USE AND LIMITATIONS OF CRASH DATA IN DETERMINING THE PRIORITY FOR TREATING SITES WITH LOW SKID RESISTANCE

H. VINER, F. COYLE & S. BRITTAIN
TRL Ltd, Wokingham, United Kingdom
HVINER@TRL.CO.UK
L. CAUDWELL
Highways Agency, United Kingdom
Louise.caudwell@highways.gsi.go.uk

ABSTRACT

In assessing the treatment of sites with low skid resistance, the relative priority of lengths marginally below a skid resistance threshold where there have been a number of recent crashes must be balanced with that of lengths substantially below the threshold with no recent crash history. The judgement is complicated by the highway authority having a duty of care to maintain the road in a safe and serviceable condition, so it is not acceptable to let the skid resistance deteriorate indefinitely, even if there have been no crashes. Furthermore, the analysis lengths are generally short and the number of crashes is generally small, leading to a high degree of uncertainty in estimating the underlying risk. And finally, as there can be a large number of sites that require investigation within any particular jurisdiction, there is a need for a simple method otherwise the whole process of site investigation demands a disproportionate level of staff resource.

This paper describes an accident risk model that has been created to provide a simple and consistent method for rating the priority of treatments at locations with low skid resistance and the dilemma for determining how to incorporate crash data within decisions on treatment priority.

1. INTRODUCTION

Previous work has shown that crash history at different locations can be extremely variable, even when the locations are grouped into similar types and have similar skid resistance [1,2]. Figure 1 shows that the variation in risk for sites with the same site category (in this case, UK single carriageway trunk roads with no events such as bends or junctions) is only partly explained by the variation in skid resistance. Specifying any intervention level for treatment will inevitably exclude a number of high-risk sites just above the threshold, as well as including a number of low, or zero-risk[*] sites below the threshold.

For this reason, thresholds for low skid resistance in the UK are treated as Investigatory Levels, rather than intervention levels, and locations with low skid resistance are subject to a detailed investigation to assess the surface condition and accident risk at each site before determining the priority for resurfacing. The objective of the investigation is to attempt to target treatments at the sites most likely to deliver a safety benefit, while monitoring the condition of sites where there is currently little risk. However, this strategy leads to two difficulties, firstly in the level of staff resource needed to thoroughly investigate all the sites below the Investigatory Level (IL) each year and, secondly, with the accuracy with which the level of risk can be determined.

---

[*] Zero-risk as defined by the recent observed accident history

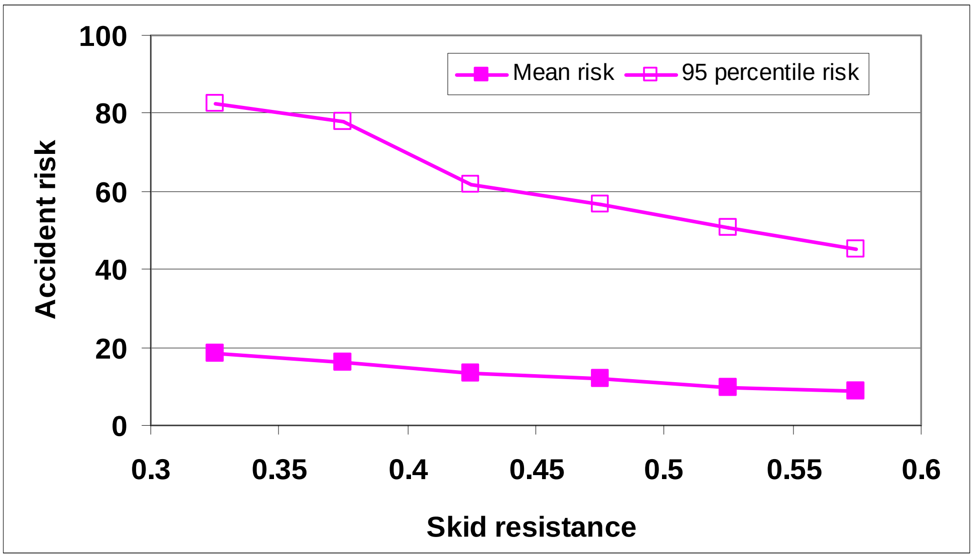

Figure 1 – Mean and 95 percentile accident risk for single carriageways non-event lengths, reproduced from [1]

This paper discusses the effect of statistical uncertainty of crash data in section 2 and, in section 3, describes the derivation of an accident risk model that has been created to provide a consistent method for rating the loss of skid resistance, history of crashes and the nature of a site during the investigation of locations with low skid resistance. This work was supported by the Highways Agency to assist highway engineers in implementing the UK skid resistance standard for trunk roads [3].

2. EFFECT OF STATISTICAL UNCERTAINTY IN CRASH DATA

All measured data are subject to error and understanding the extent and implications of these errors is important for reaching sound conclusions. The distribution of probabilities for accident data has been traditionally modelled using the Poisson distribution [4]. This distribution is suitable for describing the probability of random, independent events that occur within a unit of time or space and where the number of events that occur is theoretically unlimited. The probability distribution is given by [5]:

$$P(x) = e^{-\mu} \mu^{x} / x! \qquad \text{Equation 1}$$

Where P(x) is the probability of x number of occurrences within an interval of time or space and μ is the average (or expected) number of occurrences in the specified interval.

Using this knowledge, it is possible to calculate the uncertainty around the number of crashes reported. From the authors' experience of site investigations on UK trunk roads, the majority of sites are below 500m in length. An analysis of 3 years of personal injury accident data for the low skid resistance sites in one Area showed that the average number of accidents per year is zero for 40% of sites, one or less for a further 40% of sites and is very unlikely to be greater than 4.

Over the 3 year period typically used for analysis, this translates to typical accident numbers between 0 and 3, occasionally rising to double figures. Although, in this analysis, the numbers of accidents will also be influenced by the short and unequal length of these sites, it is of interest to consider the confidence limits that are typically associated with accident numbers of this order. The lower and upper bounds for the 95% confidence interval derived from the Poisson probability distribution, as a function of the mean level of accidents is given in Table 1. This confidence interval represents the range of values for which there is a 95% confidence that the average crash count for that site lies within, for the specified count period.

Table 1 – Confidence limits calculated from Poisson distribution

Number of crashes reported	95% confidence limit	
	Lower limit	Upper limit
0	0	3.7
1	0	5.6
2	0.2	7.2
3	0.6	8.8
4	1.1	10.2
5	1.6	11.7
6	2.2	13.1
7	2.8	14.4
8	3.5	15.8
9	4.1	17.1
10	4.8	18.4

It can be seen that there is a large uncertainty with regards to determining the underlying accident trends on short sites where the number of accidents is low. This leads to there being little or no significance attached to differences in the crash counts reported for different sites. For example, if site A had no reported crashes over a 3 year period and site B had 3 crashes in the same period, it would initially be thought that site B has a higher risk of crashes. However, examining the 95% confidence intervals shows that the average number of crashes for site A is likely to lie between 0 and 3.7, and for site B it is likely to lie between 0.6 and 8.8. This significant overlap of the confidence intervals means that there is a reasonable probability that site A could have a higher crash risk than site B.

3. DEVELOPMENT OF TRL ACCIDENT MODEL

The two previous sections have demonstrated that neither skid resistance data nor the number of accidents observed on a short site with low skid resistance, are entirely reliable as a means of judging the safety benefits of treatment to restore the surface treatment for individual sites. Nevertheless, there is a demonstrable link between skid resistance and accident occurrence for larger datasets, even though this relationship explains only a fairly low proportion of the overall variation in accident rates observed. Therefore, a means of balancing these priorities is needed and, given the large number of sites that typically require investigation in any one year, the method needs to be straightforward to implement by highway engineers.

For this purpose, an accident model has been designed that uses information about the loss of skid resistance, history of crashes and the nature of a site to produce a rating, based on an estimate of the number of accidents that can potentially be saved by a treatment to improve the skid resistance.

3.1. Overview of the accident model

The concept behind the model is to combine the various sources of information about a particular site as illustrated in Figure 2:

- It uses an algorithm to predict the number of future accidents at the site location, based on the number of past accidents observed.
- It estimates the reduction in the number of future accidents that would be achieved by improving the skid resistance.
- It multiplies the estimated saving in accident numbers by the average accident cost and returns a rating for the predicted cost saving at the site.

The following sections describe the derivation of each of these components of the model and explain the modifications made as a result of sensitivity tests.

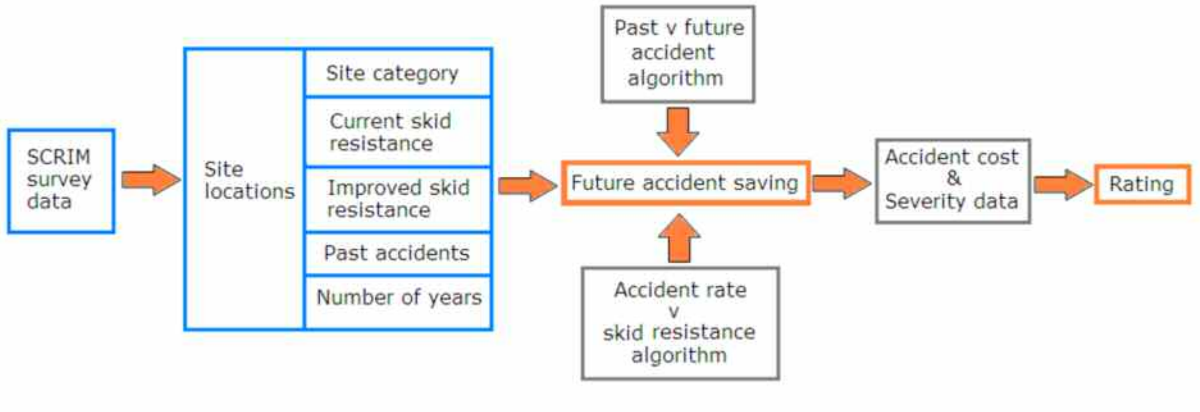

Figure 2 – Accident model flow chart

3.2. Prediction of future accident risk

It has been seen that the accident risk for different individual sites can be extremely variable (Figure 1). An analysis was carried out to assess whether the occurrence of a particular level of "past" accidents is a good guide to the number of "future" accidents that could be expected at a particular site or, alternatively, whether the behaviour is more random in nature. Random behaviour would be expected to give rise to the phenomenon of regression to the mean [6]. This a known statistical phenomenon whereby, if an individual case is chosen from a population because it is an outlier to a distribution (e.g. choosing a site with an unusually large number of accidents from a pool of equivalent sites) then (even with no intervention taken) that case is likely to appear closer to the mean of the distribution the next time it is sampled if the variation is a result of random rather than systematic influences.

For this analysis, personal injury accident data was obtained from the English trunk road network over a six year period from the start of 1999 to the end of 2006. The network was divided into continuous lengths with consistent site categories, as defined by the UK skid resistance standard [3]. The lengths used were selected to be typical of the lengths encountered during site investigations, and are shown in Table 2. There is a complication in that accident records are not positively associated with a particular direction of travel so, where the site category was different for the two sides of a single carriageway, the site category associated with this length was assigned in the order of preference: category K > Q > other event category > non-event category.

The accident data for each analysis length were summed within two periods: "past" (1999 – 2002) and "future" (2003 – 2006). These lengths were then analysed by the number of past and future accidents observed, as shown in Table 3 for category A lengths (mainline motorway). The objective of the analysis was to determine the likelihood of future accidents given the past history, for each site category and so Table 4 shows the category A data presented as a percentage of the total number of site locations in each row. Each row thus shows the distribution of future accidents for a specified number of past accidents. Changes in the distribution of future accidents can be seen as the number of past accidents increases, by comparing the rows. The maximum of each distribution is shaded in Table 3, indicating the most common number of future accidents given the number of past accidents observed.

Table 2 – Summary of analysis lengths

Site category and definition			Analysis length (m)
Non-event sites	A	Motorway	500
	B	Dual carriageway non-event	200
	C	Single carriageway non-event	
Event sites	Q	Approaches to and across major and minor junctions, approaches to roundabouts	As defined in Highways Agency Pavement Management System subject to maximum 200m and minimum 5m
	K	Approaches to pedestrian crossings and other high risk areas	
	R	Roundabout	
	G1, G2	Gradient	
	S1, S2	Bend <500m	

Table 3 – Distribution of Category A analysis lengths by number of accidents observed

N. analysis lengths	Future Accidents							
Past Accidents	0	1	2	3	4	5	>5	Total
0	1001	610	290	128	47	22	19	2117
1	584	567	329	190	72	45	46	1833
2	278	347	261	166	84	45	56	1237
3	139	197	166	130	79	54	62	827
4	47	94	121	89	61	41	83	536
5	22	41	47	43	41	34	75	303
>5	17	57	72	72	80	66	279	643
Total	2088	1913	1286	818	464	307	620	7496

Table 4 – Distribution of Category A analysis lengths by accidents observed

% of analysis lengths	Future Accidents							
Past Accidents	0	1	2	3	4	5	>5	Total
0	47.3%	28.8%	13.7%	6.1%	2.2%	1.0%	0.9%	100%
1	31.9%	30.9%	18.0%	10.4%	3.9%	2.5%	2.5%	100%
2	22.5%	28.0%	21.1%	13.4%	6.8%	3.6%	4.5%	100%
3	16.8%	23.8%	20.1%	15.7%	9.6%	6.5%	7.5%	100%
4	8.8%	17.5%	22.6%	16.6%	11.4%	7.7%	15.5%	100%
5	7.3%	13.5%	15.5%	14.2%	13.5%	11.2%	24.8%	100%
>5	2.6%	8.9%	11.2%	11.2%	12.4%	10.2%	43.4%	100%

The following observations can be made:

- Irrespective of the number of accidents observed in the past period, anywhere between 0 and 5+ accidents can be observed for an individual length in the future period.
- For lengths with zero accidents in the past period, it is more likely than for other lengths that zero accidents will be observed in the future period. It is also unlikely that a large number of accidents will be observed where there were zero previous accidents. However, for approximately 50% of these sites, at least one accident is observed in the future period.
- As the number of accidents in the past period increases, there is a progressively increasing likelihood that a larger number of accidents will be observed in the future period. It also becomes increasingly unlikely that zero or only one accident will be observed in the future.
- The maximum in the distribution of future accidents generally occurs at a lower number of accidents than was observed in the past period, which is consistent with the principle of regression to the mean.

It therefore appears that the number of accidents in the past period provides a guide to the future accident risk, but that the future trend is likely to be less extreme than might be suggested by the past observations. This is shown in Table 5: for lengths with zero or one past accident, the most likely situation is that more accidents will observed in the future period than in the past whereas, for lengths with two or more past accidents, fewer accidents are likely to be observed in the future period. Again, the most likely situation for each case has been highlighted to emphasise this trend.

Table 5 – Summary of future accidents for Category A site locations

% analysis lengths	Future Accidents		
Past Accidents	Less than past accidents	Same as past accidents	More than past accidents
0	-	47.3%	52.7%
1	31.9%	30.9%	37.3%
2	50.5%	21.1%	28.3%
3	60.7%	15.7%	23.6%
4	65.5%	11.4%	23.2%
5	64.0%	11.2%	24.8%
>5	56.5%	43.4%	-

To enable a general relationship to be developed between past and future accidents, the numbers of accidents were divided to give an accident count per year. As the site locations in categories A, B and C were of a set length, the number of accidents per year per metre could be found. This is useful since it enables site locations of different lengths to be handled with the reasonable assumption that the exposure is proportional to the site length. The regression from the graphs of future vs. past accidents per year, per metre (Figure 3), provides an algorithm to predict the number of future accidents at a location when the number of past accidents is known.

For the other site categories such as junctions and bends, which are generally shorter and more variable in length, the count is the number of accidents per year at the entire location, therefore Figure 4 and Figure 5 show the number of future accidents per site plotted against the past accidents for categories K and Q, and G, R, S1 and S2 (K and Q are shown separately for clarity).

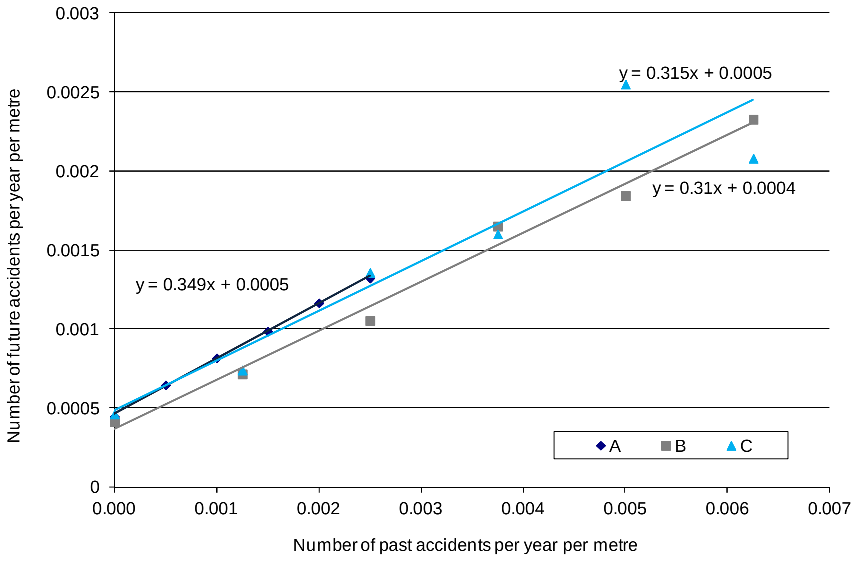

Figure 3 – Future vs. past accidents for site categories A, B and C

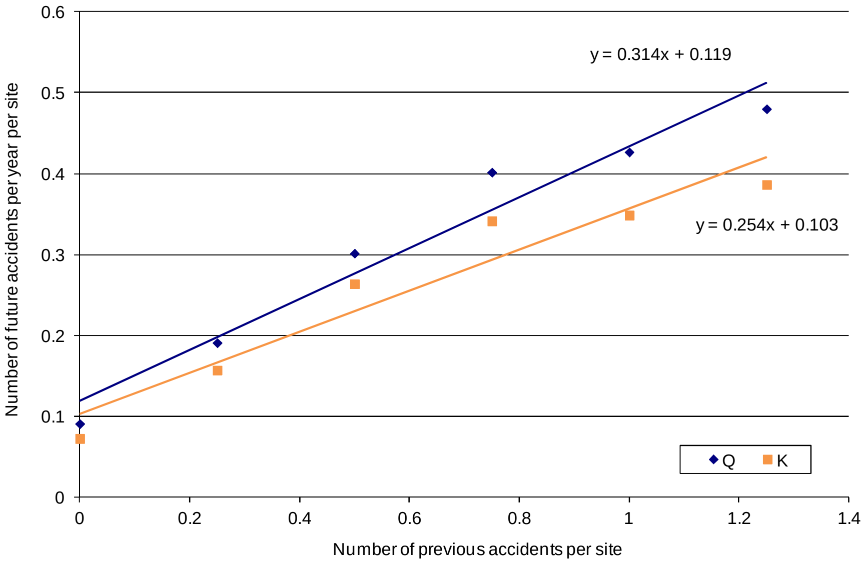

Figure 4 – Future vs. past accidents for site categories K and Q

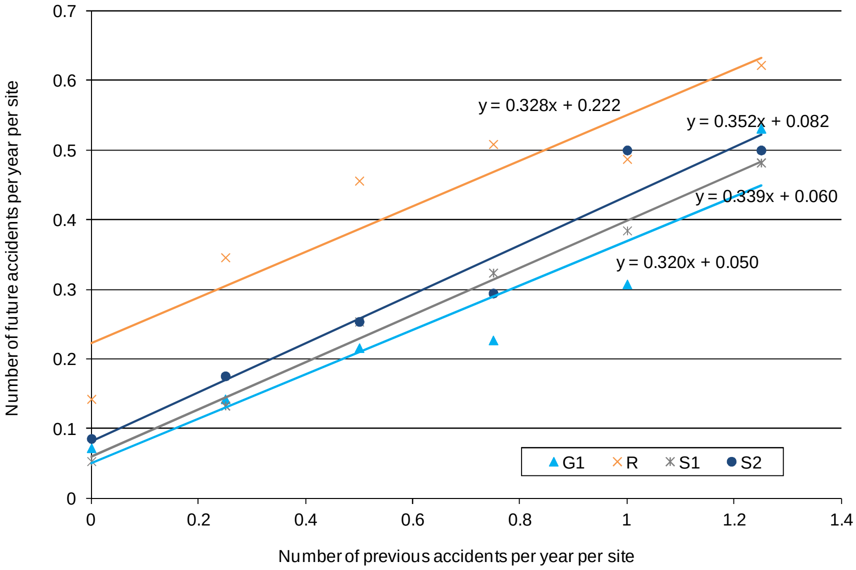

Figure 5 – Future vs. past accidents for site categories G, R, S1 and S2

The algorithm derived for each site category from linear regression is shown on the corresponding graph. Although the data for several of the categories suggest a non-linear relationship, the relatively small amount of data associated with the outermost data points in each case was felt not to justify the use of a more complex relationship.

3.3. Reduction of risk as a result of improving skid resistance

As well as predicting the future likelihood of accidents, the model also needs to estimate the benefit that will be derived from improving the skid resistance. This depends on the type of site; for example, the accident risk on single carriageway non-event sites appears to be more sensitive to changes of skid resistance than for mainline motorways. Therefore, the relationship between skid resistance and accident rate for each site category is used to calculate the percentage difference between the accident rates for the current and improved levels of skid resistance. This percentage difference is then applied to the predicted number of future accidents to find the possible accident saving.

Previous analysis of accident data and skid resistance from trunk roads in England [2], found linear relationships between the mean accident rate (the number of accidents per 10^8 vehicle kilometres) and skid resistance, for each of the site categories. However, analysis on accident data for three county councils found a power relationship between accident rate and skid resistance for a number of the site categories [7,8]. A comparison showed that both types of relationship provide a good description of the data over the central range of skid resistance where there is a large volume of data. However, they produce significantly different predictions when extrapolating to progressively lower skid resistance, which is a region where there is less data and therefore less confidence in the shape of the trend.

There are two aspects of the power trend that have desirable features for the accident model: firstly, it predicts an extreme worsening of accident risk at very low skid resistance. For the very small number of cases where the skid resistance is that low, it is important that a high priority is assigned to it, which is achieved with the power model. Secondly, the power model predicts that a limited benefit is gained by continuing to increase the skid resistance to very high levels. For practical purposes it is not possible to achieve extremely high skid resistance over a very large proportion of the network and it is therefore desirable that the model will reflect this.

Within the middle of the skid resistance range, which is where the majority of cases will lie, both the power and the linear model provide a good description of the data. It was therefore decided to use the power regression.

Examination of the power relationships for each category showed some anomalies. For example, site category Q (approach to junction) had a more gradual increase in accident rate with decrease in skid resistance than site category C (single carriageway non-event). For some site categories, insufficient data were available to define a reliable relationship. It was therefore decided to group categories together due to similarities in the trends and the lack of available data: site categories A and B were grouped, site category C was left on its own and all the event site categories were grouped together.

Figure 6 shows the data trends established for these new category groups, AB, C and event. (Following standard UK practice, skid resistance is expressed here as SCRIM coefficients after application of the Index of SFC [3].) These relationships have been implemented in the model to estimate the average reduction in accident risk that could be expected as a result of improving the skid resistance from its current level to a new, higher level.

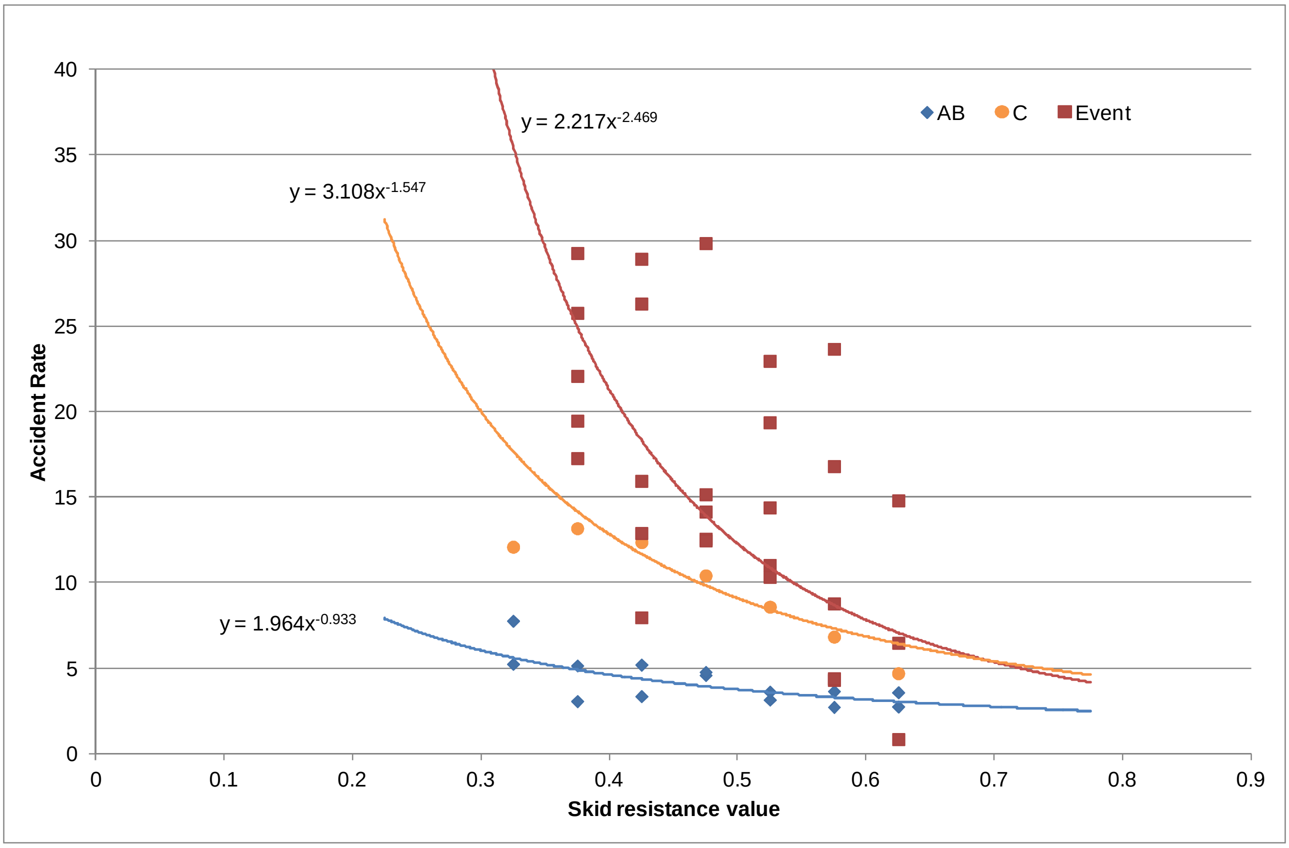

Figure 6 – Accident rate models for site categories AB, C and Event

3.4. Accident severity and cost

The previous sections have given details of the prediction of “future accident saving”, based on the likelihood of future accidents, and the reduction expected as a result of improving the skid resistance. Finally, this value is converted to a rating, based on the economic value of the accidents saved.

If the loss of skid resistance were to result in an overall increase in accident severity then this clearly needs to be incorporated in the accident model. However, a study on accident data from county councils found a lack of sensitivity between accident severity and skid resistance [8]. This is in keeping with the understanding that there are many factors that influence the severity of an accident, for example the pre-impact speed and orientation of the vehicles, with 60% of fatal accidents occurring in frontal collisions and a further 25% in side impact collisions [9]. The make and model of the vehicle can also affect the severity of the injuries sustained and the occupants can influence the outcome of the collision by their use or miss-use of the safety features provided within the vehicle (such as seatbelts and airbags).

The influence of skid resistance on severity was therefore disregarded. The proportion of fatal, serious and slight accidents occurring for each site category, together with published values for the economic value assigned for an accident of each severity [10] were used to calculate a rating by multiplying the predicted future accident saving by the average cost per accident, taking into account the average severity of crashes at that site category.

3.5. Sensitivity testing

Analysis of the ratings returned by the accident model found that lengths with very low current skid resistance and zero past accidents were being rated as low priority. The implication of this is that these lengths would not be prioritised, even where the skid resistance was very low, until a crash had occurred that resulted in injury. Although consistent with the data trends, it was felt that this was not consistent with the need for a highway authority to exercise duty of care in the maintenance of the highway. Therefore, various artificial means were explored in the model to increase the weighting given to sites with low skid resistance but no previous crashes.

The first weighting option tested simply added a portion of an accident to the past accident count if there had been zero past accidents. Adding 0.5 or 0.8 to the past accident count did give an improved priority to the sites with very low skid resistance and zero past accidents. However, this weighting reduces much of the difference between locations with zero and one past accidents. An alternative, of adding a decreasing weighting to all past accidents counts as follows was tested; zero accidents become 0.8, one accident becomes 1.6, two accidents become 2.4, three accidents become 3.2 and four or more accidents remain the value entered by the operator. This weighting method also prioritised the sites logically.

A different approach is to add on a number of crashes that depends on the difference between the current measured value and a nominal level of skid resistance for the site. If the deficit was ≥0.4 then 0.8 was added to the past accident count, if ≥0.35 then 0.7 was added, if ≥0.3 then 0.6 etc. (These values correspond to a three year accident count period; they are implemented in the model as the appropriate number of accidents per year.) Being based on the skid resistance, this approach more directly reflects the duty of care argument that it is necessary to maintain a road with poor skid resistance, irrespective of whether crashes have already resulted in injury. Therefore, this weighting method has been used in the model.

To test the sensitivity of the revised model, 132 hypothetical combinations of site category, skid resistance and accident history were rated. For each site category, the ideal skid resistance was set to 0.05 units above the investigatory level and the current skid resistance was varied from 0.2 up to the investigatory level in steps of 0.1. This was repeated for 0, 1, 2 and 3 past accidents over a 3 year count period. An extract from the tabulated output, sorted by the accident model rating is shown below as Table 6, which shows ratings from the middle of the prioritised list (positions 54 – 93).

It can be seen from this extract that sites with zero accidents but a significant deficit in their skid resistance can be given priority over sites with past accidents but a higher skid resistance. Therefore, weighting the past accident count according to the skid resistance deficit as discussed above is producing a satisfactory result.

Table 6 – Extract from prioritised list of sites

Site ID	Site category	Current skid resistance	Ideal skid resistance	Difference	Past Accidents	Rating
43	G	0.3	0.55	0.25	1	93.8
94	R	0.4	0.55	0.15	2	88.7
107	S1	0.3	0.55	0.25	1	86.4
113	S1	0.5	0.55	0.05	3	86.3
34	C	0.4	0.45	0.05	3	82.4
60	K	0.2	0.55	0.35	1	81.2
75	Q	0.3	0.55	0.25	1	79.3
62	K	0.4	0.55	0.15	2	78.5
81	Q	0.5	0.55	0.05	3	71.2
91	R	0.3	0.55	0.25	1	67.2
122	S2	0.4	0.55	0.15	1	66.0
20	B	0.3	0.45	0.15	2	60.4
125	S2	0.5	0.55	0.05	2	57.6
8	A	0.3	0.45	0.15	2	56.7
59	K	0.3	0.55	0.25	1	56.4
97	R	0.5	0.55	0.05	3	54.5
42	G	0.4	0.55	0.15	1	51.2
65	K	0.5	0.55	0.05	3	50.0
29	C	0.3	0.45	0.15	1	48.0
106	S1	0.4	0.55	0.15	1	47.4
18	B	0.2	0.45	0.25	1	46.6
45	G	0.5	0.55	0.05	2	46.3
74	Q	0.4	0.55	0.15	1	44.6
6	A	0.2	0.45	0.25	1	44.0
109	S1	0.5	0.55	0.05	2	42.3
120	S2	0.2	0.55	0.35	0	41.9
31	C	0.4	0.45	0.05	2	40.2
90	R	0.4	0.55	0.15	1	38.5
77	Q	0.5	0.55	0.05	2	36.8
22	B	0.4	0.45	0.05	3	34.9
10	A	0.4	0.45	0.05	3	32.6
58	K	0.4	0.55	0.15	1	31.8
40	G	0.2	0.55	0.35	0	30.7
72	Q	0.2	0.55	0.35	0	30.5
93	R	0.5	0.55	0.05	2	29.7
104	S1	0.2	0.55	0.35	0	28.9
88	R	0.2	0.55	0.35	0	28.7
61	K	0.5	0.55	0.05	2	26.0
56	K	0.2	0.55	0.35	0	22.0
119	S2	0.3	0.55	0.25	0	21.7

4. SUMMARY AND IMPLEMENTATION OF THE MODEL

While the overall trend between accident risk and skid resistance is significant, when assessing how to prioritise the treatment of individual sites with low skid resistance, both the skid resistance data and the past accident data have limitations in determining which sites will provide the greatest safety benefit. It has been seen that setting an intervention threshold for skid resistance will always exclude high risk sites that fall just above the threshold, whilst including low risk sites immediately below it. Likewise, the uncertainty associated with small accident counts on short lengths means that it is not possible to distinguish between many sites at the preferred 95% significance level.

Against this background, the TRL accident model has been developed to make use of both sets of information to provide a consistent way of assessing individual sites, and which is straightforward for engineers to apply. This assessment is considered to be robust in as much as it reflects the trends observed in large data sets. However, while the model output could be considered to represent the balance of probabilities, it should not be accorded a spurious degree of precision which does not represent the actual precision of assessing short, individual sites with low accident numbers.

For this reason, this approach being implemented within the UK standard is to use this method to carry out an initial sift of the large numbers of sites that require investigation because the skid resistance falls below the Investigatory Level. It will be used in a slightly modified form that takes into account the difference between the observed accident record and the “worst case” number of crashes represented by the upper confidence limit. On this basis, low risk sites will be excluded from requiring a detailed physical investigation on the grounds that it is unlikely that they represent a high priority for treatment. This has the benefit of quickly reducing the number of sites requiring investigation and giving some guidance to engineers about the relative priority of the different sites, while still requiring a full investigation of the higher priority sites that may uncover other features of the road condition and layout that need to be considered.

REFERENCES

1. Viner, H; Sinhal, R and Parry, T (2004) Review of UK skid resistance policy - Proceedings of the 5th Symposium on Pavement Surface Characteristics of Roads and Airports, Canada, June 2004
2. Parry, A and Viner, H (2005). Accidents and the skidding resistance standard for strategic roads in England (TRL622). Wokingham: TRL Limited.
3. Design Manual for Roads and Bridges HD28/04 – Skid resistance (DMRB 7.3.1). London: The Stationery Office.
4. Hauer, E. (1997). Observational before-after studies in road safety. Pergamon.
5. Fleming MC and Nellis JG (2000). Principles of applied statistics. Thomson Learning, London.
6. Wonnacott TH and Wonnacott RJ (1977). Introductory statistics, 3rd edition. John Wiley & Sons.
7. Kennedy, C. and Donbavand, J. (2008). Relationship between skid resistance and accidents on local roads in the south west of England. 2008 Safer Roads conference, Cheltenham, UK.
8. Donbavand, J. (2008). Personal communication.
9. Wood, D. P. (1995). Determination of speed from crush, the damage solution and its application in the energy-equivalent-speed model. Forensic accident investigation: motor vehicles. Bohan T. L., Damask A. C., Butterworth M., (ed). Law publishers, Charlottesville Virginia.
10. Department of Transport (2009) www.dft.gov.uk/excel/173025/221412/221549/227755/2856721/article2costdatatables.xls