The relationship between speed and accidents on rural single-carriageway roads

Prepared for Road Safety Division, Department for Transport, Local Government and the Regions

M C Taylor, A Baruya and J V Kennedy

TRL Report TRL511

First Published 2002
ISSN 0968-4107

This report has been produced by TRL Limited, under/as part of a contract placed by the Department for Transport, Local Government and the Regions. Any views expressed in it are not necessarily those of the Department.

CONTENTS

Executive Summary

Introduction

The Government's review of speed policy, published in March 2000, emphasised the need for a greater understanding of the role of speed in accidents on rural roads. A research programme at TRL over the last decade has demonstrated beyond doubt that the faster drivers choose to travel, the more likely they are to be involved in an accident, and that higher speeds on roads with otherwise similar characteristics are associated with more accidents. The programme included an EU-funded project known as MASTER, under which a speed-accident relationship (the EURO model) was derived for European rural single-carriageway roads.

The complexities involved in analyses of this kind, coupled with the limited data available in the MASTER project, meant that the effect of speed in the EURO model was particularly difficult to interpret. Only a limited amount of the data upon which the model was based was from the UK. Because of these limitations, the Road Safety Division of the Department for Transport, Local Government and the Regions commissioned TRL to carry out a more extensive investigation of the relationship between speed and accidents on rural single-carriageway roads in England. This report describes that study. It involved:

- site selection;
- the collection and analysis of data from a total of 174 road sections across the country, comprising injury accident data, traffic flow and vehicle speed data, and a wide range of details of road characteristics, geometry and layout;
- the application of statistical techniques to classify the road sections into relatively homogeneous groups in respect of their speed-accident characteristics;
- statistical modelling to relate accident frequency to other factors such as traffic flow, vehicle speed, and characteristics of the road itself.

The sites were all on roads with a 60 miles/h speed limit. The sample was stratified to cover all road classes and to provide a good geographical distribution, a wide range of flow levels, and degrees of hilliness, bendiness and junction/access frequency. A wide range of mean speeds (26 to 58 miles/h) and accident rates (0 to 271 per 100 million vehicle-kilometres) was observed.

Results

The homogeneous groups into which the road sections were successfully classified were defined by a set of 6 variables: accident rate, mean speed, minor junction density, bend density, access density (i.e. the density of private drives and other accesses joining the road) and hilliness. These together reflect the operational characteristics of the road, or 'road quality', and can be described as follows:

Group 1: Roads which are very hilly, with a high bend density and low traffic speed. These are *low quality* roads.

Group 2: Roads with a high access density, above average bend density and below average traffic speed. These are *lower than average quality* roads.

Group 3: Roads with a high junction density, but below average bend density and hilliness, and above average traffic speed. These are *higher than average quality* roads.

Group 4: Roads with a low density of bends, junctions and accesses and a high traffic speed. These are *high quality* roads.

The models developed relating accident frequency to other factors explained a high proportion of the variability in the data and the effects of the key variables were found to be strong, plausible and very stable.

The models show that:

- Accident frequency for all categories of accident increased rapidly with mean speed – the total injury accident frequency increased with speed to the power of approximately 2.5 – thus indicating that a 10% increase in mean speed results in a 26% increase in the frequency of all injury accidents.
- The relationship between accident frequency, traffic flow and link section length mirrored that typically found in other similar studies.
- Accident frequency varied between the Road Groups defined above. It was highest on the Group 1 roads, and about a half, a third and a quarter of this level on roads in Groups 2, 3 and 4 respectively.
- The frequency of total injury accidents was also found to increase rapidly with two further measures: these were the density of sharp bends (those with a chevron and/or bend warning sign) and the density of minor crossroad junctions. These increased accidents by 13% and 33% respectively, per additional bend/crossroad per kilometre. Single vehicle accidents were particularly strongly affected by the density of sharp bends (34% increase in accident frequency per additional sharp bend per kilometre.)
- The effect of mean speed was found to be particularly large for junction accidents; these accidents were roughly proportional to the 5th power of speed, suggesting substantial potential for accident reduction from strategies designed to reduce speeds at junctions.
- No other measures of speed were found to influence accident frequency as strongly as, or in addition to, mean speed.
- The percentage reduction in accident frequency per 1 mile/h reduction in mean speed implied by the relationship developed for total accidents depends on the mean speed. It ranges from 9% at a mean speed of 27 miles/h to 4% at a mean speed of 60 miles/h.

- The effect of speed on fatal and serious accidents was stronger (but not statistically significantly so) than for all accidents taken together. A 10% increase in mean speed would be expected to result in a 30% increase in the frequency of fatal and serious accidents.

Discussion

The models presented in this report differ from the previous EURO model in a number of ways. The present models are substantially more robust, being based on a more structured, extensive and relevant database. They predict a stronger effect of speed on accidents than did the EURO model. However, in terms of accident reduction potential, speed management policies applied to urban roads are still likely to provide the greatest benefits. This is because of the vastly greater number (and more concentrated distribution) of accidents occurring on those roads.

There is a lot more work to be done to develop the basis for speed management policies on rural single-carriageway roads. The issues to be addressed, raised comprehensively in the Government's review of speed policy, include:

- the need to define a rural road hierarchy according to road function;
- the need to establish what are appropriate speeds for the different types of roads in this hierarchy;
- the need to identify means of achieving these appropriate speeds;
- the need to define a policy for setting appropriate speed limits, taking account of the hierarchy and of the appropriate speeds to be achieved.

The present study has provided a basis from which to progress these issues. The classification of roads into groups reflecting road quality was fundamental to the study and this Road Group classification has the potential to contribute to defining a road hierarchy.

Conclusions

1 The study has achieved its objective of developing a speed-accident relationship for English rural single-carriageway roads which is straightforward to interpret and apply. The analytical process successfully overcame the difficulty inherent in this type of study of de-coupling the effects of inter-correlated variables.

2 The resulting predictive relationship for total injury accidents shows that accident frequency rises rapidly with the mean traffic speed on a given road, and quantifies this effect. The relationship can be used to estimate the change in accident frequency resulting from a change in mean speed on a given road and, if applied to local or national accident statistics, to estimate the effects of different speed management strategies.

3 The classification of roads into groups reflecting road quality, which underpinned the analysis, has the potential to contribute towards the development of a road hierarchy for rural single-carriageway roads.

1 Introduction

1.1 Background

The Government's review of speed policy (Department of the Environment, Transport and the Regions, 2000) emphasised the need for a greater understanding of the role of speed in accidents on rural roads. About 20% of all road casualties in Great Britain are on rural single-carriageway roads, which represents about two thirds of the casualties on all rural roads. For fatally and seriously injured casualties, the corresponding figures are about 30% and 75% respectively.

Research at TRL over the last decade has substantially increased our knowledge of the relationship between driver speed and road accidents. The programme has comprised:

- A review of 'before and after' studies, largely from abroad, of the effects on accidents of changing the posted speed limit (Finch *et al.,* 1994; Taylor *et al.,* 2000).
- Empirical road-based studies designed to establish relationships between the accident frequency on road sections and the traffic and pedestrian flow, vehicle speed and road geometry for those sections (Taylor *et al.,* 2000).
- Empirical driver-based studies designed to establish relationships between the accident involvement of individual drivers and their individual characteristics, particularly their typical speed behaviour in relation to other drivers (Quimby *et al.,* 1999; Taylor *et al.,* 2000).
- An analysis of the involvement of speed in fatal accidents (Taylor, 2001).

These studies have demonstrated beyond doubt that the faster drivers choose to travel, the more likely they are to be involved in an accident, and that higher speeds on roads with otherwise similar characteristics are associated with more accidents.

1.2 Road-based studies

The road-based studies comprised separate studies of urban roads and rural single-carriageway roads. The latter was undertaken under the MASTER project (MAnaging Speeds of Traffic on European Roads) (Baruya, 1998). The resulting 'EURO' model relating speed and accidents (Taylor *et al.,* 2000) was based on data from Sweden, the Netherlands and England; however, the quantity of English data used in the development of this model was rather limited and related mainly to A and B class roads in South East England.

The processes involved in the development of speed-accident models such as those in these road-based studies is far from straightforward. The complex inter-relationships between the variables means that extensive databases and sophisticated statistical analysis techniques are required. A particular problem is that the key variables sometimes interact in such a way that the effect of interest (here the association between speed and accidents) is masked by correlations between these variables and a third variable (the 'masking' variable). This issue is discussed again later in the report and is explained more fully in Taylor *et al.* (2000). In the urban road study, pedestrian flow was found to be such a masking variable and its identification and quantification in the speed-accident model enabled a clear interpretation of the model to be made with respect to the effect of speed. In the European rural road study, however, a corresponding masking variable was not found, probably due to limitations in the data, and the resulting speed-accident relationship proved to be particularly difficult to interpret.

1.3 The new study

Because of the limitations of the MASTER study, the Road Safety Division of the Department for Transport, Local Government and the Regions (DTLR) commissioned TRL to carry out a more extensive investigation of the relationship between speed and accidents on rural single-carriageway roads in England. This report describes that study. It involved:

- site selection;
- the collection of data from a total of 174 road sections across the country;
- the application of statistical techniques to classify the road sections into homogeneous groups in respect of their speed-accident characteristics;
- statistical modelling to relate accident frequency to other factors such as traffic flow, vehicle speed, and characteristics of the road itself.

Section 2 of this report details the site selection and data collection procedures used and Section 3 summarises the key descriptive statistics arising from the data collected. Section 4 describes the methodology used in the analysis. Section 5 describes the results of the process to classify the road sections into groups and Section 6 details the results of the statistical modelling. In Section 7 the practical implications of the resulting models are discussed. Section 8 summarises and discusses the overall findings and the key conclusions are given in Section 9.

2 Data collection

2.1 Introduction

An expanded database was built up from the data available under the MASTER project. Although only 38 UK sites were used in the development of the MASTER EURO model, the MASTER database included a number of additional sites with more limited data. This existing database was expanded by adding further sites, so as to include C and unclassified roads and to give a wider geographical spread. The data required for all sites for the study included:

- Accident data (for a defined 5 year period).
- Traffic flow data.
- Vehicle speed data.
- Road characteristics.
- Geometric and layout data.

Of the 174 sites used in the present study, 74 had previously been studied (including those in the original MASTER database) and are referred to as 'old' sites in what follows, and 100 were 'new' sites.

2.2 Site selection

The 'new' sites selected were to be between 1 km and 10km long, to be subject to the national 60 miles/h speed limit throughout their length, and to contain no major junctions (where vehicles have to give way to other traffic). The 'old' sites also satisfied these criteria.

The main criteria for selection of the final sample of new sites were location and road class. As far as possible, the sample was stratified to include about one-third each of A, B and C or unclassified roads. Sites were to be distributed fairly evenly across the DTLR regions; since however, the intention was to complement the existing MASTER sample, no sites were selected from the South East (which was already over-represented). Approximately equal numbers of sites that were straight/bendy, flat/hilly and had high/low numbers of minor junctions were sought, insofar as this was possible given the other aims of the site selection process.

2.3 Speed and flow data

Measurements of speeds and vehicle flows in both directions at the 'new' sites were made using automatic equipment at one location within each site, away from junctions, on a straight section of road (or at a representative point if the road was bendy). Speed/flow data were collected over two or more days, including at least one full 24-hour weekday period from midnight to midnight. Any incomplete days were excluded from the analysis.

Vehicle speeds were classified in 5 miles/h bins from 0 to 100 miles/h for the full 24 hours and also for separate time periods of the day. A range of variables was developed from this information, including:

- traffic flow;
- mean speed;
- 85th percentile speed;
- standard deviation of speed;
- coefficient of variation of speed (ratio of the standard deviation to the mean);
- percentage of vehicles exceeding the speed limit;
- mean excess speed (mean speed of those vehicles exceeding the limit).

Similar variables were derived from the data for the 'old' sites.

The 24 hour vehicle flows at the 'new' sites were scaled to annual average daily totals (AADTs) according to vehicle type and the day, month and year of measurement to give an average over the period for which accidents were considered (see Section 2.5 below). Those for the 'old' sites were scaled in the same way as far as possible.

2.4 Site characteristic and geometric data

2.4.1 Introduction

The overall purpose of the study was to identify factors which, in addition to traffic flow and speeds, are relevant in determining the level of accidents on different rural single-carriageway roads. As explained at the start of this report, it was likely that this would involve identifying a 'masking' factor which is strongly correlated with both speeds and accidents. For example, roads of lower 'quality' – narrow, winding country lanes - are likely to have lower speeds than roads of higher 'quality', but it is known that they have a high accident rate. Conversely, well-designed major roads which are wider and straighter, with fewer junctions and generous sight-lines, have relatively high speeds compared to other roads, but relatively low accident rates. A simple analysis which ignored road 'quality' would thus indicate that high speeds are associated with low numbers of accidents. However, both common sense and experience suggests that it is unlikely that, *on either of these types of road taken as a group,* accidents would actually decrease if speeds increased. What the present study is thus trying to ascertain is what happens to accidents when vehicle speeds increase or decrease, if *everything else (particularly the geometrics of the road and the traffic levels on it) remains constant*. The principle is discussed further in Section 4.1.

There are two purposes, then, in collecting data about road characteristics and geometry. The first is to enable a classification of the link sections into homogeneous groups to be made so that the true speed-accident relation within each can be established. A broad descriptor of this classification would be 'road quality' and we need to establish what physical features of link sections (perhaps bendiness, visibility, quality of road markings, and so on) best define road quality in the sense of providing the best differentiator between the classes. Ideally, the national road classification system of A, B, C and unclassified roads would provide this grouping, but B class roads, in particular, cover a wide range of functions and design quality so that these administrative classifications are unlikely to be suitable in the present context.

The second purpose of collecting road layout data is that having classified the link sections by road quality, layout variables may well contribute significantly in an accident predictive model.

It was therefore important to measure the main variables that seemed likely to affect accidents directly or indirectly through being determinants of road quality. A brief review of the available literature was undertaken to identify variables that had been used in similar studies in the past, notably Walmsley *et al.* (1998), who studied accidents on rural trunk roads, and Lee and Brocklebank (1993), who investigated mean speed on rural roads. Variables which would not necessarily be expected to influence accidents but which were easy to measure were also included. The aim was to assemble at a reasonable cost an extensive and reliable data set covering all types of rural single-carriageway road.

2.4.2 Drive-through video recordings

Drive-through video recordings were made for all sites for the purpose of establishing much of the site characteristic and geometric data. This technique allowed a range of descriptive variables to be readily obtained. The following information was extracted.

Discrete data

These variables included:

- type of junction at each end of the section (if any);
- number and type of minor junctions;
- number and type of accesses within the section;
- number of bends, classified as follows:
  - sharp (marked by chevrons and/or with a warning sign);
  - medium;
  - slight.

'Minor junctions' comprised marked T-junctions and crossroads within each section, including those with no-through roads. 'Accesses' included unmarked junctions, public accesses and private drives (entrance to a farm, factory, driveway, track, filling station or public house, etc.), and laybys.

Semi-continuous data

These variables included:

- lighting;
- reflecting road studs;
- kerbing;
- number of lanes;
- road markings;
- land use.

Lighting, reflecting road studs, kerbs and white edge lines were recorded as being present or absent along most of the link. This therefore refers to the condition prevailing over most of the site. The number of lanes was also taken to be that over most of the link length. Centre line variables were derived which simply indicated the presence somewhere on the link of double white lines, solid lines, broken lines and centre hatching, and the number of times the type of marking changed.

The overall percentages of each category of land use adjacent to the road were estimated by the observer. The main categories were residential, farming, wooded, open and industrial.

Continuous data

The following variables were recorded:

- visibility;
- verge width and type;
- roadside type.

Values of these variables were sampled from the video at frequent intervals.

Forward visibility was estimated in three categories: good (observer would be prepared to overtake in the absence of an oncoming vehicle), average (observer would overtake slow vehicle in the absence of an oncoming vehicle) or poor (observer would not overtake). The percentage of each site with, for example, good visibility was estimated as the number of samples classified in the 'good' category multiplied by 100 and divided by the total number of samples for visibility.

A similar procedure was used for the other continuous variables. Verge width was estimated as being less than 1m, between 1m and 2m, or 2m or greater. The main categories of verge type were grass verge, pavement, low bank, ditch, or none. This referred to the feature *immediately adjacent* to the road. Roadside type was the dominant vertical feature *nearest* to the road, for example trees (overhanging the road or not), hedge, high bank, fence ('closed' or 'open'), wall, open land, reflector posts or buildings.

2.4.3 Other geometric data

A measurement of link length was made for all sites, either from maps or on-site.

Road width was measured for all sites. In most cases it related to a point where the speed/flow measurements were made, but for some it related to a representative point away from junctions or bends and for others the average of 3 such points.

Hilliness was measured from maps by counting the number of (10m) contour lines crossed either up or down, multiplying by 10 to give the total change in height, and dividing by the link length.

2.4.4 Data checking and processing

Data entry was carefully checked and extreme values investigated further to ensure they were genuine. The discrete variables were either retained as whole link 'binary' factors (i.e. the feature was either present or absent) or were divided by the link length, to give a 'density' value - for example, bend density as the number of bends per kilometre.

2.5 Accident data

Details of personal injury accidents were obtained from TRL's copy of the STATS19 national accident database from 1992 to 1996 for the 'old' sites and from 1994 to 1998 for the 'new' sites. The numbers of fatal, serious and slight accidents which occurred on the defined link section were obtained. Minor junctions at the ends of link sections were included in the section, but not major junctions. The numbers of accidents were disaggregated by location (junction/non-junction), by vehicle involvement (single vehicle/multiple vehicle) and by severity (fatal/serious/slight).

3 Site characteristics

3.1 Location and road class

The distribution of the 174 sites by region and road class is shown in Table 1. The sample had a wide geographical spread, although the North West was under-represented. 47% of sites were A class roads; 31% were B class and 22% were C class or unclassified (U).

Table 1 Distribution of sites by region and by road class

Region	*A*	*B*	*C / U*	*Total*
East	9	9	10	28
East Midlands	9	9	8	26
North West	4	2	1	7
South East	27	8	6	41
South West	10	9	4	23
West Midlands	13	9	4	26
Yorkshire & Humberside	9	8	6	23
Total	81	54	39	174

3.2 Geometric data

The following summarises some of the main features.

Road width. Road width varied from 2.0m to 10.2m with a mean value of 6.5m.

Number of lanes. The number of lanes was defined as the dominant characteristic of the link. Four sites were either mainly, or entirely, single track roads and all but 3 of the rest were two-lane roads.

Link length. Link length ranged from 1.0km to 7.0km, with a mean length of 3.0km.

Bendiness. Approximately half the sites had 3 or more bends per km. About a third had at least 1 severe bend per km.

Hilliness. Just under half the sites were relatively flat, with a total rise and fall of less than 10m per km. The maximum hilliness value was 67m per km.

Number of junctions. Over 80% of sites had at least one minor junction (including end junctions where applicable). About a quarter had one or more crossroads and three-quarters had one or more T-junctions. Total junction density ranged from 0 to 6.0 per km (mean 1.1 per km).

Number of accesses. Access density (including public and private accesses and laybys) ranged from 0 to 57 per km (mean 8.6 per km).

3.3 Traffic flow data

The annual average daily total (AADT) traffic flow ranged from 106 to 25750. As Table 2 illustrates, on average, it was highest on the A roads (9128) and lowest on the C/U roads (1755).

Table 2 Traffic flow by road class

	A	*B*	*C/U*	*All*
Minimum	1732	862	106	106
Mean	9128	4340	1755	5990
Maximum	25750	17365	6566	25750

3.4 Accident data

Table 3 shows the numbers of personal injury accidents, their severity, accident frequencies and rates by road class, for all sites. The accident rates shown have been calculated from the accident and flow data to reflect the number of accidents per unit of *exposure*. The unit of exposure used is *100 million vehicle-kilometres*.

The overall proportion of fatal and serious accidents (27%) was the same as that found by Barker *et al.* (1998) for all rural single-carriageway roads in Great Britain.

The overall mean accident rate (*calculated as the total number of accidents divided by the total number of vehicle-kilometres*) was 42 accidents per 100 million veh-km. Rates were highest on the C/U roads and lowest on the A roads. They were higher than the equivalent figures in Barker *et al.* (1998) for all rural single-carriageway roads in Great Britain: 37 accidents per 100 million veh-km compared with 30 on A roads, 51 compared with 42 on B roads and 62 compared with 45 on C/U roads. Barker *et al.*, 1998 figures relate to all accidents including those at major junctions and might therefore be expected to be *higher* than those recorded here. However, it is possible, even likely, that the present sample is an unrepresentative cross section of road types – for example, there may be more bendy and hilly sites in the sample than in the national population. Such sites were deliberately included in the present sample to allow the effects of these features to be properly examined.

Table 4 shows the same data by region for all sites combined. Rates ranged from 37 accidents per 100 million veh-km in the South East (where A class roads were strongly represented – Table 1) to 65 per million veh-km in the North West.

Table 5 shows the numbers of accidents according to whether or not they occurred at or within 20m of a minor junction. The percentage of non-junction accidents was higher on the lower class roads; the C/U roads had the lowest non-junction accident density, but the highest rate.

Table 6 shows the number of single vehicle accidents. The C/U roads had a higher percentage of single vehicle accidents than the A or B roads; they also had the highest single vehicle accident rate, but the lowest single vehicle accident density.

3.5 Summary of speed, flow and accident data

To be consistent with previous studies (Taylor *et al.*, 2000), the speed data used in the analyses reported here are those observed in the *daytime off-peak period* (09:00 to 16:00). In this period, speeds can be considered to relate to free-flowing vehicles under typical flow conditions. (The correlation coefficient between the mean off-peak speeds used here and the equivalent values based on all 24 hours of data was in fact very high indeed - 0.99.)

Table 7 shows summary statistics for the speed, flow and accident data, including some of the key speed variables defined in Section 2.3. In this table, the mean accident frequency and rates are calculated as the average over all sites; they are therefore *unweighted* by link length, unlike the figures tabulated in Section 3.4.

The table shows the average 85th percentile speed to be almost 8 miles/h higher than the average mean speed (which is slightly more than 1 standard deviation). The 85th percentile speed varied widely from site to site with a maximum of 67 miles/h, even though the speed limit was 60 miles/h.

Table 3 Accidents by road class

Road class	No. of accidents				Severity (% fatal or serious)	No. of sites	Accident frequency (accs/site/ year)	Length (km)	Accident density (accs/ year/ km)	Annual average two way vehicle flow AADT	Veh-km ($x10^8$)	Accident rate (per 10^8 veh-km)
	Fatal	Serious	Slight	Total								
A	53	310	964	1327	27.4	81	3.28	222.6	1.19	9128	35.4	37.4
B	21	146	452	619	27.0	54	2.29	168.1	0.74	4340	12.1	51.3
C/U	4	46	168	218	22.9	39	1.12	127.9	0.35	1754	3.5	61.5
All roads	78	502	1584	2164	26.8	174	2.49	518.3	0.84	5990	51.1	42.4

Table 4 Accidents by region

Region	No. of accidents				Severity (% fatal or serious)	No. of sites	Accident frequency (accs/site/ year)	Length (km)	Accident density (accs/ year/ km)	Annual average two way vehicle flow AADT	Veh-km ($x10^8$)	Accident rate (per 10^8 veh-km)
	Fatal	Serious	Slight	Total								
East	8	94	241	343	29.7	28	2.45	93.7	0.73	5698	8.54	40.2
East Mid	12	83	267	362	26.2	26	2.78	95.0	0.76	4136	6.74	53.7
N West	5	17	52	74	29.7	7	2.11	12.7	1.17	5540	1.13	65.3
S East	16	115	450	581	22.6	41	2.83	98.3	1.18	9039	15.52	37.4
S West	11	37	156	204	23.5	23	1.77	74.5	0.55	3951	4.99	40.9
W Mid	12	58	153	223	31.4	26	1.69	55.9	0.80	5486	5.11	43.6
Y&H	14	98	265	377	29.7	23	3.28	88.2	0.85	5748	9.05	41.7
All roads	78	502	1584	2164	26.8	174	2.48	518.3	0.84	5990	51.1	42.4

Table 5 Non-junction accidents by road class

Road class	No. of accidents			% of accidents which are non-junction	No. of sites	Non-junction accident frequency (accs/site/ year)	Length (km)	Non-junction accident density (accs/ year/ km)	Annual average two way vehicle flow AADT	Veh-km ($x10^8$)	Non-junction accident rate (per 10^8 veh-km)
	Non-junction	Junction	Total								
A	786	541	1327	59	81	1.94	222.6	0.71	9128	35.4	22.2
B	449	170	619	73	54	1.66	168.0	0.53	4340	12.1	37.2
C/U	175	43	218	80	39	0.90	127.9	0.27	1754	3.5	49.3
All roads	1410	754	2164	65	174	1.62	518.3	0.54	5990	51.1	27.6

Table 6 Single vehicle accidents by road class

Road class	No. of accidents			% of accidents which are single vehicle	No. of sites	Single vehicle accident frequency (accs/ site/ year)	Length (km)	Single vehicle accident density (accs/ year/km)	Annual average two way vehicle flow AADT	Veh-km ($x10^8$)	Single vehicle accident rate (per 10^8 veh-km)
	Single vehicle	Multi-vehicle	Total								
A	290	1037	1327	22	81	0.72	222.6	0.26	9128	35.4	8.2
B	200	419	619	32	54	0.74	168.0	0.24	4340	12.1	16.6
C/U	85	133	218	39	39	0.44	127.9	0.13	1754	3.5	24.0
All roads	575	1589	2164	27	174	0.66	518.3	0.22	5990	51.1	11.3

Table 7 Summary statistics for speed, flow and accident data

	Mean	*Standard deviation*	*Minimum*	*Maximum*
Mean speed (miles/h)	44.2	5.85	26.0	57.6
85%_ile speed (miles/h)	51.9	6.69	29.5	67.2
Coefficient of variation	0.17	0.03	0.13	0.29
% exceeding 60miles/h	5.0	6.5	0.0	38.1
Mean excess speed (miles/h)	4.7	2.5	0.0	12.5
AADT (flow/day)	5990	4828	106	25750
Accident frequency (per year)	2.5	2.01	0	8.8
Accident rate (per km per year)	0.9	0.72	0	4.0
Accident rate (per 10^8 veh-km)	52.6	37.9	0	271

4 Methodology for accident analysis

4.1 Introduction

The main aim of the accident analysis was to determine whether (and by how much) accidents would change on a given link section if everyone drove faster than now, all else remaining constant. A typical speed-accident plot for data from a range of heterogeneous roads tends to show a *negative* relationship between accidents and speed. In the earlier study of urban link sections referred to in Section 1.3 (Taylor *et al.,* 2000), a simple speed-accident plot exhibited a negative relationship of this kind. But when the links were classified into homogeneous groups, a positive relationship was found within each group, showing that higher speeds are associated with more accidents. The level of pedestrian crossing activity was an important determinant of the group classification, such that when the pedestrian effect was accounted for directly in the modelling, the underlying speed-accident relationship was shown to be positive. As already mentioned in Section 1.3, the pedestrian effect had been masking the true effect of speed. Appendix A illustrates the point more fully.

The speed and accident data from the present sample of 174 links, when taken as a single plot, produces a highly significant, *negative* regression coefficient [–1.37 (±0.34)] between log(accident rate) and log(mean speed), suggesting a strong negative relationship. This is contrary to what we would expect based on the results of extensive 'before and after' studies, namely that if traffic speed decreases on a given road then accident frequency also decreases. The main source of this 'perverse' result is likely to be a masking factor caused by heterogeneity of the links.

4.2 The analytic procedure adopted

In approaching the analysis of this data, the assumption has been made that the heterogeneous sample of link sections consists of several relatively homogeneous groups. The masking problem was therefore addressed using a two-stage process:

- the links were first classified into relatively homogeneous groups;
- the relationship between speed and accidents was then examined *within* groups.

The first step, *road classification,* was addressed by applying a suitable multivariate technique that can take account of the correlations between the large number of descriptive variables. The method used was *principal components analysis.* This algebraic technique is a powerful tool for identifying those combinations of variables which carry the bulk of the variance in the data. The aim was to extract a principal component (or components) which was as appropriate as possible for classifying the link sections by road type. The process awards 'scores' to individual links which, in the context of the present study, reflect those road characteristics which can provide a satisfactory basis for such a classification. On the basis of these scores the links can then be classified into reasonably homogeneous groups.

The principal component analysis was conducted using all those variables that describe the main characteristic features of the road (and which satisfy certain appropriate statistical criteria), without making any distinction between dependent and independent variables. The list included accident rate, which is an important element in road quality determination. Once groups had been formed in this way, the variables used to form them were statistically evaluated to ascertain their role in group discrimination. Not all variables will necessarily contribute usefully to this discrimination and an 'optimum' set must generally be chosen which are justified statistically. To determine the best *discriminating variables* a *stepwise discriminant analysis* was employed.

Once the homogeneous road groups had been defined using principal components and discriminant analysis, *multivariate regression models* were developed between accident frequency as the dependent variable and a set of explanatory or independent variables. The group membership was used in this analysis as an (explanatory) categorical multilevel factor. The effect of speed variation on accident frequency within the groups was estimated using the mean speed as an explanatory variable, alongside traffic flow, link length, road geometry and environmental variables where appropriate.

The *multiple regression* method used was *Generalised Linear Modelling (GLM)* (see Section 4.3). The effects of the explanatory variables and of the road type factor were estimated using a multiplicative Poisson model fitted to the personal injury accident counts for the 5 year period considered.

4.3 Multiple regression analysis

4.3.1 Model forms

Modelling was undertaken at two levels – Level 1 (termed the 'core' model which included basic variables only) and Level 2 (which included a wider range of variables).

The mathematical form of the Level 1 model was:

$$\text{Accident Count} = (YR).k.\ Q^a\ L^b\ V^\alpha \exp[\Sigma g_i Y_i] \qquad (1)$$

where: YR = number of years of accident data
Q = AADT flow (per day)
L = link length (km)
V = mean speed (miles/h)
Y_i = a dummy variable for group 'i' (= 1 if in ith group, 0 otherwise)

and the parameters, k, a, b, α, and g_i were to be estimated from the data. The group factor was estimated by $G_i = \exp[g_i]$ for the ith group for which $Y_i = 1$. Using AF as 'accident frequency per year', defined by (Accident Count / YR), the log-linear version of the model is:

$$\text{Ln(AF)} = \ln(k) + a.\ln(Q) + b.\ln(L) + \alpha \ln(V) + \Sigma [g_i Y_i] \quad (2)$$

At Level 2, the core model was extended to include additional terms involving any geometry or feature variables that were found to have an additional, significant effect on accident frequency.

The Level 2 model was:

$$\text{Accident Count} = (YR).k.\ Q^a\ L^b\ V^\alpha \exp[g_i Y_i] \,.\, \exp[\Sigma c_i X_i] \quad (3)$$

where X_i = ith geometry/feature variable and the c_i are coefficients to be determined.

The log-linear version is:

$$\text{Ln(AF)} = \ln(k) + a.\ln(Q) + b.\ln(L) + \alpha \ln(V) + \Sigma [g_i Y_i] + \Sigma [c_i X_i] \quad (4)$$

4.3.2 Procedure

Regression analysis is a powerful tool for identifying the variables that affect accidents. Maher and Summersgill (1996) cautioned against using it blindly. In developing the accident models, the method of forward selection and backwards elimination was used, applying the following criteria:

i The level of statistical significance. This was the dominant criterion. The measure used was the *scaled deviance difference* resulting from the inclusion/exclusion of the variables, taking account of the appropriate *scale factor* attributable to over-dispersion. The scale factor was estimated from the resulting residual deviance, as in other similar studies (for example, Taylor *et al.*, 2000). No variables were accepted at less than the 5% level of significance or rejected at the 1% level or better without very careful consideration.

ii The stability of the model. Where explanatory variables are correlated with each other, introducing one tends to affect the model parameters of the other(s). Any such instability was carefully investigated at each stage, particularly with respect to traffic flow, speed and link length.

iii The credibility of the effect. It is desirable that the effect of a variable is understandable and that the models should have a logical structure. Models with estimated coefficients of the 'wrong' sign (ie. opposite to what common sense would indicate) were examined carefully to check the robustness of the finding.

iv The size of the effect and its ease of measurement. Variables with a large effect on accidents in relation to their range and which are straightforward for the engineer to measure were preferred.

Even though a large number of variables were available for examination, all were subjected to rigorous scrutiny, both individually and collectively, and none satisfying the above criteria were excluded at Level 2.

All of the continuous variables were treated as continuous explanatory variables in the regression model. The categorical variables were used as binary or multi-level factors depending on their nature. For example, 'centre line marking', 'hatching', 'lighting', 'reflecting road studs' etc. were used as binary variables, indicating whether such features were present or absent.

4.3.3 Accident categories

Models were developed initially for All accidents (2164 accidents). Additional models were developed for different STATS19 categories of accidents separately, as follows (the figures in brackets are the number of accidents in the category and the corresponding percentage of All accidents):

- Junction accidents (754 - 35%)
- Non-junction accidents (1410 - 65%)
- Accidents involving fatal/serious injury (KSI) (580 - 27%)
- Slight injury accidents (1584 - 73%)
- Single vehicle accidents (575 - 27%)
- Multiple vehicle accidents (1589 - 73%)

5 Classification of road links into groups

5.1 Identification and definition of Road Groups

5.1.1 Principal components analysis

Eighteen variables were found to be suitable for use in the principal components analysis – see Appendix B (B1). The strongest components extracted were investigated more thoroughly in terms of their statistical power and their meaning. The first two, components 1 and 2, were of particular interest; they explained respectively, 20.1% and 16.1% of the variance in the data (see Appendix B(B2)).

Component 1 had a high score for roads which are relatively wide and which have a high traffic speed together with a low accident rate, low bend density, and good visibility. If a reasonably straight, wide road with good forward visibility can sustain a high traffic speed with lower than average accident rate then such a road is likely to be better designed and of better quality than one that cannot. Thus the score of component 1 may be considered as an indicative measure of road quality; this first principal component is therefore a natural choice as the road quality factor. Any grouping based on this component will be primarily related to speed and to those geometric characteristics which are associated with speed.

Component 2 on the other hand was largely related to traffic flow and network characteristics, which are not essentially a characteristic of the road – though flow and geometry will inevitably be associated to some extent. In the context of a relationship between speed and accidents, however, the first principal component was the most appropriate one for classifying road types. The effect of variations in traffic flow on accidents can be reliably handled by incorporating flow as a predictive term in the accident model. In the subsequent analysis, therefore, component 1 alone was used for the purpose of classification.

The scores of component 1 based on all 18 variables (see Appendix B (B1)) were used to divide the link sections into four groups with equal intervals of road quality. Since these scores are standardised their mean value is zero and standard deviation unity. The four groups were formed as follows:

Group 1:	Score < -1.0	(24 links; 14%)
Group 2:	-1.0 < Score < 0.0	(58 links; 33%)
Group 3:	0.0 < Score < +1.0	(65 links; 37%)
Group 4:	Score > +1.0	(27 links; 16%)

Table 8 shows the range of traffic flow in each Road Group to be broad and virtually independent of the Road Group.

Table 8 AADT flow by Road Group

	Flow (per day)		
	Mean	*Minimum*	*Maximum*
Group 1	2767	106	13283
Group 2	5967	229	22745
Group 3	7233	857	25750
Group 4	5908	680	16077
All	5990	106	25750

5.1.2 Discriminant analysis

For practical application of the road type grouping, it is necessary to reduce considerably the number of variables used for classification. A stepwise discriminant analysis applied to the 18 variables was used to determine which of the ones included in this solution were worth retaining; the analysis suggested that only 6 out of the 18 variables made an important contribution to the Group discrimination. These were:

Mean speed	(over the period 0900 – 1600).
Accident rate	(per 100 million veh-km).
Junction density	(no. of minor junctions per km).
Bend density	(no. of bends per km).
Access density	(no. of public/private accesses and laybys per km).
Hilliness	(total rise and fall in metres per km).

It would not necessarily be expected as a result of this process that the variables that dominated component 1 in the principal components analysis would emerge as the most important discriminating variables. This is because of the complex correlations between the variables. The discriminant analysis maximises the between-group variance compared to the within group variance, whereas the principal components analysis operates on the data set as a whole. In fact it can be seen that only three of the variables above (mean speed, accident rate and bend density) were amongst those with the highest loading scores for component 1 (Appendix B (B2)).

Encouragingly, the six variables above are logical and plausible. The three density variables and the hilliness variable, which describe the type of road, encompass all the primary descriptors that would be natural choices for road quality.

These six variables, then, are the ones that will be needed to classify links into Road Groups in future applications. The use of this reduced set is statistically robust, and the amount of mis-classification occurring, compared to that obtained if all 18 variables were used, will be minimal.

In the present study the allocation of links into the Groups defined in Section 5.1.1 and used in the analyses reported in Section 5.2 onwards (including the multiple regression analysis described in Section 6), is based on the full set of 18 variables.

5.2 Characteristics of Road Groups

Table 9 shows for each Group the mean values and the standard deviations of the 6 key variables defining the Road Groups.

It can be seen that the primary variation between the Groups is in mean speed, accident rate and bend density. The remaining three variables contribute significantly, but to a lesser degree. This is also clear from the box-plots of the variables shown as Figure 1. In these plots, the boxes span the 25^{th} to 75^{th} percentile of the plotting variable, the solid horizontal lines representing the median values. The circles and asterisks are outliers and extreme values and the 'whiskers' represent the range of the remaining data.

Table 10 summarises the statistics in Table 9 and Figure 1 for the Group discriminating variables by comparing Group averages with overall sample averages.

Thus the Groups can be broadly described as follows:

Group 1: Roads which are very hilly, with a high bend density and low traffic speed. These are *low quality* roads.

Group 2: Roads with a high access density, above average bend density and below average traffic speed. These are *lower than average quality* roads.

Group 3: Roads with a high junction density, but below average bend density and hilliness, and above average traffic speed. These are *higher than average quality* roads.

Group 4: Roads with a low density of bends, junctions and accesses and a high traffic speed. These are *high quality* roads.

Table 11 shows the cross-tabulation of the link sections by road class and Road Group. It shows an interesting picture. Whilst there is, as would be expected, some correlation between road class and Road Group (more than half of the 'higher quality' Group 3 and 4 roads are A class roads while more than half of the 'low quality' Group 1 roads are C/U roads), this correlation is far from perfect. Road class is clearly not an adequate differentiator of road quality.

Table 12 shows the distribution of accidents in the categories defined in Section 4.3.3, by Road Group. It can be seen that Groups 1 and 4 tended to have more severe accidents, more non-junction accidents and more single vehicle accidents. This suggests that excessive speed for the conditions may be more dominant on roads in these two Groups. The proportion of accidents at junctions was highest on roads in Group 3, which have a high junction density; the proportions of single vehicle accidents were inversely related to the proportions of junction accidents, as would be expected.

Table 9 Mean and standard error of the Group discriminating variables

	Mean (standard error)				
(n = no. of links)	*Group 1 (n=24)*	*Group 2 (n=58)*	*Group 3 (n=65)*	*Group 4 (n=27)*	*All (n=174)*
Mean speed (miles/h)	35.1 (0.86)	41.2 (0.28)	47.2 (0.33)	51.7 (0.47)	44.2 (0.44)
Accident rate (per 10^8 veh-km)	107.8 (12.2)	48.6 (3.32)	41.3 (2.64)	39.2 (4.38)	52.6 (2.64)
Junction density (per km)	1.2 (0.24)	1.2 (0.10)	1.3 (0.10)	0.6 (0.07)	1.1 (0.06)
Bend density (per km)	5.1 (0.52)	3.5 (0.26)	2.2 (0.20)	1.6 (0.23)	3.0 (0.16)
Access density (per km)	7.9 (0.57)	10.3 (1.20)	8.4 (0.96)	5.8 (0.81)	8.6 (0.56)
Hilliness (rise + fall, m/km)	15.3 (2.80)	14.5 (2.03)	12.7 (1.75)	15.0 (1.70)	14.0 (1.05)

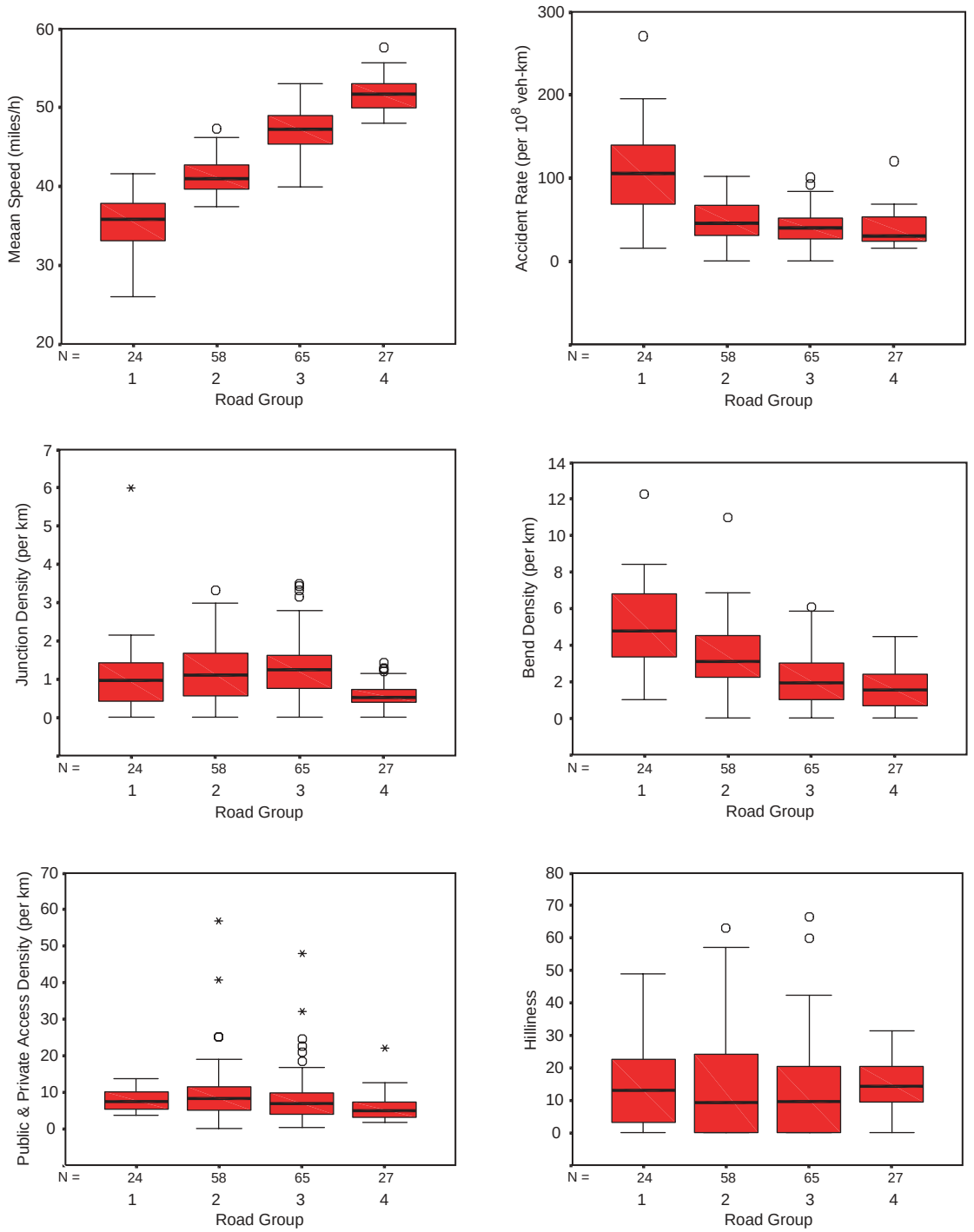

Figure 1 Box-plots for group discriminating variables

Table 10 Description of the Road Groups

	Group 1	*Group 2*	*Group 3*	*Group 4*
Mean speed	Low*	Below average*	Above average*	High*
Accident rate	High*	Average	Below average*	Low*
Junction density	Average	Average	High	Low*
Bend density	High*	Above average	Below average*	Low*
Access density	Below average	High	Average	Low*
Hilliness	High	Average	Below average	Above average

** indicates statistically significantly different from Average (at 5% level at least)*

Table 11 Number of link sections by road class and Road Group

	Road Group				
Road class	*Group 1*	*Group 2*	*Group 3*	*Group 4*	*Total all Groups*
Class A	4	24	39	14	81
Class B	7	19	19	9	54
Class C/U	13	15	7	4	39
All classes	24	58	65	27	174

Table 12 Percentage of accidents in each accident category, by Road Group

Accident category	*Group 1*	*Group 2*	*Group 3*	*Group 4*	*All Groups*	*No. of accidents*
Slight	70	76	73	70	73	1584
KSI	30	24	27	30	27	580
Junction	25	32	41	27	35	754
Non-junction	75	68	59	73	65	1410
Single vehicle	34	29	22	30	27	575
Multi vehicle	66	71	78	70	73	1589
ALL – Number	203	642	968	351		2164
(% of total)	(9%)	(30%)	(45%)	(16%)	(100%)	

6 Accident modelling

This Section presents the results of the Generalised Linear Modelling procedure described in Section 4.3. The aim was to develop accident-predictive models using the Road Grouping identified in Section 5 as a factor indicating road quality, and other measures such as traffic flow and speed as explanatory variables.

6.1 Level 1 models (Core models)

6.1.1 All accidents

The most important explanatory variables in this model were found to be the AADT flow and the link length. Road Group was also found to have a significant effect on accident frequency. With Road Group in the model, the relationship between mean traffic speed and accident frequency was significant and *positive*. In other words, Road Group has had the effect of unmasking the true speed-accident relationship. This road quality factor can therefore be considered in the present context to be the masking variable sought. The resulting speed effect in the model represents the 'within-Group' effect.

Encouragingly, the effects of this set of variables and factors were found to be remarkably stable throughout the modelling process.

Equations (1) and (3) (Section 4.3) are based on the implicit assumption that the within-Group speed effect is constant for each of the Road Groups. The possibility that this was not the case was tested by allowing the parameter α to take a different value (α_i) for each Group. The result was that the α_i's were not statistically significantly different from each other and so the assumption of a common effect was valid.

The resulting Level 1 model for All accidents was:

$$AF = (3.281 \times 10^{-7}).\ Q^{0.727}\ .L^{1.000}\ .V^{2.479}\ .G_i \qquad (5)$$

where, G_i = 1.000 for Group 1
= 0.539 for Group 2
= 0.364 for Group 3
= 0.253 for Group 4

The estimated model parameters (for the log-linear equation (2)), and their standard errors, are presented in Appendix C (C1)). The model explains about 77% of the total variation in the accident data attributable to non-Poisson sources.

The power of mean speed, V, in Equation (5) is 2.48 (±0.60) so we are 95% confident that the value lies between 1.28 and 3.68. On the assumption that Equation (5) provides the best estimate of the effect of speed on accidents, it follows that a 10% increase in mean speed will result in a 27% increase in accident frequency. The figure of 2.48 is not dissimilar to the coefficient of 2.25 found for urban roads (Taylor *et al.,* 2000), although the two are not strictly comparable because the urban model contained an additional speed parameter.

The power of link length, L, is 1.00 (±0.09), indicating that accident frequency is directly proportional to link length. In the earlier MASTER rural road study (Taylor *et al.,* 2000) the power was 0.85 (±0.07) with the number of junctions also included as a parameter in the model. In a study of accidents on modern rural trunk roads (Walmsley *et al.,* 1998), the power of L ranged between about 0.8 and about 1.0, depending on the type of model.

The power of traffic flow, Q, is 0.73 (±0.05), which indicates that if the flow is doubled then accident frequency will be increased by 65%. This is not significantly different from the power (0.75 (±0.06)) found in the MASTER study and is typical of results from other studies.

The Group factors are estimated relative to Group 1, which has a default value of unity. The factors represent a progressively decreasing accident *frequency* when moving from Group 1 to Group 4 (which parallels the decreasing accident *rate* which contributes to the identification of these Groups). This result suggests that compared to a Group 1 road, all else being equal, the accident frequency on a Group 2 road is 46% lower; on a Group 3 road it is 64% lower, and on a Group 4 road it is 75% lower.

6.1.2 Accident categories

Separate models were developed for the six categories of accident defined in Section 4.3.3. Table 13 summarises the coefficients as defined in Equation (1), including the All accidents result for comparison.

All of the models explain a high proportion of the non-Poisson variability in the accident data with the exception of the model for Single vehicle accidents for which the number of accidents was relatively small. All effects were statistically significant.

The results show that the effect of flow is quite different for different types of accident. At one extreme, Single vehicle accidents are almost proportional to the square root of flow (coefficient of 0.47 (±0.08)), while at the other extreme, Junction accidents are directly proportional to flow (coefficient of 1.03 (±0.11). It is intuitively sensible that Junction and Multiple vehicle accidents should show the strongest flow dependence. Walmsley *et al.* (1998) also found a stronger flow dependence for Multiple vehicle accidents than for Single vehicle accidents on modern rural trunk roads.

The power of L (link length) varies from 0.73 (±0.18) for Junction accidents to 1.17 (±0.10) for Non-junction accidents. Again this is intuitively reasonable as junction accidents will be less sensitive to link length than link accidents will be.

The power of V (mean speed) varies from 1.31 (±0.65) for Non-junction accidents to 5.11 (±1.25) for Junction accidents. The result for Junction accidents is a particularly important one - it supports the growing evidence that if speeds can be reduced on links then there will be a substantial beneficial safety effect at the junctions on those links as well. It also implies the potential for significant benefits to be achieved from strategies that slow traffic at junctions.

There is also a suggestion that the effect of speed on the more serious (KSI) accidents (power of 2.67 (±0.85)) is greater than that on slight accidents (power of 2.41 (±0.72)), but this difference is far from being statistically significant. Andersson and Nilsson (1997) suggest that injury accidents are proportional to the square of speed and that fatal/serious injury accidents are proportional to the cube of speed. The equivalent powers of speed here are 2.48 (±0.60) and 2.67 (±0.85).

6.1.3 Effect sizes

The size of the effect on accidents of flow and of mean speed implied by each of the models, across the observed ranges of these variables, has been examined. Details are given in Appendix C (C2). The effect size of flow – a measure of the change in the predicted accident frequency associated with the lowest observed flow and the largest observed flow - is vastly greater than that of mean speed or Road Group. Moreover, *within* each accident category the effect sizes are the largest for Group 1 roads.

Across the accident categories, the effect sizes are the largest in all Groups for Junction accidents - the effect size for Junction accidents on Group 1 roads being particularly large. This Group includes roads which are very hilly and bendy, and accident frequencies on these roads (including accidents at junctions) are particularly sensitive to increases in both traffic flow and vehicle speed. Speed management on these roads would therefore be expected to be particularly beneficial in road safety terms. As indicated in Section 5.2, roads in Group 1 tend to have more Non-junction and more Single vehicle accidents so measures addressing these accidents in particular, for example at bends, will be important.

6.2 Level 2 models

Level 2 models are the Level 1 models presented in Section 6.1 which have been extended by adding geometric variables and variables related to other road features where appropriate. They have the form of Equation (3) in Section 4.3.1.

Bend density (D_BENDS), based on all types of bends, did not have a significant direct effect on accidents (although it was one of the variables that determined Road Grouping). However, when bends were disaggregated by

Table 13 Coefficients in Level 1 models for different accident categories (Core Models – Equation (1))

	Slight injury	*KSI*	*Junction*	*Non-junction*	*Single vehicle*	*Multiple vehicle*	*All accidents*
Constant (x 10^{-7})	2.530	0.762	6.577 x 10^{-6}	339.7	16.09	0.511	3.281
Q	0.748	0.670	1.034	0.613	0.465	0.840	0.727
L	0.985	1.043	0.726	1.166	0.944	1.020	1.000
V	2.408	2.666	5.105	1.309	2.330	2.616	2.479
Group 1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Group 2	0.583	0.437	0.398	0.629	0.545	0.538	0.539
Group 3	0.382	0.325	0.251	0.428	0.312	0.381	0.364
Group 4	0.258	0.238	0.101	0.388	0.274	0.242	0.253
%Exp*	71%	77%	55%	74%	41%	77%	77%
No of accidents	1584	580	754	1410	575	1589	2164

** % variability explained by the model*

severity (into sharp, medium and slight) it was found that sharp bend density (D_SHRPBN) had a significant positive effect on All accidents. This was also true for several of the other accident categories. (A sharp bend was defined as one having a chevron and/or bend warning sign.)

Similarly, overall junction density (D_NJS) did not have a significant effect on All accidents, but when junctions were disaggregated by type (into crossroads and T-junctions), crossroad density (D_XRDS) had a significant effect on All accidents. This was also true for three other accident categories, including Junction accidents. T-junction density (D_TJS) also contributed significantly to predicting Junction accidents.

No other road feature available in the data (including hilliness and access density, which play a significant role in the definition of the Road Groups) was found to have a direct effect on accidents in any category which was both statistically significant and plausible (i.e. likely to be causal).

6.2.1 All accidents

The extended (Level 2) model for All accidents was:

$$AF = (3.152 \times 10^{-7}).Q^{0.728}.L^{1.039}.V^{2.431}.G_i \cdot e^{[0.121*DS + 0.286*DX]} \quad (6)$$

Here, the abbreviation DS is being used for sharp bend density (D_SHRPBN) and DX for crossroad density (D_XRDS). The Group factors (G_i's) were:

G_i = 1.000 for Group 1
= 0.558 for Group 2
= 0.391 for Group 3
= 0.285 for Group 4

The estimated model parameters (for the log-linear Equation (4)), and their standard errors, are presented in Appendix C (C3). This model explains 80% of the total variation arising from non-Poisson sources, that is about 3% more than the core (Level 1) model (Equation (5)). If the parameter values for flow (Q), link length (L) and mean speed (V) are compared with those for the Level 1 model it can be seen that there is hardly any change in them. The Group factors are also very similar. This suggests that the effects of the additional variables in the model are stable and do not interact with the flow and speed variables. Equation (6) implies that a 10% increase in mean speed results in a 26% increase in accidents, if all else is constant; this result is virtually the same as that predicted by the core model.

The predicted effects of the two additional variables are such that, on a one-kilometre long section of road, each additional sharp bend would be expected to increase the accident frequency by about 13%, while each additional crossroad junction would be expected to increase it by about 33%.

6.2.2 Accident categories

Table 14 summarises the coefficients as defined in Equation (3) of the models for the separate accident categories, including the All accidents results for comparison.

The table shows that sharp bends have a very substantial effect indeed on Single vehicle accidents, each additional sharp bend per kilometre increasing the Single vehicle accident frequency by 34%.

The effect of crossroad density is greatest for Junction accidents, which is only to be expected. A comparison of the coefficients of crossroad density (D_XRDS) and T-junction density (D_TJS) suggests that, on the single-carriageway rural roads included in this study, an uncontrolled crossroad is about 3 times as dangerous as an uncontrolled T-junction. On a typical link 5km long with 5 crossroad junctions (average density 1 per km), if one of the junctions is changed from a crossroad to a T-junction, then junction accidents will change by a factor $e^{(0.287-1.395)/5}$ = 0.80 i.e. a 20% decrease.

The effects of the core variables – flow, mean speed and link length are broadly similar to those in the Level 1 models. The effect of mean speed in the model for KSI accidents is such that a 10% increase in mean speed would be expected to result in a 30% increase in fatal/serious accidents.

6.2.3 Effect sizes

Effect sizes of the explanatory variables have been estimated for the Level 2 models in the same way as described for Level 1 models in Section 6.1.3. The effect

Table 14 Coefficients in the Level 2 models for different accident categories (Equation (3))

	Slight injury	*KSI*	*Junction*	*Non-junction*	*Single vehicle*	*Multiple vehicle*	*All accidents*
Constant x 10^{-7}	2.881	0.382	1.550 x 10^{-4}	216.6	4.944	1.231	3.152
Q	0.747	0.680	0.978	0.619	0.476	0.828	0.728
L	1.024	1.083	0.842	1.203	1.060	1.026	1.039
V	2.316	2.792	4.114	1.387	2.537	2.372	2.431
Group 1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Group 2	0.608	0.439	0.592**	0.633	0.559	0.558	0.558
Group 3	0.416	0.329	0.431	0.435	0.327	0.414	0.391
Group 4	0.299	0.245	0.240	0.400	0.297	0.280	0.285
D_SHRPBN	0.116	0.143	–	0.123	0.292	–	0.121
D_XRDS	0.360	–	1.395	–	–	0.432	0.286
D_TJS	–	–	0.287	–	–	–	–
%Exp*	74%	79%	67%	75%	49%	78%	80%
No of accidents	1584	580	754	1410	575	1589	2164

** % variability explained by the model*
*** not significantly different from 1.0*

sizes for the additional variables (sharp bends, crossroads and T-Junctions) have also been included. Details are given in Appendix C (C4).

The effect sizes for the additional variables are similar in size to those for mean speed, the range effect of T-junctions on the Group 1 roads being the largest. The observed density of sharp bends on roads in Group 1 was as high as 5 per km. On such link sections accident frequency is 83% higher than on sections in the same Group with no sharp bends.

6.3 Other speed parameters and their effects on accidents

During the modelling process a number of speed parameters other than the mean speed were tested in the Level 2 models, both individually and in combination with other variables. These parameters were:

- Standard deviation of speed.
- Coefficient of variation of speed (Cv) (ratio of the standard deviation to the mean speed).
- Percentage exceeding the speed limit of 60 miles/h (P).
- Mean excess speed over the 60 miles/h speed limit (mean speed of those vehicles exceeding 60 miles/h).

The result was that none of these parameters was found to improve the explanatory power of the models. However, there are two points worth noting.

In the earlier study of urban roads (Taylor *et al.,* 2000) Cv was found to have a significant (positive) effect on urban accidents (and was included with V in Model U1 of Taylor *et al.,* 2000). In that study, Cv was strongly negatively correlated with mean speed (V), the regression coefficient of Cv on V being –0.0078. In the present, rural road study, Cv had a negative, but statistically non-significant effect on accidents, for all categories of accident. Cv was slightly (but significantly) negatively correlated with V (regression coefficient of Cv on V of –0.0012). The effect of V remained stable when Cv was introduced into the models, and Cv was of no added benefit. The conclusion is that speed variability as indicated by Cv did not influence accidents in the present sample. The difference between the urban and rural results reflects the different characteristics of the speed distributions in these two situations.

In a model for All accidents, with lnV replaced by ln(P+c) (where c is a small constant added to avoid ln(0)), ln(P+c) was found to be significantly and positively related to accidents, with an estimated coefficient of 0.1137. The result implies, for example, that if the percentage of vehicles exceeding the limit is reduced from 20% to zero, accidents would be roughly halved. However, the model explains less of the accident variability than the model with mean speed, and has less practical use, so is not presented here.

It is also worth remarking on the fact that the results presented in Sections 6.1 and 6.2 involve mean speed based only on the off-peak period (0900-1600). As indicated earlier, the correlation between mean speed calculated for this period and for the full 24-hour period was very high indeed; a simple linear regression between the two variables indicated the 24-hour mean speed was on average about half a mile/h faster. A set of Level 2 models (i.e. one for each accident category) was developed in which the mean speed, V, was replaced by the 24-hour mean speed, V_{24}. All of the results were very similar indeed to the off-peak models in terms of the parameter values and goodness of fit. The models based on V are preferred for two reasons: firstly for their consistency with previous work, and secondly because 24-hour speeds are effectively made up of several different speed distributions each reflecting a different period of the day – they are thus 'hybrid' distributions, which probably lead to the slightly weaker speed effects found in the 24-hour models.

6.4 Speed limit

Since the link sections from which the models were developed were all on roads with the national 60 miles/h limit, the effect of a change in speed limit (for example, a reduction to 50 miles/h) cannot be directly assessed. However, previous work (Finch *et al.,* 1994) indicated that a reduction in speed limit, all else remaining unchanged, can be expected broadly to result in a reduction in mean speed of about a quarter of the difference between the two limits. Using this 'rule of thumb', a 2.5 miles/h reduction in mean speed would be expected to be achieved from a change in speed limit from 60 to 50 miles/h. The corresponding reduction in All injury accidents predicted by Equation (6) would be 12% assuming that the mean speed of traffic on the road before the change was 50 miles/h. This compares with 8% tentatively deduced in Taylor *et al.* (2000). Additional measures would clearly be required, however, to achieve full compliance with the new limit and reduce accidents further.

6.5 Road width

Road width has been found in other studies to influence accidents. In particular, increased road width was associated with fewer accidents in the earlier MASTER study of rural single-carriageway roads (Taylor *et al.*, 2000). Road width was not, however, a useful explanatory variable in the present study. This can be explained by the fact that it is highly correlated with traffic flow. This feature of the data was also apparent in the earlier principal component analysis (Section 5.1) in which road width was a dominant variable in the second component, component 2, with almost the same loading as flow.

7 Practical implications of the models

7.1 Speed–accident effect by Road Group

Figure 2 shows the speed-accident relationship for the four Road Groups. It is based on a constant average vehicle flow of 6000 per day for each Group (this was the average flow across all links and is well within the range covered by each Group). For each Group the range of mean speed plotted is the observed range. The relatively steep curve for Group 1 means that larger safety benefits would be expected to result if traffic speeds could be reduced on this

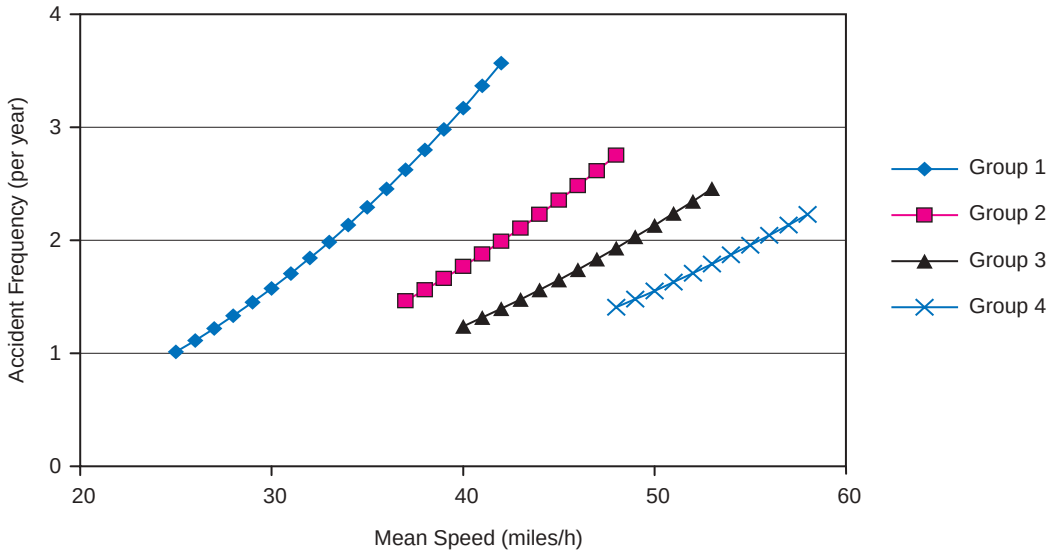

Figure 2 Speed-accident relationship by Road Group (Level 2 model)
(Accident frequency calculated with Q=6000veh/day, L = 2km, sharp bend density = 0.5/km, crossroad density = 0.14/km.)

Group of roads, compared to the same *absolute* reduction in mean speed in the Groups of faster, better quality roads. This result is analogous to that obtained for urban roads (Taylor *et al.*, 2000), where speed reductions on the slowest roads were predicted to produce greater accident benefits than the same speed reduction on faster roads. (In both cases, a constant *proportional* reduction in mean speed gives the same accident reduction on roads in the different Groups, because the power of mean speed in the predictive equations is constant.)

7.2 Accident savings per mile/h reduction in mean speed

In the previous studies of the speed-accident relationship on urban and rural roads (Taylor *et al.*, 2000), the results from the models developed were also presented in the form of predicted accident savings per 1mile/h reduction in mean traffic speed. This was to provide a more detailed understanding of an earlier broad-brush result (Finch *et al.*, 1994) that '*a 5% reduction in injury accidents is associated with each 1 mile/h reduction in mean traffic speed*'.

The present result has been used here to update the rural relationship in Figure 9 of Taylor *et al.* (2000), as shown in Figure 3. Each curve represents the predicted injury accident savings arising per 1 mile/h reduction in the mean traffic speed (V), for different values of V. Two curves are shown for the present data, one relating to the All injury accident model (Level 2) and the other to the Level 2 KSI model (for fatal/serious accidents); the figure also shows two additional curves based on the earlier work - one for the EURO model and one for the urban model (U1). The ranges of speed plotted reflect the range observed in the respective studies. The plot also shows the 5% saving line reported by Finch *et al.* (1994). The accident savings were calculated using the following formulae, converting the results to percentages:

$\Delta (AF_{U1}) / AF_{U1} = [2.252 / V - 0.046] . \Delta V$

for urban roads (model U1 of Taylor *et al.*, 2000)

$\Delta (AF_{EURO}) / AF_{EURO} = [1.536 / V] . \Delta V$

for European rural roads (EURO model of Taylor *et al.*, 2000)

$\Delta (AF_{Rural}) / AF_{Rural} = [2.431 / V] . \Delta V$

for English rural roads (present model, Equation (6))

$\Delta (AF_{Rural\ KSI}) / AF_{Rural\ KSI} = [2.792 / V] . \Delta V$

for English rural roads (KSI Level 2 model)

where AF is accident frequency, V is mean speed, Δ represents a small change and the constants are the coefficients obtained from the respective models.

It can be seen from Figure 3 that the new result for all accidents on rural roads is somewhat different from the previous result for rural roads (the EURO model), which was based largely on European data, with a limited range of UK road types. As discussed earlier, the results from that study were particularly difficult to interpret.

The present result is considered to be a much more reliable finding as far as English rural single-carriageway roads are concerned since the study covers a wide variety of rural single-carriageway roads, with mean speeds ranging from 26 to 58 miles/h. (The lower speeds on these roads were lower even than those on many of the roads in the urban study.)

For the present rural model (Level 2), the percentage reduction in All accidents per 1 mile/h reduction in the mean speed (V) is [2.431/V] x 100. This means that if V is 50 miles/h then the reduction is a 4.86% per 1 mile/h reduction in the mean speed. Table 15 shows the range of reduction to be expected for each Road Group (each is effectively an overlapping 'segment' of the curve in Figure 3).

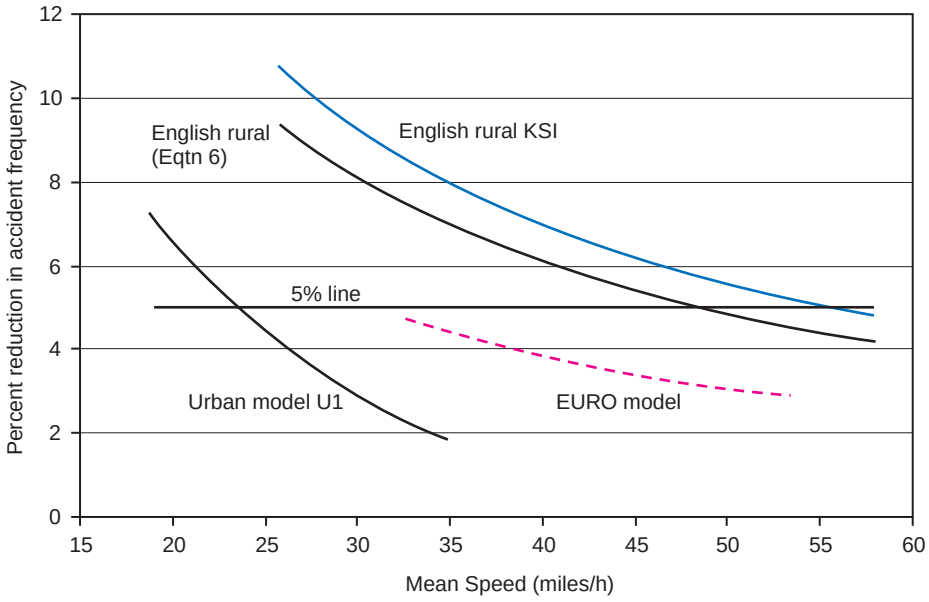

Figure 3 Predicted accident savings per 1 mile/h reduction in mean speed

Table 15 Accident reduction per 1mile/h reduction in mean speed (Level 2 model for All accidents)

	Range of mean speed (miles/h)	*Reduction in AF (%)*
Group 1	26.0 – 41.6	9.4 – 5.9
Group 2	37.4 – 47.3	6.5 – 5.1
Group 3	40.0 – 53.0	6.1 – 4.6
Group 4	48.1 – 57.6	5.1 – 4.2
All Groups	26.0 – 57.6	9.4 – 4.2

7.3 Identifying priorities for speed management

In Taylor *et al*. (2000) the speed-accident relationships quantified for urban and for rural roads were applied to national accident numbers to estimate the potential safety benefits from speed reduction strategies applied to different road types. A number of reasonable assumptions were made about the proportion of accidents that might be affected on different types of road, and on the typical speed reductions that might be achievable in each case. The assumptions take into account the speed reduction potential of available remedial measures and also the fact that resources are not unlimited. Table C1 from Taylor *et al.* (2000) is reproduced here as Table 16, replacing the rural road results based on the EURO model with results based on the present Level 2 model for All accidents.

The table shows the number of accidents occurring nationally on roads of different type and speed limit. Using a range of appropriate measures, it is assumed that speed reductions of varying degrees can be brought about which will affect certain proportions of these accidents (4th and 5th columns). For example, it is assumed that on minor urban roads, 15% of accidents can be affected by introducing 20 miles/h zones giving mean speed reductions of 10 miles/h, and that a further 30% can be affected by lesser measures resulting in a 5 miles/h reduction in mean speed. Applying figures for the percentage reduction in accidents per mile/h reduction in mean speed calculated from the relevant accident prediction models gives the estimated annual accident reductions shown in the final column. The mean speeds observed in the present study for rural single-carriageway A and B/C/U roads were used to determine the appropriate percentage reduction in accident frequency per mile/h reduction in mean speed (5% and 5.5% respectively) for these roads.

The overall estimate is that more than 24,000 accidents can be saved nationally. The new results for rural roads mean that a higher proportion of the total accident saving would be expected to come from these roads (13% compared to the 9% suggested in Taylor *et al.*, 2000). Whilst this represents quite a substantial difference for rural roads, the overall conclusion remains the same - that speed management policies have a greater potential to reduce accidents and casualties on urban than on rural roads.

If only the most serious accidents are considered, the importance of rural roads for accident reduction is greater. The assumptions above applied to accidents involving fatal/serious injury lead to an estimate of an annual reduction of 801 serious casualties and of 104 fatalities on rural roads, using the same casualty per accident figures as in Taylor *et al.*, 2000. (The corresponding figures for urban roads are 3,144 and 173.)

7.4 Model application

The models developed here apply to sections of rural single-carriageway road (60 miles/h speed limit) between 1km and 7km in length, which do not pass through any major junctions. The ranges of the other variables in the data on which the models are based are given in Appendix C (C5). These are the ranges over which application of the models can be considered to be valid. Outside these ranges the model predictions can be considered to be less robust.

Since the proportional effect of mean speed, V, on accident frequency, AF, is the same for each Road Group

Table 16 Potential national accident reductions (based on 1998 data) with the assumptions shown

Road type	*Speed limit (current: miles/h)*	*All accidents per year (AF)*	*Assumed proportion of accidents affected*	*Assumed mean speed reduction (miles/h) (V)*	*% reduction in AF per 1 mile/h reduction in V*	*Total reduction in accidents (per year)*
Urban	20	289	0	–	–	0
Minor urban	30	74,390	0.15	10	6	6,695
			0.3	5	6	6,695
Main urban	30	80,173	0.3	5	3	4,810
			0.3	2.5	3	2,405
Main urban	40	19,109	0.3	3	2	344
			0.3	1.5	2	172
Rural	50	3,818	0	–	–	0
Rural single-carriageway (A)	60	23,217	0.3	4	5	1,393
			0.3	2	5	697
Rural single-carriageway (other)	60	21,494	0.3	2	5.5	709
			0.3	1	5.5	355
Other	60/70	16,433	0	–	–	0
All		238,923				24,275

(eg. AF $\propto V^{2.48}$ for All accidents), the models developed here can be used directly to determine the *proportional change* in accidents from a known change in mean speed, without classifying the road by Group.

However, if the *absolute* value of predicted accident frequency is required, then it is necessary first to allocate the link section to the appropriate Road Group. The method by which this should be done is shown in Appendix D. It is important to note that, in using the models in this way, the Road Group must be unchanged when predicting the change in accident frequency arising from a change in mean speed.

8 Summary and discussion

8.1 Summary

A comprehensive study has been undertaken of the relationship between vehicle speeds and accidents on rural single-carriageway roads in England. This study follows an earlier investigation of this relationship, which was based on limited data from several European countries.

An extensive database was developed relating to 174 sections of English rural single-carriageway road with a 60 miles/h speed limit, a number of which had been used in the previous study. Each section was between 1km and 7km long and included minor junctions (where vehicles on the road section do not have to give way) within the link section but excluded major junctions (where there is a change of priority). The data collected for each section comprised: injury accident data (for 5 years), traffic flow and vehicle speed data (for a 24-hour weekday period), and a wide range of details of road characteristics, geometry and layout.

The sample was stratified so as to include all road classes and to provide a good geographical distribution, a wide range of flow levels, and degrees of hilliness, bendiness and junction/access frequency. A wide range of mean speeds (26 to 58 miles/h) and accident rates (0 to 271 per 100 million vehicle-kilometres) was observed.

The statistical methods of principal component and discriminant analysis were used to classify the link sections into 4 relatively homogeneous Road Groups within which the speed-accident relationship could be investigated. In this very effective classification, the Groups were defined by a set of 6 variables: accident rate; mean speed; minor junction density; bend density; access density; and hilliness. These together reflect the operational characteristics of the road, or ‘road quality’, which can be described as follows:

Group 1: Roads which are very hilly, with a high bend density and low traffic speed. These are *low quality* roads.

Group 2: Roads with a high access density, above average bend density and below average traffic speed. These are *lower than average quality* roads.

Group 3: Roads with a high junction density, but below average bend density and hilliness, and above average traffic speed. These are *higher than average quality* roads.

Group 4: Roads with a low density of bends, junctions and accesses and a high traffic speed. These are *high quality* roads.

The multiple regression technique of Generalised Linear Modelling was used to develop models relating accident frequency to speed, flow, Road Group and geometric variables for a number of different accident categories (All accidents; accidents involving fatal/serious injury and those involving slight injury; Single vehicle and Multiple vehicle accidents; and Junction and Non-junction accidents).

The models developed explained a high proportion of the variability in the accident data and the effects of the key variables were found to be strong, plausible and stable.

The main results were as follows:

- Accident frequency in all categories increased rapidly with mean speed – the All accident frequency increased with speed to the power of approximately 2.5 - thus indicating that a 10% increase in mean speed results in a 26% increase in the frequency of all injury accidents.
- The relationship between accident frequency and traffic flow and link section length mirrored that typically found in other similar studies.
- Accident frequency varied between the Road Groups defined above. It was highest on the Group 1 roads, and about a half, a third and a quarter of this level on roads in Groups 2, 3 and 4 respectively.
- The frequency of All injury accidents was also found to increase rapidly with two further measures: these were the density of sharp bends (those with a chevron and/or bend warning sign) and the density of minor crossroad junctions. These increased accidents by 13% and 33% respectively for each additional bend/crossroad per kilometre. Single vehicle accidents were particularly strongly affected by the density of sharp bends (34% increase in accident frequency per additional sharp bend per kilometre.)
- The effect of mean speed was found to be particularly large (power of about 5) for junction accidents, suggesting substantial potential for accident reduction from strategies designed to reduce speeds at junctions.
- No other measures of speed were found to influence accident frequency as strongly as, or in addition to, mean speed.
- The percentage reduction in accident frequency for a 1 mile/h reduction in mean speed implied by the 'All accidents' relationship depended on the mean speed. It ranged from 9% at a mean speed of 27 miles/h to 4% at a mean speed of 60 miles/h.
- The effect of speed on fatal and serious accidents was greater than its effect on All accidents taken together, though the difference was not statistically significant. A 10% increase in mean speed would be expected to result in a 30% increase in the frequency of fatal/serious accidents.

8.2 Discussion

The models presented in this report differ from the rural speed-accident model previously reported (Taylor *et al.*, 2000) in a number of ways. The present models are substantially more robust, being based on a more structured, extensive and relevant database. They predict a stronger effect of speed on accidents than did the previous model; the reduction in accidents for a 1 mile/h reduction in mean speed was 4.5% to 7.5% in the present study compared with 3% to 5% across the mean speed range covered in the earlier study.

Despite the size of the speed effect established here for rural roads, it has been shown that in terms of accident reduction potential, speed management policies applied to urban roads are still likely to provide the greatest benefits. This is because of the vastly greater number (and more concentrated distribution) of accidents occurring on those roads. The assumption in this assessment is that speed affects (major) junction and link section accidents in a similar way and does so similarly on urban and rural roads. The present study indicated a much stronger speed effect on accidents at minor junctions compared to that on the sections between them. This *suggests* that the effect of speed on accidents at the major junctions may also be large. If this rural speed effect does extend to major junctions, then the potential for speed management policies to reduce speed on these roads may be underestimated here.

The classification of roads into Groups reflecting road quality was fundamental to the study. An alternative simple classification by road class did not allow a satisfactory speed-accident model to be developed. Similarly, the use of a set of variables which excluded accident rate in the classification proved to be unsatisfactory. The classification approach adopted was essential to generating the 'masking' (Group) factor sought.

There is a lot more work to be done to develop the basis for speed management policies on rural single-carriageway roads. The issues to be addressed are raised comprehensively in the Government's review of speed policy (DETR, 2000). They include:

- the need to define a rural road hierarchy according to road function;
- the need to establish what are appropriate speeds for the different types of roads in this hierarchy;
- the need to identify means of achieving these appropriate speeds;
- the need to define a policy for setting appropriate speed limits, taking account of the hierarchy and of the appropriate speeds to be achieved.

The present study has provided a basis from which to progress these issues. The Road Group classification developed here has the potential to contribute to defining a road hierarchy. Hierarchy will be linked strongly to the existing road class, but B class roads in particular have a wide variety of geometry and function. In the present classification they were well distributed across the four Road Groups (Table 11). Whilst those in Groups 3 and 4 are, broadly, able to sustain higher speeds safely, those in Groups 1 and 2 are not designed to do so.

The Group classification allocated several A class roads to the low quality Group 1 and many to Group 2. A casual observer might well describe a number of these roads as of 'high quality' in the general sense of the phrase, because they have high quality signing and marking – a large amount of white paint. But in fact they are roads that have an inherently high accident risk because of their tendency to be hilly and bendy and/or to have a large number of accesses; on top of this they carry high traffic flows. Whilst it is possible that the signing and marking prevents these roads from being even more dangerous than they are, it is clear that even when this signing/marking is of the highest quality, it cannot overcome the risk implicit in the general (and largely fixed) alignment and nature of the road. These roads present a particular problem in treatment, because whilst lower speeds are needed to reduce risk, the longer journey times that result and the high levels of traffic flow involved may bring significant economic disbenefits.

9 Conclusions

1 The study has achieved its objective of developing a speed-accident relationship for English rural single-carriageway roads which is straightforward to interpret and apply. The analytical process successfully overcame the difficulty inherent in this type of study of de-coupling the effects of mutually correlated variables.

2 The resulting predictive relationship for All injury accidents shows that accident frequency rises rapidly with the mean traffic speed on a given road, and quantifies this effect. The relationship can be used to estimate the change in accident frequency resulting from a change in mean speed on a given road and, if applied to local or national accident statistics, to estimate the effects of different speed management strategies.

3 The classification of roads into groups reflecting road quality which underpinned the analysis has the potential to contribute towards the development of a road hierarchy on rural single-carriageway roads.

10 Acknowledgements

The work described in this report forms part of the programme of research undertaken by TRL for the Road Safety Division of DTLR.

The authors wish to acknowledge the very valuable advice provided by Mr G Maycock.

11 References

Andersson G and Nilsson G (1997). *Speed management in Sweden*. Linköping: Swedish National Road and Transport Institute VTI.

Barker J, Farmer S and Nicholls D (1998). *Injury accidents on rural single carriageway roads 1994-95: an analysis of STATS19 data*. TRL Report TRL304. Crowthorne: TRL Limited.

Baruya A (1998). *MASTER: speed-accident relationship on European roads*. Working Paper R1.1.3. Brussels: European Commission.

Department of the Environment, Transport and the Regions (2000). *New directions in speed management – a review of policy*. London: DETR.

Finch D J, Kompfner P, Lockwood C R and Maycock G (1994). *Speed, speed limits and accidents*. Project Report PR58. Crowthorne: TRL Limited.

Lee B H and Brocklebank P J (1993). *Speed/flow/geometry relationships on single-carriageway roads*. Contractor Report CR319. Crowthorne: TRL Limited.

Maher M J and Summersgill I (1996). *A comprehensive methodology for the fitting of predictive accident models.* Accident Analysis and Prevention, Vol 28, No.3, pp 281-296.

Quimby A, Maycock G, Palmer C and Buttress S (1999). *The factors that influence a driver's choice of speed – a questionnaire study*. TRL Report TRL325. Crowthorne: TRL Limited.

Taylor M C, Lynam D A and Baruya A (2000). *The effects of drivers' speed on the frequency of road accidents*. TRL Report TRL421. Crowthorne: TRL Limited.

Taylor M C (2001). *The speeds of vehicles which are involved in fatal accidents*. Traffic Engineering and Control 42(2).

Walmsley D A, Summersgill I and Binch C (1998). *Accidents on modern rural single-carriageway trunk roads*. TRL Report TRL336. Crowthorne: TRL Limited.

Appendix A: Illustration of the 'masking' effect

The following example has been constructed to illustrate the ‘masking’ problem faced in the analysis of the data from the present study of rural roads.

A sample of data showing a relationship between two variables X and Y has been synthesised. Four groups of data (G1 to G4) were generated, each with a linear relationship of the form $Y = a_i + b\,X$, with a common slope ‘b’ but different values of the intercept ‘a’. The values of a were –8, –9, –10 and –11; the value of b was 2.0. A random ‘error’ was added to each function value to make it resemble an observation with error. These ‘error’ values were drawn from a Normal population with mean zero and standard deviation 0.25. The values of the independent variable X were chosen to overlap between the four groups and the overall range of X was 3.0 to 4.5.

Figure A1 is the plot for the four groups separately, each showing the positive linear relationship between the X and Y values. The relationships are parallel, since the slope ‘b’ is common to the four groups. Figure A2 shows a plot of the same data, ignoring the groups. It also shows a linear relationship, with a regression equation which has a highly statistically significant, negative slope. The simple correlation coefficient is highly significant. It would be easy from this plot alone to conclude that a negative linear relationship between Y and X describes these data well, but this conclusion would not reflect the underlying pattern in the data.

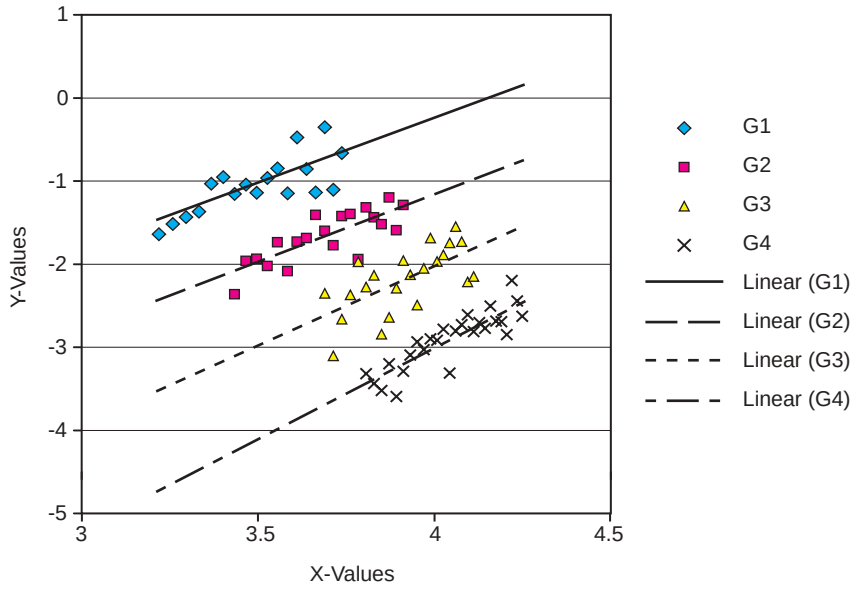

Figure A1 X-Y plot of the generated group data

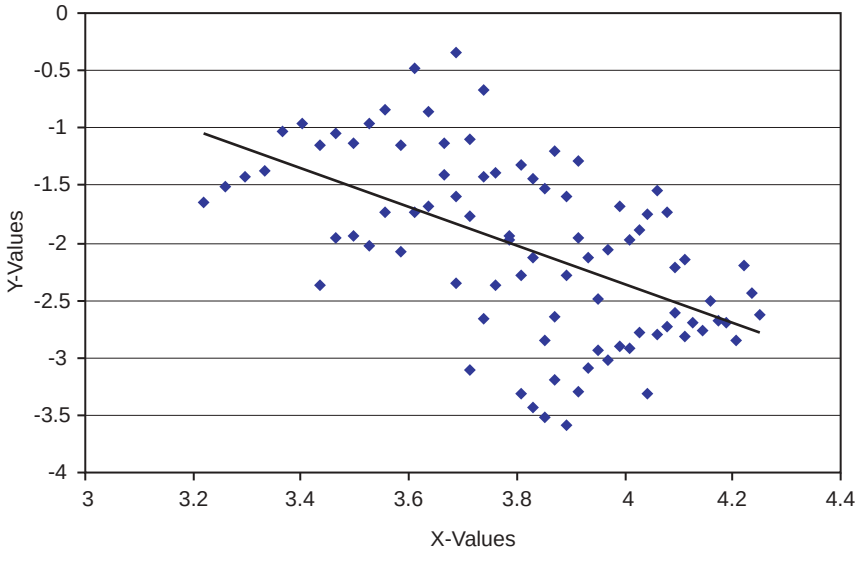

Figure A2 X-Y plot ignoring groups

Appendix B: Identification of Road Groups

B1 Variables used

The following 18 variables were found to be suitable for use in the principal components analysis. The figures in square brackets are the variable means.

AADT	Annual average daily traffic.	[5990]
ACC_RATE	Accident rate per 100 million vehicle kilometres.	[52.6]
D_ACCESS	Access density (No. of accesses and laybys per km).	[8.6]
D_BENDS	No. of (sharp/medium/slight) bends per km.	[3.0]
D_BUS	No. of bus stops/markings/bays per km.	[0.4]
D_CNTMRK	No. of changes in centre road marking per km.	[1.5]
D_NJS	No. of minor junctions (T-junctions/X-roads) per km.	[1.1]
D_RDMRK	No. of road markings (slow/arrows/ change lanes etc) per km.	[2.2]
D_RDSIGN	No. of (order/direction/information/ warning) signs per km.	[5.0]
D_TURNLN	No. of separate right/left turning lanes per km.	[0.2]
HILLINESS	Total rise and fall (m per km).	[14]
MEAN_SPD	Mean traffic speed (miles/h).	[44.2]
PC_85ILE	85th percentile speed.	[51.9]
PC_HEDGE	% of link with hedge on the roadside.	[52.0]
PC_OVHNG	% of link with overhanging trees.	[4.0]
PC_VERGW	% of link with wide verge (>2m).	[11.8]
PC_VISGD	% of link with good forward visibility.	[35.6]
WIDTH	Road width (m).	[6.5]

B2 Principal components

Principal components analysis algebraically converts the observed data matrix (expressed in terms of cases and variables) into a matrix in which the components are defined as linear combinations of the original variables X_i (standardised) as:

$$F_j = \Sigma_i \, a_{ij} \, X_i$$

The first principal component is F_1 and it is that linear combination of variables which has the greatest sample variance; F_2 has the second greatest variance, and so on. The a_{ij} are known as the variable 'loadings'.

Table B1 presents the variable loadings for the 2 statistically strongest principal components extracted, showing their association with the 18 variables in B1 above. Variables which have the highest loading make the greatest contribution. The first two principal components explained 20.1% and 16.1% respectively of the variance in the data. The third and subsequent principal components explained much smaller proportions of the variance in the data than did the first two components.

Table B1 Variable loadings for the first two principal components

	Component	
	1	*2*
AADT	0.269	**0.749**
ACC_RATE	**-0.548**	-0.273
D_ACCESS	-0.138	-0.039
D_BENDS	**-0.555**	-0.320
D_BUS	-0.047	**0.541**
D_CNTMRK	0.097	0.274
D_NJS	-0.113	**0.586**
D_RDMRK	-0.008	0.144
D_RDSIGN	-0.155	**0.543**
D_TURNLN	0.130	**0.551**
HILLINESS	-0.074	-0.101
MEAN_SPD	**0.930**	-0.041
PC_85ILE	**0.928**	-0.087
PC_HEDGE	0.122	-0.091
PC_OVHNG	-0.040	-0.211
PC_VERGW	0.194	0.026
PC_VISGD	0.428	0.339
WIDTH	0.390	**0.647**

*Loadings greater than 0.5 are shown in **bold**.*

The mean speed and the 85th percentile of speed have a high positive loading on component 1, indicating that it is primarily a 'speed' component. This component is also negatively correlated with accident rate (ACC_RATE) and bend density (D_BENDS), as indicated by their negative loading, and includes positive contributions arising from the proportion of road with good forward visibility and from road width.

Thus scores on component 1 will be strongly correlated with traffic speed and to a lesser extent with other geometric features of the road. A lower than average speed will produce a low score and a higher than average speed will produce a higher score for this component. The negative loading for accident rate means that a lower than average accident rate will produce a higher score and a higher than average accident rate will produce a low score. The role of bend density is very similar to that of accident rate; a high proportion of sections with good forward visibility and relatively wide roads will also increase the score on this component.

Component 2 is dominated by AADT flow and road width (WIDTH), with other characteristics such as the density of minor junctions (D_NJS), road signs (D_RDSGN), bus stops and turning lanes also featuring. Hence this component has been identified with high traffic flows and a relatively dense road network.

Appendix C: Model equations, effect sizes and data ranges

C1 Log-linear equation for Level 1 (Core) model for All accidents

The log-linear equation for the Level 1 models (Equation (2) in the main text) is:

$$\ln(AF) = \text{Constant} + a.\ln(Q) + b.\ln(L) + \alpha.\ln(V) + \Sigma [g_i Y_i]$$

where g_i is the appropriate parameter for the ith Group.

Table C1 Poisson regression model for All accidents (Level 1)

Variable	*Coefficient*	*se(coefficient)*	*t*	*Sig-t*
Constant	-14.93	2.215	-6.74	<0.0001
ln(Q)	0.7268	0.0483	15.05	<0.0001
ln(L)	1.0000	0.0912	10.97	<0.0001
ln(V)	2.479	0.6039	4.10	<0.001
Group1	0.000	–	–	–
Group2	-0.6176	0.1561	-3.96	<0.001
Group3	-1.010	0.2090	-4.83	<0.001
Group4	-1.376	0.2628	-5.24	<0.001

se = standard error t = studentised t-value
sig-t = significance level of t (probability)

The initial Deviance was 1341 (df 173) and the Residual Deviance after fitting the model was 433 (df 167). This gives a 'scale factor' (σ^2) of 2.6. The model explains 77% of the variation arising from non-Poisson sources. (2164 accidents analysed.)

C2 Effect sizes for Level 1 (Core) models

The effect size of a variable is defined here as the *muliplicative factor* by which the accident frequency would be expected to increase or decrease if the variable varied from the *lowest (Min)* to the *highest (Max)* in the observed range, assuming everything else is constant. Table C2 gives the *Min* and *Max* values for flow and mean speed by Group; Table C3 shows the effect sizes within each Group and for each of the 7 accident category models

Table C2 Flow and mean speed variation by group

	AADT Flow (per day)		*Mean speed (miles/h)*	
	Minimum (Min)	*Maximum (Max)*	*Minimum (Min)*	*Maximum (Max)*
Group 1	106	13283	26.0	41.6
Group 2	229	22745	37.4	47.3
Group 3	857	25750	40.0	53.0
Group 4	680	16077	48.1	57.6
All Groups	106	25750	26.0	57.6

in terms of the ratio $(Max/Min)^x$, where 'x' is the estimated power of flow (a) or mean speed (α) in Equation (1) of the main text.

C3 Log-linear equation for Level 2 model for All accidents

The log-linear equation for the Level 2 models (Equation (4) in the main text) is:

$$\ln(AF) = \text{Constant} + a.\ln(Q) + b.\ln(L) + \alpha.\ln(V) + \Sigma [g_i Y_i] + \Sigma [c_i X_i]$$

where g_i is the appropriate parameter for the ith Group and the X_i's are any other explanatory variables.

In Table C4, the variables that augment the Level 1 models are D_SHRPBN (sharp bend density) and D_XRDS (crossroad density).

C4 Effect sizes for Level 2 models

The effect size of a variable is defined here, as in section C2, as the *muliplicative factor* by which the accident frequency would be expected to increase or decrease if the variable varied from the *lowest (Min)* to the *highest (Max)* in the observed range, assuming everything else is constant. *Min* and *Max* values for flow and mean speed by Group were given in Table C2.

Table C3 Effect sizes for flow, mean speed and Group variation (Core models)

Accident category	*Slight injury*	*KSI*	*Junction*	*Non-junction*	*Single vehicle*	*Multiple vehicle*	*All accidents*
AADT flow							
Coeff (a)	*0.7479*	*0.6705*	*1.034*	*0.6135*	*0.4647*	*0.8399*	*0.7268*
Group 1	37.08	25.51	147.68	19.37	9.44	57.82	33.48
Group 2	31.16	21.83	116.13	16.80	8.47	47.57	28.28
Group 3	12.74	9.79	33.74	8.07	4.86	17.43	11.86
Group 4	10.65	8.34	26.32	6.96	4.35	14.25	9.96
Mean speed							
Coeff (α)	*2.408*	*2.666*	*5.105*	*1.309*	*2.33*	*2.616*	*2.479*
Group 1	3.10	3.50	11.02	1.85	2.99	3.42	3.21
Group 2	1.76	1.87	3.32	1.36	1.73	1.85	1.79
Group 3	1.97	2.12	4.22	1.45	1.93	2.09	2.01
Group 4	1.55	1.62	2.53	1.27	1.53	1.61	1.57
Group							
Coeff of Group *4*	*-1.353*	*-1.436*	*-2.290*	*-0.946*	*-1.293*	*-1.418*	*-1.376*
Group 1 to 4	0.26	0.24	0.10	0.39	0.27	0.24	0.25

Table C4 Poisson regression model for All accidents (Level 2)

Variable	*Coefficient*	*se(coefficient)*	*t*	*Sig-t*
Constant	-14.97	2.192	-6.83	<0.0001
ln(Q)	0.7282	0.0477	15.24	<0.0001
ln(L)	1.039	0.0903	11.51	<0.0001
ln(V)	2.431	0.5945	4.09	<0.001
Group 1	0.000	–	–	–
Group 2	-0.5840	0.1519	-3.84	<0.001
Group 3	-0.9390	0.2043	-4.60	<0.001
Group 4	-1.256	0.2585	-4.86	<0.001
D_HRPBN	0.1213	0.0485	2.50	<0.02
D_XRDS	0.2865	0.1324	2.16	<0.05

se = standard error
t = studentised t-value
sig-t = significance level of t (probability)

The initial Deviance was 1341 (df 173) and the Residual Deviance after fitting the model was 402 (df 165). This gives a 'scale factor' (σ^2) of 2.4. The model explains 80% of the variation arising from non-Poisson sources. (2164 accidents analysed.)

Table C5 Effect sizes for Level 2 Models

		Max-min range	*Slight injury*	*KSI*	*Junction*	*Non-junction*	*Single vehicle*	*Multiple vehicle*	*All accidents*
Flow, Mean speed and Group variation									
AADT	*Coeff (a)*		*0.7471*	*0.6801*	*0.9778*	*0.6187*	*0.4764*	*0.8283*	*0.7282*
flow	Group 1	*	36.93	26.72	112.57	19.86	9.99	54.67	33.71
	Group 2	*	31.04	22.81	89.68	17.20	8.94	45.10	28.46
	Group 3	*	12.71	10.12	27.86	8.21	5.06	16.75	11.92
	Group 4	*	10.62	8.59	22.04	7.08	4.51	13.73	10.01
Mean	*Coeff (α)*		*2.316*	*2.792*	*4.114*	*1.387*	*2.537*	*2.372*	*2.431*
speed	Group 1	*	2.97	3.71	6.91	1.92	3.29	3.05	3.13
	Group 2	*	1.72	1.93	2.63	1.39	1.82	1.75	1.77
	Group 3	*	1.92	2.20	3.19	1.48	2.05	1.95	1.99
	Group 4	*	1.52	1.66	2.11	1.29	1.58	1.54	1.55
Group	*Coeff of Group 4*		*-1.207*	*-1.408*	*-1.427*	*-0.9163*	*-1.214*	*-1.273*	*-1.256*
	Group 1 to 4		0.30	0.24	0.24	0.40	0.30	0.28	0.28
Geometry variation									
Sharp	*Coeff (c)*		*0.1161*	*0.1428*		*0.1230*	*0.2922*		*0.1213*
bend	Group 1	0.0 - 5.00	1.79	2.04		1.85	4.31		1.83
density	Group 2	0.0 - 3.00	1.42	1.53		1.45	2.40		1.44
	Group 3	0.0 - 4.17	1.62	1.81		1.67	3.38		1.66
	Group 4	0.0 - 0.75	1.09	1.11		1.10	1.25		1.10
X-Road	*Coeff (c)*		*0.3600*		*1.3950*			*0.4323*	*0.2865*
density	Group 1	0.0 - 1.00	1.43		4.03			1.54	1.33
	Group 2	0.0 - 1.11	1.49		4.70			1.62	1.37
	Group 3	0.0 - 1.00	1.43		4.03			1.54	1.33
	Group 4	0.0 - 0.57	1.23		2.21			1.28	1.18
T-Junction	*Coeff (c)*				*0.2875*				
density	Group 1	0.0 - 6.00			5.61				
	Group 2	0.0 - 2.00			1.78				
	Group 3	0.0 - 2.50			2.05				
	Group 4	0.0 - 1.29			1.45				

** See Table C2*

C5 Data ranges

Table C6 Data ranges for use with Level 1 and Level 2 models

Explanatory variable	*A class roads Mean (range)*	*B class roads Mean (range)*	*C/U class roads Mean (range)*	*All roads Mean (range)*
Q (AADT)	9128 (1732 – 25750)	4340 (862 – 17365)	1755 (106 – 6566)	5990 (106 – 25750)
V (miles/h)	46.0 (32.7 – 57.6)	44.5 (35.3 – 54.7)	40.1 (26.0 – 52.4)	44.2 (26.0 – 57.6)
Sharp bend density / km	0.45 (0.00 – 5.00)	0.62 (0.00 – 3.20)	0.44 (0.00 – 4.17)	0.50 (0.00 – 5.00)
Crossroad density / km	0.18 (0.00 – 1.11)	0.11 (0.00 – 0.77)	0.10 (0.00 – 1.00)	0.14 (0.00 – 1.11)
T-Junction density / km	0.81 (0.00 – 6.00)	0.63 (0.00 – 2.14)	0.37 (0.00 – 1.50)	0.65 (0.00 – 6.00)

Appendix D: Allocation of link sections to Road Groups

For the purposes of accident prediction, link sections must be allocated to their appropriate Road Group, to enable application of either the Core (Level 1) or Level 2 model. This can be done as follows, using *Fisher's classification function* applied to the six key variables identified in the discriminant analysis (Section 5.1.2).

Fisher's classification function

In any classification process a misclassification error occurs if a case is allocated to a wrong group. This is a particular problem when the parent populations overlap. The probability of such an error can be minimised by defining the population boundaries according to certain rules. Fisher's classification function is a linear combination of the variables derived to achieve this:

$$X = b_0 + \Sigma\, b_j\, x_j$$

where b_0 is a constant, the b_j's are *classification function coefficients* and the x_j's are the values (raw data) of the variables.

Table D1 presents the *classification function coefficients* estimated here for the four Road Groups.

Table D1 Classification function coefficients

	Road Group			
Variable	*Group 1*	*Group 2*	*Group 3*	*Group 4*
Mean speed	5.743	7.151	8.220	9.112
Accident rate	0.0243	-0.064	-0.091	-0.113
Junction density	-0.909	-2.188	-2.668	-4.324
Bend density	-0.316	-1.392	-2.092	-2.536
Access density	0.691	0.944	1.066	1.140
Hilliness	0.418	0.481	0.531	0.593
(Constant)	-108.525	-151.533	-196.711	-239.687

Using these coefficients, four function values (X_1, X_2 , X_3 , X_4) need to be calculated, one for each of the four Groups using the values of the 6 variables for the link under consideration. The link is then assigned to the Group for which the X value is the largest. Using the variables listed in the first column of Table D1 the procedure for classification is thus as follows:

i Calculate for Group 1, using column 1

X_1 = -108.525 + 5.743*(Mean speed) + … + 0.418*(Hilliness).

ii Calculate for Group 2, using column 2

X_2 = -151.533 + 7.151*(Mean speed) + … + 0.481*(Hilliness).

iii Similarly calculate X_3 , X_4 for Group 3 and 4.

iv Compare the values of X_1, X_2, X_3 and X_4 and choose the Group with the largest value.

v Allocate the link to that Group.

Abstract

A research programme at TRL over the last decade has demonstrated beyond doubt that the faster drivers choose to travel, the more likely they are to be involved in an accident, and that higher speeds on roads with otherwise similar characteristics are associated with more accidents. The programme included a European study of speed and accidents on rural single-carriageway roads, but this involved a relatively small proportion of UK roads.

The report describes a more extensive investigation of the relationship between speed and accidents on rural single-carriageway roads in England. The study involved: site selection; the collection and analysis of data from 174 road sections; the application of statistical techniques to group the sections; and statistical modelling to relate accident frequency to factors such as traffic flow, vehicle speed and other characteristics of the road.

The report presents the resulting speed-accident relationships and discusses their implications for rural speed management.

Related publications

TRL452 *Changes in accident frequency following the introduction of traffic calming in villages* by A H Wheeler and M C Taylor. 2000 (price £25, code E)

TRL440 *The characteristics of speeders* by D C Webster and P A Wells. 2000 (price £25, code E)

TRL421 *The effects of drivers' speed on the frequency of road accidents* by M C Taylor, D A Lynam, and A Baruya. 2000 (price £35, code H)

TRL336 *Accidents on modern rural single-carriageway trunk roads* by D A Walmsley, I Summersgill and C Binch. 1998 (price £50, code L)

TRL325 *Accidents on modern rural dual-carriageway trunk roads* by D A Walmsley, I Summersgill and A Payne. 1998 (price £50, code L)

TRL304 *Injury accidents on rural single-carriageway roads, 1994-95: an analysis of STATS19 data* by J Barker, S Farmer and D Nicholls. 1998 (price £25, code E)

Prices current at February 2002

For further details of these and all other TRL publications, telephone Publication Sales on 01344 770783, or visit TRL on the Internet at www.trl.co.uk.