Skid resistance and road crashes

1 Introduction

We wish to relate the probability of a road crash to skid resistance. Two approaches are as follows:

1. Divide the road network into, say, 200m sections. Then relate the number of road crashes in each section over a specified period to the road geometry, amount of traffic and road surface characteristics including skid resistance. Most likely, the statistical analysis method will be Poisson regression. An example of this was the study by Wanty et al (1995). At that time, skid resistance was not available and so was not included in the model.
2. Look at two periods of time, separated by, say 5 years, and compare the road surface characteristics at the two periods at crash sites. An example of this type of study is Opus Central Laboratories Report No. 00-529283.00, *A skid resistance crash database to evaluate the effectiveness of the T/10 specification*.

In this report our analysis is of the second type and is similar but not identical to the analysis given in appendix 4 of Report No. 00-529283.00. This type of analysis requires less data and is probably more robust that the first type of analysis. However, the results that can be obtained are less detailed than those that one might potentially get from the first kind of analysis.

2 The data

We have the reported location of each reported injury crash for 1995, 1999 and 2000. This data also includes information about the crash including whether the road was wet. The movement data has been used to decide whether the crash was likely to be skid related. We have the texture, skid resistance, skid site category for each 20 metre section of road for each of the three years.

Our analysis is going to do two sets of analyses. The first compares the year 2000 data with 1995 and the second the year 1999 data with the 1995 data.

For each pair of years we are going to use the crash data that satisfies the following conditions:

- it was a reported injury crash and was on the State Highway network;
- the road surface was identified as wet and the movement and cause codes suggested that it might have involved a skid;
- the data suggests that the road surface was not renewed during either of the years (we cannot fully check this);
- skid and other data used in the analysis was available for both years for the site;
- the skid site category is the same for both years;
- skid resistances were greater than 0.12 at either sites (less than 0.12 may indicate a missing observation).

3 Theory

Suppose the number of crashes at site *i* in year *j* is denoted by $Y_{i,j}$. (For example *j* = 1 for 1995 and 2 for 2000). Suppose $Y_{i,j}$ has a Poisson distribution (independent for different *i* and *j*) with expectation

$$\mu_{i,j} = \exp(c + a_i + b_j + \lambda x_{i,j}) \tag{1}$$

where

c is a constant
a_i is a site effect (this includes the AADT effect)
b_j is a year effect – common to all sites
$x_{i,j}$ is the skid resistance (measured)
λ relates the crash rate to skid resistance

The parameters c, a_i, b_j and λ are unknown parameters of the model and $x_{i,j}$ is the skip resistance which has been measured. The parameter, λ, is the crucial one that we want to estimate since it shows the effect of skid resistance on the crash rate. Because we are taking the *exp* of a sum, the individual effects are supposed to be multiplicative.

Then statistical theory says that the conditional distribution of $Y_{i,2}$ given $Y_{i,1} + Y_{i,2}$ (number of crashes at site *i* in the second year given the total number at site *i* for the two years) is binomial with probability value

$$p_i = \frac{\mu_2}{\mu_1 + \mu_2} = \frac{\exp(c + a_i + b_2 + \lambda x_{i,2})}{\exp(c + a_i + b_1 + \lambda x_{i,1}) + \exp(c + a_i + b_2 + \lambda x_{i,2})}$$

$$= \frac{\exp\{b_2 - b_1 + \lambda(x_{i,2} - x_{i,1})\}}{1 + \exp\{b_2 - b_1 + \lambda(x_{i,2} - x_{i,1})\}} \tag{2}$$

and with sample size $n_i = Y_{i,1} + Y_{i,2}$.

This is just the binomial logistic regression model where we are trying to fit p_i using the difference of the skid resistances $x_{i,2} - x_{i,1}$. So we can use a logistic regression program to estimate the parameter λ which tells us the dependence of the crash rate on the skid resistance. Terms where n_i is zero provide no information so we are interested only in sites where crashes have occurred. The a_i parameters have cancelled, so we don't need, for example, the AADT at each site.

A critical assumption is the presence of a year effect and site effect in equation (1) but no interaction term apart from that introduced by the skid resistance deficit. In other words as we move from year 1 to year 2 the crash rate must change by the same factor apart from the effect introduced by the skid resistance.

If the AADT changed by different amounts at different sites then this would have to be introduced into the model. Different weather conditions in the two years at the different sites could also cause a problem. To partly allow for these effects it is worth allowing for a regional effect in equation (2).

We can simplify the model a little further. Imagine the sites are made very small (of length dx using the calculus notation) so that only 0 or 1 crashes occur at any site. Then the binomial model always has $n = 1$ (since the $n = 0$ sites don't come into the model). To apply the model to the real data, all we need to do is find the skid resistance at each crash site for *both* of the years being considered. Calculate the difference of the skip resistances and use this to predict Y using the binomial logistic regression model. Here, $Y = 1$ if the crash was in the second year and $Y = 0$ if it was in the first year. We can include other variables in the model if appropriate.

4 The analyses

4.1 Comparing years 1995 and 2000 data

The set of variables being fitted are:

- region – the regions are the seven TNZ database regions and the Waikato PMSC region;
- the year 2000 skid resistance minus the year 1995 skid resistance – i.e. the change in skid resistance;
- the year 2000 texture minus the year 1995 texture – i.e. the change in texture
- the average skid resistance over the two years – this is included so that the next term makes sense
- the interaction between the change in skid resistance and the average skid resistance – to see if the effect of change of skid resistance is different when the average of the two values is high compared with when it is low.

We need estimates of the skid resistance at the accident sites. It is not appropriate to take the skid resistance just at the site because the site is not recorded very accurately. So three estimates were tried:

The median of the skid resistance found from all 20 metre sections whose centres were no more than 200 metres from the recorded accident site. Sections from both sides of the road were used since it is not clear on which side the skid occurred on and past experience has been that the side of the road is not always recorded correctly.

The median of the skid resistance from 20 metres sections no more than 100 metres from the recorded site.

The median of the skid resistance from 20 metres sections no more than 50 metres from the recorded site.

Since the lower values of skid resistance may be important the analyses were repeated with lower quartiles rather than medians.

The texture was calculated similarly but only medians were used.

Medians were also used for calculating the surface dates and skid site categories used for selecting accidents as described in section 2.

Here are the analysis of variance tables for the six situations as the terms are added to the model.

Analysis using road sections up to 200 metres distant and medians of skid resistance

	Df	Deviance	Resid. Df	Resid. Dev	Pr(Chi)
NULL			763	1044.928	
region	7	7.02718	756	1037.901	0.4260552
skid.diff	1	49.33913	755	988.561	0.0000000
tex.diff	1	1.30181	754	987.260	0.2538830
skid.av	1	0.36979	753	986.890	0.5431181
interaction	1	4.79455	752	982.095	0.0285500

Analysis using road sections up to 200 metres distant and lower quartiles of skid resistance

	Df	Deviance	Resid. Df	Resid. Dev	Pr(Chi)
NULL			763	1044.928	
region	7	7.02718	756	1037.901	0.4260552
skid.LQ.diff	1	56.94325	755	980.957	0.0000000
tex.diff	1	0.68296	754	980.274	0.4085693
skid.LQ.av	1	0.09810	753	980.176	0.7541245
interaction	1	3.31929	752	976.857	0.0684714

Analysis using road sections up to 100 metres distant and medians of skid resistance

	Df	Deviance	Resid. Df	Resid. Dev	Pr(Chi)
NULL			756	1036.441	
region	7	4.85517	749	1031.585	0.6776319
skid.diff	1	42.53988	748	989.045	0.0000000
tex.diff	1	2.79388	747	986.252	0.0946248
skid.av	1	0.17236	746	986.079	0.6780241
interaction	1	4.34159	745	981.738	0.0371922

Analysis using road sections up to 100 metres distant and lower quartiles of skid resistance

	Df	Deviance	Resid. Df	Resid. Dev	Pr(Chi)
NULL			756	1036.441	
region	7	4.85517	749	1031.585	0.6776319
skid.LQ.diff	1	46.99907	748	984.586	0.0000000
tex.diff	1	1.67031	747	982.916	0.1962168
skid.LQ.av	1	0.10056	746	982.815	0.7511609
interaction	1	1.74340	745	981.072	0.1867088

Analysis using road sections up to 50 metres distant and medians of skid resistance

	Df	Deviance	Resid. Df	Resid. Dev	Pr(Chi)
NULL			701	959.4529	
region	7	9.79476	694	949.6582	0.2005058
skid.diff	1	34.97814	693	914.6800	0.0000000

```
    tex.diff  1  3.10418       692   911.5759 0.0780916
     skid.av  1  0.43031       691   911.1456 0.5118369
 interaction  1  4.82473       690   906.3208 0.0280542
```

Analysis using road sections up to 50 metres distant and lower quartiles of skid resistance

```
              Df Deviance Resid. Df Resid. Dev   Pr(Chi)
        NULL                    701   959.4529
      region  7  9.79476       694   949.6582 0.2005058
skid.LQ.diff  1 32.91008       693   916.7481 0.0000000
    tex.diff  1  2.04646       692   914.7017 0.1525610
  skid.LQ.av  1  0.10605       691   914.5956 0.7446879
 interaction  1  1.58129       690   913.0143 0.2085752
```

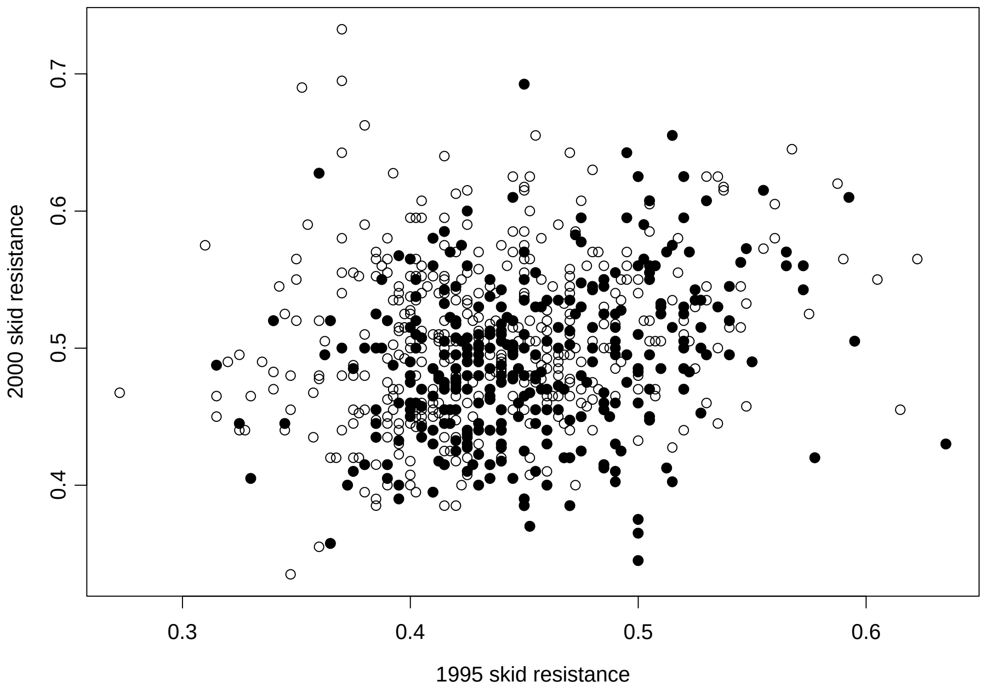
Interaction means the interaction between the change in skid resistance (*skid.diff* or *skid.LQ.diff*) and the average of the skid resistance (*skid.av* or *skid.LQ.av*) – i.e. the extent to which the *skid.diff* effect changes as *skid.av* changes.

In all of the analyses, the *change in skid resistance* is very highly statistically significant ($\mathrm{Pr(Chi)}$ is very small), and *region*, *change in texture* and *average skid resistance* are not statistically significant ($\mathrm{Pr(Chi)} > 0.05$). The interaction term is significant in some of the analyses.

The statistical significances should be worked out when the order of the variables being fitted is changed so that the variable is question is fitted last. However, doing this does not change the overall results.

Based on the change in the residual deviance the model using lower quartiles and the 200 metre maximum displacement fits best.

The following graph shows the values of the median of skid resistance for the year 2000 versus the value for the year 1995 with the 200 metre maximum displacement.



The filled points show the points where the accident was in the year 2000. If skid resistance is important the filled points should tend to be towards the lower right corner – high skid resistance in 1995, low skid resistance in 2000.

For estimating the size of the effects I included only *region* and *skid resistance difference*. Although the interaction term is just statistically significant in some of the analysis it is difficult to interpret estimates when interactions are present so I have omitted this term. I have kept *region* because it does not interfere with the analysis and provided some protection against trends in AADT and driver behaviour being different in different regions. Here are the estimates of the effect of change of skid resistance for the six versions of the models.

Analysis using road sections up to 200 metres distant and medians of skid resistance

	Value	Std. Error	t value
skid.diff	-8.00	1.21	-6.61

Analysis using road sections up to 200 metres distant and lower quartiles of skid resistance

	Value	Std. Error	t value
skid.LQ.diff	-8.43	1.20	-7.03

Analysis using road sections up to 100 metres distant and medians of skid resistance

	Value	Std. Error	t value
skid.diff	-7.24	1.17	-6.17

Analysis using road sections up to 100 metres distant and lower quartiles of skid resistance

	Value	Std. Error	t value
skid.LQ.diff	-7.32	1.13	-6.47

Analysis using road sections up to 50 metres distant and medians of skid resistance

```
                 Value   Std. Error    t value
   skid.diff     -6.54     1.16        -5.63
```

Analysis using road sections up to 50 metres distant and lower quartiles of skid resistance

```
                 Value   Std. Error    t value
skid.LQ.diff     -6.17     1.12        -5.48
```

A t-value with absolute value greater than 2 is statistically significant at about the 5% level. Looking the at second analysis we get an approximate 95% confidence interval of –10.8 to –6.0 for the value of λ in equation (1). That is a reduction in the skid related accident rate by a factor between (1.8 and 2.9) for each 0.1 increase in skid resistance. Of course we can't say that this holds for the whole range of skid resistances.

The interaction term which was included in the earlier analyses came out positive suggesting that the skid resistance effect is stronger for the lower average values of skid resistance.

We could repeat the analyses for differences in skid resistance deficiency. However, we would get close to the same answers since we are looking at sites at which the skid site category doesn't change so that the difference in skid resistance deficiency is equal to the difference in skid resistance. However, it might be thought that only negative and perhaps small positive values of skid resistance deficiency would be important. So I repeated the analysis with the values of skid resistance deficiency truncated above. I was unable to get a noticeable improvement in fit doing this suggesting that skip resistance is important even when the skid resistance deficiency is positive (i.e. the skid resistance is up to specifications).

4.2 Comparing years 1995 and 1999 data

I repeated the analysis with the 1995 and 1999 data. Here the skid resistance effect was much weaker although still statistically significant. There was also a region effect. The strongest skid resistance effect was for the 100 metre maximum displacement and the medians. Here is the analysis of variance table.

```
                  Df Deviance Resid. Df Resid. Dev   Pr(Chi)
           NULL                     607   840.4904
         region  7 17.66779         600   822.8226 0.0135623
      skid.diff  1  9.79662         599   813.0260 0.0017483
       tex.diff  1  0.73125         598   812.2948 0.3924799
        skid.av  1  1.53563         597   810.7591 0.2152691
    interaction  1  0.84480         596   809.9143 0.3580281
```

Fitting just the region and change in skid resistance we get the following estimates

```
                 Value Std. Error    t value
  skid.diff      -3.45    1.12       -3.09
```

so the 95% confidence interval for λ is –5.7 to –1.2. This corresponds to a reduction in skid related accidents by a factor of between 1.13 and 1.8 per 0.1 increase in skid resistance.

4.3 Comparing years 1995 and 1998 data

The analysis in report 00-529283.00 compared the 1995 and 1998 data. It found an estimate of λ for equation (1) of -3.7 ± 1.8. This corresponds to a multiplicative decrease in the skid accident rate of 1.4 (confidence interval 1.2 to 1.7) for an increase in skid resistance of 0.1.

5 Conclusion

We found a strong effect of change in skid resistance when comparing the 1995 and 2000 year data and a somewhat weaker effect when comparing the 1995 and 1999 data. The 1995 to 1999 effect is similar to that found when comparing 1995 and 1998 in the previous study.

6 References

Wanty, D., McLarin, M.W., Davies, R.B. & Cenek, P.D. (1995). *Application of the Road Geometry Data Acquisition System (RGDAS).* 7th World Conference on Transport Research, Sydney, Australia, 16th-21st July (unpublished).