT)	ADT range	Length (km)
1	<200	70
2	≥200, <500	644
3	≥500, <1000	2126
4	≥1000, <2000	2608
5	≥2000, <5000	2518
6	≥5000, <10000	1377
7	≥10000, <20000	426
8	≥20000	79

As one would expect there is an upwards trend: higher ADT roads tend to have higher carriageway widths but with quite a bit of variation. However the general trend means that a carriageway effect will be masked, at least to an extent by the ADT effect that we have included in the model.

9.6 Crossfall and curvature

Figure 61 and Figure 62 show boxplots relating crossfall to curvature. See section 9.5 for and explanation of box plots. The curvature categories are listed in Table 60.

Figure 61

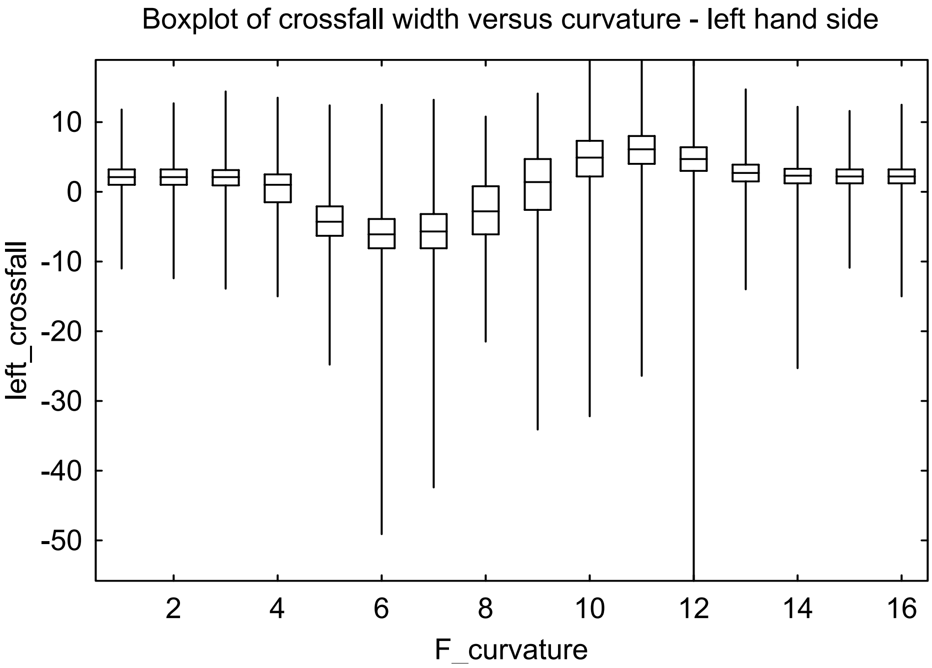


Figure 62

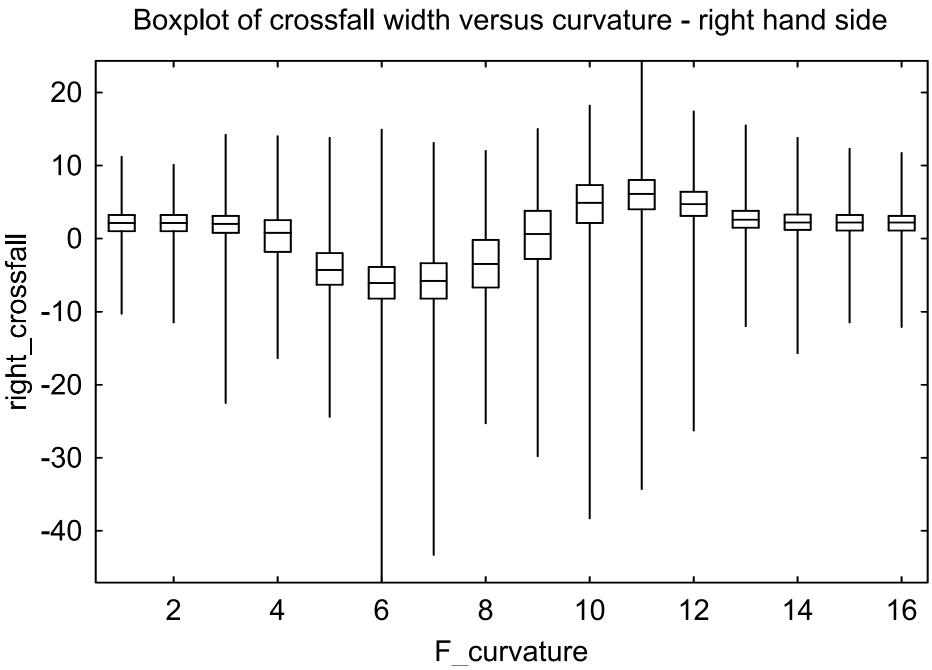


Table 60: Grouping of curvature values

Curvature group (F_curvature)	Curvature range
1	< -30,000
2	-30,000 to -10,000
3	-10,000 to -3,000
4	-3,000 to -1,000
5	-1,000 to -300
6	-300 to -100
7	-100 to -30
8	-30 to 0
9	0 to 30
10	30 to 100
11	100 to 300
12	300 to 1,000
13	1,000 to 3,000
14	3,000 to 10,000
15	10,000 to 30,000
16	> 30,000

These graphs confirm that crossfall should be of the same sign as curvature for moderate curves and is slightly positive for near straight roads. It also shows that there is less crossfall for very sharp curves and that in some sections of road, crossfall differs substantially from the usual values.

Further analyses of the effect of crossfall perhaps should divide the curvatures into three or four categories and then try to model the effect of crossfall on crash-rate in each of these categories, after adjusting for the sign of the curvature.

10 Postscript

These are some additional comments added in 2006.

Gradient: The apparent problems noted with gradient are probably just due to accumulation of random error and are not due to a problem with the equipment.

Skid-site: Skid-site 3 should include a gradient effect. So in this analysis, skid-site 3 should be been separated into sites where there is a junction and sites where there is a gradient effect. The failure to do this probably accounts for the peculiar effects concerning gradient found in this analysis and in some of the preliminary analyses that are not reported here. The overall predictive power of the model should not have been greatly affected by this, but interpretation of the skid-site and gradient effects will have been confused.