AN EXAMINATION

OF RELATIONSHIPS BETWEEN

ROAD ACCIDENTS AND TRAFFIC FLOW

David Ronald Dickson McGuigan

Thesis submitted for the Degree of
Doctor of Philosophy

University of Newcastle upon Tyne
1987

TABLE OF CONTENTS

ABSTRACT

In this thesis it is suggested that the cost-effectiveness of road safety expenditure on low cost engineering remedial works could be improved because the currently adopted methods for assessing expenditure priorities do not necessarily identify those sites at which the greatest potential for accident reduction exists.

An alternative method for the generation of more cost-effective programmes of works is proposed and justified. This method adopts the rationale of identifying those sites at which accidents are occurring in higher numbers than would otherwise be expected for such sites with equivalent traffic volumes and locations.

The justification for the method involves detailed statistical analyses of over 10,000 accidents occurring in Lothian Region for the years 1979-1982 which demonstrate that there are significant relationships between accidents and traffic volumes and location details (eg junction type, form of junction control, adjacent roadside development and carriageway type). On this basis, models for accident occurrence have been determined. The analyses show that the temporal distribution conforms with a Poisson process and that the spatial distribution is negative binomial.

It is shown - for both links and junctions - that whilst there are significant differences between the models for different accident types, they do not, in aggregate, produce significantly better models for all accidents than simple all accident models.

In addition, the importance of regression-to-mean has been established as an effect which should be accounted for not just at the monitoring stage of completed schemes but as an integral part of the initial site selection process.

Finally, it is demonstrated that the proposed method, which is called Potential Accident Reduction (PAR), may provide an improvement of cost-effectiveness of road safety expenditure of up to 29% over the currently adopted methods.

ACKNOWLEDGEMENTS

This project would not have been a practical proposition without the kind permission of the Chief Constable of Lothian and Borders Police to allow access to the Region's road accident data. Nor would it have been possible without the assistance of the staff, in particular Archie Fairley, of the Council's Computer Services who took the trouble to interrogate the accident files held on computer and provide the raw data files on magnetic tape in appropriate format for input to SIR.

I am indebted to a number of people for their contributions to the project. They are:

the Director and members of staff of the Department of Highways of Lothian Regional Council for their encouragement and interest in the work;

Geoff Cohen of the Statistics Department of the University of Edinburgh for his help and guidance on many statistical aspects and in particular for his introduction to the computer program GLIM;

Dr Steve Lawson for his advice, interest and encouragement throughout the whole project;

Dr Andrew Metcalfe of the Department of Engineering Mathematics of the University of Newcastle upon Tyne for his considerable help and, in particular, for guiding me through some of the finer points of the use of GLIM; and

my wife, Elizabeth, for her forebearance and patience during the five years.

I am, however, especially indebted to the Trustees of the Rees Jeffreys Road Fund for their generous financial backing and kind understanding throughout the work.

CHAPTER 1 INTRODUCTION

1.1. Background

Under the terms of Section 8 of the Road Traffic Act 1974, each local authority in Great Britain has a duty to promote road safety and

> ".....shall prepare and carry out a programme of measures designed to promote road safety, and shall have power to make contributions to the cost of measures for promoting road safety taken by other authorities or bodies."

In fulfilling these responsibilities, local authorities invest public money in projects designed to improve road safety, the effectiveness of which can be assessed by determining some measure of accident reduction per unit of expenditure. In practice, accidents are assigned financial values and the effectiveness of expenditure is generally expressed in terms of a first year rate of return (FYRR).

It is through the application of low cost engineering measures that local authorities generally discharge their Section 8 responsibilities. There are four alternative approaches available which are not mutually exclusive: local authorities can and do mix their strategies to achieve local objectives. These four alternative approaches are:

Single Site:	treatment of individual locations identified as hazardous road locations by some numerical or statistical criterion. They are more often referred to as **blackspots** or **blacksites.**
Route Action:	treatment of lengths of road identified as hazardous.
Area Action:	treatment of areas or neighbourhoods identified as hazardous.
Mass Action:	treatment of particular accident types (eg skidding accidents) at a number of individual sites throughout the whole or selected part of a local authority area.

Many local authorities maintain a very high profile in their road safety work. Hertfordshire County Council [1984], for example, has claimed a FYRR of 197% in saving some 284 accidents per year at 187 treated sites at an average cost for remedial work of £8,000 per site. The sites were identified and ranked as blacksites on the basis of three-year accident totals.

Recently, attention has been given to two statistical aspects of accident analysis, both having a direct bearing on the work of local authorities: they are the topics of "regression-to-mean" and an alternative blackspot ranking criterion called "potential accident reduction".

The site selection criterion of high accident total does not necessarily provide a set of sites at which the greatest potential for cost-effective accident reduction exists. An alternative selection criterion of potential accident reduction (PAR) has been suggested by the author (McGuigan [1982]) and earlier by Jorgensen [1972] in which sites are ranked according to the difference between the observed number of accidents and the expected number for the type of site and its level of traffic flow. The use of PAR has been examined by Maher and Mountain [1988] who concluded that its effectiveness as a ranking criterion depended on the accuracy of the prediction of the expected number of accidents.

In addition, Abbess et al [1981], Hauer [1980] and McGuigan [1985] have focused in on the statistical phenomenon of "regression-to-mean" (alternatively referred to as "bias by selection"), the effect of which would suggest that claims by local authorities may considerably overstate the true accident reductions achieved. The line of argument is that the numbers of accidents occurring at sites selected for remedial work on the basis of high accident totals, will not be representative of their long term average numbers. Accordingly, sites selected by this method will be biased towards those at which accidents have occured, in the short term, at higher than expected levels on the basis of chance alone. In general, such sites will, in future years, maintain their average long term level of accident occurrence and would, accordingly, demonstrate accident reductions even if no remedial works were undertaken.

Indeed, the author in another context, but on the basis of the current study data, has shown a 33% reduction in accidents at 61 of Lothian Region's worst untreated junction sites (McGuigan [1985] Table IIB refers). Whilst it would be uncharitable to suggest that Hertfordshire County Council has not engineered some real long term reductions in accidents, the effect of regression-to-mean should form part of the early considerations associated with site selection. There exists, therefore, some possible scope for improving the methods of determining priorities for road safety expenditure on blackspot treatment and monitoring its effectiveness.

The current Department of Transport's Accident Investigation Manual (Department of Transport [1986]) recognises the importance of the regression-to-mean effect and provides some guidance on the calculation of its magnitude. The Manual, surprisingly, suggests that the regression-to-mean effect be considered only during the monitoring of the effectiveness of a treatment <u>after</u> completion of the remedial road works. The fact that the effect may have a significant role to play in the determination of priorities for treatment has, therefore, either been overlooked or ignored (the latter perhaps for practical reasons).

1.2. Objectives of the Thesis

The objectives can be broken down into two main themes which are:

1. to model the relationships between accidents and traffic at sites classified by a number of factors, including the type of site and its location; and

2. to examine the justification for PAR as a ranking criterion by determining the effectiveness of the model in predicting accidents.

1.3. Study Method

The study is based on the data available for those personal injury road accidents reported to the Police which occurred on the Lothian Region's computerised node and link model of the major road network during the four year period 1979-1982. The data are held on a locational basis together with relevant traffic flow information. The data have not been the subject of any sieving process (with the sole exception of the exclusion of sites which were the subject of road improvements during the study period) and, therefore, constitute a source of information comprising the whole range of location types encountered in Lothian Region, whether or not they were recognised as blackspots.

This method of data collection differs, therefore, from that of many other studies in which the data have been collected according to sets of tightly defined criteria. Such processes are quite validly designed to eliminate or minimise the effects on the variation of accident occurrence caused by factors not forming part of the study. The results of such studies, however, may not be appropriate for wider application. As one of the purposes of road safety remedial work is to identify and treat blackspots - whatever the cause - the Lothian Region data provide a unique source of location based information which has been the subject of neither sieving nor modification.

It is stressed that the purpose of the study was not to gather detailed road layout information (eg, for junctions; entry widths, entry radii, gradients on approaches and sight distances) about individual locations to determine their effect on accident occurrence. Indeed, quite the opposite is true, as the purpose was to model accident occurrence with a view to highlighting those locations at which a higher than expected hazard exists, perhaps indicating some poor feature of the road layout which might respond to remedial treatment. The former approach, as indicated above is a perfectly valid technique in determining the difference between good and bad design standards but is not relevant to the local authority with a responsibility to promote road safety on a road network perhaps comprising many thousands of kilometres.

Independent databases have been set up for link and junction locations, each of which included a considerable number of locations with zero accident totals which were retained throughout the study. The inclusion of these zero accident sites was considered to be an important feature in determining the statistical distribution of accident occurrence between sites to assess the most appropriate form for the model.

Having assessed the best form for the model, a data splitting exercise was undertaken to determine:

- the consistency of the model with time;
- its predictive accuracy; and
- the justification for the use of PAR as a blackspot ranking criterion.

1.4. The Structure of the Thesis

The thesis starts with a detailed description of the individual data items (Chapter 2) and is followed by a general breakdown of the numbers of accidents disaggregated by type of location and other site specific factors (Chapter 3).

Following some preliminary observations on the distribution of accidents (Chapter 4) there follow detailed analyses of both the temporal distribution (Chapter 5) and the spatial distibution of accidents (Chapter 6).

On the basis of the analyses of accident distribution an appropriate form of model is developed (Chapter 7) and used, in turn, to examine the junction accident data (Chapter 8) and the link accident data (Chapter 9) with a view to determine appropriate models for accident occurrence.

The predictive power of the models at individual sites is determined (Chapter 10) and, further, the justification for the use of PAR as a ranking criterion, in preference to other criteria, is examined (Chapter 11).

CHAPTER 2 DESCRIPTION OF THE DATA

2.1. Introduction

The purpose of this work is to explore the relationships that may exist between road accidents and traffic flows on different parts of the road network with a view to determining the extent to which accident occurrence can be explained by traffic flow and other site specific parameters.

To obtain appropriate data for this type of work it is necessary to have both road accident and traffic flow information related to individual elements of the road network.

The Lothian Regional Council's road accident recording and analysis system, LORASS (LOthian Road Accident Statistics System), provides a comprehensive source for such data (Lothian Regional Council [1979]).

2.2. Background to LORASS

Following Scottish local government re-organisation in 1975 the newly formed Lothian Regional Council inherited a range of different accident recording systems whose combination did not prove to be entirely satisfactory. At much the same time the Department of Transport's Standing Committee on Road Accident Statistics was beginning to "firm up" on the format of the new STATS 19 document which was – at latest – to take effect from 1st January 1978.

It was decided that this presented an opportunity for a complete review of the accident statistics system and a working party was set up in the latter part of 1975 comprising officers of the Lothian and Borders Police Force and the Region's Departments of Highways and Finance (Computer Division).

It was through this working party that the concepts behind LORASS were first formulated and subsequently implemented by 1st January 1978.

2.3. A Description of LORASS

The backbone of LORASS is a node and link network which comprises all Motorway and A classified roads in the Region together with:

- selected B classified roads in the rural areas of the Region; and
- all B classified and major distributor roads in the City of Edinburgh.

It was not considered practicable at the inception of LORASS to set up a network for all roads in the Region. The above network, however, was specifically selected to capture about 75% of the accidents occurring in the Region.

The computer program TRAMS (Transport Referencing and Mapping System) was installed by the Region to support this model of the road network. TRAMS was developed by TRRL (Transport and Road Research Laboratory) and is based on a computerised description of the road network as a series of straight lines representing the centre line of each constituent road. Such a network which comprises some 14,000 nodes and 13,000 segments was digitised to a notional one metre accuracy. This network is known as the Accident Network.

In the Accident Network, the centre line of each carriageway of all dual-carriageway roads has been independently digitised so that accidents can be assigned to the appropriate carriageway for retrieval and analysis purposes.

To improve accident retrieval and "black-spot" analysis techniques, an early decision was made to store accidents by three categories:

- at (or within 20 metres) of junctions on the Accident Network;
- on links between junctions on the Accident Network; and
- at locations not on the Accident Network.

TRAMS itself is a relatively unsophisticated referencing system which neither recognises junctions nor enables the user to define a link as being the sum of a set of adjacent segments. Accordingly, a Junction File and a Link File were were created to represent a higher level description of the TRAMS network. The following four definitions will make the relationships clear.

NODE	a point defined by a twelve digit Ordnance Survey Grid Reference. It may be attached to one or more SEGMENTS.
SEGMENT	a line joining two NODES.
JUNCTION	a NODE connected to three or more SEGMENTS. There are some 3,250 JUNCTIONS on the Accident Network.
LINK	a series of one or more SEGMENTS connecting two adjacent JUNCTIONS. There are some 2,700 LINKS on the Accident Network to which accidents can logically be assigned (ie LINKS of sufficient length to have stretches outside the influence of JUNCTIONS).

2.4. Annual Traffic Flows

Both the Link and Junction Files are cross-referenced to the Region's traffic count system which is interrogated on a regular basis to provide estimates of annual average traffic flow for each link and annual number of vehicles negotiating each junction. The large number of links on the accident network precludes actual measurement of flow for each link over the a whole year. It was, therefore, necessary to adopt a reliable method of estimating annual traffic volumes from short-term counts (such counts might typically range from a few hours manual count to a many weeks ad hoc automatic count). As a working rule, manual counts are normally updated on a three-year cycle.

In Lothian Region there are 28 permanent automatic traffic counter sites which continuously monitor traffic on a cross-section of road types and geographical areas. Annual traffic volumes for these permanent sites can be easily calculated.

Each link on the Accident Network is cross-referenced to both a short-term count and to the most appropriate permanent counter site (known as the Reference Station) and an estimated annual traffic volume is calculated thus:

$$A_e = A_r.(s_e/s_r) \qquad (2.1)$$

where:

A_e is the estimated annual flow for the link with short-term count s_e;

A_r is the observed annual flow for the reference station;

s_e is the observed short-term count for the link; and

s_r is the matching observed short-term count for the Reference Station (ie for the identical time period as s_e).

On the basis of unpublished research on Lothian Region data it has been shown that where the Reference Sites are representative of the short-term count location, 95% of short-term counts of as low as two hours duration can provide estimates for annual flow to within 20% of the true annual flow.

2.5. Accident Data

All STATS 19 forms (which have been specifically modified for use with LORASS) are forwarded by Lothian and Borders Police to the Department of Highways for appropriate coding for assignment to the Accident Network by the addition of a ten digit Ordnance Survey Grid Reference and - where appropriate - network identifiers which are vetted by LORASS for continuity before being added to the appropriate file.

Accidents with no network identifiers are placed solely on the Grid Reference Accident File, whilst accidents with network identifiers are placed on either the Link File or the Junction File as appropriate and on the Grid Reference Accident File. Over and above the routine STATS 21 (Department of Transport et al [1977b]) checks, the accident data are fully vetted for continuity with the Accident Network both in terms of grid reference and locational details.

Accidents are, in this way, accumulated at individual network junctions and links and can be related to estimated traffic flow figures.

2.6. Accident, Road Network and Traffic Data for the Current Study

2.6.1. General

For the purposes of this study it was decided to use data unaffected by changes in legislation likely to have major effects on accident frequency. This precluded the use of data beyond 31st January 1983 which saw the start of the compulsory seat-belt wearing legislation in the United Kingdom. For practical reasons, data for the month of January 1983 were not used. In view of the difficulties associated with reconstituting previously archived data prior to 1979, it did not prove possible to obtain data prior to that year. In any event, in 1978 there were some changes in legislation which may have had an effect on accident occurrence: they were the making permanent of the 60 and 70 miles per hour speed limits and new rules relating to the number of hours which may be worked by goods vehicle drivers. There were no such major relevant changes in legislation until the seat belt legislation in 1983. Consequently, data for the four year period 1979-1982 (inclusive) were used for the study.

The LORASS Junction and Link Files on the Region's mainframe computer were interrogated to provide the data required for this study on magnetic tape for transfer to the Edinburgh Regional Computing Centre (ERCC) where a comprehensive range of computing facilities was avaliable.

The data were read into and stored in hierarchically structured databases using SIR (Scientific Information Retrieval, Robinson et al [1980]). Entirely separate databases were constructed for the Link and the Junction data.

SIR is a well documented database management system ideally suited to the needs of a road accident database which has a heirarchical structure. In the case of the study data, a site (ie a junction or a link) "owns" nil, one or more accidents, with each accident "owning" one or more vehicles and each vehicle, in turn, "owning" nil, one or more casualties. Data structured like this can be the subject of comprehensive interrogation by procedures incorporating high level SIR retrieval commands. In addition, SIR can generate SPSS (Statistical Package for the Social Sciences, Nie [1983]) system files and either SIR or SPSS is capable of producing raw data files for subsequent analysis by other packages such as GLIM (Generalised Linear Interactive Modelling, Baker and Nelder [1978]).

Before progressing further, all links and junctions which had been the subject of either remedial works (other than normal maintenance) or significant changes in trafic flow (perhaps, for example, as a result of the opening of a new bypass) during the study period (ie 1979-1982) were deleted from the databases to provide source information for sites enjoying as near homogeneous conditions throughout the study period as was reasonably possible. This deletion process, which was undertaken before the databases were set up for use, was based on information provided by Lothian Region's Road Accident Investigation Team. In this way, 138 junctions were deleted from the 3,250 which were then "live" on the Accident Network and, similarly, 172 links were deleted from the "live" total of 2664 links.

2.6.2. Junction Data

For the purposes of defining the LORASS Accident Network, a junction was defined as a place where two or more public roads meet, including those parts of the roads within 20 metres of the extended kerblines of adjacent entries (this is the definition adopted by the Department of Transport for road accident purposes). Accordingly, all junctions (excluding private drives and accesses) were modelled in the Accident Network irrespective of the magnitude of the side road traffic flow. This was achieved by the use of "dummy" links to delineate the centre line of non-Accident Network side roads. These "dummy" links do not constitute Accident Network links on the Link File. Each junction is identified by its LORASS node number (JREFRNCE).

The junction details adopted for the present study are similar to those used on the STATS 19 form but with some important differences. These differences relate to one of the objectives of LORASS which was to assign accidents to logical elements of the road network which would aid accident retrieval and analysis work. This objective led to the development of a network which did not simply represent a complex junction as a single node but as a series of nodes each representing particular traffic conflict points. In this way a roundabout, for example, was coded as a series of nodes with each node representing the intersection of the roundabout's circulating carriageway and an approach carriageway.

The STATS 19 form (Department of Transport et al [1977a]) enables the following junction types to be coded:

1. Roundabouts
2. Mini-roundabouts
3. T- or staggered junctions
4. Y-junctions
5. Sliproads
6. Crossroads
7. Multiple junctions
8. Private drives
9. Other junctions.

These types have been adopted with the following exceptions and clarifications:

- the manner in which the node and link network was applied to the road system has meant that roundabouts may be coded according to one of two different methods according to whether or not the approaches are channelised. Figure 2.1 shows the two methods and it should be noted that each node within the roundabout is treated as a separate junction for accident purposes. It can be seen that roundabout junction details can take one of three basic forms to allow for the following manoeuvres:

    - diverging and merging;
    - diverging only; and
    - merging only;

- T-junctions are those conventional 3-way junctions whose intersection angles exceed 60 degrees;
- Y-junctions are those conventional 3-way junctions whose intersection angles are less than 30 degrees;
- crossroads are those conventional 4-way junctions where the alignments of both roads are uninterrupted irrespective of their intersection angle;
- private drives are not considered to be junctions for the purpose of the present study and are included in the link accident data;
- slip roads and multiple junctions have been recoded as other junctions; and
- other junctions include:

    - conventional 3-way junctions whose intersection angles exceed 30 degrees but are less than 60 degrees;
    - slip roads;
    - multiple junctions; and
    - all nodes within complex junctions (but not roundabouts); eg consider a T-junction on a dual-carriageway (see Figure 2.2) which, dependent on its layout, may have 5 internal locations - each coded as an independent node - where there is traffic conflict perhaps involving different manoeuvre patterns.

In this way, nine junction types (JUNCTYPE) were identified:

1. Roundabout Merge
2. Roundabout Diverge
3. Roundabout Other
4. T-junction
5. Y-junction
6. Crossroads
7. Other 3-way
8. Other 4-way
9. Other 4^{+}-way

In addition, for each junction a location code (JLOCATION) was attached to allow a distinction to be made between (1) rural, (2) suburban, (3) local centre and (4) Edinburgh city centre junctions. A further code (JCONTROL) distinguished between (1) priority and (2) signal controlled junctions.

The Edinburgh city centre code was reserved for those junctions lying within the city's central area controlled parking zone. Local centre codes were used for junctions associated with roads providing a comprehensive range of "high street" shopping facilities. The suburban code was adopted for all other junctions for which a speed limit of 40 miles per hour or less applied. Rural codes were used for all junctions for which a speed limit of over 40 miles per hour applied.

Large complex junctions (eg roundabouts) were defined as series of "internal" junctions or nodes to allow accidents to be assigned to logical spatial units, thus, enabling a more rational means of site by site comparison to take place. For example, simple comparison of a conventional three-way roundabout with a large six-way roundabout would not be sensible without a whole series of qualifications. However, if the roundabouts are "dismantled" to provide, in total, nine merging and nine diverging junctions, a more satisfactory comparison could be carried out on the operation of the individual elements of the roundabouts.

These codes give rise to 72 (ie 9x4x2) junction type categories which when combined with three traffic flow bands (see Table 2.1) provide 216 potential junction disaggregations. Table 2.2 shows how sitecodes have been generated to simplify reference to these disaggregations.

Lawson [1986] has shown - for accidents on radial roads in Birmingham, UK - that the frequencies of certain accident types (eg pedestrian accidents and single non-pedestrian accidents) are significantly affected by adjacent roadside land use. It was, therefore considered logical to categorise accidents by type and, because of the difference between junction and link accident patterns, the accident types for two vehicle accidents at junctions are different to those adopted for the link accidents. It should, however, be noted that the definitions of the accident types adopted for the present study do vary from those in other studies (eg Lawson [1986]). For example, because a PSV accident type has been defined, this has the effect of removing such accidents from from the single, two and multi- vehicle definitions. The accident types, for junctions, are described as follows:

<u>Single vehicle non-pedestrian accidents</u>

JSVNP — Accidents involving a single moving vehicle with one or more casualties but where neither pedestrian nor solely PSV casualties were involved.

<u>Two vehicle non-pedestrian accidents</u>

Accidents involving two moving vehicles with one or more casualties but where neither pedestrian nor solely PSV casualties were involved categorised by the following manoeuvre patterns (see Table 2.3):

JTVNP1 — Diverging.

JTVNP2 — Merging.

JTVNP3 — Non-conflicting manouevres.

JTVNP4 — Rear-end shunts.

JTVNP5 — Both vehicles going straight ahead - at a 90^0 angle to each other.

JTVNP6 Right turn from nearside of a straight ahead vehicle.

JTVNP7 Right turn from offside of a straight ahead vehicle.

JTVNP8 Both vehicles going straight ahead – head on.

JTVNP9 Other turns involving conflicting manoeuvres including U-turns.

Multiple vehicle non-pedestrian accidents

JMVNP Accidents involving three or more moving vehicles with one or more casualties but where neither pedestrian nor solely PSV casualties were involved.

Pedestrian accidents

JPED Accidents involving at least one pedestrian casualty.

Public Service Vehicle accidents

JPSV Accidents involving at least one Public Service Vehicle (PSV) passenger casualty and no other road users.

The sum of these thirteen accident types provides the total number of accidents for each junction (JACCTOT).

Table 5.1, in another context, shows the annual accident totals for junctions disaggregated by the above accident types.

2.6.3. Link Data

As the definition of a junction for the purposes of this study includes what is generally understood to be all junctions, it is, therefore, axiomatic that the definition of a link for this study excludes all junctions. This definition differs from that of other studies where minor junctions are included as part of the link definition (eg Chapman [1978] and Maltby and Bennett [1986]). In addition, this link definition implies variable link lengths which – in the study data – range from 1 metre to over 8 kilometres in length. Each link is identified by a reference number (LREFRNCE) which comprises the route number together with its start or "A" node.

The length of each link (LINKLENGTH) is calculated as the distance in metres between two adjacent junctions but exclusive of the junctions (see Figure 2.3).

Chapman [1978], in his study of accident and development data for urban arterial roads concluded that accident rate varied according to the roadside development.

The author, in another context (McGuigan [1982]), has similarly shown, with Lothian Region link accident and development data, that accident rates do vary according to the link categories defined by Chapman [1978]. In addition to the categories adopted by Chapman, McGuigan included the further two development types of "commercial" and "rural". The possibility of adopting further categories was considered but rejected because the overlap in the observed accident rates for the different development types outlined by McGuigan did not suggest that any new development types would produce accident rates statistically independent from those already established.

The adopted development type codes codes are:

1. Urban: shops
2. Urban: Commercial Premises (non-industrial)
3. Urban: Industrial Premises
4. Urban: Residential Premises
5. Urban: Open Ground (urban recreational etc, but not rural)
6. Rural
7. Urban: Other (not capable of being coded 1-5)

The urban categories refer to links which are subject to speed limits of 40 miles per hour or less and rural categories to links subject to speed limits of 60 or 70 miles per hour (there being no 50 miles per hour speed limits in Lothian Region).

The roadside development codes did not prove capable of absolute definition as some degree of subjectivity was necessary in a small minority of instances. For codes 1-5, the relevant development type was used where it was observed along at least 75% of the length of the adjacent ground-level roadside. Where the development type was mixed (ie where none of the specified development types exceeded 75% of the roadside length) or predominantly of an unspecified development type (eg school or hospital) a code 7 was adopted. Code 6 was used to define rural links (ie links subject to a speed limit of over 40 miles per hour) irrespective of the development type.

When both sides of the road are considered, these codes provide 22 potential development categories (LDEVTYPE) (the 6 'starred' categories are illogical because of the speed limit distinction between urban and rural links; they are, therefore, excluded) viz:

11	12	13	14	15	16*	17
	22	23	24	25	26*	27
		33	34	35	36*	37
			44	45	46*	47
				55	56*	57
					66	67*
						77

In addition to the adjacent roadside development, the carriageway type (LCWYTYPE) was recorded, where:

1 = dual-carriageway
2 = single carriageway
3 = one-way street

giving rise to 66 (ie 22x3) link categories which when combined with three traffic flow bands gives rise to 198 potential link disaggregations. Table 2.2 shows how sitecodes have been generated to simplify reference to these disaggregations.

Each accident in the link file has been categorised to enable analyses of different accident types to be undertaken (the accident types for links are different to those adopted for junction accidents). The accident types are described as follows:

<u>Single vehicle non-pedestrian accidents</u>

LSVNP — Accidents involving a single moving vehicle with one or more casualties but where neither pedestrian nor solely PSV casualties were involved.

Two vehicle non-pedestrian accidents

Accidents involving two moving vehicles with one or more casualties but where neither pedestrian nor solely PSV casualties were involved categorised by the following manoeuvre patterns :

LTVNP1 Rear-end shunts or two vehicles travelling in the same direction.

LTVNP2 Head-on collisions or two vehicles travelling in opposing directions.

LTVNP3 Turning at a private drive.

LTVNP4 Turning not at a private drive (including U-turns).

Multiple vehicle non-pedestrian accidents

LMVNP Accidents involving three or more moving vehicles with one or more casualties but where neither pedestrian nor solely PSV casualties were involved.

Pedestrian accidents

LPED Accidents involving at least one pedestrian casualty.

Public Service Vehicle accidents

LPSV Accidents involving at least one Public Service Vehicle (PSV) passenger casualty and no other road users.

The sum of these eight accident types provides the total number of accidents for each link (LACCTOT).

Table 5.3, in another context, shows the annual accident totals for the link data disaggregated by the above accident types.

2.6.4. Traffic Data

As outlined above, the traffic flow measurement in the LORASS Junction File is that of annual number of vehicles negotiating the junction. For the purposes of this study, however, the Junction File was interrogated in parallel with the Link File in order to provide - for each junction - the annual traffic volumes entering and leaving on each arm. This provided for a greater flexibility in calculating alternative traffic flow measurements at junctions.

On this basis it was possible to determine the major road flow (JMAINFLOW) and the side road flow (JSIDEFLOW) for each junction. Both JMAINFLOW and JSIDEFLOW are measured in units of annual vehicle flow $\div 10^6$ (MV). JMAINFLOW was calculated as the average of the combined flows (or one-way flows where appropriate) for the two busiest arms and JSIDEFLOW as the average combined flow for the remaining arms. From the LORASS files it was not possible to ensure that the major and side road flows conformed with the layout of the site and the disposition of the traffic flows: it is possible, therefore that in a number of instances JMAINFLOW and JSIDEFLOW do not accord with the priorities in existance. From these flows it was possible to calculate the product of JMAINFLOW and JSIDEFLOW (JPRODUCT) and its square root (JROOTPROD).

For links, the traffic flow (LINKFLOW) expressed in units of MV has been multiplied by the link length (LINKLENGTH) expressed in kilometres to provide an annual travel distance estimate (LMVKM) measured in units of MVkm.

At a majority of junctions, the contribution of turning traffic to flows on adjacent sections of road is negligible. Consequently, it has not been considered necessary - for LORASS purposes - to count traffic on every link on the Accident Network and, therefore, individual counts have been used to assess traffic flows on series of adjacent links. Side roads at many minor junctions were initially designated nominal traffic flows on the basis of local experience and although a programme to update these flows is continuing, not every side road has been the subject of a specific count.

To determine, at a preliminary level, whether or not the variation in accident totals is related to traffic use, both link and junction sites have been further disaggregated into three traffic use bands (LMVKMBAND and JROOTBAND) entitled Low, Medium and High.

These bands have been specifically selected to divide the data into three approximately equal parts. The band widths are are shown in Table 2.1. Table 3.12, for example, in another context, shows how these traffic flow bands together with other site related characteristics have been combined to enable disaggregations of the data. Although, potentially, there are some 198 link and 216 junction disaggregations, it will be shown later that there is an insufficient volume of data to generate meaningful sample sizes for a majority of the disaggregations. In addition, in later chapters the junction and the link data have been grouped to form contingency tables in which JMAINFLOW and JSIDEFLOW for junctions and LINKFLOW and LINKLENGTH for links were used as classification criteria. To do this, these four variables were distilled into four new variables (ie JMAINBAND, JSIDEBAND, LFLOWBAND and LENGTHBAND) each of which divides the data into three approximately equal parts. These band widths are also shown in Table 2.1.

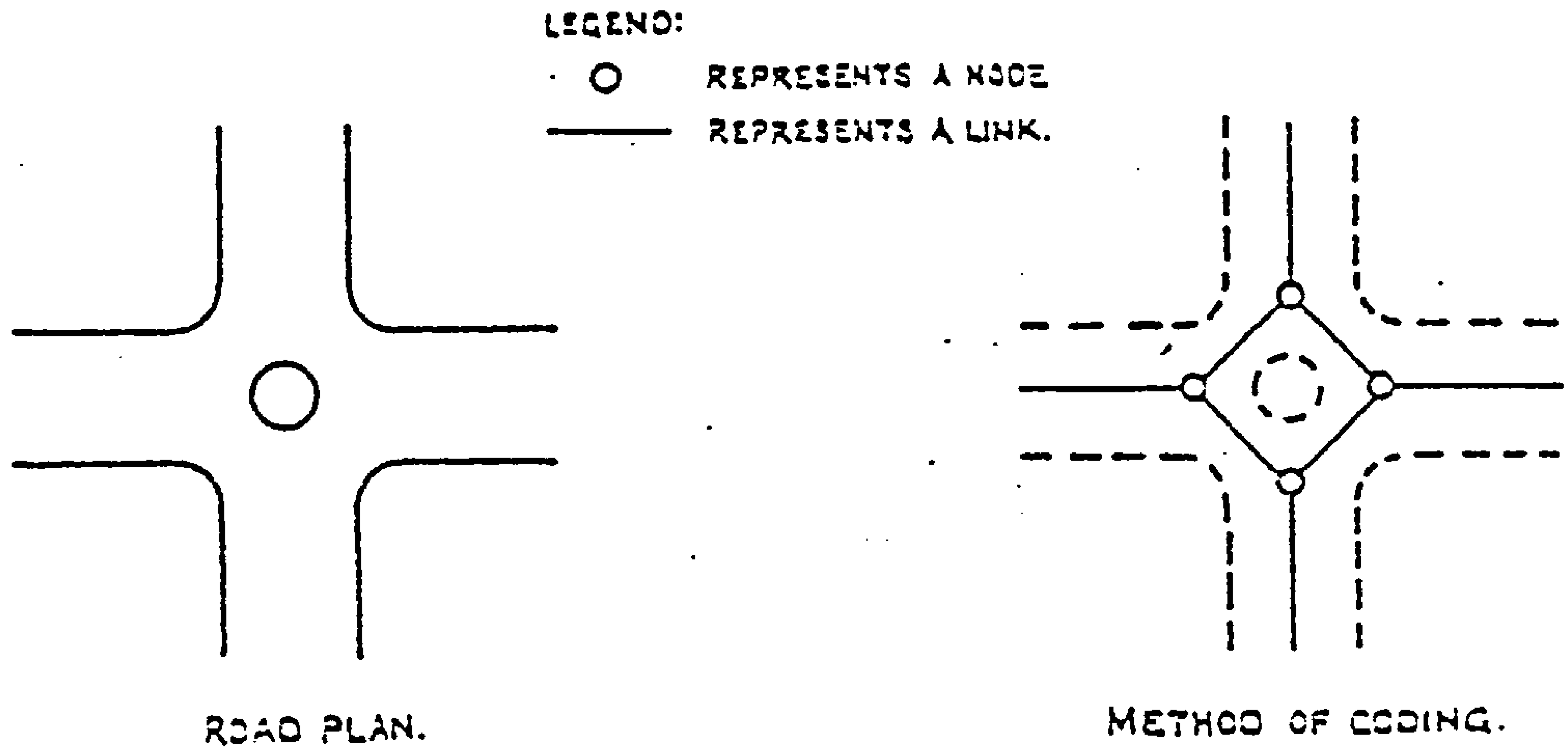
Appendix I provides a glossary of the variable names together with definitions.

2.6.5. Sitecode Numbering Convention for Tabulations

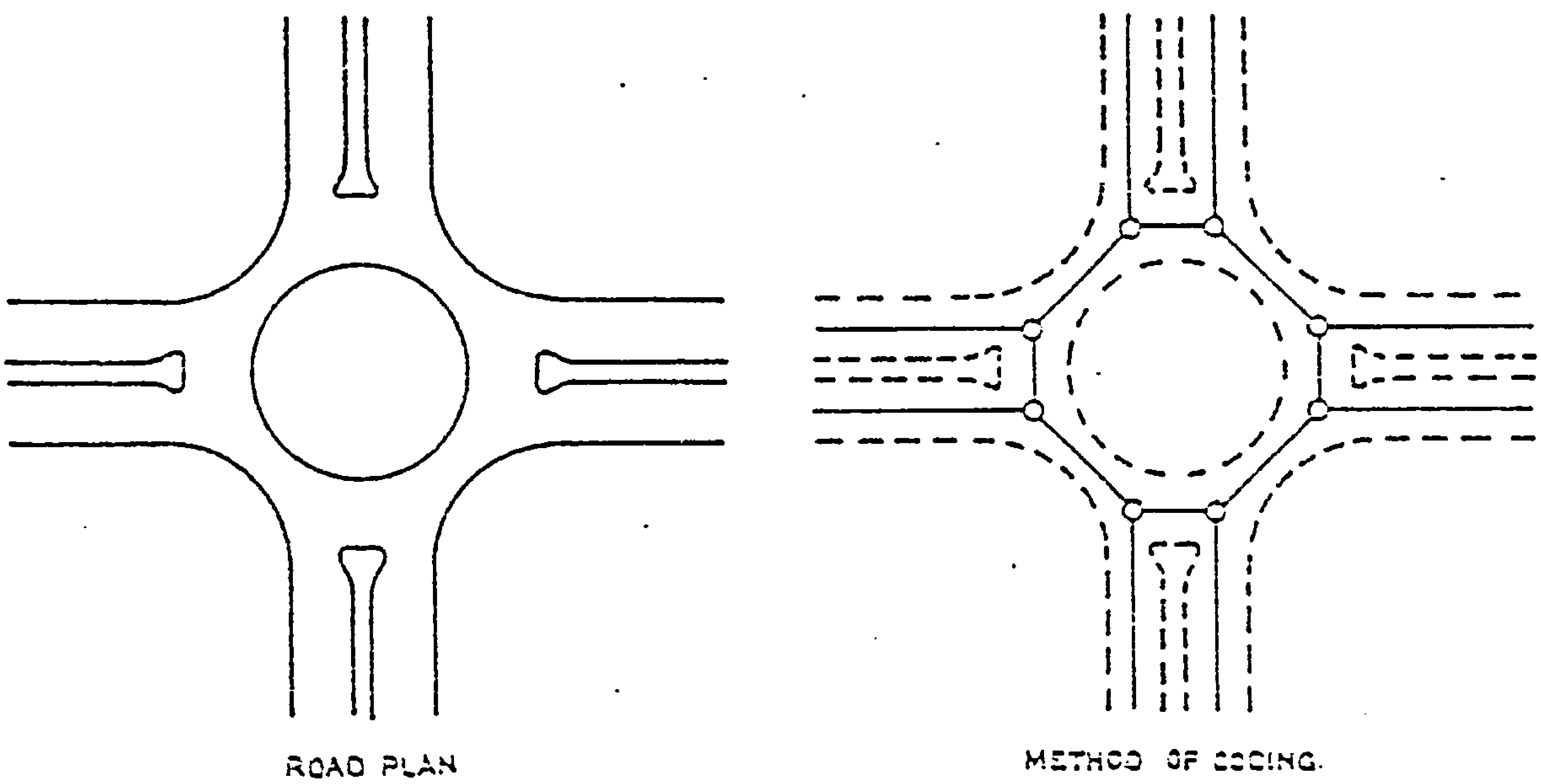
For tabulation purposes the sitecode numbering convention for re-aggregated data uses value ranges or comma separated values contained within brackets; eg J4(2-4)1(2,3) refers to non-rural (ie suburban + local centre + city centre) priority controlled T-junctions with either medium or high traffic flow charateristics. In addition, for ease of reference a code "X" has been used to indicate all values for a specific part of the sitecode; eg J4X1X refers to all priority controlled T-junctions (irrespective of location and/or traffic flow).

In a number of the tabulations "wild-card" characters (eg "#" or "£") have been adopted where appropriate to simplify cross-tabulation of the data.

FIGURE 2.1: NETWORK DEFINITION OF A ROUNDABOUT

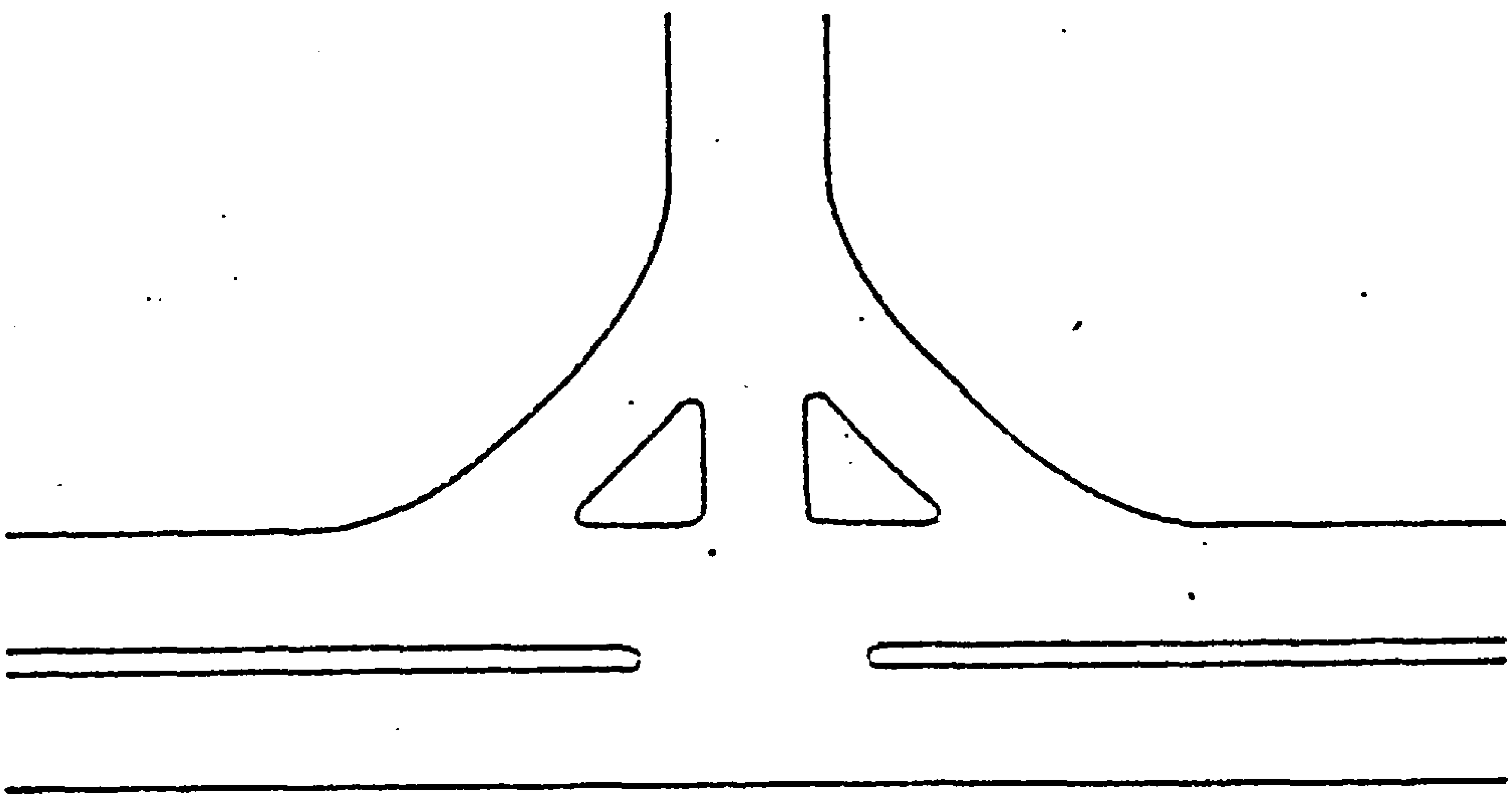


(a) UNCHANNELLISED APPROACHES.

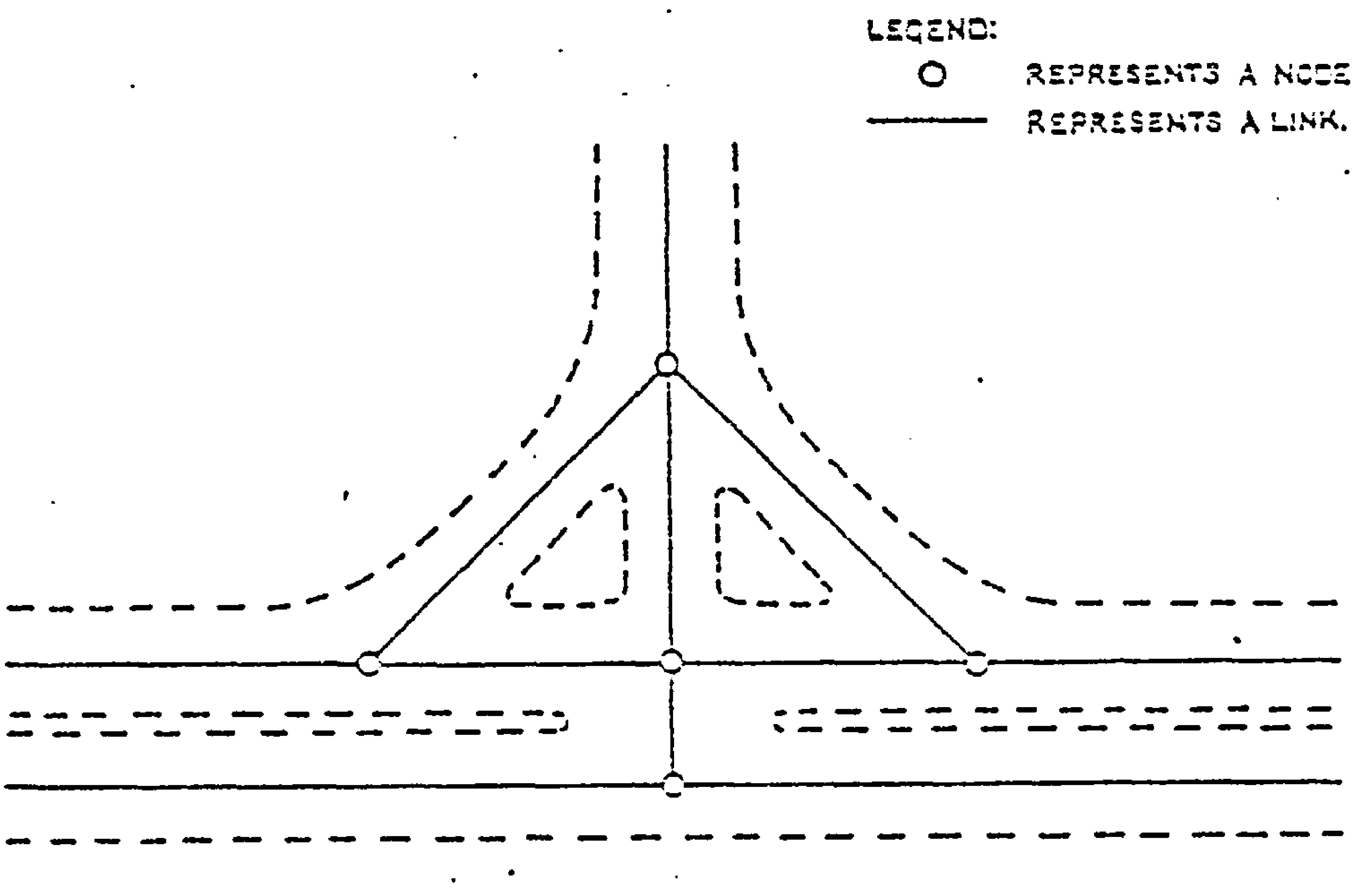


(b) CHANNELLISED APPROACHES.

FIGURE 2.2: NETWORK DEFINITION OF AN OTHER JUNCTION



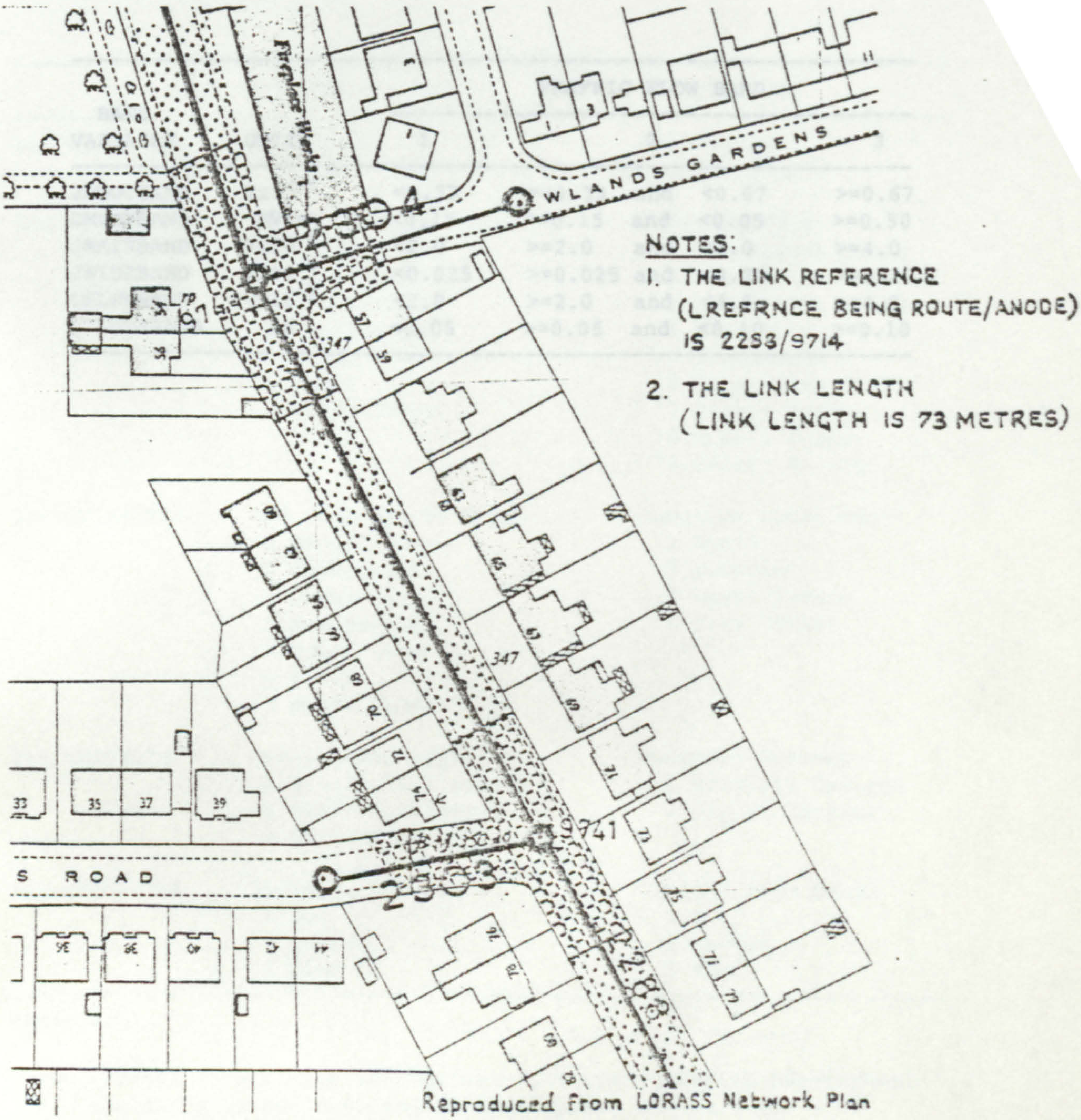
ROAD PLAN.



METHOD OF CODING.

EXAMPLE OF METHOD OF CODING A T-JUNCTION AT
A DUAL CARRIAGEWAY

FIGURE 2.3: NETWORK DEFINITION OF A LINK



KEY:

——→	DIRECTION OF ROUTE
(hatched box)	AREA OF JUNCTION
(dotted box)	AREA OF LINK
2288	ROUTE NUMBER
9741	NODE NUMBER

TABLE 2.1: TRAFFIC FLOW RANGES FOR BAND VARIABLES

BAND VARIABLE	UNITS	TRAFFIC FLOW BAND 1	2	3
JROOTBAND	$Vx10^{-6}$	<0.33	>=0.33 and <0.67	>=0.67
LMVKMBAND	MVkm	<0.15	>=0.15 and <0.05	>=0.50
JMAINBAND	$Vx10^{-6}$	<2.0	>=2.0 and <4.0	>=4.0
JSIDEBAND	$Vx10^{-6}$	<0.025	>=0.025 and <0.20	>=0.20
LFLOWBAND	$Vx10^{-6}$	<2.0	>=2.0 and <4.0	>=4.0
LENGTHBAND	km	<0.05	>=0.05 and <0.10	>=0.10

TABLE 2.2: SITECODE DEFINITIONS

	LINK	JUNCTION
1st CHARACTER	Site Type: L Link	Site Type: J Junction
2nd CHARACTER	1st Development Type: 1 Shops 2 Commercial 3 Industrial 4 Residential 5 Urban Open 6 Rural 7 Urban Other	Junction Type: 1 Roundabout Merge 2 Roundabout Diverge 3 Roundabout Other 4 T-Junction 5 Y-Junction 6 Crossroads 7 Other: 3-way 8 Other: 4-way 9 Other: 4+-way
3rd CHARACTER	2nd Development Type: 1 Shops 2 Commercial 3 Industrial 4 Residential 5 Urban Open 6 Rural 7 Urban Other	Junction Location: 1 Rural 2 Suburban 3 Local Centre 4 City Centre
4th CHARACTER	Carriageway Type: 1 Single Carriageway 2 Dual-carriageway 3 One-way Street	Junction Control: 1 Priority Control 2 Signal Control
5th CHARACTER	Traffic Flow Band: 1 Low 2 Medium 3 High	Traffic Flow Band: 1 Low 2 Medium 3 High

Examples:

L1412 refers to a single carriageway link with shopping development on one side and residential development on the other with medium traffic flow characteristics.

J42P3 refers to a suburban priority controlled T-junction with high traffic flow characteristics.

TABLE 2.3: DEFINITION TABLE FOR TWO VEHICLE NON-PEDESTRIAN ACCIDENT CATEGORIES FOR JUNCTIONS[1]

VEHICLE 2	VEHICLE 1			
	FROM TO	FROM TO	FROM TO	FROM TO
FROM TO	N N	N E	N S	N W
N N	(JTVNP) 9	9	9	9
N E	9	4	1	1
N S	9	1	4	1
N W	9	1	1	4
E N	9	3	6	9
E E	9	9	9	9
E S	9	3	2	3
E W	9	3	5	2
S N	9	3	8	7
S E	9	2	7	9
S S	9	9	9	9
S W	9	3	3	2
W N	9	3	3	3
W E	9	2	5	6
W S	9	3	2	9
W W	9	9	9	9

[1]EXPLANATORY NOTE.

To simplify the categorisation, all two vehicle accidents have been adjusted to a four point compass reference (rather than the eight point one adopted for the STATS 19 document). To further simplify this work, the first vehicle has in all cases been assumed to have been travelling from the north. The actual geographical disposition of the vehicles is irrelevant in the context of this study: it is the relative movements of the vehicles that is of importance here. These relative movements are not affected by this assumption.

CHAPTER 3 OVERVIEW OF THE DATA

3.1. Comparison of Study Data with Lothian Data

Table 3.1 shows, in aggregate terms, the total numbers of accidents in the Region together with those accidents which comprise the study data. It can be seen that some 70% of the Region's accident total form the study data. This is slightly lower than the 75% figure quoted above for the percentage of road accidents occurring on the Accident Network by virtue of the accidents associated with the deleted links and junctions. Nevertheless, the study data are representative of a clear majority of accidents in the Region.

3.2. Comparison of Lothian Data with Great Britain Data

Table 3.2 shows a comparison of the split between the study link and junction accidents and the equivalent data for Great Britain. The study data have a marginally higher proportion of accidents at junctions than those for Great Britain. This observation may be explained by the relatively high proportion of urban road miles in Lothian Region (see below), suggesting a higher than average number of junctions per road kilometre. In addition, the year by year variations in the accident totals do not indicate that there are any reasons to suspect that the study data are out of step with the national data.

For Great Britain, there is an almost equal split between urban and rural road travel. Table 3.3 shows Government statistics for 1979-1982 which, in total, indicate that 49.4% of those years' travel was on built-up roads and 50.6% was on non built-up roads. Whereas Table 3.4 indicates, for the study data, that some 59% of travel was on urban roads (ie 1,468 MVkm on urban roads from a total of 2,476 MVkm).

The overall accident rate for the study data (ie junctions and links for the years 1979 to 1982) is estimated to be 1.01 accidents/MVkm (see Table 3.4) which is some 15% in excess of the equivalent Great Britain rate of 0.875 (see Table 3.7). However, when the rates are calculated independently for rural and urban accidents, better agreement between the Great Britain and study data is observed. For the study data, the rural and urban rates are 0.38 and 1.44 accident/MVkm respectively which can be compared with the equivalent Great Britain rates of 0.41 and 1.35 accidents/MVkm (see Tables 3.5 and 3.6). The study rural rate is some 7% lower and the urban rate some 6% higher than the Great Britain rates and are, accordingly, in fairly good agreement. The higher urban rate, for example, could be explained by what may be in Lothian Region (cf Great Britain) a relatively high proportion of travel on roads with adjacent commercial and shopping development which, it will be seen later, gives rise to relatively higher accident rates. This, however, has not been the subject of formal verification; such verification (not necessarily using Lothian Region data) might prove to be a profitable topic for further research. In any event, it may be that there is a systematic difference between the Great Britain road travel estimates and those for Lothian Region.

3.3. Disaggregation of Study Data

Tables 3.8 and 3.9 show the observed frequencies of accident occurrence for all links and all junctions respectively in the study data. For both junctions and links it is seen that over 40% of sites have a zero accident total.

Table 3.10 provides data for link accidents disaggregated by link category. It can be seen that the higher accident rates (expressed as accidents/MVkm) are generally associated with the more highly urbanised link categories. This is not to say that it is the roadside development itself that causes the accidents but that it is indicative, for example, of the greater vehicle/pedestrian conflict, greater proliferation of street furniture and the more unpredictable nature of traffic conditions associated with more urbanised roads. (All the links which have been coded as category 152 are on Edinburgh's main city centre street, Princes Street, and, consequently, the high accident rate is unlikely to be representative - in any wider context - of such a category).

Table 3.11 shows the junction accident data disaggregated by junction type, form of traffic control and location. As with the link data, it can be seen, generally, that as the sites become more urbanised the frequency of accidents per junction increases

Both the link and junction accident data are "unbalanced" in that there are great differences between the numbers of sites falling into the individual categories. A number of categories contain few or no sites and are, therefore, not likely to be representative of their respective categories. In subsequent analysis the data will be aggregated as appropriate for the purpose.

In Table 3.12 the link data have been disaggregated into the three traffic flow bands.

It can be seen that in all but a few cases the accident frequency per junction increases with increasing traffic flow band.

Tables 3.13 and 3.14 show similarly disaggregated data for junction accidents. As with the link data the accident frequency is seen to increase with increasing traffic flow band.

3.4. Conclusions

This chapter, although brief, indicates that there are quite different accident frequencies associated with different location types. It was not purpose of this chapter to quantify these differences - for this will be considered later - but rather to indicate their strong presence. In addition, it is shown that the data may be considered as being reasonably representative of the national picture.

TABLE 3.1: COMPARISON OF REGIONAL ACCIDENT TOTALS WITH STUDY DATA ACCIDENT TOTALS

YEAR	ALL ACCIDENTS IN REGION	LINK FILE ACCIDENTS	JUNCTION FILE ACCIDENTS	ALL STUDY ACCIDENTS
1979	3603	1018 (28%)[1]	1466 (41%)	2484 (69%)
1980	3594	1006 (30%)	1570 (44%)	2576 (72%)
1981	3634	955 (26%)	1492 (41%)	2447 (67%)
1982	3444	995 (29%)	1509 (44%)	2504 (73%)
TOTAL	14275	3974 (28%)	6037 (42%)	10011 (70%)

[1] Figures in brackets represent %age of all accidents in Lothian Region (ie Column 2)

TABLE 3.2: COMPARISON OF LINK AND JUNCTION ACCIDENTS WITH NATIONAL STATISTICS

	LOTHIAN REGION				GREAT BRITAIN[1]			
YEAR	LINK	(%)	JUNCTION	(%)	LINK	(%)	JUNCTION	(%)
1979	1018	41.0%	1466	59.0%	109758	43.1%	145101	56.9%
1980	1006	39.1%	1570	60.9%	104554	41.7%	146386	58.3%
1981	955	39.0%	1492	61.0%	103939	41.9%	144328	58.1%
1982	995	39.7%	1509	60.3%	105449	41.2%	150529	58.8%
ALL	3974	39.7%	6037	60.3%	423700	41.9%	586344	58.1%

[1] Department of Transport [1981a], [1981b], [1982] and [1983] (Tables 20, 17, 16 and 18 respectively)

TABLE 3.3: ANNUAL TRAFFIC (MVkm) FOR GREAT BRITAIN 1979-1982

	BUILT-UP ROADS		NON BUILT-UP ROADS		ALL ROADS	
YEAR	MVkm	%	MVkm	%	MVkm	%
1979	139988	50.0	140222	50.0	280210	100.0
1980	143912	49.6	146066	50.4	289978	100.0
1981	135782	48.1	146679	51.9	282461	100.0
1982	150357	49.7	151891	50.3	302248	100.0
TOTAL	570039	49.4	584858	50.6	1154897	100.0

Sources:
Department of Transport et al [1980] (Table 2.4)
Department of Transport et al [1981a] (Table 2.4)
Department of Transport et al [1982a] (Table 2.2)
Department of Transport et al [1983a] (Table 2.2)

TABLE 3.4: ESTIMATED ACCIDENT RATES FOR URBAN AND RURAL ROADS FOR THE STUDY DATA

	LINKS	JUNCTIONS	TOTAL
URBAN ROADS (Built-up):			
Number of sites	1860	2507	4367
Road kilometres	205.89	200.56	406.45
Annual MVkm	743.75	724.50[1]	1468.25
4-year accident total	2884	5590	8474
Accident rate	0.97	1.93	1.44
RURAL ROADS (Non built-up):			
Number of sites	632	605	1237
Road kilometres	412.88	48.40	461.28
Annual MVkm	901.65	105.70[1]	1007.35
4-year accident total	1090	447	1537
Accident rate	0.30	1.06	0.38
ALL ROADS:			
Number of sites	2492	3112	5604
Road kilometres	618.77	248.96	867.73
Annual MVkm	1645.40	830.19	2475.59
4-year accident total	3974	6037	10011
Accident rate	0.60	1.82	1.01

[1] Ratio of road kilometres to annual MVkm for links is assumed to apply.

TABLE 3.5: ACCIDENTS, TRAVEL AND ACCIDENT RATES FOR NON BUILT-UP ROADS FOR GREAT BRITAIN FOR THE YEARS 1979-1982

ROAD TYPE		1979	1980	1981	1982	TOTAL
Motorways:						
	Accidents	4044	4076	4034	4277	16431
	MVkm	27168	28629	28926	29242	113965
	Rate	0.149	0.142	0.139	0.146	0.144
A roads:						
	Accidents	33793	32406	32587	32800	131586
	MVkm	71247	74092	73669	76158	295166
	Rate	0.474	0.437	0.442	0.431	0.446
A and M roads:						
	Accidents	37837	36482	36621	37077	148017
	MVkm	98415	102721	102595	105400	409131
	Rate	0.384	0.355	0.357	0.352	0.362
B roads:						
	Accidents	7837	7972	7842	8222	31873
	MVkm	12987	13162	13037	14466	53652
	Rate	0.603	0.606	0.602	0.568	0.594
Other roads:						
	Accidents	14578	14251	14762	14759	58350
	MVkm	28819	30185	31047	32025	122076
	Rate	0.506	0.472	0.475	0.461	0.478
All non built-up roads:						
	Accidents	60252	58705	59225	60058	238240
	MVkm	140221	146068	146679	151891	584859
	Rate	0.430	0.402	0.404	0.395	0.407

Sources:
Department of Transport et al [1980] (Table 2.4)
Department of Transport et al [1981a] (Table 2.4)
Department of Transport et al [1982a] (Table 2.2)
Department of Transport et al [1983a] (Table 2.2)
Department of Transport et al [1983b] (Table 4)

TABLE 3.6: ACCIDENTS, TRAVEL AND ACCIDENT RATES FOR BUILT-UP ROADS FOR GREAT BRITAIN FOR THE YEARS 1979-1982

ROAD TYPE		1979	1980	1981	1982	TOTAL
A roads:						
	Accidents	90426	88858	86600	90237	356121
	MVkm	66560	67640	65161	65615	264976
	Rate	1.359	1.314	1.329	1.375	1.344
B roads:						
	Accidents	22824	22524	21994	23109	90451
	MVkm	21324	21741	17372	22315	82752
	Rate	1.070	1.036	1.266	1.036	1.093
Other roads:						
	Accidents	81360	80858	80447	82570	325235
	MVkm	52106	54531	53248	62427	222312
	Rate	1.561	1.483	1.511	1.323	1.463
All built-up roads:						
	Accidents	194610	192240	189041	195916	771807
	MVkm	139990	143912	135781	150357	570040
	Rate	1.390	1.336	1.392	1.303	1.354

Sources:
Department of Transport <u>et al</u> [1980] (Table 2.4)
Department of Transport <u>et al</u> [1981a] (Table 2.4)
Department of Transport <u>et al</u> [1982a] (Table 2.2)
Department of Transport <u>et al</u> [1983a] (Table 2.2)
Department of Transport <u>et al</u> [1983b] (Table 4)

TABLE 3.7: ACCIDENTS, TRAVEL AND ACCIDENT RATES FOR ALL ROADS FOR GREAT BRITAIN FOR THE YEARS 1979-1982

ROAD TYPE		1979	1980	1981	1982	TOTAL
All built-up roads:						
	Accidents	194610	192240	189041	195916	771807
	MVkm	139990	143912	135781	150357	570040
	Rate	1.390	1.336	1.392	1.303	1.354
All non built-up roads:						
	Accidents	60252	58705	59225	60058	238240
	MVkm	140221	146068	146679	151891	584859
	Rate	0.430	0.402	0.404	0.395	0.407
All roads:						
	Accidents	254862	250945	248266	255974	1010047
	MVkm	280211	289980	282460	302248	1154899
	Rate	0.910	0.865	0.879	0.847	0.875

Sources:
Department of Transport et al [1980] (Table 2.4)
Department of Transport et al [1981a] (Table 2.4)
Department of Transport et al [1982a] (Table 2.2)
Department of Transport et al [1983a] (Table 2.2)
Department of Transport et al [1983b] (Table 4)

TABLE 3.8: LINK ACCIDENT DATA FREQUENCIES OF FOUR-YEAR ACCIDENT TOTALS

OBSERVED ACCIDENT TOTAL (y)	OBSERVED FREQUENCY OF LINKS WITH y ACCIDENTS (n)	RELATIVE FREQUENCY (%)	OBSERVED ACCIDENT TOTALS FOR LINKS WITH y ACCIDENTS (y.n)
0	1124	45.10	0
1	557	22.35	557
2	313	12.56	626
3	173	6.94	519
4	99	3.97	396
5	61	2.45	305
6	45	1.81	270
7	28	1.12	196
8	15	0.60	120
9	22	0.88	198
10	12	0.48	120
11	10	0.40	110
12	4	0.16	48
13	5	0.20	65
14	3	0.12	42
15	7	0.28	105
16	4	0.16	64
17	1	0.04	17
18	1	0.04	18
19	2	0.08	38
20	3	0.12	60
29	1	0.04	29
31	1	0.04	31
40	1	0.04	40
TOTAL	2492	100.00	3974

TABLE 3.9: JUNCTION ACCIDENT DATA FREQUENCIES OF FOUR-YEAR ACCIDENT TOTALS

OBSERVED ACCIDENT TOTAL (y)	OBSERVED FREQUENCY OF JUNCTIONS WITH y ACCIDENTS (n)	RELATIVE FREQUENCY (%)	OBSERVED ACCIDENT TOTALS FOR JUNCTIONS WITH y ACCIDENTS (y.n)
0	1260	40.49	0
1	689	22.14	689
2	382	12.28	764
3	248	7.97	744
4	144	4.63	576
5	96	3.08	480
6	72	2.31	432
7	46	1.48	322
8	40	1.29	320
9	34	1.09	306
10	18	0.58	180
11	21	0.67	231
12	14	0.45	168
13	8	0.26	104
14	8	0.26	112
15	5	0.16	75
16	5	0.16	80
17	6	0.19	102
18	2	0.06	36
19	4	0.13	76
20	2	0.06	40
21	2	0.06	42
23	1	0.03	23
24	1	0.03	24
25	1	0.03	25
28	1	0.03	28
29	2	0.06	58
TOTAL	3112	100.00	6037

TABLE 3.10: FREQUENCIES OF LINK TYPE WITH SELECTED ACCIDENT AND TRAFFIC STATISTICS

SITE CODE[1]	FREQUENCY OF LINK CATEGORY	AGGREGATE LINK LENGTH (km)	AGGREGATE 4-YEAR ACCIDENT TOTAL (A)	AGGREGATE ANNUAL TRAFFIC (MVkm)	LINK CATEGORY ACCIDENT RATE A/(4xMVkm)
L111X	161	9.90	515	46.43	2.77
L112X	14	1.44	51	3.98	3.20
L121X	15	0.90	22	3.94	1.40
L141X	71	3.85	111	15.20	1.83
L151X	38	1.93	82	7.87	2.61
L152X	16	1.68	130	4.80	6.77
L171X	22	1.37	73	6.74	2.71
L221X	19	1.41	33	6.42	1.28
L241X	40	4.13	66	13.78	1.20
L251X	42	4.39	86	17.18	1.25
L331X	17	3.69	35	15.16	0.58
L341X	32	3.00	29	8.02	0.90
L351X	17	2.48	19	9.68	0.49
L441X	574	55.42	494	175.05	0.71
L442X	15	0.86	13	1.31	2.48
L451X	342	43.15	403	156.27	0.65
L452X	17	1.85	13	3.51	0.93
L471X	32	3.07	59	13.75	1.07
L551X	177	34.95	330	140.24	0.59
L552X	60	10.85	67	37.22	0.45
L571X	21	5.13	57	20.90	0.68
L661X	438	306.25	880	605.39	0.36
L662X	194	106.63	210	296.26	0.18
L771X	69	7.36	130	28.77	1.13
L772X	17	1.60	41	2.82	3.64
L773X	32	1.81	25	4.71	1.33
LXXXX	2492	618.77	3974	1645.40	0.60

[1] See Table 2.1 for description of link categories.

TABLE 3.11: JUNCTION ACCIDENT FREQUENCIES OF ACCIDENTS BY JUNCTION TYPE AND LOCATION

SITE CODE[1]		PRIORITY JUNCTIONS ($=1)				SIGNAL JUNCTIONS ($=2)			
		RURAL (£=1)	SUB-URBAN (£=2)	LOCAL CENTRE (£=3)	CITY CENTRE (£=4)	RURAL (£=1)	SUB-URBAN (£=2)	LOCAL CENTRE (£=3)	CITY CENTRE (£=4)
	N[2]	35	43	2	1	0	0	0	0
J1£$X	A[3]	10	46	1	5	0	0	0	0
	A/N	0.29	1.07	0.50	5.00	-	-	-	-
	N	37	42	3	0	0	0	0	0
J2£$X	A	16	25	5	0	0	0	0	0
	A/N	0.43	0.60	1.67	-	-	-	-	-
	N	11	67	11	1	0	0	0	0
J3£$X	A	16	79	18	2	0	0	0	0
	A/N	1.45	1.18	1.64	2.00	-	-	-	-
	N	247	1159	211	93	0	11	11	8
J4£$X	A	154	1467	534	310	0	61	84	43
	A/N	0.62	1.27	2.53	3.33	-	5.55	7.64	5.38
	N	36	47	17	3	0	1	0	0
J5£$X	A	25	64	46	7	0	8	0	0
	A/N	0.69	1.36	2.71	2.33	-	8.00	-	-
	N	39	127	19	10	2	30	23	19
J6£$X	A	80	370	49	71	14	263	209	260
	A/N	2.05	2.91	2.58	7.10	7.00	8.77	9.09	13.68
	N	166	204	33	44	0	28	14	39
J7£$X	A	94	220	71	112	0	36	21	170
	A/N	0.57	1.08	2.15	2.55	-	1.29	1.50	4.36
	N	29	78	15	27	0	11	10	25
J8£$X	A	33	157	80	181	0	41	57	242
	A/N	1.14	2.01	5.33	6.70	-	3.73	5.70	9.68
	N	3	6	0	2	0	3	7	2
J9£$X	A	5	19	0	27	0	25	76	28
	A/N	1.67	3.17	-	13.50	-	8.33	10.86	14.00
	N	603	1773	311	181	2	84	65	93
JX£$X	A	433	2447	804	715	14	434	447	743
	A/N	0.72	1.38	2.59	3.95	7.00	5.17	6.88	7.99

GRAND TOTALS (JXXXX) N=3112 A=6037 A/N=1.94

1 See Table 2.1 for description of SITECODE categories.
2 N is the number of junctions in the cell.
3 A is the four year accident total (1979-1982) for the cell.

TABLE 3.12: LINK ACCIDENT DATA BY LINK CATEGORY, LOCATION AND TRAFFIC FLOW BAND

SITE CODE[1]	TRAFFIC FLOW BAND (LMVKMBAND) LOW (£=1) N[2]	A[3]	A/N	MEDIUM (£=2) N	A	A/N	HIGH (£=3) N	A	A/N
L111£	67	79	1.18	74	264	3.57	20	172	8.60
L112£	4	6	1.50	8	36	4.50	2	9	4.50
L121£	8	2	0.25	5	14	2.80	2	6	3.00
L141£	40	33	0.83	23	41	1.78	8	37	4.63
L152£	5	13	2.60	6	22	3.67	5	95	19.00
L151£	19	17	0.89	17	63	3.71	2	2	1.00
L171£	10	13	1.30	9	30	3.33	3	30	10.00
L221£	10	8	0.80	6	12	2.00	3	13	4.33
L241£	17	11	0.65	15	18	1.20	8	37	4.63
L251£	11	8	0.73	16	23	1.44	15	55	3.67
L331£	0	0	-	4	6	1.50	13	29	2.23
L341£	13	2	0.15	16	17	1.06	3	10	3.33
L351£	8	0	0.00	1	2	2.00	8	17	2.13
L441£	267	103	0.39	202	169	0.84	105	222	2.11
L442£	13	11	0.85	2	2	1.00	0	0	-
L451£	117	50	0.43	146	145	0.99	79	208	2.63
L452£	14	13	0.93	2	0	0.00	1	0	0.00
L471£	10	4	0.40	14	26	1.86	8	29	3.63
L551£	54	24	0.44	52	55	1.06	71	251	3.54
L552£	16	6	0.38	20	18	0.90	24	43	1.79
L571£	3	1	0.33	8	9	1.13	10	47	4.70
L661£	50	8	0.16	90	52	0.58	298	820	2.75
L662£	71	7	0.10	55	12	0.22	68	191	2.81
L771£	23	8	0.35	28	46	1.64	18	76	4.22
L772£	6	8	1.33	11	33	3.00	0	0	-
L773£	20	16	0.80	11	7	0.64	1	2	2.00
LXXX£	876	451	0.51	841	1122	1.33	775	2401	3.10
ALL TRAFFIC FLOW BANDS (LXXXX)							2492	3974	1.59

[1] See Table 2.1 for description of SITECODE categories.
[2] N is the number of junctions in the cell.
[3] A is the four year accident total (1979-1982) for the cell.

TABLE 3.13: ACCIDENT FREQUENCIES BY JUNCTION TYPE, LOCATION AND TRAFFIC FLOW BAND FOR PRIORITY JUNCTIONS

SITE CODE[1]		RURAL (£=1)			SUBURBAN(£=2)			LOCAL CENTRE(£=3)			CITY CENTRE(£=4)		
		LOW ($=1)	MED ($=2)	HIGH ($=3)	LOW ($=1)	MED ($=2)	HIGH ($=3)	LOW ($=1)	MED ($=2)	HIGH ($=3)	LOW ($=1)	MED ($=2)	HIGH ($=3)
	N[2]	4	13	18	7	4	32	0	1	1	0	0	1
J1£1$	A[3]	1	4	5	3	2	41	0	1	0	0	0	5
	A/N	0.25	0.31	0.28	0.43	0.50	1.28	-	1.00	0.00	-	-	5.00
	N	4	13	20	6	7	29	0	2	1	0	0	0
J2£1$	A	0	4	12	0	2	23	0	2	3	0	0	0
	A/N	0.00	0.31	0.60	0.00	0.29	0.79	-	1.00	3.00	-	-	-
	N	0	1	10	5	11	51	0	3	8	0	0	1
J3£1$	A	0	0	16	5	5	69	0	10	8	0	0	2
	A/N	-	0.00	1.60	1.00	0.45	1.35	-	3.33	1.00	-	-	2.00
	N	197	29	21	582	381	196	79	76	56	17	20	56
J4£1$	A	80	20	54	426	555	486	134	222	178	29	39	242
	A/N	0.41	0.69	2.57	0.73	1.46	2.48	1.70	2.92	3.18	1.71	1.95	4.32
	N	19	10	7	18	12	17	2	7	8	1	1	1
J5£1$	A	2	9	14	12	16	36	7	15	24	1	3	3
	A/N	0.11	0.90	2.00	0.67	1.33	2.12	3.50	2.14	3.00	1.00	3.00	3.00
	N	23	11	5	47	51	29	8	5	6	0	3	7
J6£1$	A	20	27	33	70	148	152	24	12	13	0	12	59
	A/N	0.87	2.45	6.60	1.49	2.90	5.24	3.00	2.40	2.17	-	4.00	8.43
	N	46	45	75	87	49	68	5	13	15	2	9	33
J7£1$	A	13	18	63	54	61	105	4	18	49	0	15	97
	A/N	0.28	0.40	0.84	0.62	1.24	1.54	0.80	1.38	3.27	0.00	1.67	2.94
	N	7	12	10	22	20	36	0	6	9	1	3	23
J8£1$	A	5	11	17	15	44	98	0	48	32	3	5	173
	A/N	0.71	0.92	1.70	0.68	2.20	2.72	-	8.00	3.56	3.00	1.67	7.52
	N	0	0	3	0	0	6	0	0	0	0	0	2
J9£1$	A	0	0	5	0	0	19	0	0	0	0	0	27
	A/N	-	-	1.67	-	-	3.17	-	-	-	-	-	13.5
	N	300	134	169	774	535	464	94	113	104	21	36	124
JX£1$	A	121	93	219	585	833	1029	169	328	307	33	74	608
	A/N	0.40	0.69	1.30	0.76	1.56	2.22	1.80	2.90	2.95	1.57	2.06	4.90

GRAND TOTALS (JXX1X) N=2868 A=4339 A/N=1.53

[1] See Table 2.1 for description of SITECODE categories.
[2] N is the number of junctions in the cell.
[3] A is the four year accident total (1979-1982) for the cell.

TABLE 3.14: ACCIDENT FREQUENCIES BY JUNCTION TYPE, LOCATION AND TRAFFIC FLOW BAND FOR SIGNAL CONTROLLED JUNCTIONS

SITE CODE[1]		RURAL(£=1) LOW ($=1)	RURAL(£=1) MED ($=2)	RURAL(£=1) HIGH ($=3)	SUBURBAN(£=2) LOW ($=1)	SUBURBAN(£=2) MED ($=2)	SUBURBAN(£=2) HIGH ($=3)	LOCAL CENTRE(£=3) LOW ($=1)	LOCAL CENTRE(£=3) MED ($=2)	LOCAL CENTRE(£=3) HIGH ($=3)	CITY CENTRE(£=4) LOW ($=1)	CITY CENTRE(£=4) MED ($=2)	CITY CENTRE(£=4) HIGH ($=3)
	N[2]	0	0	0	0	2	7	0	0	11	0	0	8
J4£2$	A[3]	0	0	0	0	3	58	0	0	84	0	0	43
	A/N	-	-	-	-	1.50	6.44	-	-	7.64	-	-	5.38
	N	0	0	0	0	0	1	0	0	0	0	0	0
J5£2$	A	0	0	0	0	0	8	0	0	0	0	0	0
	A/N	-	-	-	-	-	8.00	-	-	-	-	-	-
	N	0	0	2	1	1	28	0	1	22	0	0	19
J6£2$	A	0	0	14	3	4	256	0	1	208	0	0	260
	A/N	-	-	7.00	3.00	4.00	9.14	-	1.00	9.5	-	-	13.7
	N	0	0	0	3	2	18	0	1	13	0	0	39
J7£2$	A	0	0	0	1	6	29	0	0	21	0	0	170
	A/N	-	-	-	0.33	1.17	1.61	-	0.00	1.62	-	-	4.36
	N	0	0	0	0	0	11	0	0	10	0	0	25
J8£2$	A	0	0	0	0	0	41	0	0	57	0	0	242
	A/N	-	-	-	-	-	3.73	-	-	5.70	-	-	9.68
	N	0	0	0	0	0	3	0	0	7	0	0	2
J9£2$	A	0	0	0	0	0	25	0	0	76	0	0	28
	A/N	-	-	-	-	-	8.33	-	-	10.9	-	-	14.0
	N	0	1	1	1	22	61	0	12	53	1	12	80
JX£2$	A	0	6	8	3	49	382	0	32	415	3	44	696
	A/N	-	6.00	8.00	3.00	2.23	6.26	-	2.67	7.83	3.00	3.67	8.70

GRAND TOTALS (JXX2X) N=244 A=1638 A/N=6.71

[1] See Table 2.1 for description of SITECODE categories.
[2] N is the number of junctions in the cell.
[3] A is the four year accident total (1979-1982) for the cell.

CHAPTER 4 REGRESSION AND THE STATISTICAL DISTRIBUTION OF ROAD ACCIDENTS

The purpose of the present study is to determine the strength of any relationships that may exist between road accidents and traffic flow at different location types. These relationships can be examined by the statistical technique of regression in which the purpose is to find (or model) relationships between a dependent variable (eg number of accidents per given time period) and one or more independent or explanatory variables (eg location type, traffic volume, adjacent land use etc).

Such a statistical model may be thought of as comprising two elements, these being:

1. the systematic component; and
2. the error component.

Consider an equation of the form:

$$y = f(a_0 + a_1x_1 + a_2x_2 + \ldots\ldots + a_ix_i + \ldots\ldots + a_nx_n) + e \qquad (4.1)$$

where:

y	is the dependent variable;
x_i	is the independent or explanatory variable;
a_i	is the regression coefficient;
e	is the error term or error component; and
$f(a_0+\ldots\ldots+a_nx_n)$	is the systematic compenent.

A requirement of least-square regression is that observed values for the dependent variable are drawn from a population with a mean value equal to the predicted (or expected) value and with a variance which remains constant over the whole range of the data. Where observations are drawn from Normally distributed data this requirement is not violated. On the other hand, if the observations are drawn from a non-Normal data source (eg Poisson distributed data), the assumption of constant variance is not met. Under such conditions it would be incorrect to use standard least-square regression techniques to model such data.

Accordingly, to determine an appropriate form of model for the prediction of accident frequency it is necessary to examine in some detail the statistical distribution of accidents.

An early distinction is made between (1) the temporal and (2) the spatial distribution of accidents viz:

1. accidents at the same site over different non-overlapping time periods; and

2. accidents at different sites over the same time period.

This distinction is crucially important for it will be shown that the two distributions are essentially different.

The following two chapters describe detailed analyses of the data to determine the nature of both the temporal and spatial distribution of accidents and in Chapter 7 the form of the statistical model is returned to and discussed in detail in the light of the results of the analyses of accident distribution.

CHAPTER 5 THE TEMPORAL DISTRIBUTION OF ROAD ACCIDENTS

5.1. Literature Review

The literature in the field of road safety is by no means conclusive in this respect. It is generally assumed – and perhaps not without good reason – that the distribution of road accidents (perhaps disaggregated by accident type) at a site (or group of sites) with respect to time is Poissonic.

Poisson assumed

Maycock and Hall [1984] state (Appendix 5) that ".....it is usually assumed that accidents occur randomly, and that therefore the total number occurring at a specific site in a particular junction period is a Poisson variate."

Abbess et al [1981] assume that ".....accidents occur in a Poisson process....."

Gipps [1980] states that the ".....number of accidents which occur at a particular site in the course of a fixed time period may safely be assumed to be Poisson distributed."

Hakkert and Mahalel [1978] discuss the use of the Poisson distribution in more detail and indicate their reservations on its applicability because ".....one of the Poisson assumptions becomes doubtful, namely the assumption that the process is stationary over time." They were, however, unable to verify its applicability although they did adopt it in their work.

Hauer and Persaud [1983a] ".....assume, as is common, that accident occurrence obeys the Poisson probability law."

These references demonstrate that the use of the Poisson distribution in this respect is largely based on an assumptive premise.

Poisson examined

The review of the literature has indicated that whilst conformity of the temporal distribution of accidents with a Poisson process has been shown on relatively few occasions (eg Zahavi [1962] and Chapman [1970]); more recently, some researchers have reported that there would appear to be little support for it.

Zahavi [1962], using monthly accident data for railroad/road vehicle accidents in Israel for 1959-1961, concluded that ".....agreement with the Poissonic process.....is high."

Chapman [1970], using daily accident data for State Highway 2 in New South Wales, Australia for 1958-1963, demonstrated that ".....the observed distribution is well fitted by Poisson". He also stresses ".....that although a theoretical distribution may be fitted to a set of data, one cannot infer that the conditions upon which the theoretical model is built are satisfied in the observed data".

Jorgensen [1972] refers to work on Danish and Swedish accident data which suggests that the numbers of accidents per day over large parts of national road networks are Poisson distributed. Jorgensen reports that the Poisson law also applies to the number of accidents per year on shorter road sections.

Hutchinson and Mayne [1977] - on the basis of an examination of annual accident totals for sixty classes of accident for the years 1969-1972 - observed that there ".....appeared to be more variability (*in annual accident totals*) than would be expected on the hypothesis of accidents occurring randomly in different years (the Poisson Law)." The authors recognised that this was perhaps not very surprising as there was so much variation because weather, legislation, the road network and vehicles on it change with time.

In a study into daily accident conflict totals, Hauer [1978] observed ".....it is apparent, that the Poisson hypothesis does not hold." Hauer does, however, suggest that conflict occurrence may change from day to day because of changes in weather, vehicular flow, pedestrian volumes etc and possibly through the subjectivity of conflict identification by observers.

Work by the author in another context (McGuigan [1980]) on monthly data for accidents on a length of road in Edinburgh for 1976–1978 demonstrates the ability of the Poisson distribution to model temporally distributed accident data.

More recently, Nicholson [1985] recognising that there has been little empirical work to support the Poisson assumption concludes – on the basis of time series accident data for thiry-five intersections in Auckland, New Zealand – that ".....there are grounds for doubting the general validity of the Poisson assumption." However, in view of the lengthy duration of some of the individual time series which ranged from five to thirty-five years, the Poisson condition of an underlying consistency of accident rate is itself doubtful because of possible changes in legislation and general traffic levels. Nicholson did recognise this and 'controlled' for the possibility of inter alia a changing underlying accident rate by determining – through a least-square regression analysis – a linear trend and calculating the variance associated with each time series from the 'fitted' annual accident totals rather than the observed totals. This form of analysis did not alter his conclusions that there was ".....a great deal of variation in the variance/mean ratio, whether a linear trend is taken into account or ignored."

One criticism of the 'control' procedure adopted by Nicholson involves the assumption that a changing underlying accident rate could be modelled by a simple linear trend. It is quite possible that the changes in weather conditions, legislation, road network and possibly even the rules regarding accident reporting could create conditions which give rise to a non-linear trend for accident rate.

Having brought into doubt the general validity of the Poisson assumption, Nicholson goes on to suggest that the choice of a probability distribution should vary as the observed variance/mean ratios differ from unity. It is, accordingly, suggested by Nicholson that where the variance/mean ratio is less than unity a binomial distribution is used and where this ratio is greater than unity a negative binomial distribution is used. This suggestion, however, is unsatisfactory in that it precludes the need for an underlying theoretical justification for the use of a particular distribution and relies quite simply on the value of an arithmetical expression.

The point stressed by Chapman [1970] above regarding how ".....one cannot infer that the conditions upon which the theoretical model is built are satisfied in the observed data....." is poignant because of his later contribution to the topic. Chapman [1973] describes the concepts of accident exposure and accident propensity as:

exposure:	"the number of opportunities for accidents of a certain type to occur in a given time in a given area (ie it is the possible number of accidents of that type which could occur in that time in that area)"; and
propensity:	"the conditional probability that an accident occurs given the opportunity for one";

where ".....the two definitions are connected by a simple equation: the number of accidents is equal to the exposure multiplied by the propensity."

Chapman did not consider the statistical distribution that these concepts would generate. In theory, Chapman's definitions suggest the applicability of the positive binomial disitribution (more commonly referred to simply as the binomial distribution but because of later use of the negative binomial distribution the adjective "positive" is used to avoid confusion).

In practice, however, because of (a) the extreme rarity - in statistical terms - of an accident and (b) the very large number of opportunities for accidents to occur the positive binomial distribution converges on the Poisson distribution.

In addition, Hutchinson and Mayne [1977] and Hauer [1978] noted that not only was there more variability in temporally distributed data than would be expected under a Poisson assumption but that this variability increased with the magnitude of the mean value. Hutchinson and Mayne [1977] fitted a curve of the form:

$$\text{standard deviation} = (\text{mean} + k^2.\text{mean}^2)^{1/2}$$

to their data which is suggestive of a negative binomial distribution for the underlying data (see Section 7.3.3).

Scott [1983], in an examination of the monthly variation in two-vehicle accident frequencies for 1970-1978, also observed this propensity for greater variability to be associated with greater values of mean accident frequency.

The foregoing points to the lack of agreement among researchers about the usefulness of the Poisson model and, indeed, raises arguments in favour of using other models which have variances both greater and less than the mean. Where the Poisson assumption was questioned, most researchers have reported on reasons why this may have occurred. The possibility remains that had these reasons been 'controlled', a Poisson assumption may have proved valid; this, however, is pure speculation. In view of the conflicting evidence regarding this issue, the first logical step, here, was to examine the temporal distribution of the study data.

5.2. Examination of the Study Data

In the context of road safety, accident data to which the Poisson distribution may apply must satisfy four conditions:

1. underlying rate of accident occurrence is constant over time;
2. accident occurrence in one time period is independent of that in another time period;
3. accidents must be singly occurring with no upper limit to their number; and
4. accidents must occur at random points in time.

Strictly speaking, as traffic volumes vary throughout the day, the first of these conditions cannot be met on an hour-by-hour basis. However, if each day is considered to be made up of a series of time intervals, each of which, over a series of days, is assumed to satisfy the four conditions then – because a variate which is an aggregation of individual Poisson distributed variates is itself Poisson (Moran [1968]) – the accidents per day would be Poisson. This assumes that the apportionment of traffic and weather conditions by time interval on a daily basis is constant, which, of course, it is not, because of the variations associated with not only the days of the week but also the seasons of the year. If, however, a complete annual cycle is considered as the time period over which the individual time periods are aggregated then these sources of variation are largely controlled. Nonetheless, it has to be accepted that the underlying true rate of accident occurrence is not absolutely constant from year to year. However, as it has been shown that there may be some relaxation of this first condition before the Poisson assumption is invalidated

(Cox and Millar [1965]), the above lack of consistency of the underlying accident rate may not be of any practical significance.

On top of this, the third and fourth conditions are not wholly satisfied. Firstly, there is an upper limit to the number of accidents that can occur at a location during a given time period: it is the number of vehicles negotiating the location during the time period. Secondly, as accidents can only occur when vehicles are present on the road, it is not strictly correct to assume that accidents occur at at random in time. This suggests that accident risk is associated with the passage of vehicles through a location rather than a truly random time dependent phenomenon and, as such, gives support to the concept of an underlying positive binomial explanation for accident occurrence. As indicated above, however, the positive binomial distribution, within the context of this study, would converge on a Poisson series.

It was, therefore, decided to examine the four annual accident totals in much the same way as Hutchinson and Mayne [1977] reported.

Because of the low numbers of accidents occurring at individual sites it was not considered sensible to disaggregate the data to an individual site level. In view of the manner in which the SIR databases were set up, it was not practicable to combine individual adjacent sites in such a way as to generate annual accident totals for contiguous geographical areas. Accordingly, groups of non-adjacent sites were combined according to their type, thus providing a set of non-contiguous geographical areas.

The non-contiguous nature of the geographical areas does not affect the rationale of this exercise since the concept of a consistent geographical area (albeit non-contiguous) over each of the four years is maintained.

The link and junction accident data were disaggregated according to the same two criteria, viz:

- Vehicle movements associated with the accident; and
- Location type at which the accident occurred.

To avoid working with groups of data comprising low numbers of accidents, some of the disaggregations generating low accident totals were "re-aggregated" according to higher order criteria to provide - where sensible - accident totals averaging not generally less than 25 accidents per year.

For each disaggregation the annual average accident total has been calculated together with the variance (unbiased). From these values the Index of Dispersion (IOD) was calculated which is chi-squared (χ^2) distributed with n-1 degrees of freedom where:

$$IOD = \sigma^2.df/\mu \qquad (5.1)$$

with:

σ^2 is the variance of the accident totals;

df is the number of degrees of freedom; and

μ is the mean value of the accident totals.

Agreement of the data with the Poisson distribution is accepted if IOD lies between the tabulated values of χ^2 for the adopted level of significance and number of degrees of freedom. Throughout the study, two-tailed tests at the 5% level of significance have been adopted. Spearman's rank correlation coefficient was also calculated to determine the relationship between the average annual accident total and the associated IOD.

The Spearman's rank correlation coefficient (see, for example, Sokal and Rohlf [1967]) has been used here because the distributions of the data which are dependent on the classification criteria are not Normal and, accordingly, a non-parametric index of association was considered to be more appropriate than the usual Pearson product-moment correlation coefficient.

Tables 5.1 to 5.4 summarise the analysis from which the following observations can be made:

- of the 21 disaggregations by Vehicle Movement (ie 8 link + 13 junction disaggregations), 2 provided an IOD significantly different from an underlying random assumption. The values of the Spearman's rank correlation coefficient were not significant; and

- of the 53 disaggregations by Location Type (ie 26+27 disaggregations), 1 provided an IOD significantly different from an underlying random assumption. The values of the Spearman's rank correlation coefficient were not significant.

5.3. Conclusions

These results, which suggest the applicability of the Poisson assumption, are contrary to those reported by Hutchinson and Mayne [1977]. In another important respect, the results also differ from those of Hutchinson and Mayne and others in that no significant relationship appeared to exist between the magnitude of the mean accident values and the associated variances.

Reasons to explain the above incompatability between the study data and those of Hutchinson and Mayne have not been pursued in detail, although possible explanations could be:

1. that Hutchinson and Mayne's and others' data encompassed significant changes in legislation which the current study data do not; and

2. that the current study data are for a selection of sites at which there was no (or little) change in traffic and road conditions over the study period.

It should be stressed, however, that although the Poisson distribution appears to fit the data well, it does not prove that the data satisfy the conditions upon which the Poisson distribution is built. Indeed, an underlying positive binomial model cannot be discounted on this evidence.

TABLE 5.1: ANALYSIS OF TEMPORAL DISTRIBUTION OF JUNCTION ACCIDENT DATA DISAGGREGATED BY NON-SPATIAL CRITERIA

ACCIDENT TYPE	ANNUAL ACCIDENT TOTAL 1979	1980	1981	1982	ANNUAL MEAN	RANK R_1	VAR'NCE	VARIANCE TO MEAN RATIO	RANK R_2	IOD
JPED	431	472	457	467	456.75	13	333.58	0.73	6	2.19
JPSV	151	119	120	141	132.75	10	250.92	1.89	12	5.67
JSVNP	166	178	156	180	170.00	12	125.33	0.74	7	2.21
JTVNP1	75	95	90	60	80.00	5	250.00	3.13	13	9.38[1]
JTVNP2	39	39	38	39	38.75	4	0.25	0.01	1	0.02[1]
JTVNP3	9	17	8	10	11.00	1	16.67	1.52	10	4.55
JTVNP4	100	107	119	116	110.50	8	75.00	0.68	5	2.04
JTVNP5	104	116	92	88	100.00	6	160.00	1.60	11	4.80
JTVNP6	97	115	127	109	112.00	9	156.00	1.39	8	4.18
JTVNP7	149	156	139	151	148.75	11	50.92	0.34	3	1.03
JTVNP8	32	25	26	23	26.50	3	15.00	0.57	4	1.70
JTVNP9	14	24	18	13	17.25	2	24.92	1.44	9	4.33
JMVNP	99	107	102	112	105.00	7	32.67	0.31	2	0.93
JACCTOT	1466	1570	1492	1509	1509.25		1952.92	1.29		3.88

[1] Indicates a value for the IOD which does not lie within the tabulated χ^2 value for 3 degrees of freedom at the 5% level of significance (ie $\chi^2 < 0.216$ or > 9.348).

The Spearman's rank correlation coefficient (r_s) is:

$r_s = 1-6.\Sigma D^2/(N^3-N) = 1-6x406/(2197-13) = -0.115$

where $D^2 = (R_1-R_2)^2$.

TABLE 5.2: ANALYSIS OF TEMPORAL DISTRIBUTION OF JUNCTION ACCIDENT DATA DISAGGREGATED BY SPATIAL CRITERIA

SITECODE	ANNUAL ACCIDENT TOTAL 1979	1980	1981	1982	ANNUAL MEAN	RANK R_1	VAR'NCE	VARIANCE TO MEAN RATIO	RANK R_2	IOD
J(1,2)X1X	29	21	30	28	27.00	4	16.67	0.62	14	1.85
J3X1X	31	25	27	32	28.75	6	10.92	0.38	10	1.14
J411X	35	32	35	52	38.50	10	83.00	2.16	23	6.47
J4211	95	125	106	100	106.50	25	172.33	1.62	21	4.85
J4212	123	145	145	142	138.75	27	112.25	0.81	15	2.43
J4213	113	131	125	117	121.50	26	65.00	0.53	13	1.60
J4311	33	33	32	36	33.50	7	3.00	0.09	2	0.27
J4312	48	58	46	70	55.50	16	121.00	2.18	24	6.54
J4313	44	50	41	43	44.50	12	15.00	0.34	8	1.01
J441X	74	86	76	74	77.50	22	33.00	0.43	11	1.28
J4XSX	49	53	38	48	47.00	15	40.67	0.87	17	2.60
J5X2X	35	30	43	42	37.50	9	37.67	1.00	18	3.01
J61XX	24	26	21	23	23.50	2.5	4.33	0.18	3	0.55
J62PX	105	104	88	73	92.50	24	229.67	2.48	25	7.45
J621X	67	63	65	68	65.75	21	4.92	0.07	1	0.22
J632X	64	58	65	71.	64.50	20	28.33	0.44	12	1.32
J64XX	80	102	88	61	82.75	23	292.92	3.54	27	10.62[1]
J711X	17	30	25	22	23.50	2.5	29.67	1.26	19	3.79
J72XX	62	70	64	60	64.00	19	18.67	0.29	5	0.87
J73XX	27	25	17	23	23.00	1	18.67	0.81	16	2.43
J741X	29	27	25	31	28.00	5	6.67	0.24	4	0.71
J742X	31	42	54	43	42.50	11	88.33	2.08	22	6.24
J8(1,2)XX	54	60	54	63	57.75	17	20.25	0.35	9	1.05
J83XX	29	25	37	46	34.25	8	86.25	2.52	26	7.55
J841X	58	42	42	39	45.25	14	74.25	1.64	21	4.92
J8423	65	58	63	56	60.50	18	17.67	0.29	6	0.88
J9XXX	45	49	40	46	45.00	13	14.00	0.31	7	0.93
JXXXX	1466	1570	1492	1509	1509.25	-	1952.92	1.29	-	3.88

[1] Indicates a value for the IOD which does not lie within the tabulated χ^2 value for 3 degrees of freedom at the 5% level of significance (ie $\chi^2 < 0.216$ or > 9.348).

The Spearman's rank correlation coefficient (r_s) is:

$r_s = 1-6.\Sigma D^2/(N^3-N) = 1-6x2838.5/(19683-27) = 0.134$

where $D^2 = (R_1-R_2)^2$.

TABLE 5.3: SUMMARY OF ANALYSIS OF TEMPORAL DISTRIBUTION OF LINK ACCIDENT DATA DISAGGREGATED BY NON-SPATIAL CRITERIA

REF	ANNUAL ACCIDENT TOTAL 1979	1980	1981	1982	ANNUAL MEAN	RANK R_1	VAR'NCE	VARIANCE TO MEAN RATIO	RANK R_2	IOD
LPED	305	329	295	316	311.25	8	213.58	0.69	4	2.06
LPSV	95	88	77	86	86.50	6	55.00	0.64	7	1.91
LSVNP	266	254	251	292	265.75	7	348.25	1.31	5	3.93
LTVNP1	101	89	93	98	95.25	3	28.25	0.30	8	0.89
LTVNP2	81	62	73	62	69.50	2	85.67	1.23	1	3.70
LTVNP3	57	66	75	56	63.50	1	79.00	1.24	6	3.73
LTVNP4	29	30	23	27	27.25	5	9.58	0.35	3	1.06
LMVNP	84	88	68	58	74.50	4	195.67	2.63	2	7.88
LACCTOT	1018	1006	955	995	993.50	-	747.00	0.75	-	2.26

The Spearman's rank correlation coefficient (r_s) is:

$r_s = 1-6.\Sigma D^2/(N^3-N) = 1-6x80/(512-8) = 0.048$

where $D^2 = (R_1-R_2)^2$.

TABLE 5.4: SUMMARY OF ANALYSIS OF TEMPORAL DISTRIBUTION OF LINK ACCIDENT DATA DISAGGREGATED BY SPATIAL CRITERIA

SITE CODE	ANNUAL ACCIDENT TOTAL 1979	1980	1981	1982	ANNUAL MEAN	RANK R_1	VAR'NCE	VARIANCE TO MEAN RATIO	RANK R_2	IOD
L111X	147	144	118	106	128.75	25	399.58	3.10	26	9.31
L112X	13	8	14	16	12.75	10	11.58	0.91	15	2.73
L121X	7	6	4	5	5.50	4	1.67	0.30	3	0.91
L141X	25	30	31	25	27.75	18	10.25	0.37	5	1.11
L151X	20	17	24	21	20.50	16	8.33	0.41	6	1.22
L152X	24	27	34	45	32.50	19.5	87.00	2.68	25	8.03
L171X	17	17	19	20	18.25	15	2.25	0.12	1	0.37
L221X	5	9	8	11	8.25	7	6.25	0.76	12	2.27
L241X	12	22	16	16	16.50	13	17.00	1.03	17	3.09
L251X	15	29	19	23	21.50	17	35.67	1.66	22	4.98
L331X	10	6	8	11	8.75	8	4.92	0.56	8	1.69
L341X	9	9	4	7	7.25	6	5.58	0.77	13	2.31
L351X	7	4	5	3	4.75	3	2.92	0.61	10	1.84
L441X	132	127	115	120	123.50	24	56.33	0.46	7	1.37
L442X	2	4	4	3	3.25	1.5	0.92	0.28	2	0.85
L451X	109	101	96	97	100.75	23	34.92	0.35	4	1.04
L452X	7	2	2	2	3.25	1.5	6.25	1.92	23	5.77
L471X	17	11	14	17	14.75	12	8.25	0.56	8	1.68
L551X	79	94	77	80	82.50	22	60.33	0.73	11	2.19
L552X	19	19	11	18	16.75	14	14.92	0.89	14	2.67
L571X	10	18	18	11	14.25	11	18.92	1.33	20	3.98
L661X	213	202	227	238	220.00	26	248.67	1.13	19	3.39
L662X	67	51	39	53	52.50	21	131.67	2.51	24	7.52
L771X	39	34	31	26	32.50	19.5	29.67	0.91	16	2.74
L772X	9	9	6	16	10.25	9	15.58	1.25	21	4.56
L773X	4	6	10	5	6.25	5	6.92	1.11	18	3.32
ALL	1018	1006	955	995	993.50	-	747.00	0.75	-	2.26

The Spearman's rank correlation coefficient (r_s) is:

$$r_s = 1-6.\Sigma D^2/(N^3-N) = 1-6\times 2400/(17576-26) = 0.179$$

where $D^2 = (R_1-R_2)^2$.

CHAPTER 6 THE SPATIAL DISTRIBUTION OF ROAD ACCIDENTS

6.1. Literature Review

Very little practical work on the locational (or spatial) distribution of road accidents has been reported in the relevant literature. This is perhaps rather surprising because a considerable amount of the literature outlines research on the formulation of statistical models which, for instance, relate accidents to road geometry parameters and/or traffic volumes.

There is, however, a considerable body of literature on research into accidents (not necessarily road accidents) occurring to individual people and into the statistical distribution of these accidents. Here the basic unit of analysis - hereinafter referred to as the "case" - is the individual person in which accidents sustained by individuals are summed over given time periods.

This case structure is different from the location-based case structure used in the present study in which accidents occurring at geographically static individual road locations are summed over given time periods. Individual persons, however, are not similarly static.

Nonetheless, because of the depth of research into accidents occurring to individuals and because of the strong statistical parallels between the two case structures, a brief review of this body of literature is outlined first. See Kemp [1970] for a largely bibliographical account of this research work.

Throughout this literature the concept of accident proneness is alluded to and it is this concept, in particular, which provides a fruitful parallel with the current study data. By all accounts, whilst there is a certain amount of circumstantial evidence to support the use of the accident proneness concept as a factor in the explanation of accident occurrence at an individual person level it has yet to gain universal acceptance. Whilst the accident proneness concept will be developed with respect to a location-based case structure, conclusions regarding personal accident proneness are neither offerred nor implied.

The concept of the complete randomness of accidents underlay the earliest work on their distribution and led to the use of the Poisson distribution to model the frequency of accident occurrence.

Greenwood and Woods [1919] and Greenwood and Yule [1920] contradicted the concept of randomness with data on industrial accidents in munition factories during the First World War.

Greenwood and Woods [1919] described three possible hypotheses concerning the distribution of accident data, viz:

Pure Randomness:	where each individual has an equal chance of having an accident and where this chance is not altered by the occurrence of an accident.
True Contagion:	where each individual has an equal chance of having an accident but where this chance is altered by the occurrence of an accident.
Apparent Contagion:	where each individual has a constant chance of having an accident but where this chance differs between individuals.

The term contagion is derived from the theory of infectious diseases, in which the occurrence of a disease in an individual is more likely if other individuals in the vicinity have previously contracted or carry the disease.

The Poisson distribution can be used to model data conforming with the hypothesis of pure randomness.

However, for the two contagion hypotheses, the varying chances or non-randomness of accident occurrence between individual people has to be recognised and accommodated in any statistical model. This non-randomness is caused by a "clustering" of the data about individual "accident prone" people which leads to what is often referred to as over-dispersed data (over-dispersion, although, not a well defined statistical term, may be used to describe data for which the variance is significantly greater than the mean value). To model this, an appropriate compounding distribution is required where independent distributions are assumed for each of the within-case distributions and the between-case distribution as the compounding agent.

The compounding distribution could take any form, although a number of distributions would be theoretically inappropriate for accident research work. For example, the Normal distribution could lead to the concept of negative accidents and imply Poisson distributions with negative parameters – an illogical state of affairs.

Greenwood and Yule [1920] wanted a physically meaningful model and chose the non-negative gamma distribution as their compounding agent on the grounds of mathematical convenience. Having adopted the Poisson distribution to model the within case accident occurrence; this yielded the negative binomial distribution.

In the context of the present study, Greenwood and Yule's approach can be paralleled as follows. If the number of accidents at a site in some fixed period of time follows a Poisson distribution with a mean value which differs from site to site, the distribution of accident totals over all sites will not be Poisson. The gamma distribution provides a plausible and mathematically convenient model for the parameter (ie the mean) of the Poisson distribution. The variation in this parameter may be regarded as a measure of accident proneness.

The hypothesis of true contagion with respect to a location-based case structure does not seem, intuitively, to be sound. For true contagion to exist the occurrence of an accident at a location would alter the likelihood of the occurrence of a further accident. If this were so, either:

1. the location itself is attributed with both a memory and the means to enable the memory to alter future accident propensities. This is a fanciful notion and can be discounted; or

2. the occurrence of an accident at a location becomes a part of the road users' collective knowledge engendering a change in collective behaviour at the location which, in turn, brings about an increased collective propensity to generate accidents. This notion is somewhat less fanciful and is worthy of consideration.

The concept of accident proneness is associated with the hypothesis of apparent contagion where each case has a different propensity to incur accidents. However, as the theories behind both true and apparent contagion lead to negative binomial distributions (see, for example, Kemp [1970] and Moran [1968]), conformity of data with a negative binomial distribution, in itself, cannot allow a distinction to be made between the two contagion hypotheses. Kemp [1970], however, describes a statistical method - which will be used later - to allow this distinction to be made.

From the 1920's through to the late 1940's the literatue suggests that the accident proneness concept became accepted as "established fact" (Farmer and Chambers [1939]). Smeed [1949], however, suggested that the variation in accident occurrence to individuals may not be related to accident proneness but to accident liability. The distinction between accident proneness and liability had been made earlier by

Farmer and Chambers [1926], in which accident proneness was defined as a narrower term than accident liability relating to a ".....personal idiosyncracity predisposing the individual who possesses it in a marked degree to a relatively high accident rate". Accident liability, however, was defined to include all factors determining accident rate. For example, an individual may have a high accident liability not because he is accident prone but because he regularly uses roads on which the risk of an accident is higher.

The term proneness will be used in the present study to describe that idiosyncratic predisposition of a site to generate a relatively high accident rate. Implicit in this, is the assumption that those non-proneness factors associated with accident liability do not vary from site to site. That is to say, that factors such as the distribution of driver ability is assumed to be uniform throughout the Lothian Region Accident Network.

A further hypothesis has been put forward by Cresswell and Froggatt [1963] which was shown to reproduce successfully observed data for acccidents experienced by individual bus drivers. They rejected the concept of accident proneness and suggested that individual drivers are liable to have "spells" - perhaps associated with bad health - during which accidents may occur.

In this respect, a study by Palmer [1979] researched the relationships between biorythms and accident risk. Palmer described biorythms as:

> ".....three supposedly predictable cycles beginning from the moment of birth and applying to all human beings which govern their physical, emotional and intellectual 'ups' and 'downs' and hence influence their performance of various activities, including the driving of a motor vehicle."

She concluded, on the basis of data for 112,560 drivers obtained from insurance companies, that no convincing evidence existed to support the biorythm theory. Such a theory, however, should not be of concern to the present study because it can be reasonably assumed that the distribution of biorythm cycles is not biased by both time and space: that is to say, the proposition that road users at certain times in their biorythm cycle will tend to use only certain parts of the road network is discounted. Indeed, such a proposition begs the question that there is, in the first instance, a relationship between biorythms and accident risk.

In the context of location-based data, a "spell" can be considered as a period of time only during which accidents can occur and no accident can occur outside a "spell". In statistical terms, a "spell" is a rare event which occurs at random and the number of "spells" in any one time period is independent of the number in any other time period. Further, all sites are equally likely to experience a "spell" and that the probability of an accident occurring during a "spell" is constant and independent of the probability of a "spell".

The adoption of a "spells" hypothesis with respect to location-based accident data would require it to be accepted that for individual sites there is some reason why accidents occur during one time period and not during another. Some such reason could relate to the presence of road works or other environmental factors (eg vigorous seasonal growth of roadside vegetation reducing sight-lines) or perhaps even some site-selective, corporate, non-contagious, road-user phenomenon (eg short term alterations to driver behaviour following public information films on the topic of, say, procedures at roundabouts).

The difference between the "spells" and the true contagion hypotheses is that with true contagion it is the very fact that an accident has occurred that leads to the potential generation of further accidents, whereas a "spells" hypothesis merely states that accidents have constant probability of occurring during a "spell" but not outside. In addition, the difference between the "spells" and apparent contagion hypotheses is that with apparent contagion the risk of an accident occurring at a site is constant with time (albeit the risk varies from site to site) whereas the "spells" hypothesis suggests a non-constant risk of accidents with time.

In mathematical terms, the "spells" hypothesis can be given an interpretation similar to that for accident proneness. Consequently, agreement of the data with a negative binomial model would not, of itself, deny the "spells" hypothesis. However, because the "spells" hypothesis implies zero or very low correlation between successive time periods (Kemp [1970]), a distinction can be made between the apparent contagion and "spells" hypotheses.

Although these four hypotheses were reasoned with respect to individual people, the parallel with individual road locations is useful.

In addition to these four hypotheses, a fifth exists. It is the opposite phenomenon to that of contagion and is known as repulsion. Two possible contextural interpretations of repulsion involve:

1. the idea of a compensatory process where, when accidents do occur at a location, a degree of immunity to further accidents is

generated perhaps encouraged by road users becoming more careful for a while following the occurrence of an accident: this would require, of course, that the occurrence and location of an accident was collectively known to road users (note that this concept is the opposite to that of true contagion); or

2. the idea that there are only so many accidents which can occur at a location and that the probability of any one of these accidents occurring is constant.

The repulsion hypothesis would lead to a regular distribution of data for which the positive binomial distribution is an approximate model (see Elliott [1977] and Nicholson [1985]).

Early consideration of these five hypotheses is essential to the appropriateness of one of the prime purposes of the current study; that is to examine the possibility of predicting road accidents and the subsequent use of the predictions to identify hazardous locations (or blackspots) for possible remedial treatment. Without the acceptance of the accident proneness concept it would have to be concluded that accidents occur according to either the regular, random or the true contagion hypotheses. The consequence of this is to deny the possibility that individual locations have different underlying chances of incurring accidents and, in turn, deny the theoretical justification for directing resources to the identification of accident prone locations thus rendering impotent the whole concept of identification and treatment of hazardous road locations. On the other hand, if the apparent contagion hypothesis were to be accepted, the theoretical justification for the concept of blackspot identifaction and treatment would remain intact. It is, for the above reasons, crucial to the present study that early consideration of the likely controlling hypothesis is examined. The importance of this cannot be overstated.

Little help is afforded by the existing literature on this topic. Although there is a limited amount of work - outlined below - on the spatial distribution, no references have been found in which an attempt has been made to determine which of the two contagion hypotheses - indeed, if any - may be appropriate to location-based road accident data. A number of references, however, are known which examine the spatial distribution with respect to real data.

Chapman [1969] examined numbers of accidents per quarter mile section of 29 miles of the Pacific Highway in New South Wales and observed that there was a marked difference between the observed and the Poisson expected distribution of accidents. Chapman concluded ".....that some sections of road have more accidents than could be expected by chance, and since length, flow, periods of light, darkness, wetness and dryness are the same for all sections, some design feature may well be responsible." Chapman made no attempt to fit his data to other distributions (see, however, Table 6.14 in which Chapman's data are examined favourably for goodness of fit with a negative binomial distribution).

Abbess et al [1981] examined the effectiveness of the negative binomial distribution to model spatially distributed accident data at blacknodes in Hertfordshire for 1975-1979. Overall, Abbess et al found the fit to be sufficiently close to justify the use of this distribution.

Hauer and Persaud [1983b] examined data for 2,736 highway ramps in Ontario, Canada and concluded that ".....there is good support....." for the negative binomial assumption.

Maycock and Hall [1984] examined the the use - in a regression analysis on different accident types at 84 4-arm roundabouts in the United Kingdom - of both the negative binomial and Poisson errors. They reported ".....the negative binomial analyses did not generate values (for the regression coefficients) which were significantly different from the Poisson analyses for any of the accident types."

The author in another context (McGuigan [1985]) in an examination into the theory of "accident migration", which made use of the current study data, suggested that the fit of accident occurrence at junctions with the negative binomial distribution ".....is not an unreasonable one....."

Andreassen and Hoque [1986], in a study using spatially distributed accident data for the Melbourne Metropolitan Area, Australia, reported that a truncated negative binomial distribution did not suit the data but that the logarithmic series distribution was found to describe the data adequately. The logarithmic series distribution has one parameter and can only normally be used when zero values are missing. It is, therefore, particularly useful when events are not reported unless they occur at least once. One reported example by Ross [1980] of its use is to describe data such as numbers of parasites per host: which provides a rather colourful parallel to the current study topic.

Maher [1987], however, has shown, in comments on Andreassen and Hoque [1986], that the truncated negative binomial distribution fits the Melbourne data better than the log series distribution. This was disputed by Andreassen [1987] in a response to Maher's comments; however, the analysis presented by Maher is the more cogent and compelling. Indeed, Andreassen - a recognised researcher in this field - surprisingly seemed unaware of the role of the negative binomial distribution in the literature of this aspect of road accident research.

In conclusion, there would - on balance - appear to be a majority view that the negative binomial distribution provides an approximate model for the spatial distribution of accidents where the zero accident sites are known and this, therefore, provided a logical starting point for the study analysis.

6.2. The Distribution of Between-site Mean Accident Totals

Before determining the effectiveness of the negative binomial distribution to model spatially distributed accident data, a brief detour was taken to examine the nature of the distribution of between-site accident accident totals. It has been indicated above that Greenwood and Yule [1920] adopted the gamma disitribution - on the grounds of mathematical convenience - as the compounding agent to represent the distribution of the true mean accident liability associated with each individual person. Its use with the Poisson distribution to model the within site temporal distribution leads to a negative binomial distribution.

Abbess et al [1981] and Hauer and Persaud [1983b] have extended this concept to location-based road accident data where the gamma distribution is used to model the between site variation of accident proneness. Quimby et al [1986] in a study of accident frequency for individual drivers indicate that the gamma distribution ".....is not too unrealistic a representation....." of the variation of accident liability between individual drivers. The gamma assumption, however, has not been the subject of analysis insofar as it relates to location-based accident data.

At individual locations it has been shown for the study data that the within site temporal distribution of accident occurrence can be modelled by a Poisson process. For individual locations, therefore, the accuracy with which an observed accident total reflects the true mean accident total depends on the magnitude of the observed accident total. The observed four-year accident totals associated with individual locations, as defined in this study, are small and, accordingly, the confidence limits for the true mean four-year accident totals are large. Table 6.1 shows 95% Poisson confidence limits for a range of counts.

As the observed accident total increases, the accuracy of the estimate of the true mean is improved. It is possible to aggregate locations to obtain larger observed accident totals and, thus, obtain more accurate true mean estimates. It would then be possible to examine the distribution of these true mean estimates eliminating - to a greater or lesser extent - the sampling error asociated with the small observed accident totals at individual sites.

For example, Column 6 of Table 3.10 shows the overall accident rate for each link category expressed as accidents/LMVKM. These rates are aggregated over a number of sites where the observed aggregated 4-year accident totals on which the rates are calculated average just over 150, ranging from 13 to 880 accidents per category with a 25th percentiles of 29 and a 50th percentile of 66 for the 26 categories. Accordingly, the category accident rates are subject to varying degrees of error which, to a large extent, undermines the exercise.

The goodness of fit of these 26 accident rates with both the Normal and gamma distributions were analysed individually by the FIT FREQUENCY module in MLP (ROSS [1980]) - Maximum Likelihood Program. This module requires as input data the individual accident rates together with information to enable the data to be grouped according to the user's needs. The output from this module provides maximum likelihood estimates for the distribution parameters together with appropriate χ^2 values. With only 26 values it was observed that the analysis was sensitive to even minor adjustments of the grouping criteria, to the extent that significant changes in χ^2 values could be achieved. On the other hand, on the basis of conventional hypothesis testing, the gamma distribution - irrespective of the grouping criteria applied - always provided an acceptable fit. However, in view of the small number of aggregations and the arbitrary nature of the data classification used, no conclusions could reasonably be drawn from the χ^2 analysis.

From a theoretical point of view, however, the data would not support the use of the Normal distribution as a model for distribution of between site accident rates. The sample mean of 1.584 with a standard deviation of 1.192 provides a 95% confidence interval which would indicate - for a Normal model - that some 20% of the link categories would have a negative accident rate. For junction categories, a similar analysis yielded broadly similar results. The concept of negative accident rates is illogical. The gamma distribution, on the other hand, does not model negative accident rates and, therefore, on purely theoretical considerations would be preferred.

6.3. Analysis of the Study Data

As outlined in earlier chapters, the weight of evidence would suggest that the spatial distribution of accidents might be well modelled by the negative binomial distribution and, accordingly, it is the effectiveness of this distribution that is first examined.

In mathematical terms the negative binomial distribution is a two parameter distribution defined by the mean μ, and the exponent, k. The variance, σ^2, of the distribution is equal to $\mu+\mu^2/k$ from which it can be seen that as k becomes large, the distibution converges to a Poisson series (ie as σ^2 approaches μ). The reciprocal of k, $1/k$, is a measure of the excess variance in a sample and is sometimes referred to as the "clumping" factor describing the propensity of the data to form into groups rather than be dispersed randomly.

It has been shown in earlier chapters that the accident frequencies do vary according to site type and traffic flow characteristics and, therefore, the overall spatial distribution of accident totals will be dependent inter alia on the relative frequency of site type and traffic flow characteristic. Accordingly, the spatial distribution of accidents for mixed groups of site types would not necessarily be expected to follow any known distribution because the composition of a group is purely a function of the data available for the study. For example, the spatial distribution of a composite group comprising J41P1, J53P2 and J64S3 sites would be largely dependent on the relative frequencies of the site types, particularly where one site type predominates. In statistical terms such distributions are said to be multi-modal in nature. To eliminate the effect of the multi-modal nature of the data they have been disaggregated – as far as is sensible – into groups of similar characteristics.

The disaggregation process has been undertaken in a hierarchical process where higher level disaggregations have been analysed first before proceeding to lower level disaggregations.

For each disaggregation the data have been analysed initially in an identical manner to that adopted for the temporal distribution analysis in the previous chapter where the value of the IOD is used to determine whether or not the data conform with the Poisson distribution. Where the data are shown not to be in agreement with the Poisson distribution they have been further analysed to determine whether or not they conform with the negative binomial distribution.

To check for agreement with the negative binomial distribution the computer program MLP (Ross [1980]) has been used. This program calculates maximum likelihood estimates for the parameters of user selected distributions and, on the basis of these parameters, expected values for distributions are calculated. For each selected distribution, the program calculates a χ^2 statistic and, if its value lies between the relevant tabulated upper and lower 95% confidence limits, conformity with the distribution will be accepted for the purposes of this study.

This process provides four possible interpretations:

INTERPRETATION 1: The data conform with the Poisson distribution and, therefore, do not require the extra parameter of the negative binomial distribution to generate a satisfactory fit.

INTERPRETATION 2: The data do not agree with the Poisson distribution but do agree with the negative binomial distribution.

INTERPRETATION 3: The data agree with neither the Poisson distribution nor the negative binomial distribution.

INTERPRETATION 4: The data do not agree with the Poisson distribution but there is insufficient data to check for agreement with the negative binomial distribution.

Examples of the analysis process are shown in Table 6.2 for data conforming with INTERPRETATION 1 and in Table 6.3 similarly for INTERPRETATION 2 data.

Tables 6.4 and 6.5 show the analyses for priority and signal controlled junction data respectively, disaggregated by junction type and location for those disaggregations with 10 or more constituent sites. Of the 37 analyses, 6 and 27 lead to INTERPRETATIONS 1 and 2 respectively with 4 leading to INTERPRETATION 4. In addition, it can be noted when the data are disaggregated simply by either location or junction type INTERPRETATION 2 prevails.

Analyses for 27 selected disaggregations by junction type, location and ROOTBAND are shown in Table 6.6, of which 8 conform with INTERPRETATION 1 and all 19 remaining analyses agree with INTERPRETATION 2.

For links, Table 6.7 summarises the analyses of 26 disaggregations by sitecode of which 5 conform with INTERPRETATION 1 and all 21 remaining analyses agree with INTERPRETATION 2. For all 2,491 links combined INTERPRETATION 4 is observed.

Table 6.8 shows for the 21 more detailed disaggregations by link sitecode and LMVKMBAND that INTERPRETATIONS 1, 2 and 4 occur 5, 14 and 2 times respectively.

These results are summarised in Table 6.9.

It is worthy of note that the occurrences of INTERPRETATION 1 are generally associated with sitecodes experiencing lower annual mean accident totals often related to lower levels of traffic use and/or to less urbanised location types.

Two hypotheses cannot be sustained by the results of the above analysis. They are that of (a) pure randomness because the Poisson distribution does not provide a robust model in all but a few of the disaggregations, and (b) repulsion because the variance of observed accident totals for nearly all disagregations is greater than the mean (for repulsed data the variance is expected to be less than the mean).

The foregoing suggests, therefore, that accident data may conform with either of the two contagion hypotheses or with the "spells" hypothesis. To determine which, a method attributed by Kemp [1970] to Maritz [1950], Arbous and Kerrich [1951], and Bates and Neyman [1952] has been applied to the data.

The method involves splitting the data into two non-overlapping time periods and then correlating the number of accidents occurring at sites during the two time periods. Under the "spells" (and also pure randomness) hypothesis there should be no correlation, but for the true and apparent contagion hypotheses correlation would be expected to occur. For true contagion, however, the removal of sites experiencing no accidents during the first time period should increase the correlation, whereas this would not be the case for the apparent contagion hypothesis.

The all accident JACCTOT and LACCTOT data were split into the simple time periods of 1979-1980 and 1981-1982 and disaggregated to the sitecode groupings as used in Tables 6.4 and 6.5 for junctions and Table 6.7 for links. The new variables so created have been ascribed the labels JACCTOT12 and LACCTOT12 for the 1979-1980 junction and link accidents respectively. Similarly, the variables JACCTOT34 and LACCTOT34 refer to 1981-1982 accidents. For each sitecode, Pearson product-moment correlation coefficients were calculated for the relationships between the first 2-year and second 2-year accident totals for all sites and also only for those sites with non-zero accident totals for the first 2-year period. The use of the non-parametric Spearman's Rank Correlation Coefficient was not considered to be appropriate because of the very large number of tied ranks associated with the data.

Tables 6.10 and 6.11 show the values of the Pearson correlation coefficients for junctions and links respectively. Of the 63 disaggregations, 15 show increased values for the coefficients following the removal of the first period zero-accident sites and

46 show reduced values (for 2 disaggregations there were insufficient data to calculate coefficients for the second time period or there were no zero accident sites for the first time period). This analysis, therefore, does not provide any evidence to support the "spells" and true contagion hypotheses; it does, however, leave the apparent contagion hypothesis intact.

There still, however, remains the possibility that the exposure to accident risk at individual locations varies by virtue of the traffic volumes at the locations giving rise to an accident proneness type of interpretation. This possibility has been paralleled by McKenna [1983] in relation to accidents occurring to individual people and their likely different levels of exposure to risk. To examine this in rather more detail, the data for sitecode J42PX have been disaggregated into ten narrower traffic flow bands each representing approximately 10% of the 1,159 J421X sites.

The distribution of JROOTBAND within these ten disaggregations are summarised in Table 6.12 where it can be seen that for the four disaggregations at the tails (the first one and the final three); the JROOTBAND ranges are relatively wider than the remaining six and the variance to mean ratios are correspondingly greater. The six central ranges contain values for JROOTBAND which are reasonably evenly distributed across the range and the low values for the variance to mean ratios are evidence of this. Further, in Table 6.12, the the Pearson correlation coefficient is shown for the relationship between accidents and traffic flow for each range and whilst at the tails there is some evidence of correlation none is observed for the central ranges. Therefore, where the variation in traffic flow is tightly controlled (ie in the central ranges) the accident totals are, as might be expected, seen to be independent of traffic.

Table 6.13 shows an analysis - based on the same JROOTBAND ranges as adopted for Table 6.12 - of the distribution of accident totals in a manner identical to that carried out above in which it can be seen, with the exception of the first JROOTBAND range, that the data agree with INTERPRETATION 2 in that they conform with the negative binomial distribution but not with the Poisson. This indicates that accident occurrence can conform with a negative binomial distribution independently of traffic volumes and, accordingly, any argument to the effect that the negative binomial distribution is purely a manifestation of the underlying distribution of traffic volumes, cannot be realistically sustained. Accordingly, the hypothesis of apparent contagion remains the only satisfactory explanation for the spatial distribution of accidents to have been considered in the above analysis.

6.4. An Examination of Data from Other Sources

Table 6.14 shows a summary of an analysis carried out on data published by a number of different authors for a range of different accident locations and type. All of the eleven sets of data examined were found to conform with the negative binomial distribution but not with Poisson and, accordingly, add independent support for INTERPRETATION 2.

6.5. Conclusions

The evidence of this analysis lends considerable support to the hypothesis of apparent contagion as a model for road accident occurrence and – by corollary – indicates that the hypotheses of true contagion and "spells" cannot be logically sustained (indeed, the analysis provides an additional argument for rejecting the pure randomness hypothesis).

This can be viewed with some relief by members of local authority road accident investigation teams, for if there were support for one of the other competing hypotheses, much of their work would be wholly without theoretical justification.

Whilst these conclusions are not, perhaps, unexpected, the degree to which the data conform with the apparent contagion hypothesis is worthy of remark.

TABLE 6.1: 95% CONFIDENCE LIMITS FOR SINGLE COUNTS FROM POISSON DISTRIBUTED DATA

SINGLE COUNT	LOWER LIMIT		UPPER LIMIT	
	VALUE	% DIFF	VALUE	% DIFF
0	0.00		3.70	
1	0.10	-90.00	5.60	460.00
2	0.20	-90.00	7.20	260.00
3	0.60	-80.00	8.80	193.33
4	1.00	-75.00	10.20	155.00
5	1.60	-68.00	11.70	134.00
6	2.20	-63.33	13.10	118.33
7	2.80	-60.00	14.40	105.71
8	3.40	-57.50	15.80	97.50
9	4.00	-55.56	17.10	90.00
10	4.70	-53.00	18.40	84.00
15	8.40	-44.00	24.80	65.33
20	12.20	-39.00	30.80	54.00
25	16.20	-35.20	36.80	47.20
30	20.20	-32.67	42.80	42.67
40	28.60	-28.50	54.50	36.25
50	37.00	-26.00	65.90	31.80
75	58.03	-22.63	91.97	22.63
100	80.40	-19.60	119.60	19.60
150	126.00	-16.00	174.00	16.00
200	172.28	-13.86	227.72	13.86
250	219.01	-12.40	280.99	12.40
300	266.05	-11.32	333.95	11.32

TABLE 6.2: EXAMPLE OF SPATIAL DISTRIBUTION ANALYSIS OF JUNCTION DATA FOR SITECODE J1111

			NEGATIVE BINOMIAL	
NUMBER OF ACCIDENTS	OBSERVED FREQUENCY (O)	OBSERVED ACCIDENTS	EXPECTED FREQUENCY (E)	$(O-E)^2/E$
0	23	0	-	-
1	23	23	-	-
2	11	22	-	-
3	8	24	-	-
4	0	0	-	-
5	2	10	-	-
SUM	67	79	-	-

Test for agreement with a Poisson distribution:

Mean(O)=1.179 Variance(O)=1.452 Sample size=67
Index of Dispersion (O)=81.28
Degrees of Freedom (O)=66
Lower 95% Confidence Limit=45.431
Upper 95% Confidence Limit=90.349

Index of Dispersion > lower and < upper 95% Confidence Limits
Agreement with a Poisson distribution is accepted.

Test for agreement with a negative binomial distribution not carried out because of satisfactory fit with Poisson process.

TABLE 6.3: EXAMPLE OF SPATIAL DISTRIBUTION ANALYSIS OF LINK DATA FOR SITECODE L4413

			NEGATIVE BINOMIAL	
NUMBER OF ACCIDENTS	OBSERVED FREQUENCY (O)	OBSERVED ACCIDENTS	EXPECTED FREQUENCY (E)	$(O-E)^2/E$
0	31	0	30.23	0.02
1	24	24	25.20	0.06
2	17	34	17.79	0.04
3	11	33	11.81	0.06
4	11	44	7.59	1.53
5	4	20	4.78	0.13
over 5	7*	67	7.61	0.05
SUM	105	222	105.01	1.88

* Tail values are 7 8 8 9 9 11 and 15.

Test for agreement with a Poisson distribution:

Mean(O)=2.114 Variance(O)=6.564 Sample size=105
Index of Dispersion (O)=332.92
Degrees of Freedom (O)=104
Lower 95% Confidence Limit= 77.672
Upper 95% Confidence Limit=134.111

Index of Dispersion > upper 95% Confidence Limit
Agreement with a Poisson distribution is rejected.

Test for agreement with a negative binomial distribution:

χ^2=1.88 with 4 (7-3) Degrees of Freedom
Lower 95% Confidence Limit= 0.484
Upper 95% Confidence Limit=11.143

χ^2 > lower Limit and < upper 95% Confidence Limits
Agreement with negative binomial distribution is accepted.

TABLE 6.4: PRIORITY JUNCTION DATA - SUMMARY OF SPATIAL DISTRIBUTION ANALYSIS FOR CELLS WITH OVER 10 SITES

SITE CODE[1]		RURAL (£=1)	SUB-URBAN (£=2)	LOCAL CENTRE (£=3)	CITY CENTRE (£=4)	ALL LOCATIONS (£=X)
	IOD	38.99	60.57	-	-	139.20
J1£1X	df	34	42	1	0	80
	P/NB[2]	Y/-	Y/-	-	-	N/Y
	IOD	58.08	101.02	-	-	167.91
J2£1X	df	36	41	2	-	81
	P/NB	N/Y	N/Y	-	-	N/Y
	IOD	38.99	125.39	13.78	-	180.04
J3£1X	df	10	66	10	0	89
	P/NB	N/-	N/Y	Y/-	-	N/Y
	IOD	661.40	2619.7	480.76	315.83	4708.73
J4£1X	df	246	1158	210	92	1709
	P/NB	N/Y	N/Y	N/Y	N/Y	N/Y
	IOD	198.35	90.21	43.44	-	312.08
J5£1X	df	35	46	16	2	102
	P/NB	N/Y	N/Y	N/Y	-	N/Y
	IOD	113.07	384.49	57.63	42.10	686.20
J6£1X	df	38	126	18	9	194
	P/NB	N/Y	N/Y	N/Y	N/-	N/Y
	IOD	192.11	649.79	85.62	133.95	1358.42
J7£1X	df	165	203	32	43	446
	P/NB	N/Y	N/Y	N/Y	N/Y	N/Y
	IOD	62.79	201.89	27.45	79.00	637.60
J8£1X	df	28	77	14	26	148
	P/NB	N/Y	N/Y	N/Y	N/Y	N/Y
	IOD	-	-	-	-	59.21
J9£1X	df	2	5	-	1	10
	P/NB	-	-	-	-	N/-
	IOD	1618.19	4661.13	807.56	730.75	9396.16
JX£1X	df	602	1772	310	180	2867
	P/NB	N/Y	N/Y	N/Y	N/Y	N/Y
JXXXX		IOD = 14942.87		df = 3111		P/NB = N/N

[1] See Table 2.1 for description of SITECODE categories.
[2] P/NB indicates whether the distribution conforms with either the Poisson or negative binomial distribution respectively where "Y" = Yes, "N" = No and "-" = not applicable.

TABLE 6.5: SIGNAL-CONTROLLED JUNCTION ACCIDENT DATA - SUMMARY OF SPATIAL DISTRIBUTION ANALYSIS FOR CELLS WITH OVER 10 SITES

SITE CODE[1]		RURAL (£=1)	SUB-URBAN (£=2)	LOCAL CENTRE (£=3)	CITY CENTRE (£=4)	ALL LOCATIONS (£=X)
	IOD	-	57.48	9.24	-	74.34
J4£2X	df	-	10	10	7	29
	P/NB[2]	-	N/-	Y/-	-	N/Y
	IOD	-	-	-	-	-
J5£2X	df	-	0	-	-	0
	P/NB	-	-	-	-	-
	IOD	-	143.86	98.36	48.09	298.73
J6£2X	df	1	29	22	18	73
	P/NB	-	N/Y	N/Y	N/Y	N/Y
	IOD	-	96.20	21.00	108.51	289.33
J7£2X	df	-	27	13	38	80
	P/NB	-	N/Y	Y/-	N/Y	N/Y
	IOD	-	38.69	16.51	107.17	214.16
J8£2X	df	-	10	9	24	45
	P/NB	-	N/-	Y/-	N/Y	N/Y
	IOD	-	-	-	-	34.44
J9£2X	df	-	2	6	1	11
	P/NB	-	-	-	-	N/-
	IOD	-	510.34	288.94	452.25	1272.72
JX£2X	df	1	83	64	92	243
	P/NB	-	N/Y	N/Y	N/Y	N/Y

[1] See Table 2.1 for description of SITECODE categories.

[2] P/NB indicates whether the distribution conforms with either the Poisson or negative binomial distribution respectively where "Y" = Yes, "N" = No and "-" = not applicable.

TABLE 6.6: SELECTED JUNCTION ACCIDENT DATA - SUMMARY OF SPATIAL DISTRIBUTION ANALYSIS

SITE CODE[1]	NUMBER OF SITES	4-YEAR MEAN ACCIDENT TOTAL	VARIANCE OF ACCIDENT TOTALS	IOD	df	P/NB[2]	$\hat{k}$[3]
J1213	32	1.281	1.757	42.52	31	Y/-	-
J2213	29	0.793	1.956	69.06	28	N/Y	0.534
J3213	51	1.353	2.553	94.35	50	N/Y	4.685
J4111	197	0.198	0.313	309.84	196	N/Y	0.457
J4112	29	0.690	0.650	26.38	28	Y/-	-
J4113	21	2.571	7.957	61.90	20	N/Y	0.741
J4211	582	0.732	1.191	945.32	581	N/Y	1.528
J4212	381	1.457	2.544	663.50	380	N/Y	2.225
J4213	196	2.480	6.097	479.40	195	N/Y	1.811
J4311	79	1.696	5.112	235.10	78	N/Y	0.929
J4312	76	2.921	5.700	146.35	75	N/Y	2.600
J4313	56	3.179	5.568	96.33	55	N/Y	4.616
J4411	17	1.706	2.346	22.00	16	Y/-	-
J4412	20	1.950	2.155	21.00	19	Y/-	-
J4413	56	4.321	15.204	193.52	55	N/Y	1.447
J6211	47	1.489	3.212	99.23	46	N/Y	1.639
J6212	51	2.902	7.530	129.74	50	N/Y	2.220
J6213	29	5.241	12.261	65.50	28	N/Y	5.441
J7111	46	0.283	0.380	60.42	45	Y/-	-
J7112	45	0.400	0.336	36.96	44	Y/-	-
J7113	75	0.840	0.890	78.40	74	Y/-	-
J7211	87	0.621	1.401	194.02	86	N/Y	0.399
J7212	49	1.245	7.064	272.34	48	N/Y	0.623
J7213	68	1.544	3.088	134.00	67	N/Y	1.000
J8211	22	0.682	1.084	33.38	21	Y/-	-
J8212	20	2.200	4.800	41.45	19	N/Y	1.412
J8213	36	2.722	8.092	104.05	35	N/Y	0.909
JXX1X	2868	1.534	5.027	9395.31	2867	N/Y	0.738

[1] See Table 2.1 for description of SITECODE categories.
[2] P/NB indicates whether the distribution conforms with either the Poisson or negative binomial distribution respectively where "Y" = Yes, "N" = No and "-" = not applicable.
[3] Maximum likelihood estimate.

TABLE 6.7: LINK ACCIDENT DATA - SUMMARY OF SPATIAL DISTRIBUTION ANALYSIS FOR CELLS WITH 10 OR MORE CONSTITUENT SITES

SITE CODE[1]	NUMBER OF SITES	4-YEAR MEAN ACCIDENT TOTAL	VARIANCE OF ACCIDENT TOTALS	IOD	df	P/NB[2]	$\hat{k}$[3]
L111X	161	3.199	19.535	977.06	160	N/Y	1.290
L112X	14	3.643	6.709	23.94	13	Y/-	-
L121X	15	1.467	5.695	54.35	14	N/Y	1.178
L141X	71	1.563	4.935	221.02	70	N/Y	1.754
L151X	38	2.158	10.028	171.94	37	N/Y	1.776
L152X	16	8.125	95.317	175.97	15	N/Y	0.640
L171X	22	3.318	12.418	78.59	21	N/Y	6.312
L221X	19	1.737	2.760	28.60	18	Y/-	-
L241X	40	1.650	3.874	91.57	39	N/Y	1.762
L251X	42	2.048	7.022	140.58	41	N/Y	1.046
L331X	17	2.059	3.434	26.68	16	Y/-	-
L341X	32	0.906	1.701	58.20	31	N/Y	0.793
L351X	17	1.118	1.735	24.83	16	Y/-	-
L441X	574	0.861	2.358	1569.26	573	N/Y	0.616
L442X	15	0.867	1.552	25.06	15	Y/-	-
L451X	342	1.178	3.631	1051.08	341	N/Y	0.989
L452X	17	0.765	2.566	53.67	16	N/Y	0.203
L471X	32	1.844	4.652	78.21	31	N/Y	0.556
L551X	177	1.864	6.686	631.30	176	N/Y	0.822
L552X	60	1.117	3.257	172.03	59	N/Y	1.092
L571X	21	2.714	11.414	84.11	20	N/Y	2.674
L661X	438	2.009	9.135	1987.06	437	N/Y	0.618
L662X	194	1.082	8.843	1577.36	193	N/Y	0.249
L771X	69	1.884	5.780	208.62	68	N/Y	0.570
L772X	17	2.412	7.257	48.14	16	N/Y	1.042
L773X	32	0.781	0.951	37.75	31	N/Y	1.645
LXXXX	2492	1.595	7.558	11803.75	2491	N/N	0.593

[1] See Table 2.1 for description of SITECODE categories.
[2] P/NB indicates whether the distribution conforms with either the Poisson or negative binomial distribution respectively where "Y" = Yes, "N" = No and "-" = not applicable.
[3] Maximum likelihood estimate.

TABLE 6.8: SELECTED LINK ACCIDENT DATA - SUMMARY OF SPATIAL DISTRIBUTION ANALYSIS

SITE CODE[1]	NUMBER OF SITES	4-YEAR MEAN ACCIDENT TOTAL	VARIANCE OF ACCIDENT TOTALS	IOD	df	P/NB[2]	$\hat{k}$[3]
L1111	67	1.179	1.452	81.28	66	Y/-	-
L1112	74	3.568	10.057	205.76	73	N/Y	2.362
L1113	20	8.600	75.200	166.14	19	N/Y	-65.1
L4411	267	0.386	0.589	405.89	266	N/Y	0.820
L4412	202	0.837	1.451	348.45	201	N/Y	1.171
L4413	105	2.114	6.564	322.92	104	N/Y	1.440
L4511	117	0.427	0.643	174.68	116	N/Y	1.129
L4512	146	0.993	2.034	297.01	145	N/Y	1.398
L4513	79	2.633	8.082	239.42	78	N/Y	5.634
L5511	54	0.444	0.554	66.13	53	Y/-	-
L5512	52	1.057	1.742	84.05	51	N/Y	0.960
L5513	71	3.535	10.252	203.01	70	N/Y	2.517
L6611	50	0.160	0.137	41.96	49	Y/-	-
L6612	90	0.578	0.719	110.71	89	Y/-	-
L6613	298	2.752	11.453	1236.03	297	N/Y	0.953
L6621	71	0.099	0.204	144.24	70	N/-	-
L6622	55	0.218	0.285	70.60	54	Y/-	-
L6623	68	2.809	20.366	485.77	67	N/Y	1.077
L7711	23	0.348	1.146	72.45	22	N/-	-
L7712	28	1.643	2.757	45.31	27	N/Y	5.604
L7713	18	4.222	8.183	32.95	17	N/Y	2.245

[1] See Table 2.1 for description of SITECODE categories.
[2] P/NB indicates whether the distribution conforms with either the Poisson or negative binomial distribution respectively where "Y" = Yes, "N" = No and "–" = not applicable.
[3] Maximum likelihood estimate.

TABLE 6.9: SUMMARY OF RESULTS OF SPATIAL DISTRIBUTION ANALYSIS

	INTERPRETATION[1]			
	1	2	3	4
HYPOTHESES:				
Pure Randomness	Y	N	N	N
True Contagion	N	Y	N	-
Apparent Contagion	N	Y	N	-
"Spells"	N	Y	N	-
Repulsion	N	N	?	-
OBSERVATIONS:				
Junctions (Table 6.7)	5	21	0	0
Links (Table 6.8)	5	14	0	2
TOTAL OBSERVATIONS	10	35	0	2

[1] "N" = Reject hypothesis
"Y" = Accept hypothesis
"-" = Insufficient data
"?" = Undertake additional examination

TABLE 6.10: SELECTED PEARSON CORRELATION COEFFICIENTS FOR THE RELATIONSHIPS BETWEEN JACCTOT12 AND JACCTOT34

	ALL SITES		ALL SITES (JACCTOT12>0)		
SITE CODE	SAMPLE SIZE	CORRELATION COEFFICIENT (r_1)	SAMPLE SIZE	CORRELATION COEFFICIENT (r_2)	COMMENT
J111X	35	0.300	5	*****[1]	o[2]
J121X	43	0.158	14	-0.090	-[3]
J211X	37	0.232	6	-0.270	-
J221X	42	0.558	10	0.249	-
J311X	11	0.315	3	-0.500	-
J321X	67	0.208	25	0.269	+[4]
J331X	11	0.076	6	0.426	+
J411X	247	0.408	47	0.247	-
J421X	1159	0.372	438	0.383	+
J431X	211	0.218	133	0.129	-
J441X	93	0.567	62	0.531	-
J511X	36	0.790	5	0.707	-
J521X	47	0.398	19	0.131	-
J531X	17	0.137	9	0.128	-
J611X	39	0.645	23	0.665	+
J621X	127	0.467	82	0.375	-
J631X	19	0.476	10	0.174	-
J641X	10	0.784	9	0.741	-
J711X	166	0.052	37	-0.044	-
J721X	204	0.472	65	0.470	-
J731X	33	0.349	17	0.332	-
J741X	44	0.262	22	0.404	+
J811X	29	0.553	11	0.451	-
J821X	78	0.501	38	0.448	-
J831X	15	0.549	12	0.392	-
J841X	27	0.506	24	0.480	-
J422X	11	0.754	9	0.701	-
J432X	11	-0.186	10	0.069	+
J622X	30	0.650	28	0.694	+
J632X	23	0.680	21	0.640	-
J642X	19	0.517	19	0.517	o
J722X	28	0.506	9	0.099	-
J732X	14	0.403	8	0.240	-
J742X	39	0.456	27	0.485	+
J822X	11	0.799	6	0.540	-
J832X	10	0.250	7	0.042	-
J842X	25	0.743	22	0.680	-
JXXXX	3112	0.626	1333	0.601	-

[1] "*****" it was not possible to calculate r_2.
[2] "o" indicates a null comment.
[3] "–" indicates that r_2 is less than r_1.
[4] "+" indicates that r_2 is greater than r_1.

TABLE 6.11: SELECTED PEARSON CORRELATION COEFFICIENTS FOR RELATIONSHIPS BETWEEN BETWEEN LACCTOT12 AND LACCTOT34

	ALL SITES		ALL SITES (LACCTOT12>0)		
SITE CODE	SAMPLE SIZE	CORRELATION COEFFICIENT (r_1)	SAMPLE SIZE	CORRELATION COEFFICIENT (r_2)	COMMENT
L111X	161	0.517	105	0.500	–[1]
L112X	14	0.055	11	-0.171	–
L121X	15	0.695	6	0.893	+[2]
L141X	71	0.543	28	0.600	+
L151X	38	0.692	19	0.719	+
L152X	16	0.820	11	0.747	–
L171X	22	0.785	17	0.769	–
L221X	19	0.319	8	0.143	–
L241X	40	0.440	21	0.272	–
L251X	42	0.479	20	0.404	–
L331X	17	-0.164	7	-0.310	–
L341X	32	0.228	12	-0.094	–
L351X	17	0.387	7	0.471	+
L441X	574	0.412	164	0.353	–
L442X	15	0.433	5	0.000	–
L451X	342	0.482	127	0.487	+
L452X	17	0.657	4	0.440	–
L471X	32	0.527	15	0.409	–
L551X	177	0.590	84	0.557	–
L552X	60	0.309	22	0.146	–
L571X	21	0.647	14	0.639	–
L661X	438	0.582	177	0.494	–
L662X	194	0.726	45	0.763	+
L771X	69	0.548	9	0.357	–
L772X	17	0.616	8	0.683	+
L773X	32	0.022	21	0.000	–
LXXXX	2492	0.564	977	0.512	–

[1] "–" indicates that r_2 is less than r_1.
[2] "+" indicates that r_2 is greater than r_1.

TABLE 6.12: SITECODE J42PX - SUMMARY OF ANALYSIS OF CORRELATIONS BETWEEN ROOTBAND AND 4-YEAR ACCIDENT TOTALS FOR FURTHER DISAGGREGATIONS

ROOTBAND RANGE	NUMBER OF SITES	ROOTPROD VARIANCE TO MEAN RATIO	CORRELATION COEFFICIENT (r)	SIGNIFICANCE OF r (AT 5% LEVEL)
0.00-0.13	129	0.516	0.178	YES
0.13-0.16	94	0.040	-0.178	NO
0.16-0.22	122	0.172	0.032	NO
0.22-0.26	110	0.052	-0.177	NO
0.26-0.32	128	0.106	0.100	NO
0.32-0.38	115	0.068	0.019	NO
0.38-0.45	110	0.115	-0.069	NO
0.45-0.60	116	0.408	0.220	YES
0.60-0.90	116	1.060	0.171	NO
over 0.90	119	47.432	0.309	YES

TABLE 6.13: SITECODE J42PX - SUMMARY OF SPATIAL DISTRIBUTION ANALYSIS FOR FURTHER DISAGGREGATIONS

ROOTBAND RANGE	NUMBER OF SITES	4-YEAR MEAN ACCIDENT TOTAL	VARIANCE OF ACCIDENT TOTALS	IOD	df	P/NB[1]	$\hat{k}$[2]
0.00-0.13	129	0.264	0.321	155.64	128	Y/-	-
0.13-0.16	94	0.564	0.743	122.52	93	N/Y	1.467
0.16-0.22	122	0.836	1.494	216.24	121	N/Y	1.027
0.22-0.26	110	0.900	1.228	148.72	109	N/Y	1.906
0.26-0.32	128	1.086	1.701	198.92	127	N/Y	2.296
0.32-0.38	115	1.200	1.688	160.36	114	N/Y	3.174
0.38-0.45	110	1.273	2.641	226.13	109	N/Y	2.192
0.45-0.60	116	1.828	3.187	200.50	115	N/Y	2.325
0.60-0.90	116	1.784	3.301	212.79	115	N/Y	1.949
over 0.90	119	2.882	7.223	295.74	118	N/Y	2.045
ALL	1159	1.266	2.864	2619.68	1158	N/Y	1.109

[1] P/NB indicates whether the distribution conforms with either the Poisson or negative binomial distribution respectively where "Y" = Yes, "N" = No and "-" = not applicable.
[2] Maximum likelihood estimate.

TABLE 6.14: SUMMARY OF SPATIAL DISTRIBUTION OF ACCIDENTS REPORTED IN OTHER RESEARCH PAPERS

REFERENCE	NUMBER OF SITES	MEAN ACCIDENT TOTAL	VARIANCE OF ACCIDENT TOTALS	IOD	P/NB[1]	$\hat{k}$[2]
Bucks CC [1981]) Sect 3 Table 2 Sect A Col 4 All accidents	69	5.304	14.921	191.29	N/Y	2.083
Bucks CC [1981] Sect 3 Table 2 Sect A Col 4 Fat+ser accidents	69	1.826	3.146	117.15	N/Y	1.806
Herts CC [1984] Blacksite Table Col 9	189	15.984	67.941	799.11	N/Y	8.031
van Maren [1977] Appendix B Table 1B Nonsignalized Intersections Col 5 (1974)	34	2.088	11.053	174.69	N/Y	0.486
van Maren [1977] Appendix B Table 1B Nonsignalized Intersections Col 6 (1975)	34	2.677	15.559	191.80	N/Y	0.249
van Maren [1977] Appendix B Table 1B Nonsignalized Intersections Col 7 (1976)	34	2.588	19.280	245.84	N/Y	0.232
Proctor [1985] Table 5 Row 2	292	0.9144	3.144	1000.55	N/Y	0.478
Chapman [1969] Table I Col 2	116	2.621	9.420	413.32	N/Y	1.220
Hauer & Persaud [1983b] Table 3 Col 2	2736	0.3414	1.068	8555.89	N/Y	0.973
Leong [1973] Table VIII Col 4 (multiplied to 3-year total)	45	12.844	73.453	251.63	N/Y	3.852
Colgate and Tanner [1967] Table 1 Sum of Cols T1-T10	139	5.648	35.476	866.80	N/Y	1.536

[1] P/NB indicates whether the distribution conforms with either the Poisson or negative binomial distribution respectively where "Y" = Yes, "N" = No and "–" = not applicable.

[2] Maximum likelihood estimate.

CHAPTER 7 DEVELOPING A MODEL FOR ACCIDENT OCCURRENCE

7.1. Introduction

As indicated in Chapter 4, least square regression modelling procedures are not appropriate for data which do not conform with a Normal distribution. From the analyses described in Chapters 5 and 6 it is confirmed that accident data do not follow a Normal distribution. Accordingly, a generalised linear modelling technique has been adopted for the study and the computer program GLIM (Baker and Nelder [1978]) has been used. GLIM is an acronym for Generalised Linear Interactive Modelling.

7.2. Factors Affecting Accident Occurrence

Site Specific Factors

As outlined in Chapter 3 the frequency of accident occurrence is seen to vary according to the type and location of site and traffic flow. The analysis of the data will formally examine these relationships.

Many researchers including, for example, Tanner [1953], Gwynn [1967], Pfundt [1969], Gwynn and Baker [1970], Yu [1972], Brilon [1972], and Leong [1973] and, more recently, Maltby and Bennett [1986] report that the rate at which accidents happen tends to vary with traffic flow. Satterthwaite [1981] in his comprehensive review of the topic confirms this general point.

For links, this evidence suggests that single vehicle accident rates are inversely proportional to traffic flow and that multi-vehicle accidents are proportional to traffic flow. Taken together these two relationships would suggest the possibility that total accident rate varies in a U-shaped fashion with traffic flow. The hypothesised U-shaped nature of the relationship, however, has not been a consistent feature of research findings and this may have been caused by some systematic effects based on the nature of the different study data sets. Satterthwaite [1981] suggests that it is probably preferable to consider these two types of accidents separately when seeking relationships with traffic flows.

For junctions, the work of Tanner [1953], Colgate and Tanner [1967], and McDonald [1953] all suggest that accidents are broadly proportional to the square root of the product of the major and side road traffic flows. This evidence would indicate that the accident rate per vehicle negotiating a junction is inversely proportional to the flow of traffic. Satterthwaite [1981] in his review of relevant literature confirms this general point.

Chapman [1973] has described the concept of accident exposure as a measure of the opportunity for accidents to happen (see Section 5.1). This opportunity for link accidents may be modelled in terms of total distance travelled and for junctions to some appropriate function of turning traffic flows. In addition, Chapman refers to propensity as ".....the conditional probability that an accident occurs given the opportunity for one." It is apparent from the earlier research that the propensity may vary for both links and junctions according to the traffic flow.

There are a number of plausible reasons why this should be so, for example:

- as traffic flows increase, the amount of care exercised by road users increases, with the result that the chance of an accident occurring at each opportunity decreases (Tanner [1953] and Leong [1973]);
- locations with heavy traffic flows may well be built to a better design standard than lightly trafficked locations (Chapman [1973]); and
- as a result of the relationship between speed and flow (ie speed is inversely related to flow), the more heavily trafficked locations are negotiated by vehicles travelling at lower speeds than at their lightly trafficked counterparts with the result that either:
    - vehicles and/or pedestrians are more able to take avoiding action; or
    - there is a reduced chance of personal injury occurring in the event of an accident.

It may be that all three (or, indeed, others not considered here) of these mechanisms are in operation.

The relationship between traffic speed and flow varies according to the capacity of the road and it is possible, therefore, that the relationship between accident rate and propensity varies not simply with traffic flow but with some function of traffic flow and road capacity. For this reason, it may be better to introduce the concept of "traffic intensity" to describe this function and use it in preference to traffic flow.

This theme has been discussed in considerable detail by Hauer [1982] where the term "conflict" is used in a similar sense as "traffic intensity" is here. Hauer draws a distinction between conflict and exposure in which conflict ".....is a device for indirect estimation of.....safety" whereas exposure allows ".....the estimation of 'accident risk'....." which is ".....said to be a measure of the probability of a potential accident event to result in an accident....." The definition of conflict implies uniform risk per measure of conflict throughout a system but exposure underlies the differences between two systems in terms of risk where the two systems experience the same conflict.

There is a fundamental conceptual difference between the exposure to accidents on links and at junctions. Link accidents are generally considered to be related to aggregated travel distance, whereas junction accidents are related to conflict, with travel distance playing no significant role.

For links, the traffic flow may be considered to be a measure of what has been defined as traffic intensity. The concept of traffic intensity is necessary to form a model which recognises the variation in accident propensity indicated in the earlier research. Consider a simple example of three links of 1, 2 and 3 km lengths with annual traffic flows of 6, 3 and 2 million vehicles (MV) respectively; each, therefore, experiencing an identical annual accident exposure measure of 6 MVkm but with quite different levels of traffic flow or traffic intensity. It is suggested, therefore, that traffic intensity is inversely related to propensity.

For junctions, the concepts of exposure and traffic intensity are rather more difficult to visualise. Exposure, here, should relate to the opportunity for accidents to happen and it is generally accepted that some measure of the the conflict experienced at junctions modelled by a multiplicative function of major, Q^{α}, and side, q^{β}, road flow (or, indeed, of conflicting flows at individual points within junctions) provides the most robust exposure measurement. For example:

$$G = Q^{\alpha}.q^{\beta} \tag{7.1}$$

where:

G is the exposure to accidents.

Where, however, this function involves values of exponents for individual flows which are significantly different from unity (ie where both α and $\beta \neq 1$), it would not be a true measure of exposure as it could imply a relationship between traffic intensity and propensity. Another explanation for this – not necessarily associated with a relationship between traffic intensity and propensity – could lie in the fact that it is simply not possible – because of the temporal distribution of traffic – that every vehicle in traffic flow Q could negotiate a junction at the same time as every vehicle in traffic flow q suggesting values for the exponents of less than unity.

It is, however, generally understood in the context of junction accidents that measurements of exposure may involve such values for exponents and this understanding is maintained in the present study. It is true, however, that whatever multiplicative function for exposure is used, identical values for exposure can be generated from quite different dispositions of traffic flows negotiating the junction and, accordingly, traffic intensity could be considered to be either an additive, multiplicative or ratio function of such traffic flows.

Hourly v. annual traffic flows.

Much of the above early work is based on analyses of hourly traffic accident rates and hourly traffic flows. In the present study, as it is the annual accident totals that are of interest, the conclusions based on the above hourly analyses would not be relevant unless there were some relationship between annual and hourly distribution of traffic flows.

Phillips [1979] indicates that annual traffic flows can be estimated with reasonable accuracy from short period counts of as little duration as a few hours. This has been verified by the author in unpublished work based on continuous automatic traffic counter data for a number of sites in Lothian Region. The corollary of this work is that short period traffic flows may be estimated from annual flows, indicating that the hourly distribution of traffic over the course of a year is reasonably constant from site to site and independent of the annual traffic flow.

Accordingly, the annual traffic flow is an indicator of the distribution of hourly flows and the conclusions reported by Satterthwaite [1981] may be relevant to the present study. The possibility of a relationship between annual accident rate and annual traffic flow will be examined.

7.3. The Form of the Model

7.3.1. The Systematic Component

On the basis of earlier research, and, in particular, that by Maycock and Hall [1984], a multiplicative exponential model was considered to be the most appropriate.

Simply, the form of the model is:

$$A = G.r \tag{7.2}$$

where:

A is the accident total;

G is the exposure to accidents; and

r is the accident rate (or accident propensity) in relation to the exposure G.

The accident rate, r, is an exponential function of explanatory variables so that:

$$r = \exp(a_0 + a_1 X_1 + a_2 X_2 + \ldots\ldots + a_n x_n + b_1 \ln I + b_2 (\ln I)^2) \tag{7.3}$$

where:

a_i and b_i are the parameters to be estimated;

X_i are the explanatory variables;

I is the traffic intensity function; and

ln is log to the base e.

The I and I^2 terms are included in the model to determine whether or not there is a relationship between accident rate and traffic intensity where the inclusion of the I^2 term will determine if there is a quadratic effect which would provide evidence to suggest that the relationship is U-shaped.

It would have been possible, for links, to design a model where the dependent variable is defined as accidents per unit length of road (eg A/km) rather than accidents alone. This form of model has been adopted by previous researchers (see, for example, Turner and Thomas [1986]). It was decided, however, for the following reasons, not to adopt this form of model in order:

- to maintain a degree of conformity between the link and junction analyses; and

- to retain a more readily understood form for the dependent variable.

Brodsky and Hakkert [1983] highlight the problems associated with statistical models which contain a common component on both sides of the equation. This problem exists in the analysis of relationships, for example, between accident rate and traffic on links where the dependent variable would be accidents per million vehicle kilometres (A/MVkm) and the independent variable would be million vehicles (MV). The use of GLIM, however overcomes this problem by enabling the user to define an "offset" variable whose coefficient is constrained to be unity. So, in the link example, rather than examine the relationship between A/MVkm and MV it is possible to examine that between A and MVkm.MV where MVkm is defined as an "offset" variable.

Returning to the model we have:

$$A = G.\exp(a_0 + a_1 X_1 + a_2 X_2 + \ldots\ldots + a_n x_n + b_1 \ln I + b_2 (\ln I)^2) \qquad (7.4)$$

where:

G is defined as an "offset" variable.

To transform the multiplicative model into a suitable linear form for analysis by GLIM a log link function was used so that:

$$\ln(A) = \ln(G) + a_0 + a_1 X_1 + \ldots\ldots + a_n x_n + b_1 \ln I + b_2 (\ln I)^2 \qquad (7.5)$$

7.3.2. The Random Component

In Section 6.3 it has been shown that the spatial distribution of accidents is non-Poisson and that there is over-dispersion in the data which can be modelled by the negative binomial distribution. This over-dispersion has been explicitly discussed by Pickering *et al* [1986] in a study of accidents at rural T-junctions and also by Maycock and Hall [1984] who were studying accidents at 4-arm roundabouts. In both studies, however, a Poisson error structure was used and its limitations recognised. Indeed, Maycock and Hall report an analysis on the basis of a negative binomial error structure but found that the fitted parameter values did not differ significantly from those obtained from a similar analysis using a Poisson error structure.

Lawson [1986] concluded from an analysis of accidents on radial roads in the city of Birmingham, UK, that there was ".....little difference between the Poisson and negative binomial error structures in terms of their ability to describe the data....."

7.3.3. Significance Testing

As the purpose of statistical modelling is to determine which form of the systematic component of the model is most appropriate, it is necessary to be able to assess formally the statistical significance of the addition of explanatory variables to the model and their contribution to explaining the variation in the data.

In GLIM the statistic called Scaled Deviance, S, can be used to test the significance of terms added to the model. For Poisson and negative binomial error structures S is asymptotically distributed like χ^2 with $n-p-1$ degrees of freedom (where n is the number of observations and p the number of parameters fitted in the model). For a well fitting model it follows that the expected value for S, $E(S)$, will be approximately equal to the degrees of freedom, ie:

$$E(S) \approx n-p-1. \tag{7.6}$$

Accordingly, if there is over-dispersion in the data and a Poisson error structure is adopted then S should be greater than the degrees of freedom, (ie $E(S) > n-p-1$) and, if this is the case, a negative binomial error structure may be more appropriate, so that:

$$\mathrm{var}(y) = \mu+\mu^2/k \tag{7.7}$$

where:

y is the dependent variable (eg accident total);

μ is the the expected value of y (ie $E(y)$); and

k is the negative binomial shape parameter (sometimes referred to as the exponent).

Then, over a reasonable range of μ for the right-hand side of the equation, it can be assumed that:

$$\mathrm{var}(y) = \lambda.\mu \tag{7.8}$$

where:

λ is an empirically derived scale parameter > 1

and an estimate, $\hat{\lambda}$, of λ can be made thus:

$$\hat{\lambda} = S/(n-p). \quad (7.9)$$

In this way a negative binomial error structure could be approximated by a Poisson error structure combined with λ, hereafter referred to as a "scaled" Poisson model. However, where the value of k is known or can be reasonably assumed a negative binomial model could be fitted exactly by GLIM (See Baker and Nelder [1978]).

A practical result, however, of the over-dispersion, whether modelled directly by a negative binomial model or a Poisson model with λ, is that the standard errors associated with the parameter estimates will increase – in the Poisson approximation by $\sqrt{\lambda}$ – providing a basis of significance testing more appropriate to the over-dispersed nature of the data.

To put it another way: t-values based on the standard errors associated with parameter estimates for over-dispersed data calculated on the basis of a simple Poisson fit, over-estimate the significance the parameter estimates.

There was a choice, therefore, whether to have proceeded with a negative binomial or a "scaled" Poisson error structure and for reasons outlined in the following section it was decided to continue on the basis of a "scaled" Poisson model.

In GLIM it is possible to declare a "scaled" model (by the use of the GLIM directive, $SCALE, with a null or zero argument) obviating the need for further calculation of the "scaled" standard errors.

7.4. Contingency Tables

In the context of the present study, it is possible to present accident data in the form of a multi-way classification; where accidents, accident rates or mean accident totals are classified according to a set of background characteristics such as junction type, location or traffic flow band. If each characteristic is represented by a factor (or variable) which has a set of possible levels, then it is possible to aggregate accident data for each permutation of the levels of factors and to tabulate accordingly. Such a table is termed a contingency table.

In addition to determining whether or not the individual sitecode classifications have a bearing on accident occurrence, it is possible to assess how accident occurrence varies between the sub-classifications. If, for example, at junctions, it is found that the relationship between accident occurrence and traffic flow is not independent of junction type, then any statistical model for accident occurrence should be structured to recognise this.

Individual sites v. Aggregated sites as the data set

At this point it would have been possible to move directly to an analysis of data at an individual site level. This course of action was rejected in favour of the contingency table approach for the following reasons:

- to reduce the prohibitive computer costs of undertaking a large number of GLIM analyses on large data sets (the costs associated with GLIM runs where the data sets comprise 2,500 cases are high, amounting to many tens of pounds for even reasonably modest runs);
- to aggregate the data to reduce the incidence of zero values for the accident data;
- to simplify the methodology; and
- to reduce the effect of between site variation by concentrating on site type aggregations and averages.

Negative Binomial v. "scaled" Poisson as error structure

A consequence of this decision is that the error structure associated with the accident estimates will be determined by the classifications adopted to group the data. These groupings will be largely arbitrary in nature as the data are not well balanced through the classifications and, accordingly, although the data will be over-dispersed it is unlikely that they will conform with any recognisable distribution. For this reason a "scaled" Poisson model has been adopted.

This, of course, is not to say that the accident data within each classification are not negative binomially distributed.

7.5. Conclusions

The conclusions here represent a summary of the modelling techniques adopted.

In view of the problems associated with traditional regression techniques, it was decided to use the computer program GLIM to analyse the data. In GLIM terminology a log link function was used with a "scaled" Poisson error structure. A measure of the traffic exposure was defined as an "offset" variable.

The individual site by site accident data were aggregated according to a number of factors (described in the following two chapters) to produce contingency tables.

The modelling techniques are now applied to the junction and link data in Chapters 8 and 9 respectively.

CHAPTER 8 THE ANALYSIS OF JUNCTION ACCIDENTS

8.1. Introduction

The analysis of the data has been conducted within a climate of achieving one of the basic objectives of statistical modelling, which is to find as parsimonious a model as is sensibly possible: that is to say, one which requires the smallest number of parameters to explain the variation in the data.

To achieve this, the analysis was carried out in five stages.

In the first stage, the classification criteria of the data were examined to determine whether or not re-aggregation of the data could reasonably be achieved.

The second stage assessed whether or not the exposure function varied according to the accident type.

Stage three constituted fitting the parameters to the accident data to determine appropriate models for each accident type.

The fourth stage compared accident estimates based on the sum of the individual accident type models to determine whether or not the disaggregation of models by accident type improved the estimates for the all accident total (ie JACCTOT).

Finally, in the fifth stage, the year-by-year consistency of the selected model(s) was examined.

Throughout this (and, indeed, the following) chapter, the results of the analysis are shown, for clarity, in the above logical order. It should be noted, however, that in the preliminary analyses which were carried out, the consequences of the decisions made at each stage were checked against subsequent stages, where appropriate, to ensure that no significant parameter effects were lost.

8.2. Classification of the Data

The data were originally classified according to:

JUNCTYPE (with 9 factor levels)
JCONTROL (2)
JLOCATION (4)

JMAINBAND (3)
JSIDEBAND (3).

to provide 648 (ie 9x2x4x3x3) possible classifications, not all of which contained data. For each classification, the following cell values were stored in a raw data file appropriately formatted for input to GLIM:

JCELLTOT
JSVNP
JTVNP1 to JTVNP9
JPED
JPSV
JMAINFLOW
JSIDEFLOW
JPRODUCT
JROOTPROD.

Where JCELLTOT is the number of junctions per cell and the other variables are defined in Chapter 2.

Following an initial analysis, a number of junction types were combined because the parameter estimates were not significantly different from one another. In addition, because of the restricted nature of the data available for other 4^+-way junctions (ie JUNCTYPE=9) they were excluded. These combinations and exclusion provide five rather than nine factor levels which were stored in a new variable JUNCTYPEA (see Table 8.1) to give 360 (ie 5x2x4x3x3) possible classifications of which 184 contain data. Of the 176 classifications for which no data are observed, 36 refer to signal control at roundabouts (no traffic signals are located on roundabouts in Lothian Region) and 34 to rural non-roundabout signal controlled junctions (only two such junctions exist in Lothian Region) with the remainder largely accounted for by classifications matching low flows with signal control (traffic signals are not usually installed at low flow sites). In addition and for similar reasons, a number of the accident types were combined to provide 6 new accident categories, viz:

JSVNP	JSVNP (no change)
JTVNPA1	JTVNP1 to JTVNP4 & JTVNP8

JTVNPA2	JTVNP5 to JTVNP7 & JTVNP9
JMVNP	JMVNP (no change)
JPED	JPED (no change)
JPSV	JPSV (no change).

8.3. Assessment of an Exposure Function

In this, the second stage of the analysis, the aim was to determine which measure of exposure to accidents was most appropriate for each of the accident types. To do this, a Poisson model with no "offset" variable for each accident type was adopted with the following terms fitted (together with any interaction effects):

JUNCTYPEA
JCONTROL
JLOCATION
ln(JMAINFLOW)
ln(JSIDEFLOW).

to provide a model including only the significant effects.

Ignoring, for the time being, the parameter estimates for the three factor effects (ie JUNCTYPEA, JCONTROL and JLOCATION), the estimates for ln(JMAINFLOW) and ln(JSIDEFLOW) are summarised in Tables 8.2 and 8.3. It can be seen that the parameter estimates for both ln(JMAINFLOW) and ln(JSIDEFLOW) are in some cases quite different from one another suggesting that the measurement of exposure to accidents may vary by accident type. The adopted measurements are:

$JEXPOS72=Q^{0.70}.q^{0.20}$	for JACCTOT, JSVNP, JTVNPA1 and JPSV;
$JEXPOS535=Q^{0.50}.q^{0.35}$	for JTVNPA2; and
$JEXPOS815=Q^{0.80}.q^{0.15}$	for JMVNP and JPED.

In relative terms, the exponents for JTVNPA2 accidents suggest the importance of side road flow as an exposure factor whereas for JMVNP and JPED accidents, side road flow is less important but the main road flow is relatively more so.

The different exponents for the accident types are retained for the next stage of analysis and used to calculate the appropriate exposure functions, G, which are used as "offset" variables. In this context, for greater accuracy; a return was made to the database to aggregate the individual site specific values for G rather than calculate on the basis of the existing cell mean values for JMAINFLOW and JSIDEFLOW.

8.4. Determination of Appropriate Models

8.4.1. Determination of a Traffic Intensity Function

Seven potential traffic intensity functions were examined viz:

1 JMAINFLOW
2 JSIDEFLOW
3 JTHROUGHPUT (ie JMAINFLOW+JSIDEFLOW)
4 JFLOWRATIO (ie JMAINFLOW/JSIDEFLOW)
5 JPRODUCT
6 JROOTPROD
7 JPRODPERVEH (ie JPRODUCT/JTHROUGHPUT).

The first three were selected as being representative of the use of the junction in terms of absolute vehicle numbers. JFLOWRATIO was examined as a measure of the relativity of main road to side road traffic. JPRODUCT and JROOTPROD were selected as perhaps the more obvious candidates having been the subject of considerable interest to researchers in the past. JPRODPERVEH was thought worthy of consideration as it provided a measure of the relative conflict at a junction measured in terms of the number of vehicles using the junction.

Using a "scaled" Poisson model of the form (7.5) for each of the six accident types; the terms JUNCTYPEA, JCONTROL and JLOCATION (including interaction effects) were fitted together with the exposure measures, G, as calculated at the previous stage and each of the above seven candidate measures for traffic intensity, I, to provide models for accidents which included only the significant terms.

The outcome of this stage of the analysis indicated that JPRODUCT and JROOTPROD provided the most robust measures for traffic intensity. In most cases JPRODUCT provided marginal but significant improvements on JROOTPROD and for this reason it was preferred to JROOTPROD. In the interests of brevity the results of this analysis are not outlined here because it comprised 42 (ie 6 accident types x 7 traffic intensity candidates) individual GLIM runs. In the following discussion, traffic intensity is referred to as JTINT and JTINTSQ for $\ln I$ and $(\ln I)^2$ respectively in (7.5). Accordingly, in the context of the present study, JTINT=ln(JPRODUCT) and JTINTSQ=$(\ln JPRODUCT)^2$.

8.4.2. Individual Models for Different Accident Types

In this stage, a "scaled" Poisson model of type (7.5) for each accident type was fitted by the three terms JUNCTYPEA, JCONTROL and JLOCATION together with the "offset" exposure measure, G, as calculated at the second stage and with JPRODUCT as the traffic intensity function. Again, the interaction effects were examined, but none was found to be significant. These models are tabulated in Tables 8.4 to 8.9. From these tables it is noted that the significant effects on accident rate vary quite considerably between the accident types. It is stressed that with the declaration of exposure as an "offset" variable, it is the effect on the accident rate (ie accidents/exposure measure) that is modelled by fitting the variates.

The single vehicle non-pedestrian (JSVNP) accident rate was found not to vary according to any of the fitted variates (see Table 8.4). Accordingly, there is no support, with respect to junctions, for the findings of the earlier studies referred to in Section 7.2 which have suggested an inverse relationship between accident rate and traffic flow for single vehicle accidents.

Reference to Table 8.5 indicates that, for JTVNPA1 accidents, JUNCTYPEA and JLOCATION were found to be significant effects with the 4-way junction factors producing, not unexpectedly, higher accidents rates. JLOCATION, however, produced a significant effect only for suburban locations which suggested a reduced risk of JTVNPA1 accidents at such locations (this result, however, is only just significant at the 5% level).

For JTVNPA2 accidents, JUNCTYPEA and the traffic intensity function, JTINTSQ produced significant effects (Table 8.6 refers). All five junction types produce statistically independent factor effects with, as expected, the 4-way junction types being associated with the higher accident rates. The significant traffic intensity function suggests a reducing effect on accident rate with increasing accident traffic intensity. Figure 8.1 shows, in graphical form, selected relationships between accident rate and traffic intensity for this accident type.

Neither of the two two-vehicle acident types, therefore, indicate any increase in accident rate associated with increased traffic intensity and, indeed, the opposite is suggested for JTVNPA2 accidents. These results would appear to contradict those observed in the previous studies outlined in Section 7.2. These earlier studies, however, did not generally make the distinction between junction and link accidents as is the case here.

The multi-vehicle, JMVNP, accident rates were found to vary according to JUNCTYPEA with, again, all five junction types producing statistically independent effects with the higher accident rates being associated with 4-way junctions (see Table 8.7).

In Table 8.8 it is shown that the JPED accident rates vary according to all the variates with each of the junction and location types producing statistically independent effects. JCONTROL was also found to be a significant effect indicating an increased risk to pedestrians at signal controlled junctions. JLOCATION is, not unexpectedly, a highly significant determinator of the pedestrian accident rate with the more highly urbanised locations generating higher accident rates. This is not to say that it is the type of location itself or the presence of traffic signals that generate particular accident rates but rather that they may act as a "proxy" for pedestrian density. Both traffic intensity functions produce significant effects which suggest that, as is the case for JTVNPA2 accidents, the accident rate is inversely related to traffic intensity. Figure 8.2 shows selected relationships between accident rate and traffic intensity for this accident type.

Finally, for JPSV accidents, it is JUNCTYPEA and JLOCATION which produce significant effects on the accident rate (see Table 8.9). Again, in keeping with what would be expected, the 4-way junction types produce higher accident rates and likewise for the more urbanised JLOCATION factors (where each location type produces a statistically independent effect).

One interesting general observation on these results is that JLOCATION has a robust significant effect on only the JPED and JPSV accident rates (with the exception of the somewhat unexpected effect for the one factor for the JTVNPA1 accident rate) which suggests that vehicular accidents (ie those not involving solely pedestrian or public transport passengers) occur at rates which are independent of the location type. As traffic speed would, at least, be expected to vary between rural and urban locations, if not between the three urban locations, this observation is perhaps unexpected suggesting that the observed differences between accident rates may be accounted for by the superimposition of pedestrian and public service vehicle accidents.

It is possible to sum, or re-aggregate, the individual accident estimates for each of the six individual accident types to provide an estimate for the all accident total, JACCTOT. This estimate is called the 're-aggregated all accident total' and is ascribed the variable name JERACCTOT.

8.5. Comparison of an All Accident Model with a Re-aggregated All Accident Model for Junctions

Whilst the above observations on the models for the individual accident types are of considerable descriptive interest, it was considered logical, in the interests of parsimony, to determine whether or not the JERACCTOT model constituted a significantly more robust model than that for the estimates (ie JEACCTOT) based on a more simple model for all accidents.

The all accident model (see Table 8.10) was fitted in an identical manner as those for the individual accident types. The rates for all accidents were found to vary according to junction type with 4-way junctions significantly greater rates. The provision of signal control is associated with significantly higher accident rates. The location type also proved to be a highly significant effect with each location producing significantly independent rates. The effect of traffic intensity suggests an inverse relationship between accident rates and traffic. Earlier research has indicated (see Section 7.2) that a U-shaped relationship might have been expected; however, the results here suggest that this is not the case for junction accidents and, indeed, there is some robust evidence to suggest an inverse relationship which denies the presence of an upward trend in accident rate at higher levels of traffic flow.

The two estimates (ie JERACCTOT and JEACCTOT) are plotted against the observed accident totals (ie JACCTOT) in Figures 8.3 and 8.4 respectively in which a marked degree of agreement is visually evident. More convincingly, the variance-ratio (or F-test) statistic (see Section 10.2) indicates that there is no significant difference ($p<0.05$) between the variance of the residuals associated with the two estimates for JACCTOT:

variance ratio = s_1^2/s_2^2 = 120.51/105.31 = 1.144 (df=183)

where:

s_1^2 is the larger of the two variances of the residual values being for those estimates based on the simple all accident model (ie JACCTOT-JEACCTOT); and

s_2^2 is the smaller of the two variances of the residual values being for those estimates based on the re-aggregated all accident model (ie JACCTOT-JERACCTOT).

Figure 8.5 shows some selected relationships between accident rate and traffic intensity for the JACCTOT all accident model.

8.6. The Temporal Consistency of the Junction Model

The model for JACCTOT outlined in Table 8.10 is based on the aggregated accident totals for the four years covered by the study data (ie 1979-1982). This model, however, would be of limited use if it were found to vary significantly from year to year, as little or no confidence could be placed in its ability to predict future accidents.

To examine the temporal consistency of the model, the data were split to form accident totals (JACCTOT1 to JACCTOT4) for each of the four constituent years and, for each year, the JACCTOT model was fitted and the parameter estimates compared.

The comparison is shown in Table 8.11 in which it can be seen that just two parameter estimates are found to be significantly different from their counterparts. They are those for the 1980 T-junction accidents (ie where JUNCTYPEA=2) and the 1982 city centre accidents (ie where JLOCATION=4).

These two inconsistencies, however, do not constitute any sound reason to doubt the general conclusion that the model observes a high degree of temporal consistency. This result, though simply stated, is absolutely crucial to any justification for using such models to predict accidents.

8.7. Conclusions

In this chapter it has been shown that the form of accident models do vary quite markedly by accident type. In addition, however, it has been shown that, whilst these individual accident type models are of considerable descriptive interest, they do not, in aggregate, produce a significantly better model for all accidents than a simple all accident model. Accordingly, the all accident model was favoured in the interests of parsimony.

Further, the all accident model was found to be stable from year to year giving added confidence to its capability to predict future accident rates and totals.

The applicability of the modelled accident rates to other data sets might provide a profitable avenue for future research.

FIGURE 8.1: RELATIONSHIP BETWEEN JTVNPA2 ACCIDENT RATES AND TRAFFIC FOR SELECTED JUNCTION TYPES

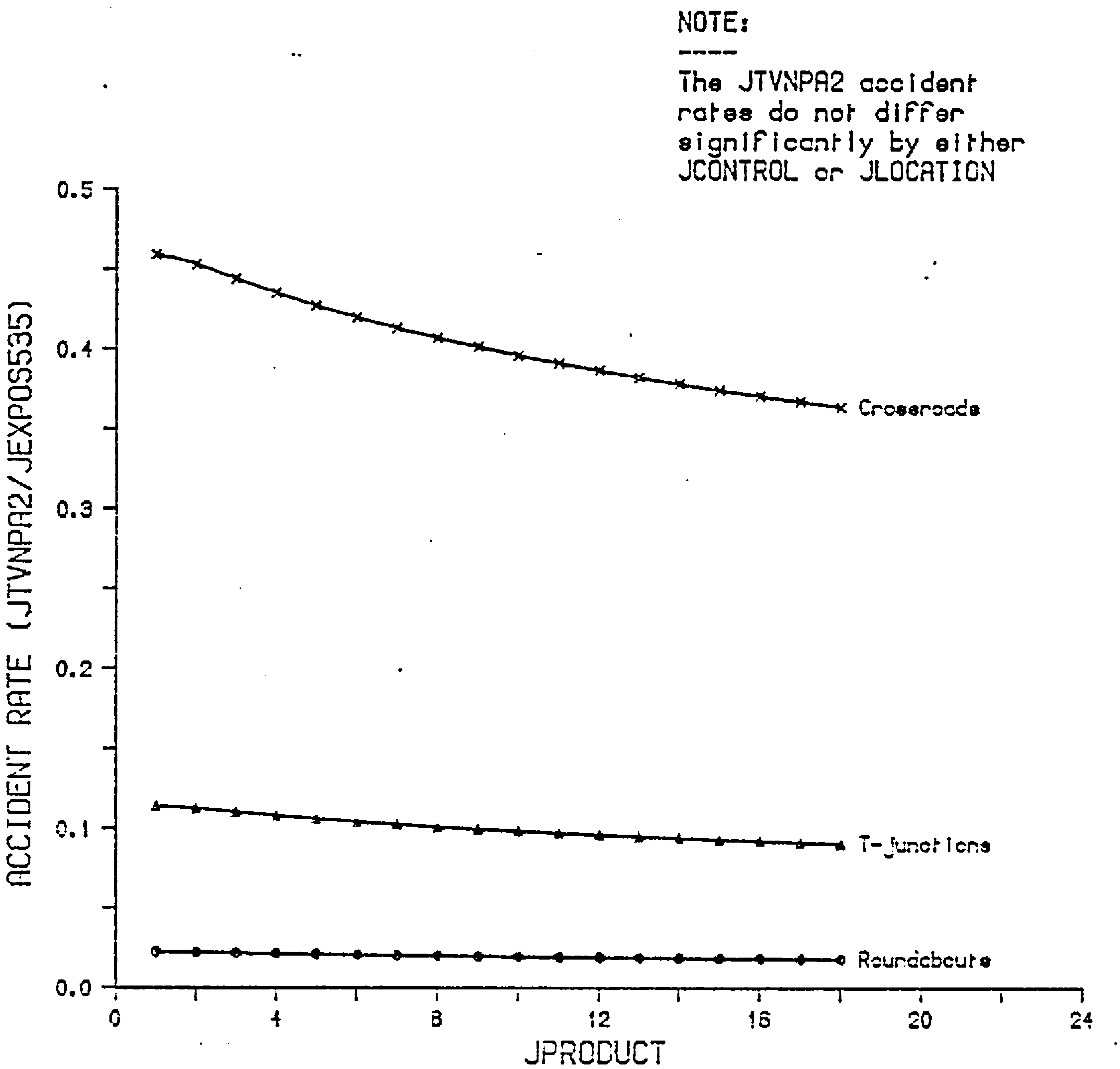


FIGURE 8.2: RELATIONSHIP BETWEEN JPED ACCIDENT RATES AND TRAFFIC FOR SELECTED PRIORITY CONTROLLED JUNCTION TYPES

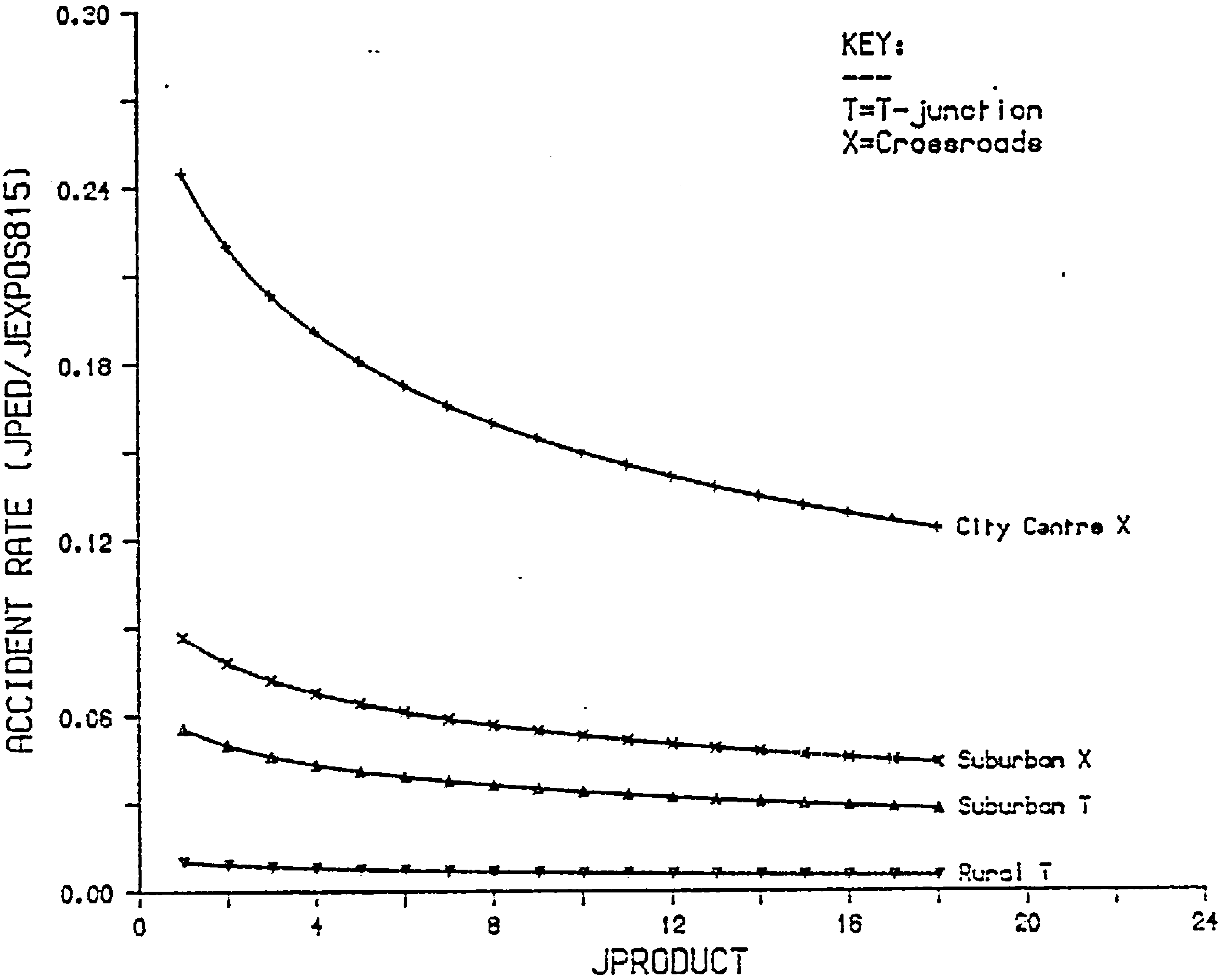


FIGURE 8.3: SCATTERPLOT OF OBSERVED v EXPECTED RE-AGGREGATED ACCIDENT TOTALS FOR JUNCTIONS

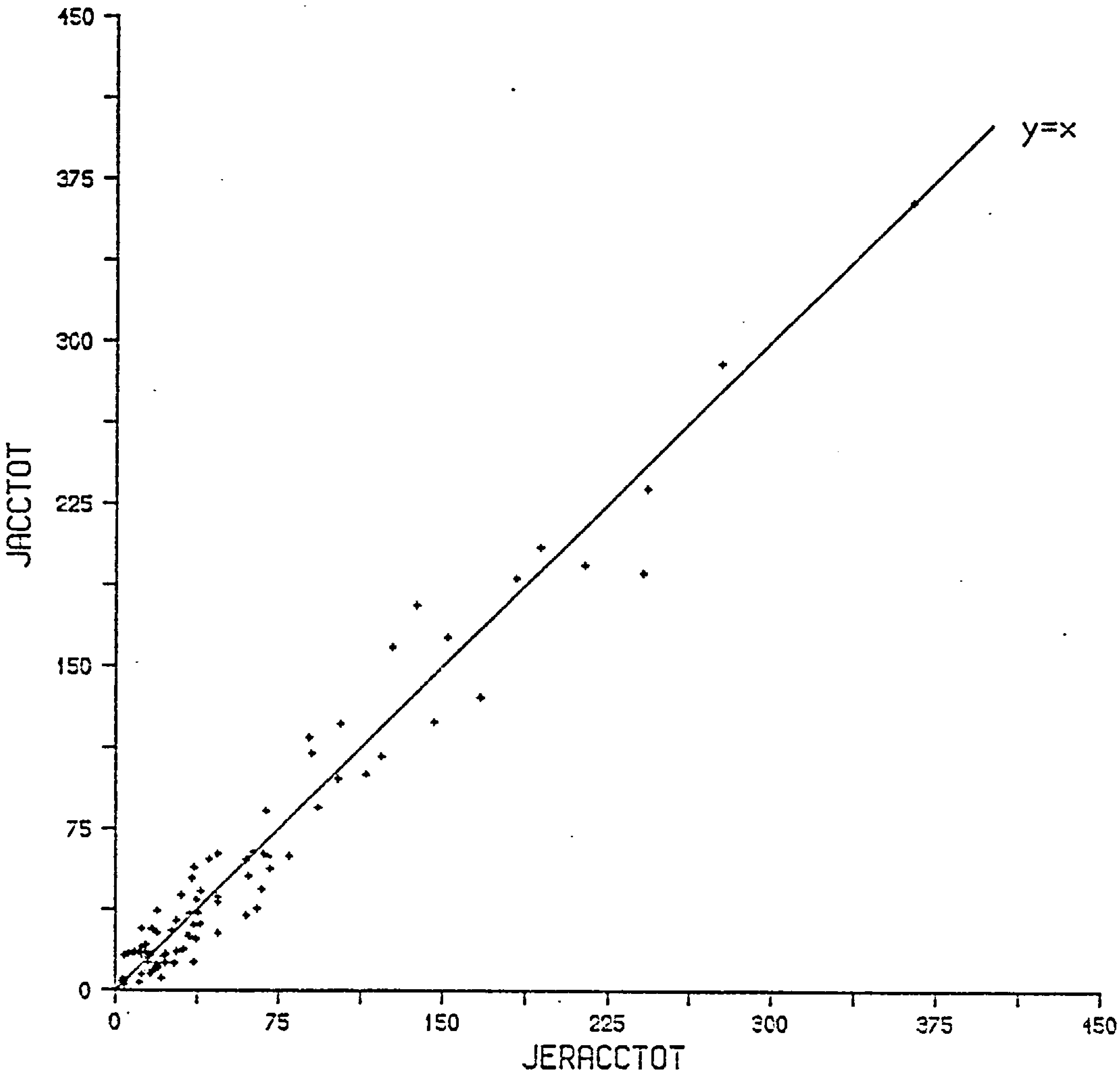


FIGURE 8.4: SCATTERPLOT OF OBSERVED v EXPECTED ALL ACCIDENT TOTALS FOR JUNCTIONS

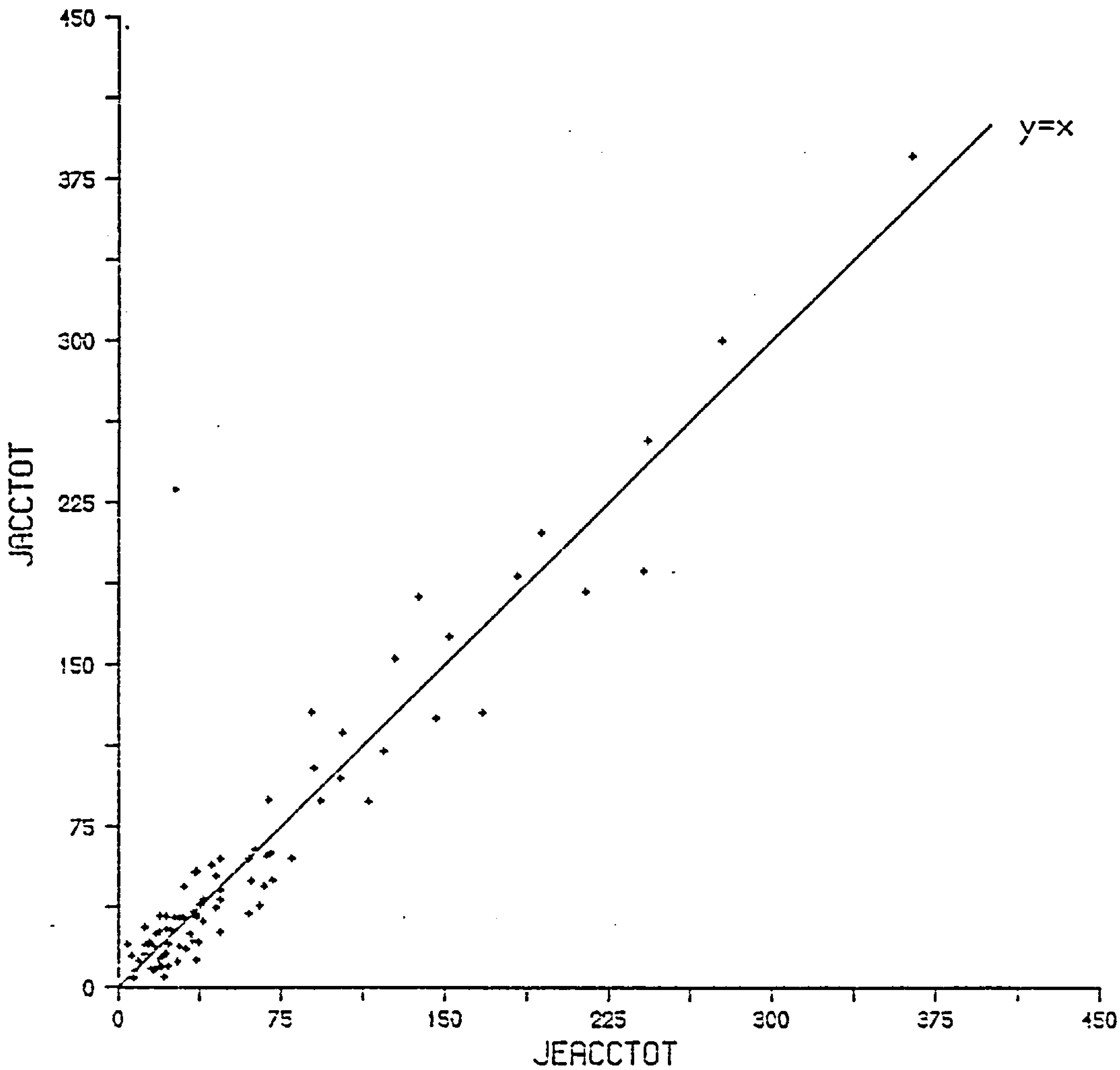


FIGURE 8.5: RELATIONSHIP BETWEEN ALL ACCIDENT RATES FOR SELECTED JUNCTION TYPES

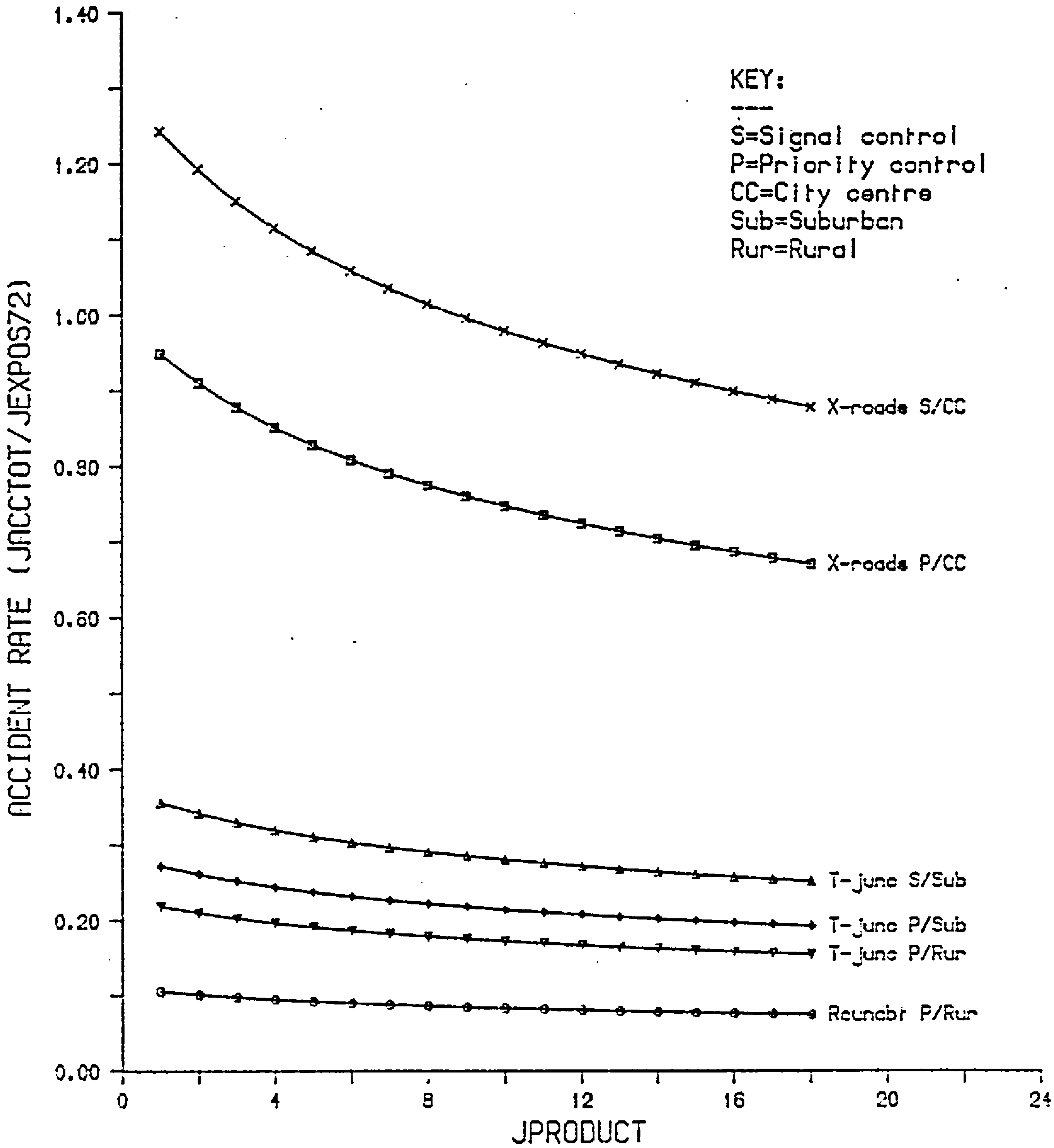


TABLE 8.1: COMBINATION OF JUNCTION TYPES INTO NEW VARIABLE JUNCTYPEA

JUNCTYPEA	JUNCTYPE	DESCRIPTION
1	1,2,3	All roundabouts
2	4,5	T- and Y- junctions
3	6	Other 3-way junctions
4	7	Crossroads
5	8	Other 4-way junctions

TABLE 8.2: SCALED AND UNSCALED CONFIDENCE LIMITS FOR JMAINFLOW

ACCIDENT TYPE	RESIDUAL DEVIANCE	df	SCALE PARAMETER λ	ESTIMATE	STANDARD ERRORS (unscaled /scaled)	95% CONFIDENCE LIMITS unscaled (LOWER/ UPPER)	95% CONFIDENCE LIMITS scaled (LOWER/ UPPER)
JSVNP	278.24	181	1.2399	0.7076	0.03178	0.6449	0.6299
					0.03940	0.7703	0.7853
JTVNPA1	274.36	177	1.2450	0.6979	0.03202	0.6347	0.6192
					0.03987	0.7611	0.7766
JTVNPA2	416.75	177	1.5344	0.5286	0.02790	0.4736	0.4441
					0.04281	0.5836	0.6131
JMVNP	177.88	177	1.0025	0.8082	0.05259	0.7044	0.7042
					0.05272	0.9120	0.9122
JPED	445.39	174	1.5999	0.7744	0.02844	0.7183	0.6846
					0.04550	0.8305	0.8642
JPSV	283.11	174	1.2756	0.7369	0.05579	0.6268	0.5965
					0.07116	0.8470	0.8773
JACCTOT	724.89	174	2.0411	0.6777	0.01528	0.6476	0.6162
					0.03119	0.7078	0.7392

TABLE 8.3: SCALED AND UNSCALED CONFIDENCE LIMITS FOR JSIDEFLOW

ACCIDENT TYPE	RESIDUAL DEVIANCE	df	SCALE PARAMETER λ	ESTIMATE	STANDARD ERRORS (unscaled /scaled)	95% CONFIDENCE LIMITS unscaled (LOWER/ UPPER)	95% CONFIDENCE LIMITS scaled (LOWER/ UPPER)
JSVNP	278.24	181	1.2399	0.1992	0.02403	0.1518	0.1404
					0.02979	0.2466	0.2580
JTVNPA1	274.36	177	1.2450	0.2051	0.02060	0.1645	0.1545
					0.02565	0.2457	0.2557
JTVNPA2	416.75	177	1.5344	0.3687	0.01838	0.3324	0.3131
					0.02820	0.4050	0.4243
JMVNP	177.88	177	1.0025	0.1744	0.03238	0.1105	0.1104
					0.03246	0.2383	0.2384
JPED	445.39	174	1.5999	0.1407	0.01623	0.1087	0.0895
					0.02597	0.1727	0.1919
JPSV	283.11	174	1.2756	0.2086	0.03271	0.1441	0.1263
					0.04172	0.2731	0.2909
JACCTOT	724.89	174	2.0411	0.2223	0.00910	0.2043	0.1857
					0.01857	0.2403	0.2589

TABLE 8.4: PARAMETER ESTIMATES[1] FOR JSVNP ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2	3	4
GM	-	-3.371	0.047	71.72			

INITIAL SCALED DEVIANCE = 271.71 (df=183)
RESIDUAL SCALED DEVIANCE = 271.71 (df=183)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in bold type.

TABLE 8.5: PARAMETER ESTIMATES[1] FOR JTVNPA1 ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2	3	4
GM	-	**-3.419**	0.183	**18.68**			
JUNCTYPEA	1	0.000	-				
	2	**0.731**	0.170	**4.307**			
	3	0.271	0.194	1.399	**3.853**		
	4	**1.092**	0.189	**5.766**	**3.287**	**5.711**	
	5	**0.861**	0.203	**4.237**	0.984	**3.723**	1.518
JLOCATION	1	0.000	-				
	2	**-0.256**	0.119	**2.164**			
	3	-0.133	0.143	0.932	1.139		
	4	-0.026	0.140	0.185	**2.120**	0.809	

INITIAL SCALED DEVIANCE = 365.32 (df=183)
RESIDUAL SCALED DEVIANCE = 266.13 (df=176)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in bold type.

TABLE 8.6: PARAMETER ESTIMATES[1] FOR JTVNPA2 ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2	3	4
GM	-	**-3.793**	0.273	**13.89**			
JUNCTYPEA	1	0.000	-				
	2	**1.622**	0.275	**5.907**			
	3	**0.996**	0.297	**3.356**	**4.345**		
	4	**3.014**	0.274	**10.98**	**15.29**	**13.95**	
	5	**2.188**	0.287	**7.629**	**4.624**	**7.183**	**6.715**
JTINTSQ	-	**-0.028**	0.011	**2.545**			

INITIAL SCALED DEVIANCE = 1353.4 (df=183)
RESIDUAL SCALED DEVIANCE = 405.57 (df=178)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in bold type.

TABLE 8.7: PARAMETER ESTIMATES[1] FOR JMVNP ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2	3	4
GM	-	**-5.534**	0.316	**17.51**			
JUNCTYPEA	1	0.000	-				
	2	**1.448**	0.324	**4.473**			
	3	**0.805**	0.355	**2.267**	**3.683**		
	4	**2.275**	0.333	**6.826**	**6.616**	**7.662**	
	5	**1.803**	0.349	**5.160**	**2.178**	**4.567**	**2.602**

INITIAL SCALED DEVIANCE = 290.82 (df=183)
RESIDUAL SCALED DEVIANCE = 180.62 (df=179)

[1] All parameter estimates are natural logarithms.
[2] Significant (p<0.05) values are shown in **bold** type.

TABLE 8.8: PARAMETER ESTIMATES[1] FOR JPED ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2	3	4
GM	-	**-5.914**	0.438	**13.50**			
JUNCTYPEA	1	0.000	-				
	2	**1.307**	0.329	**3.968**			
	3	**0.792**	0.346	**2.292**	**4.188**		
	4	**1.754**	0.340	**5.162**	**4.154**	**7.210**	
	5	**1.718**	0.341	**5.032**	**3.608**	**6.738**	**0.293**
JCONTROL	1	0.000	-				
	2	**0.572**	0.130	**4.390**			
JLOCATION	1	0.000	-				
	2	**1.716**	0.307	**5.597**			
	3	**2.539**	0.312	**8.151**	**9.220**		
	4	**2.753**	0.313	**8.790**	**10.56**	**2.153**	
JTINT	-	**-0.128**	0.028	**4.571**			
JTINTSQ	-	**-0.038**	0.012	**3.167**			

INITIAL SCALED DEVIANCE = 1484.5 (df=183)
RESIDUAL SCALED DEVIANCE = 376.17 (df=173)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in bold type.

TABLE 8.9: PARAMETER ESTIMATES[1] FOR JPSV ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2	3	4
GM	-	**-6.625**	0.613	**10.81**			
JUNCTYPEA	1	0.000	-				
	2	0.336	0.296	1.133			
	3	0.369	0.318	1.161	0.205		
	4	**0.897**	0.315	**2.845**	**3.538**	**2.804**	
	5	**1.076**	0.316	**3.408**	**4.759**	**3.912**	0.981
JLOCATION	1	0.000	-				
	2	**1.998**	0.565	**3.537**			
	3	**3.065**	0.568	**5.401**	**7.313**		
	4	**3.347**	0.564	**5.932**	**9.726**	1.982	

INITIAL SCALED DEVIANCE = 667.33 (df=183)
REDIDUAL SCALED DEVIANCE = 273.66 (df=176)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in **bold** type.

TABLE 8.10: PARAMETER ESTIMATES[1] FOR ALL ACCIDENTS (JACCTOT)[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2	3	4
GM	-	**-2.253**	0.167	**13.52**			
JUNCTYPEA	1	0.000	-				
	2	**0.730**	0.149	**4.906**			
	3	0.301	0.162	1.858	**4.945**		
	4	1.461	0.158	**9.241**	**9.869**	**12.02**	
	5	1.111	0.162	**6.858**	**4.468**	**8.007**	**3.910**
JCONTROL	1	0.000	-				
	2	**0.270**	0.094	**2.860**			
JLOCATION	1	0.000	-				
	2	**0.218**	0.102	**2.128**			
	3	**0.529**	0.115	**4.595**	**4.448**		
	4	**0.739**	0.116	**6.356**	**7.157**	**2.584**	
JTINT	-	**-0.040**	0.020	**2.042**			
JTINTSQ	-	**-0.028**	0.008	**3.402**			

INITIAL SCALED DEVIANCE = 2381.5 (df=183)
RESIDUAL SCALED DEVIANCE = 658.21 (df=173)

[1] All parameter estimates are natural logarithms.
[2] Significant (p<0.05) values are shown in bold type

TABLE 8.11: COMPARISON OF PARAMETER ESTIMATES[1] FOR ACCIDENT TOTALS FOR INDIVIDUAL YEARS

VARIATE	FACTOR LEVEL	PARAMETER ESTIMATES AND (SCALED STANDARD ERRORS) JACCTOT1	JACCTOT2	JACCTOT3	JACCTOT4	SIGNIFICANT DIFFERENCES (p<0.05)
GM	-	-2.229	-2.522	-2.235	-2.062	
		(0.1694)	(0.1833)	(0.1703)	(0.1629)	None
JUNCTYPEA	2	0.5541	1.012	0.6416	0.7276	
		(0.1491)	(0.1649)	(0.1510)	(0.1475)	79/80
	3	0.1482	0.6096	0.2357	0.2338	
		(0.1639)	(0.1768)	(0.1648)	(0.1622)	None
	4	1.446	1.788	1.358	1.260	
		(0.1579)	(0.1733)	(0.1604)	(0.1591)	None
	5	1.081	1.288	1.009	1.079	
		(0.1615)	(0.1786)	(0.1651)	(0.1613)	None
JCONTROL	2	0.1464	0.1854	0.3755	0.3746	
		(0.09760)	(0.09663)	(0.09630)	(0.09671)	None
JLOCATION	2	0.2859	0.2722	0.2889	0.03605	
		(0.1096)	(0.1039)	(0.1075)	(0.09967)	None
	3	0.6700	0.4803	0.5038	0.4677	
		(0.1219)	(0.1180)	(0.1217)	(0.1117)	None
	4	0.8756	0.7975	0.8328	0.4576	79/82 80/82
		(0.1233)	(0.1175)	(0.1214)	(0.1157)	81/82
JTINT	-	-0.04022	-0.03562	-0.05539	-0.02871	
		(0.02040)	(0.01995)	(0.02017)	(0.01995)	None
JTINTSQ	-	-0.02998	-0.02194	-0.03219	-0.02743	
		(0.008591)	(0.008275)	(0.008397)	(0.008287)	None
INIT SCALED DEV		792.43	758.47	710.99	688.26	
	df	183	183	183	183	
RESID SCALED DEV		305.72	300.14	276.61	304.04	
	df	173	173	173	173	

[1] All parameter estimates are natural logarithms.

CHAPTER 9 THE ANALYSIS OF LINK ACCIDENTS

9.1. Introduction

As with junctions, the aim of the analysis was to determine a parsimonious model which contained the smallest number of parameters to explain the variation in the data.

The analysis of link data was not as complex as that required for junctions because:

1. Only one sensible exposure function exists, that of vehicle distance travelled (ie LMVKM); and
2. Although it was theoretically possible to examine a number of traffic intensity functions for links (eg traffic flow per lane), the database for the study did not contain sufficient relevant information for this work to be undertaken.

For these reasons, the traffic intensity function adopted for all link accident models was that of traffic flow (ie LINKFLOW). In the following discussion, traffic intensity is referred to as LTINT and LTINTSQ for $\ln I$ and $(\ln I)^2$ respectively in (7.5). Accordingly, in the context of the present study, LTINT=ln(LINKFLOW) and LTINTSQ=$(\ln \text{LINKFLOW})^2$.

The analysis of the link data was carried out in four stages.

In the first stage, the classification criteria of the data were examined to determine whether or not re-aggregation of the data of the data could be reasonably achieved.

The second stage constituted fitting the parameters to the accident data to determine appropriate models for each accident type.

The third stage compared accident estimates based on the sum of the individual accident type models to determine whether or not the disaggregation of models by accident type improved the estimates for all accident data (ie LACCTOT).

Finally, in the fourth stage, the year-by-year consistency of the selected model(s) was examined.

9.2. Classification of the Data

The data were originally classified according to:

LDEVTYPE (22 factor levels)
LCWYTYPE (3)
LENGTHBAND (3)
LFLOWBAND (3)

to provide 594 (ie 22x3x3x3) possible classifications, not all of which contained data. For each classification, the following cell values were stored in a raw data file appropriately formatted for input to GLIM:

LSVNP
LTVNP1 to LTVNP4
LMVNP
LPED
LPSV
LINKFLOW
LMVKM.

Following an initial analysis, it was found possible to combine many of the link types because the parameter estimates were not significantly different from one another. This re-aggregation of the data into a new variable LDEVTYPEA which provided just three factor levels rather than the original twenty-two is shown in Table 9.1. In addition, because of the restricted nature of the data available for one-way streets (ie LCWYTYPE = 3) they were excluded from the analysis.

These re-aggregations produced 54 (ie 3x2x3x3) possible classifications of which 52 contained data.

In addition, and for similar reasons, LTVNP3 and LTVNP4 accident types were re-aggregated to provide seven accident types:

LSVNP = LSVNP	(no change)
LTVNPA1 = LTVNP1	(no change)
LTVNPA2 = LTVNP2	(no change)

LTVNPA3 = LTVNP3+LTVNP4 (all accidents involving turning)

LMNVP = LMNVP (no change)

LPED = LPED (no change)

LPSV = LPSV (no change).

9.3. Examination of Individual Models for the Different Accident Types

A "scaled" Poisson model for each accident type was fitted by the two terms LDEVTYPEA and LCWYTYPE together with the traffic intensity functions and with LMVKM as an "offset" variable. The interaction effects were examined and both LDEVTYPEA.LCWYTYPE and LDEVTYPE.LTINT were found to be significant in a number of the models. The final models (ie those including only the significant effects) are tabulated in Tables 9.2 to 9.9 and, as with the junction data, it is noted that the significant effects on accident rate vary quite considerably between the various accident types.

The single vehicle (LSVNP) accident rates are found to vary according to LDEVTYPEA, LCWYTYPE and LTINT (Table 9.2 refers). The significant factors suggest greater accident rates in urban areas with lower rates observed on dual than on single carriageway links. The effect of traffic intensity (LTINT) indicates an inverse relationship between accident rate and traffic intensity: a result which confirms the finding of earlier studies discussed in Section 7.2.

The LTVNPA1 analysis (for accidents involving rear-end shunts or two vehicles travelling in the same direction) indicates quite a number of significant effects which include two interaction terms. These results, which are shown in Table 9.3, suggest that higher accident rates are associated with greater urbanisation particularly on those dual carriageway links adjacent to shopping development (this latter observation is largely influenced by the data for Edinburgh's major street, Princes Street, which is a dual carriageway road with shopping development along one side). The effects of traffic intensity are indicative of a U-shaped relationship between accident rate and traffic intensity there being a negative estimate for LTINT and a positive estimate for LTINTSQ. Finally, there is some evidence to suggest that the effect of traffic intensity becomes more marked for urban links. The U-shaped relationship for this accident type is perhaps unexpected, as previous evidence (outlined in Section 7.2) has indicated that accident rate and traffic intensity are directly proportional to one another suggesting no initial inverse relationship between the two. Figure 9.1 shows selected relationships between accident rate and traffic intensity for this accident type.

The higher LTVNPA1 accident rate associated with dual carriageway links is perhaps somewhat surprising and it may give some weight to the argument propounded most vigorously by Adams [1985] regarding the risk compensation (or risk homoeostasis) effect. It may be argued here that drivers on dual carriageways who are not required to consider the risks associated with two way traffic flow, consume this safety benefit by maintaining their perceived level of risk and, as a consequence, pay relatively less regard to their fellow travellers on the same carriageway. Later, however, it is shown that the effect of carriageway type on overall accident rates is mixed (see Section 9.4) and that it would be difficult to sustain this risk compensation argument when all accident types are considered.

The LTVNPA2 head-on accidents (including vehicles colliding while travelling in opposing directions) are shown to occur at a lower rate on rural links (see Table 9.4). As expected for this type of accident, the accident rates associated with dual carriageway links are significantly less than those for single carriageway links. An inverse relationship between the accident rate and traffic flow is observed: this is contrary to the earlier evidence suggesting a direct relationship. As with the LTVNPA1 accidents, however, there is some evidence to suggest that the effect of traffic intensity becomes more marked on urban links.

The analysis for the remaining category of two-vehicle accidents, LTVNPA3 (involving turning vehicles and the use of private drives), indicates, not unexepectedly, that the higher accident rates are associated with urban development (see Table 9.5). The dual carriageway links show accident rates significantly less than those for single carriageway links, whilst traffic intensity did not prove to be a significant effect.

The rates at which multi-vehicle accidents (LMVNP) occur are found to be greater on urban links and there are indications that a U-shaped relationship exists between rate and traffic intensity (Table 9.6 refers).

Table 9.7 shows the pedestrian accident (LPED) rates which, again, are found to increase on the more urban links and, in addition, are significantly greater on dual carriageways particularly those adjacent to shopping development (the comments above for LTVNPA1 accidents with reference to Princes Street may also be relevant here). The rates are also shown to be inversely related to traffic intensity. Figure 9.2 shows selected relationships between accident rate and traffic intensity for this accident type.

Finally, in the model for LPSV accidents (Table 9.8 refers), the rates, as with all the previous accident types, are observed to be greater on urban links and, as with LTVNPA1 (rear-end shunts) accidents, they are greater on dual than they are on single carriageway links (this result may also be largely influenced by the Princes Street data).

9.4. Comparison of an All Accident Model with a Re-aggregated All Accident Model for Links

As with junction accidents, estimated accident totals (ie LEACCTOT) based on an all accident model were examined and compared with the re-aggregated estimates (ie LERACCTOT) for the individual accident types with no significant difference being observed between the two estimates. Indeed, for the link data, the LACCTOT model provided a marginal (but wholly insignificant) improvement on the re-aggregated model. Accordingly, the all accident model is preferred (see Table 9.9). Figures 9.3 and 9.4 show, in graphical form, the agreement between the observed accident totals and the two estimates. The model shown in Table 9.9 indicates that the different development types have highly significant and statistically independent effects on accident rate.

The effect of dual carriageways on the overall accident rate is mixed where:

1. for links adjacent to shopping/commercial development, the presence of dual carriageways leads to significantly higher accident rates than those observed on single carriageways (this effect, as indicated above, is very probably peculiar to Edinburgh);
2. for other urban links, no significant difference is observed between dual and single carriageway accident rates; and
3. for rural links, dual carriageways experience significantly lower accident rates than single carriageway roads.

This mixed response to carriageway type is determined to a large extent by the presence of pedestrians and may well be very much a local effect within Lothian Region.

As with the junction accident models, the effect of traffic intensity is suggestive of an inverse relationship with accident rate. This final observation denies the presence of a U-shaped relationship which might have been expected on the basis of the earlier research reported in Section 7.2. It is perhaps worthy of note that the effect of traffic intensity on the accident rate is more marked for links than it is for junctions.

9.5. The Temporal Consistency of the Link Model

In an identical procedure to that adopted for junction accidents in Section 8.6, the data were split to form accident totals (LACCTOT1 to LACCTOT4) for each of the constituent years and the LACCTOT model was fitted and a comparison was made of the parameter estimates.

The comparison is tabulated in Table 9.10 in which it can be seen that no significant differences exist between any of the parameter estimates and, accordingly, the model demonstrates a high degree of temporal consistency. The importance of this finding has been stressed in the previous chapter.

9.6. Conclusions

As with the junction models, it has been shown that the form of accident models do vary quite markedly by accident type but that although they are of considerable descriptive interest, they do not, in aggregate, produce a significantly better model for all accidents than a simple all accident model. Again, the all accident model was found to be stable from year to year.

The applicability of the modelled accident rates to other data sets might provide a profitable avenue for further research.

FIGURE 9.1: RELATIONSHIP BETWEEN LTVNPA1 ACCIDENT RATES AND TRAFFIC FOR SELECTED LINK TYPES

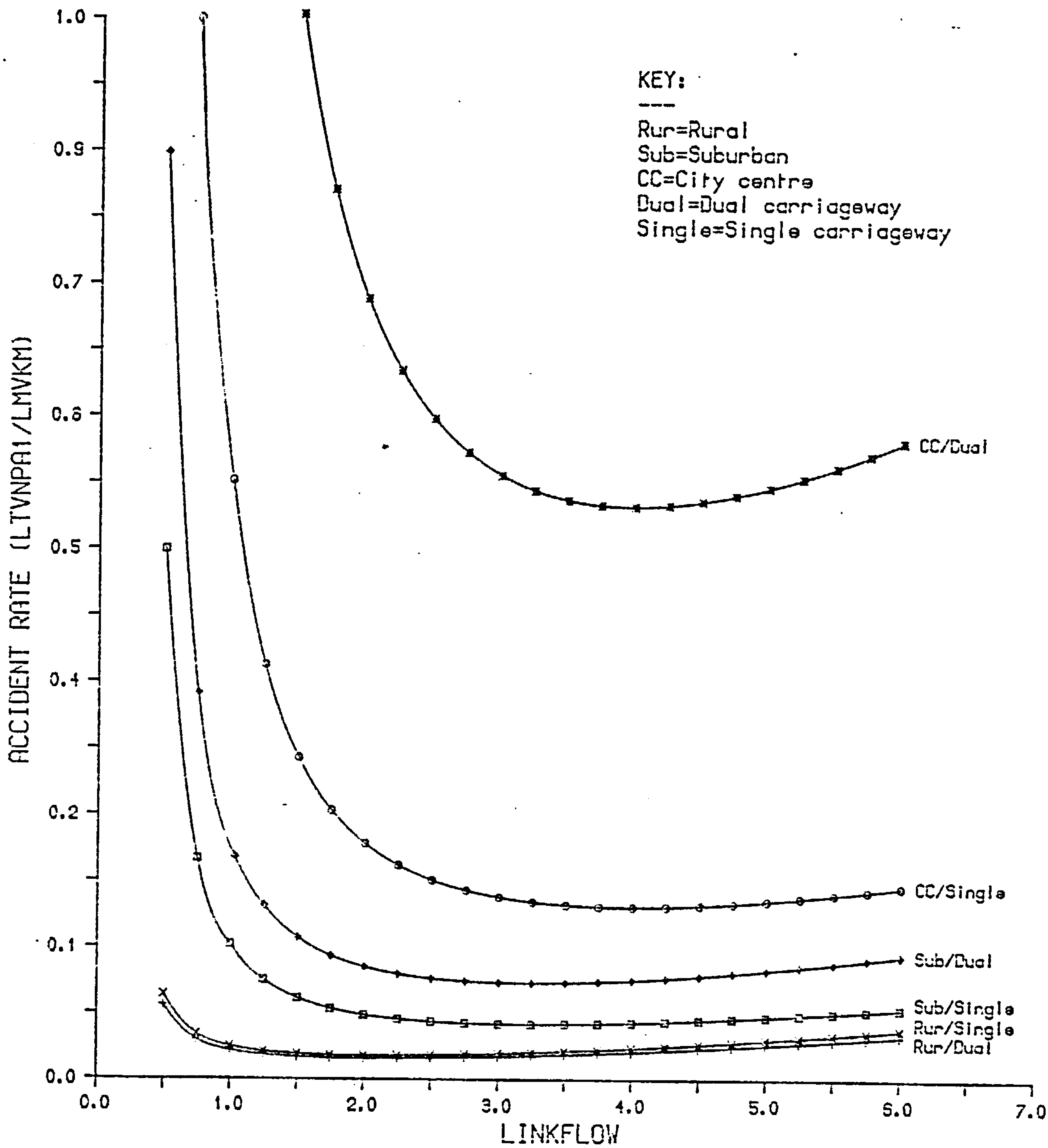


FIGURE 9.2: RELATIONSHIP BETWEEN LPED ACCIDENT RATES AND TRAFFIC FOR SELECTED LINK TYPES

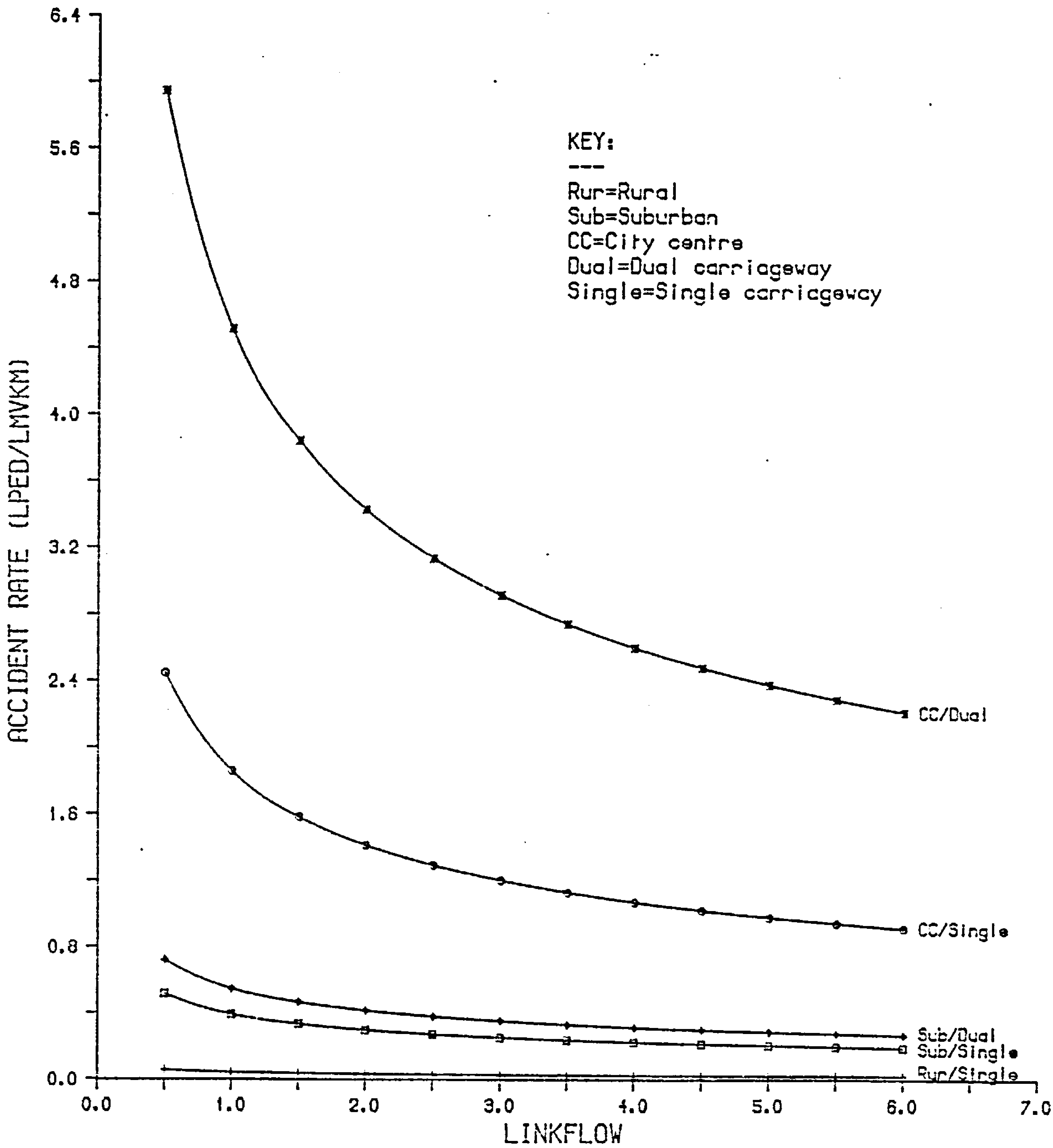


FIGURE 9.3: SCATTERPLOT OF OBSERVED v EXPECTED RE-AGGREGATED ACCIDENT TOTALS FOR LINKS

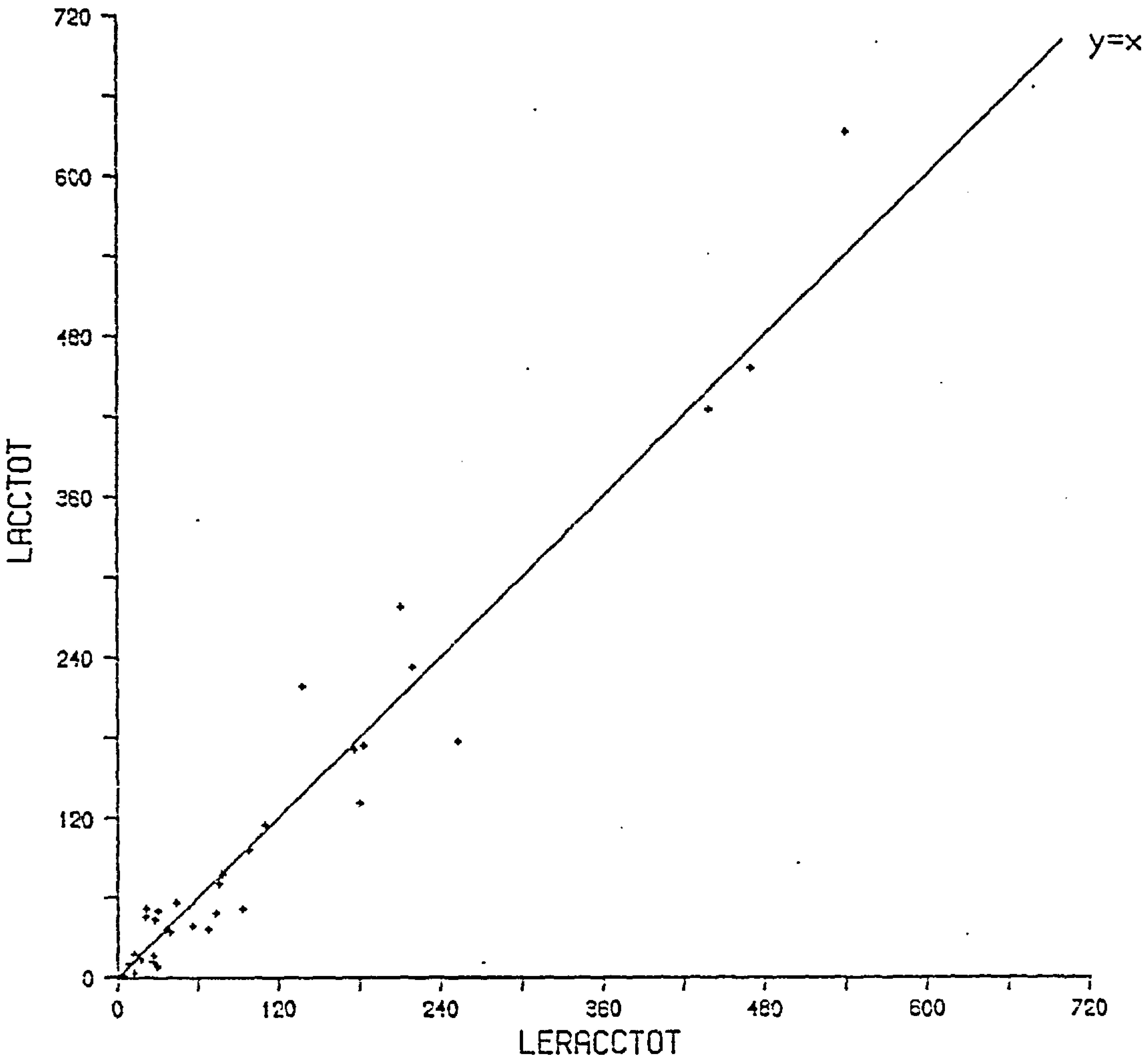


FIGURE 9.4: SCATTERPLOT OF OBSERVED v EXPECTED ALL ACCIDENT TOTALS FOR LINKS

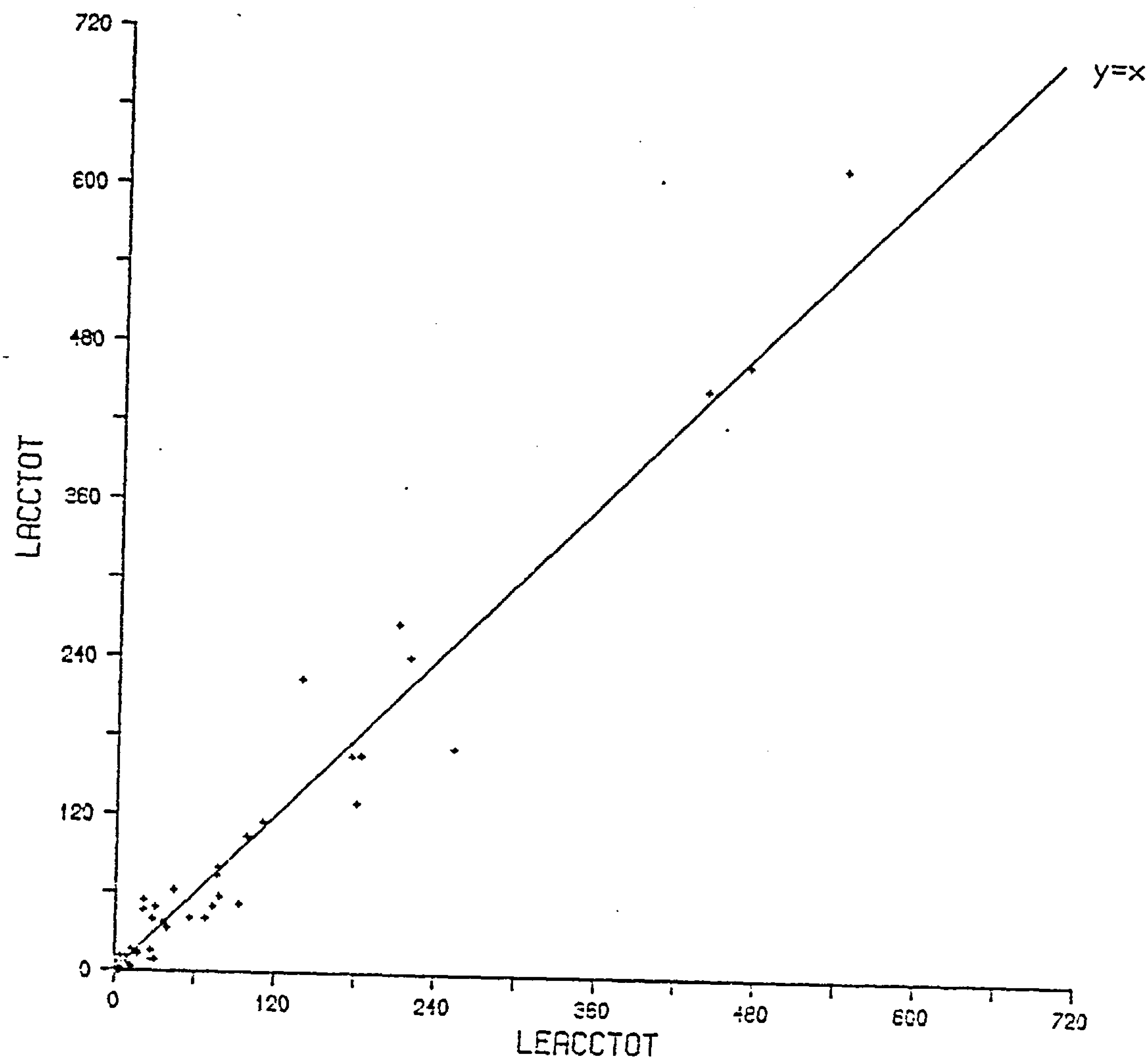


FIGURE 9.5: RELATIONSHIP BETWEEN ALL ACCIDENT RATES FOR SELECTED LINK TYPES

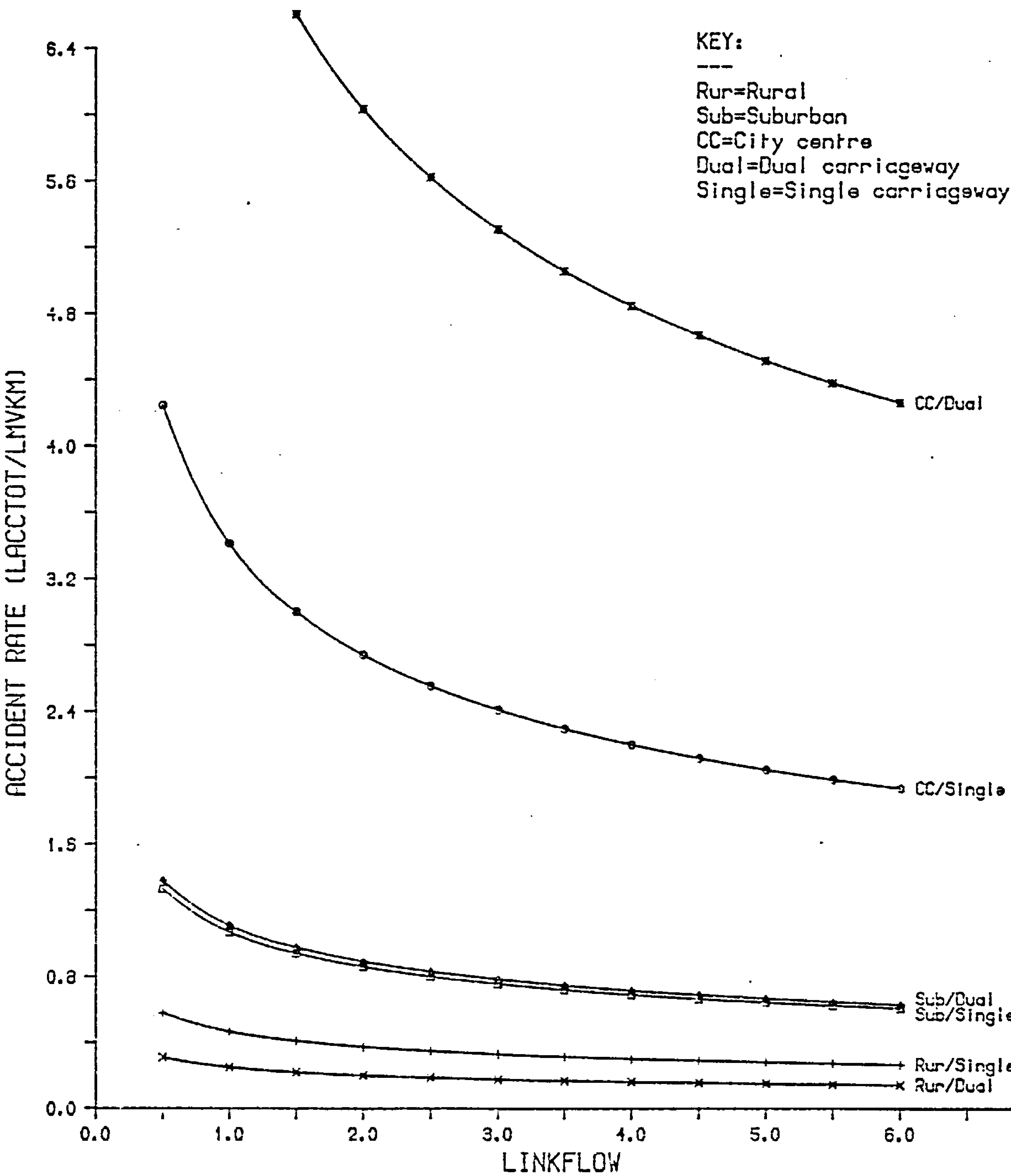


TABLE 9.1: COMBINATION OF DEVELOPMENT TYPES INTO NEW VARAIABLE LDEVTYPEA

LDEVTYPEA	LDEVTYPE	DESCRIPTION
1	66	Rural links
2	33,34,35,37,44,45 47,55,57,77	Any urban link with no adjacent shopping or commercial development
3	11,12,13,14,15,17 22,23,24,25,26	Any urban link with either shopping or commercial development on at least one side.

TABLE 9.2: PARAMETER ESTIMATES[1] FOR LSVNP ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	t-VALUES BETWEEN FACTOR LEVELS 2
GM	-	-1.355	0.072	**18.82**	
LDEVTYPEA	1	0.000	-		
	2	0.270	0.086	**3.132**	
	3	0.471	0.137	**3.443**	1.540
LCWYTYPE	1	0.000	-		
	2	-0.568	0.112	**5.074**	
LTINT	-	-0.480	0.065	**7.355**	

INITIAL SCALED DEVIANCE = 194.78 (df=51)
RESIDUAL SCALED DEVIANCE = 63.814 (df=47)

[1] All parameter estimates are natural logarithms.
[2] Significant (p<0.05) values are shown in **bold** type.

TABLE 9.3: PARAMETER ESTIMATES[1] FOR LTVNPA1 ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	t-VALUES BETWEEN FACTOR LEVELS 2
GM	-	-3.617	0.266	**13.59**	
LDEVTYPEA	1	0.000	-		
	2	1.555	0.338	**4.598**	
	3	3.045	0.440	**6.927**	**3.423**
LCWYTYPE	1	0.000	-		
	2	0.135	0.263	0.511	
LTINT	-	-0.928	0.412	**2.250**	
LTINTSQ	-	0.649	0.190	**3.409**	
LDEVTYPEA.	1.2	0.000	-		
LCWYTYPE	2.2	0.424	0.384	1.103	
	3.2	1.062	0.398	**2.671**	1.559
LDEVTYPEA.	1.LTINT	0.000	-		
LTINT	2.LTINT	-0.595	0.262	**2.271**	
	3.LTINT	-0.868	0.317	**2.740**	1.014

INITIAL SCALED DEVIANCE = 330.92 (df=51)
RESIDUAL SCALED DEVIANCE = 59.322 (df=42)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in **bold** type

TABLE 9.4: PARAMETER ESTIMATES[1] FOR LTVNPA2 ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	t-VALUES BETWEEN FACTOR LEVELS 2
GM	-	-2.478	0.095	25.99	
LDEVTYPEA	1	0.000	-		
	2	-0.040	0.198	0.201	
	3	1.400	0.355	3.940	3.752
LCWYTYPE	1	0.000	-		
	2	-1.831	0.200	9.169	
LTINT	-	-0.322	0.106	3.028	
LDEVTYPEA. LTINT	1.LTINT	0.000	-		
	2.LTINT	-0.255	0.164	1.556	
	3.LTINT	-1.130	0.285	3.964	2.992

INITIAL SCALED DEVIANCE = 130.23 (df=51)
RESIDUAL SCALED DEVIANCE = 21.659 (df=45)

[1] All parameter estimates are natural logarithms.
[2] Significant (p<0.05) values are shown in bold type

TABLE 9.5: PARAMETER ESTIMATES[1] FOR LTVNPA3 ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	t-VALUES BETWEEN FACTOR LEVELS 2
GM	-	-3.350	0.135	24.85	
LDEVTYPEA	1	0.000	-		
	2	0.794	0.164	4.857	
	3	1.683	0.189	8.924	5.450
LCWYTYPE	1	0.000	-		
	2	-1.644	0.376	4.370	

INITIAL SCALED DEVIANCE = 296.56 (df=51)
RESIDUAL SCALED DEVIANCE = 76.440 (df=48)

[1] All parameter estimates are natural logarithms.
[2] Significant (p<0.05) values are shown in **bold** type

TABLE 9.6: PARAMETER ESTIMATES[1] FOR LMVNP ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2
GM	-	-3.909	0.308	**12.70**	
LDEVTYPEA	1	0.000	-		
	2	1.776	0.377	**4.713**	
	3	3.230	0.512	**6.305**	**2.897**
LTINT	-	-0.662	0.503	1.316	
LTINTSQ	-	0.656	0.229	**2.867**	
LDEVTYPEA. LTINT	1.LTINT	0.000	-		
	2.LTINT	-0.976	0.280	**3.492**	
	3.LTINT	-1.255	0.359	**3.492**	0.861

INITIAL SCALED DEVIANCE = 208.17 (df=51)
RESIDUAL SCALED DEVIANCE = 67.375 (df=45)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in **bold** type

TABLE 9.7: PARAMETER ESTIMATES[1] FOR LPED ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	t-VALUES BETWEEN FACTOR LEVELS 2
GM	-	-3.198	0.305	**10.48**	
LDEVTYPEA	1	0.000	-		
	2	2.259	0.321	**7.044**	
	3	3.816	0.328	**11.65**	**9.768**
LCWYTYPE	1	0.000	-		
	2	-0.908	0.743	1.222	
LTINT	-	-0.399	0.131	**3.038**	
LDEVTYPEA. LCWYTYPE	1.2	0.000	-		
	2.2	1.242	0.817	1.521	
	3.2	1.796	0.795	**2.261**	1.272

INITIAL SCALED DEVIANCE = 2385.8 (df=51)
RESIDUAL SCALED DEVIANCE = 277.44 (df=45)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in **bold** type

TABLE 9.8: PARAMETER ESTIMATES[1] FOR LPSV ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	t-VALUES BETWEEN FACTOR LEVELS 2
GM	-	-6.050	0.586	10.33	
LDEVTYPEA	1	0.000	-		
	2	3.025	0.600	5.041	
	3	4.997	0.592	8.438	8.630
LCWYTYPE	1	0.000	-		
	2	0.959	0.291	3.297	

INITIAL SCALED DEVIANCE = 963.50 (df=51)
RESIDUAL SCALED DEVIANCE = 198.50 (df=48)

[1] All parameter estimates are natural logarithms.
[2] Significant (p<0.05) values are shown in bold type

TABLE 9.9: PARAMETER ESTIMATES[1] FOR ALL ACCIDENTS (LACCTOT)[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	t-VALUES BETWEEN FACTOR LEVELS 2
GM	-	-0.768	0.117	6.579	
LDEVTYPEA	1	0.000	-		
	2	0.833	0.134	6.229	
	3	1.994	0.150	13.30	9.758
LCWYTYPE	1	0.000	-		
	2	-0.634	0.224	2.827	
LTINT	-	-0.317	0.086	3.670	
LDEVTYPEA. LCWYTYPE	1.2	0.000	-		
	2.2	0.671	0.346	1.940	
	3.2	1.424	0.328	4.339	2.140

INITIAL SCALED DEVIANCE = 2800.6 (df=51)
RESIDUAL SCALED DEVIANCE = 380.68 (df=45)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in bold type

TABLE 9.10: COMPARISON OF PARAMETER ESTIMATES[1] FOR LINK ACCIDENT TOTALS FOR INDIVIDUAL YEARS

VARIATE	FACTOR LEVEL	PARAMETER ESTIMATES AND (SCALED STANDARD ERRORS) LACCTOT1	LACCTOT2	LACCTOT3	LACCTOT4	SIGNIFICANT DIFFERENCES
GM	-	-0.7498	-0.9218	-0.7761	-0.6389	
		(0.1418)	(0.1292)	(0.1317)	(0.1373)	None
LDEVTYPEA	2	0.9650	0.8988	0.7129	0.7518	
		(0.1620)	(0.1451)	(0.1511)	(0.1605)	None
	3	2.083	2.114	1.890	1.884	
		(0.1829)	(0.1595	(0.1693)	(0.1828)	None
LCWYTYPE	2	-0.3416	-0.6004	-0.9756	-0.6870	
		(0.2493)	(0.2433)	(0.2840)	(0.2706)	None
LTINT	-	-0.3869	-0.2249	-0.2637	-0.3868	
		(0.1024)	(0.09252)	(0.1005)	(0.1047)	None
LDEVTYPEA.	2.2	0.4093	0.6282	0.7427	0.9095	
LCWYTYPE		(0.3943)	(0.3689)	(0.4445)	(0.4042)	None
	3.2	0.9023	1.061	1.875	1.836	
		(0.4032)	(0.3724)	(0.3875)	(0.3764)	None
INIT SCALED DEV		714.67	755.43	730.08	766.59	
	df	51	51	51	51	
RESID SCALED DEV		141.41	107.52	119.90	141.83	
	df	45	45	45	45	

[1] All parameter estimates are natural logarithms.

CHAPTER 10 THE USE OF ACCIDENT MODELS AT INDIVIDUAL SITES

10.1. Introduction

The accident models formulated in the previous two chapters were based on an analysis of aggregated data in the form of a contingency table. In this chapter, the appropriateness of the model to predict accidents at individual sites is examined.

10.2. Methodology

The prodedure adopted here involved:

1. fitting an all accident model of type (7.5) to the JACCTOT12 and LACCTOT12 data (see Section 6.3) which comprise the all accident totals for the period 1979-80;

2. using the model to generate expected accident totals for each site (ie JEACCTOT12 and LEACCTOT12); and

3. comparing the two-year accident estimates on a site-by-site basis with the observed accidents for the following two-year period 1981-1982 (ie JACCTOT34 and LACCTOT34).

In addition, as the mean accident totals for individual sites have been shown in Chapter 6 to conform with a negative binomial distribution, the possibility of improving the predictive qualities of the model was examined with the adoption of an appropriate empirical Bayesian method. This method, which is discussed in detail by Abbess et al [1981], enables estimates for future period accidents to be based on existing information used in such a way as to take into account an estimate for the 'regression-to-mean' effect, viz:

$$\text{EB2ACCTOT12}_{lmn} = (k_{mn}+O_{lmn})/((k_{mn}/O_{mn})+1) \qquad (10.1)$$

where:

EB2ACCTOT12_{lmn} is the empirical Bayesian estimate for n-year accident totals at site l of site type m (for clarity, the prefixes J and L have been omitted throughout this section: EB2ACCTOT12, therefore, represents both JEB2ACCTOT12 and LEB2ACCTOT12);

k_{mn} is the negative binomial exponent for site type m for an n-year period;

O_{lmn} is the observed n-year accident total at site l of site type m (ie ACCTOT12); and

O_{mn} is the observed n-year site mean accident total for site type m.

For clarity the subscripts l, m and n will be dropped from future references where appropriate.

In the terms of the present study, n can be substituted by 2 and O_{mn} by EACCTOT12 and because of the nature of the data it did not prove possible to generate for all sites statistically robust estimates $\hat{k}$ and $\hat{O}_{mn}$ for k and O_{mn} respectively. Accordingly, to allow a 'regression-to-mean' effect to be calculated for each location, EACCTOT12 was substituted in (10.1) for both $\hat{k}$ and $\hat{O}_{mn}$ (ie this assumes $\hat{k}=\hat{O}_{mn}$) to generate the alternative Bayesian estimated accident total of EB1ACCTOT12. This substitution reduced (10.1) to:

$$EB1ACCTOT12 = (ACCTOT12+EACCTOT12)/2. \quad (10.2)$$

In addition, one further Bayesian estimate, EB3ACCTOT12 was generated in which EACCTOT12 was substituted for $\hat{O}_{mn}$ in (10.1), viz:

$$EB3ACCTOT12 = (\hat{k}+ACCTOT12)/((\hat{k}/EACCTOT12)+1). \quad (10.3)$$

The above assumption that $\hat{k}=\hat{O}_{mn}$ is a somewhat arbitrary generalisation and should not be seen to be universally applicable. The values for $\hat{k}$ are seen to vary quite considerably by location type and clearly, where possible, observed values for $\hat{k}$ should be adopted. However, for the purposes of the present study the assumption $\hat{k}=\hat{O}_{mn}$ is considered to lead to a simple and readily understood expression for EB2ACCTOT12. Reference to the eleven examples in Table 6.14 shows relationships between $\hat{k}$ and $\hat{O}_{mn}$ ranging from $\hat{k}=0.09\hat{O}_{mn}$ to $\hat{k}=2.85\hat{O}_{mn}$; however, all but four of the eleven have a value for $\hat{k}<0.5\hat{O}_{mn}$. Accordingly, during the preliminary analysis, two other assumptions were examined (viz $\hat{k}=0.25\hat{O}_{mn}$ and $\hat{k}=0.50\hat{O}_{mn}$) but did not generate significant improvements on the $\hat{k}=\hat{O}_{mn}$ assumption over the whole range of the data.

It is appreciated that this assumption implies that the spatial distribution conforms with a one parameter distribution because the value of k would be solely dependent on the mean value of the distribution (ie μ). Abbess et al [1981] describe the method of moments estimate for k which requires a knowledge of the between site variance and is relatively easy to calculate. For even quite low numbers of sites this method will provide a more robust estimate for $\hat{k}$ than $\hat{k}=\hat{O}_{mn}$. However, road safety practitioners - particularly those in the employ of local authorities - may not have access to appropriate spatially distributed data and, for this reason, the $\hat{k}=\hat{O}_{mn}$ assumption has been adopted as a "rule-of-thumb". There is, however, an additional problem here which relates to the sampling framework for accident locations which could prove to be a source of quite considerable bias. For example, where a local authority does not employ a location-based structure for recording its road accident data and where staff resources are limited, the sites at which accident data are known may be those with high accident totals and, as such, would not be representative of the distibution for all peer group sites. Under such conditions, values for $\hat{k}$ - which are intended to be representative of all sites - could be subject to considerable error and a $\hat{k}=\hat{O}_{mn}$ assumption may be safer.

It is perhaps strictly true to say that, as a result of the above substitutions, the models can no longer be classified as Bayesian. The term, however, has been retained in recognition of the Bayesian approach whereby predictions can be improved on the basis of "prior" information (see Section 10.3 for a short discussion on the definition of "prior").

Having generated the above four estimates for the expected accident totals, the residual values were calculated as a measure of their predictive accuracy (an additional residual value for the difference between the two observed accident totals was also calculated), viz:

$$RES1 = ACCTOT34 - ACCTOT12 \quad (10.4)$$

$$RES2 = ACCTOT34 - EACCTOT12 \quad (10.5)$$

$$RESB1 = ACCTOT34 - EB1ACCTOT12 \quad (10.6)$$

$$RESB2 = ACCTOT34 - EB2ACCTOT12 \quad (10.7)$$

$$RESB3 = ACCTOT34 - EB3ACCTOT12 \quad (10.8)$$

The analysis was conducted on two levels:

1. where EACCTOT12, EB1ACCTOT12 were calculated for all sites in the data base (with the exclusion of 4^+-way junctions and one-way street links); and
2. where EACCTOT12, EB1ACCTOT12, EB2ACCTOT12 and EB3ACCTOT12 were calculated for a subset of the data for which statistically robust estimates for k and O exist;

in which the variances of the residual values were compared by the application of the

F-test (see, for example, Sokal and Rohlf [1969]) which determines the significance of the difference between two variances by examining their ratio, viz:

$$F = s_1^2/s_2^2 \qquad (10.9)$$

where:

F is the F-statistic, the significance of which can be determined from tabulated values for the appropriate number of degrees of freedom;

s_1^2 is the larger of the variance estimates; and

s_2^2 is the smaller of the variance estimates.

It is stressed here that the estimates for the expected accident totals are all based solely on information relating to the two-year period 1979–1980 and could, in theory, have been the subject of calculation during the early hours of 1st January 1981!

It is further stressed that the expected accident totals being estimated by the models are not site specific but rather relate to the accident total expected at an average site given its type and traffic flow characteristics.

10.3. Analysis of Junction Accidents

The two-year accident model based on JACCTOT12 is detailed in Table 10.1 with the comparison of the residual values for all junctions shown in part (a) of Table 10.2.

Table 10.3 shows the maximum likelihood $\hat{k}$ values (calculated through the use of the computer program MLP (Ross [1980]) see also Section 6.3) together with the associated mean accident totals for the nine permutations of JMAINBAND and JSIDEBAND for the 1,206 suburban priority controlled T-junctions which were used to enable the JEB2ACCTOT12 and JEB3ACCTOT12 values to be calculated and compared with the other expected values. This comparison is shown in part (b) of Table 10.2.

The analysis shown in part (a) of Table 10.2 indicates that there are no significant differences ($p<0.05$) between the mean values for the residuals. There are, however, significant differences between the variances associated with JRES1 and JRES2, and JRES1 and JRESB1 indicating that the two modelled expected accident totals are more efficient predictors of future accident totals than are the prior accident totals at sites.

The term "prior" is used here because it is the generally accepted Bayesian term used to define some known parameters related to data and their distributions. In the context of road safety, the prior accident totals are usually the latest available three-year figures.

The more detailed analysis of the 1,206 suburban priority T-junctions (see Table 10.2 part (b)) indicates, however, that no significant ($p<0.05$) improvement to the JEB1ACCTOT12 model was offered by the JEB2ACCTOT12 and JEB3ACCTOT12 accident totals.

The correlation coefficient for the relationship between JACCTOT34 and JEB1ACCTOT12 for all 3,089 junctions was 0.685 and, similarly, for the 1,206 suburban priority T-junctions it was 0.432.

10.4. Analysis of Link Accidents

The two-year accident model based on LACCTOT12 is detailed in Table 10.4 with the comparison of the residual values for all links shown in part (a) of Table 10.5.

Table 10.6 shows the maximum likelihood $\hat{k}$ values together with the associated mean accident totals for the nine selected link types (comprising 1,119 links) which were used to enable the LEB2ACCTOT12 and LEB3ACCTOT12 values to be calculated and compared with the other expected values. This comparison is shown in part (b) of Table 10.5.

The result of the analysis shown in part (a) of Table 10.5 indicates that whilst there are no significant differences ($p<0.05$) between the mean values for the residuals, there are significant differences between all three variances indicating that the modelled expected accident totals are more efficient predictors of future accident totals than are the prior accident totals at sites and that LEB1ACCTOT12 is more efficient than LEACCTOT12.

As observed with the junction data, the more detailed analysis for the 1,119 selected links (see Table 10.5 part (b)) indicates that no significant ($p<0.05$) improvement to the LEB1ACCTOT12 model was offered by the LEB2ACCTOT12 and LEB3ACCTOT12 accident totals.

The correlation coefficient for the relationship between LACCTOT34 and LEB1ACCTOT12 for all 2,460 links was 0.633 and similarly for the 1,119 selected links it was 0.524.

10.5. Conclusions

The results of this analysis indicate that significant improvements to the models are achieved by the application of empirical Bayesian methods and this stresses the fundamental importance of considering the effect of regression-to-mean when predicting future accident totals.

A profitable further research topic could examine the possibility of improving the empirical Bayesian methods adopted here for the estimation of future years' accident totals.

TABLE 10.1: PARAMETER ESTIMATES[1] FOR JACCTOT12 ACCIDENTS[2]

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	2	3	4
GM	-	-2.366	0.197	**11.99**			
JUNCTYPEA	1	0.000	-				
	2	0.777	0.176	**4.427**			
	3	0.374	0.190	1.962	**4.041**		
	4	1.609	0.185	**8.688**	**9.932**	**11.66**	
	5	1.178	0.190	**6.190**	**4.075**	**6.882**	**4.201**
JCONTROL	1	0.000	-				
	2	0.167	0.109	1.527			
JLOCATION	1	0.000	-				
	2	0.278	0.120	**2.321**			
	3	0.573	0.135	**4.252**	**3.636**		
	4	0.835	0.135	**6.169**	**6.690**	**2.799**	
JTINT	-	-0.038	0.023	1.680			
JTINTSQ	-	-0.026	0.009	**2.722**			

INITIAL SCALED DEVIANCE = 1369.9 (df=183)
RESIDUAL SCALED DEVIANCE = 437.27 (df=173)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in **bold** type

TABLE 10.2: COMPARISON OF JUNCTION ACCIDENT ESTIMATES[1]

	JRES1	JRES2	JRESB1	JRESB2	JRESB3
(a) ALL JUNCTIONS:					
MEAN	-0.009	-0.008	-0.008	-	-
t-value					
wrt JRES1	-	0.030	0.030	-	-
wrt JRES2	-	-	0.030	-	-
VARIANCE	2.065	1.669	1.432	-	-
F-statistic					
wrt JRES1	-	**1.237**	**1.442**	-	-
wrt JRES2	-	-	**1.166**	-	-
(b) SELECTED 1,206 T-JUNCTIONS:					
MEAN	0.007	-0.001	0.003	0.100	0.008
t-value					
wrt JRES1	-	0.304	0.304	0.126	0.316
wrt JRES2	-	-	0.304	1.000	0.881
wrt JRESB1	-	-	-	1.117	0.817
wrt JRESB2	-	-	-	-	0.450
VARIANCE	1.307	0.851	0.839	0.813	0.798
F-statistic					
wrt JRES1	-	**1.536**	**1.558**	**1.608**	**1.638**
wrt JRES2	-	-	1.014	1.047	1.066
wrt JRESB1	-	-	-	1.032	1.051
wrt JRESB2	-	-	-	-	1.019

[1] Significant ($p<0.05$) values shown in bold type.

TABLE 10.3: MEAN ACCIDENT TOTALS AND $\hat{k}$ VALUES BY JMAINBAND AND JSIDEBAND FOR SUBURBAN PRIORITY T-JUNCTIONS

JMAINBAND	JSIDEBAND	NUMBER OF SITES	SITE MEAN JACCTOT12 ($\hat{O}$)	NEGATIVE BINOMIAL EXPONENT ($\hat{k}$)	$\hat{k}/\hat{O}$
1	1	116	0.164	1.907	11.643
1	2	157	0.389	0.763	1.963
1	3	57	0.509	1.117	2.196
2	1	155	0.361	0.883	2.445
2	2	249	0.494	1.366	2.764
2	3	103	1.218	3.723	3.057
3	1	114	0.693	1.150	1.659
3	2	190	0.963	2.275	2.362
3	3	65	1.262	1.231	0.976

TABLE 10.4: PARAMETER ESTIMATES[1] FOR LACCTOT12 ACCIDENTS

VARIATE OR FACTOR	FACTOR LEVEL	PARAMETER ESTIMATE	PARAMETER SCALED STND ERR	t-VALUES BETWEEN FACTOR LEVELS GM/1	t-VALUES BETWEEN FACTOR LEVELS 2
GM	-	-0.833	0.126	**6.640**	
LDEVTYPEA	1	0.000	-		
	2	0.933	0.142	**6.562**	
	3	2.101	0.158	**13.26**	**9.542**
LCWYTYPE	1	0.000	-		
	2	-0.461	0.228	**2.025**	
LTINT	-	-0.308	0.090	**3.412**	
LDEVTYPEA. LCWYTYPE	1.2	0.000	-		
	2.2	0.509	0.353	1.442	
	3.2	0.970	0.359	**2.704**	1.202

INITIAL SCALED DEVIANCE = 1425.8 (df=51)
RESIDUAL SCALED DEVIANCE = 212.37 (df=45)

[1] All parameter estimates are natural logarithms.
[2] Significant ($p<0.05$) values are shown in **bold** type.

TABLE 10.5: COMPARISON OF LINK ACCIDENT ESTIMATES[1]

	LRES1	LRES2	LRESB1	LRESB2	LRESB3
(a) ALL LINKS:					
MEAN	-0.032	-0.024	-0.028	-	-
t-value					
wrt LRES1	-	0.293	0.293	-	-
wrt LRES2	-	-	0.293	-	-
VARIANCE	2.137	1.631	1.411	-	-
F-statistic					
wrt LRES1	-	**1.310**	**1.515**	-	-
wrt LRES2	-	-	**1.156**	-	-
(b) SELECTED 1,119 LINKS:					
MEAN	-0.089	-0.054	-0.072	-0.089	-0.047
t-value					
wrt LRES1	-	0.656	0.656	0.000	1.783
wrt LRES2	-	-	0.656	1.038	0.231
wrt LRESB1	-	-	-	1.180	**2.510**
wrt LRESB2	-	-	-	-	**3.292**
VARIANCE	3.353	2.155	1.952	2.073	2.036
F-statistic					
wrt LRES1	-	**1.556**	**1.718**	**1.617**	**1.647**
wrt LRES2	-	-	1.104	1.040	1.058
wrt LRESB1	-	-	-	1.062	1.043
wrt LRESB2	-	-	-	-	1.018

[1] Significant ($p<0.05$) values are shown in bold type.

TABLE 10.6: MEAN ACCIDENT TOTALS AND $\hat{k}$ VALUES BY LFLOWBAND AND LENGTHBAND FOR SELECTED SINGLE CARRIAGEWAY LINKS

LDEVTYPEA	LFLOWBAND	LENGTHBAND	NUMBER OF SITES	SITE MEAN LACCTOT12 ($\hat{O}$)	NEGATIVE BINOMIAL EXPONENT ($\hat{k}$)	$\hat{k}/\hat{O}$
1	1	3	253	0.909	0.519	0.571
1	2	3	81	1.222	0.669	0.547
1	3	3	54	1.556	0.935	0.601
2	1	3	140	0.650	1.142	1.757
2	2	3	210	1.143	2.788	2.439
2	3	3	176	1.614	1.369	0.848
3	3	1	118	0.873	1.927	2.207
3	3	2	52	2.750	2.072	0.753
3	3	3	35	3.171	1.627	0.513

CHAPTER 11 APPLICATION OF MODEL TO BLACKSPOT IDENTIFICATION

11.1. Introduction

In Chapters 8 and 9 models have been developed on the basis of contingency table data for both junction and link accidents. In Chapter 10 these models were applied to individual sites and it was concluded that they provided more accurate predictions for future year accidents than the prior accident record. In addition, it was shown that the application of empirical Bayesian methods, which take account of the regression-to-mean effect, further improved the predictive qualities of the model.

This has implications for the procedures for blackspot identification. As the modelled accident totals are generally better predictors of the levels of future accidents, it is these totals, rather than the prior totals, which provide a more appropriate basis for site by site comparison.

In this chapter a comparison is made between the blackspots identified by the simple criterion of prior accident total and that of potential accident reduction (PAR). The prior accident total is often referred to as the annual accident total for which the acronym "AAT" is often substituted.

11.2. Literature Review

Silcock and Smyth [1984], in a survey of blackspot identification procedures adopted by British local authorities, report that the majority of authorities adopt an AAT approach. On the basis of 55 returns from 67 questionnaires distributed to local authorities (excluding London Boroughs), some 41 authorities indicated the use of AAT as a blackspot ranking criterion.

Maher and Mountain [1988] have suggested that the relative efficiency of the PAR approach over that of AAT is dependent on the accuracy of the estimate for the expected accident frequency for each site. They suggest that because regression models, based on data for individual sites, produce r^2 values of up to about 0.50, the expected accident frequencies are subject to a substantial degree of error. Accordingly, it is suggested that the efficiency of PAR may not be much greater than that for AAT.

The r^2 values quoted by Maher and Mountain [1988] are, however, based on data for individual sites which would be subject to between site or spatial variation which, it has been shown, conform with a negative binomial distribution (see Chapter 6). It seems, therefore, curious that they should have compared the r^2 values for regressions on an individual site basis to determine the accuracy of the estimate for expected accident total because observed accident totals at individual sites are not necessarily related to the expected number of accidents at an average site.

This is an important point because the PAR rationale is based on the assumption that road safety remedial works will reduce the long term accident total at a site to the expected accident total for such a site type and traffic flow. What is relevant here is the ability to predict accurately the expected accident total not the observed accident total at a site. The correlation coefficients referred to by Maher and Mountain [1988] are not measures of the strength of relationship between the expected accident total and the estimate of the expected accident total.

In the context of this study, the estimates for the expected accident totals have been determined on the basis of site type and traffic flow and, accordingly, are representative of average accident frequencies for site types not for individual sites.

The conclusions drawn from the earlier chapters suggest that the site mean accident totals (ie the relevant accident total, ACCTOT, divided by the number of sites per cell, CELLTOT) for the contingency table cells are rather more accurately predicted than those for individual locations. To show this formally, the modelled expected site mean accident totals based solely on the 1979-1980 (ie ACCTOT12) data for both junctions and links were calculated for each cell in the two contingency tables. These expected values (ie EB1ACCTOT12/CELLTOT) were correlated with the observed values for 1981-1982 (ie ACCTOT34). The models used are those shown in Tables 10.1 and 10.4 for junction and link accidents respectively.

For the junction accidents, a correlation coefficient, r, of 0.723 was observed for the site mean values for all 184 cells in the junction contingency table. However, as many of these cells contained very few sites (often single sites) a second value for r of 0.907 was calculated for those 106 cells with five or more constituent sites. The relationship between the expected and observed site mean accident totals for the 106 cells containing five or more junctions is shown graphically in Figure 11.1.

The correlation coefficient for all 54 link cells was 0.907 (no significance should be read into the similarity of this value with the above value for the junction data). As only 7 of the cells contained less than five links it was not considered necessary to calculate a second value for r. This relationship is shown in Figure 11.2 where three outliers (enclosed in circles) are associated with values for LCELLTOT of 2, 3 and 4 as indicated.

This level of correlation indicates that the estimate of the expected values provides a rather more accurate prediction of the expected accident frequency than that suggested by Maher and Mountain [1988].

Further evidence of the predictive accuracy of the models for expected accident totals can be gleaned from Sections 8.6 and 9.5 which relate to the temporal consistency of the models.

11.3. Blackspot Identification

The data have, as in Chapter 10, been split into the two non-overlapping time periods relating to 1979-1980 and 1981-1982 and, for both junction and link accidents, the two blackspot identification criteria have been defined as:

$$AAT = ACCTOT12 \quad (11.1)$$
$$PAR = EB1ACCTOT12 - EACCTOT12 \quad (11.2)$$

For clarity, where appropriate, the prefixes "J" and "L" have been omitted throughout this chapter.

AAT is, therefore, the blackspot ranking criterion based simply on the prior accident total and has been adopted by a considerable majority of British local authorities (Silcock and Smyth [1984]).

PAR, which takes account of the regression-to-mean effect, adopts an empirical Bayesian estimate (ie EB1ACCTOT12) for the individual site accident frequencies.

These two blackspot identification criteria have been calculated for all junctions (excluding other 4^+-way junctions) and all links (excluding one-way streets).

On the basis of the AAT and PAR values, the ranks AATRANK and PARRANK have been established. The ranks, for example for AAT, were determined by sorting individual sites in descending order of the value of AAT and storing, in AATRANK, the integer values from 1 to n denoting the position of the individual site in the sorted file. The top ranked sites for both ranking criteria are shown together with corresponding alternative rank values for both the junction and link data in Tables 11.1 to 11.4.

It was, therefore, possible to "select" sites for treatment on the basis of accident data for 1979-1980 and with a knowledge of the posterior observed (ie not modelled) accident totals for 1981-1982 at the "selected" sites, determine the relative effectiveness of the two alternative ranking criteria.

11.4. Justification of PAR as a Ranking Criterion in preference to AAT

Table 11.1 shows the forty-one top junctions ranked by JAAT. The selection of 41 junctions is explained by the fact that the junctions between the 42nd and 58th ranks (inclusive) each experienced 3.5 accidents per year for the years 1979-1980 and rather than employ a secondary ranking criterion to determine the top 50 junctions, the tabulation was restricted to the 41 junctions with over 3.5 accidents/year. Table 11.2 shows equivalent data for the top forty-one junctions ranked by JPAR. The accident values in these tables, as in Tables 11.3 and 11.4, are expressed in accidents/year.

In Table 11.1 (column 5) it can be seen that the site mean expected accident total is 2.145 accidents/junction/year and that the actual observed site mean accident totals for the following two years (ie JACCTOT34÷2) is 2.927 accidents/junction/year. On the assumtions that:

1. remedial works had been undertaken at each site on 1st January 1981;
2. the remedial works would have achieved a reduction in accident occurrence to the expected levels (ie EXPACC); and
3. the costs of remedial work do not vary by blackspot selection criterion;

an accident reduction of 0.782 accidents/junction/year (ie 2.927-2.145) would, therefore, have been observed for the junctions ranked by JAAT.

The equivalent reduction for junctions ranked by JPAR is 1.031 accidents/junction/year (ie 2.573-1.542 from columns 5 and 11 of Table 11.2) which suggests that a 32% improvement of 0.249 (ie 1.031-0.782) accidents/junction/year would have been achieved from the adoption of a PAR ranking criterion to one of AAT.

Examination of these tables indicates that thirty-two (78%) of the top ranked forty-one junctions are common to both blackspot ranking criteria. This may seem to suggest a high degree of similarity between the two lists, but when it is considered that valuable staff time and scarce financial resources may be spent on almost one in four sites, the practical relevance of the differences between the two blackspot lists becomes apparent. Of the nine non-common junctions in Table 11.1 all are 4-way junctions, eight of which are signal-controlled and eight of which are in Edinburgh's city centre. In Table 11.2 it can be seen that these nine junctions have been replaced by six 3- and three 4-way junctions, seven of which are not signal-controlled and only one of which is in Edinburgh's city centre. The PAR selected junctions, therefore, comprise a more balanced set of priorities for investigation than is the case for AAT.

Tables 11.3 and 11.4 show the top ranked thirty-two links based on the LAAT and LPAR blackspot identification criteria respectively. On the basis of the same assumptions outlined above, the AAT criterion shows an accident reduction of 0.919 (ie 2.938-2.019) accidents/link/year whereas an equivalent value for PAR of 1.184 (ie 2.344-1.160) is observed. The PAR criterion, therefore, demonstrates an increased accident saving of 0.265 (ie 1.184-0.919) accidents/link/year which would constitute a 29% improved return over that for AAT if it is assumed that the costs of the remedial works so identified are of the same order as those identified by AAT.

Of the thirty-two top ranked links, twenty (62.5%) are common to both blackspot ranking criteria. The comment above relating to the practical significance of the differences between the two blackspot lists is, therefore, even more relevant.

The above arguments are based on the assumption that all sites whether ranked by AAT or PAR would be the subject of remedial works. In practice this will not be the case because, for example, some ranked sites:

1. may be affected in some way by other road schemes in the course of design or construction;
2. may be considered to be relatively satisfactory in road safety terms in view of the high traffic volumes negotiating the site; and
3. may not offer cost-effective solutions in view of the high costs associated with proposed remedial works.

There is no reason to believe that the PAR criterion would be disadvantaged in any way by this with respect to the AAT criterion. Indeed, where the second of the above reasons is applied to AAT generated priorities, the process converges to a greater or lesser extent on PAR. It is perhaps true to say that such a converging process may, in the limit, produce much the same schemes at the end of the day for both criteria but would inevitably involve considerably more staff resources if based on initial AAT ranked priorities. However, it is possible that some sites which would have been ranked by PAR may never come into the AAT reckoning.

It is perhaps salutory, at this point, to consider the accident reduction that might have been claimed by a local authority on the basis of a simple before and after study. It is assumed, here, that such a study would examine the effectiveness of a programme of remedial road works at those sites detailed in Tables 11.1 and 11.3 all undertaken on 1st January 1981 where the before data comprise the two-year accident totals ACCTOT12 and the after data comprise the two-year accident totals EACCTOT12 (for the sake of argument it is assumed that the road works proved successful in reducing accidents to the expected values). The accident saving could, therefore, have been measured in terms of ACCTOT12-EACCTOT12 and for both junctions and links, such savings of the order of 3 accidents/site/year would have been claimed. This would have overstated the "observed" accident savings (ie ACCTOT34-EACCTOT12) by a factor of three.

Indeed, for the junction and link sites ranked on the basis of AAT in Tables 11.1 and 11.3 there were 40% reductions in accidents from the 1979-1980 levels (ie ACCTOT12) to the 1981-1982 levels (ie ACCTOT34). This suggests that had wholly ineffective remedial works been undertaken on 1st January 1981 a 40% reduction in accidents would have been incorrectly (but apparently justifiably) claimed as being attributable to the works. This observation indicates just how important it is to ensure that due account is taken of the regression-to-mean effect in this area of engineering. It is interesting to ponder on the possibility that remedial works could even increase the long-term level of accidents at a site and yet still show a short-term accident reduction which could find its way into the annals of effective road safety remedial works.

11.5. Conclusions

In this chapter it has been shown that that the use of alternative blackspot identification procedures can produce quite different lists of sites for investigation and further that there is some justification for adopting the Bayesian PAR criterion in preference to the AAT criterion. This suggests that the adoption of the PAR approach by local authorities would generate a more cost effective approach to their road safety work than that offered by AAT.

In addition, it has been shown that those methods for the before and after evaluation of remedial works which do not take regeression-to-mean into account can significantly overstate - perhaps by a factor of three - the true rate of return.

FIGURE 11.1: SCATTERPLOT OF EXPECTED v. OBSERVED SITE MEAN ACCIDENT TOTALS FOR JUNCTION ACCIDENT CONTINGENCY TABLE CELLS (WHERE NUMBER OF SITES/CELL > 5).

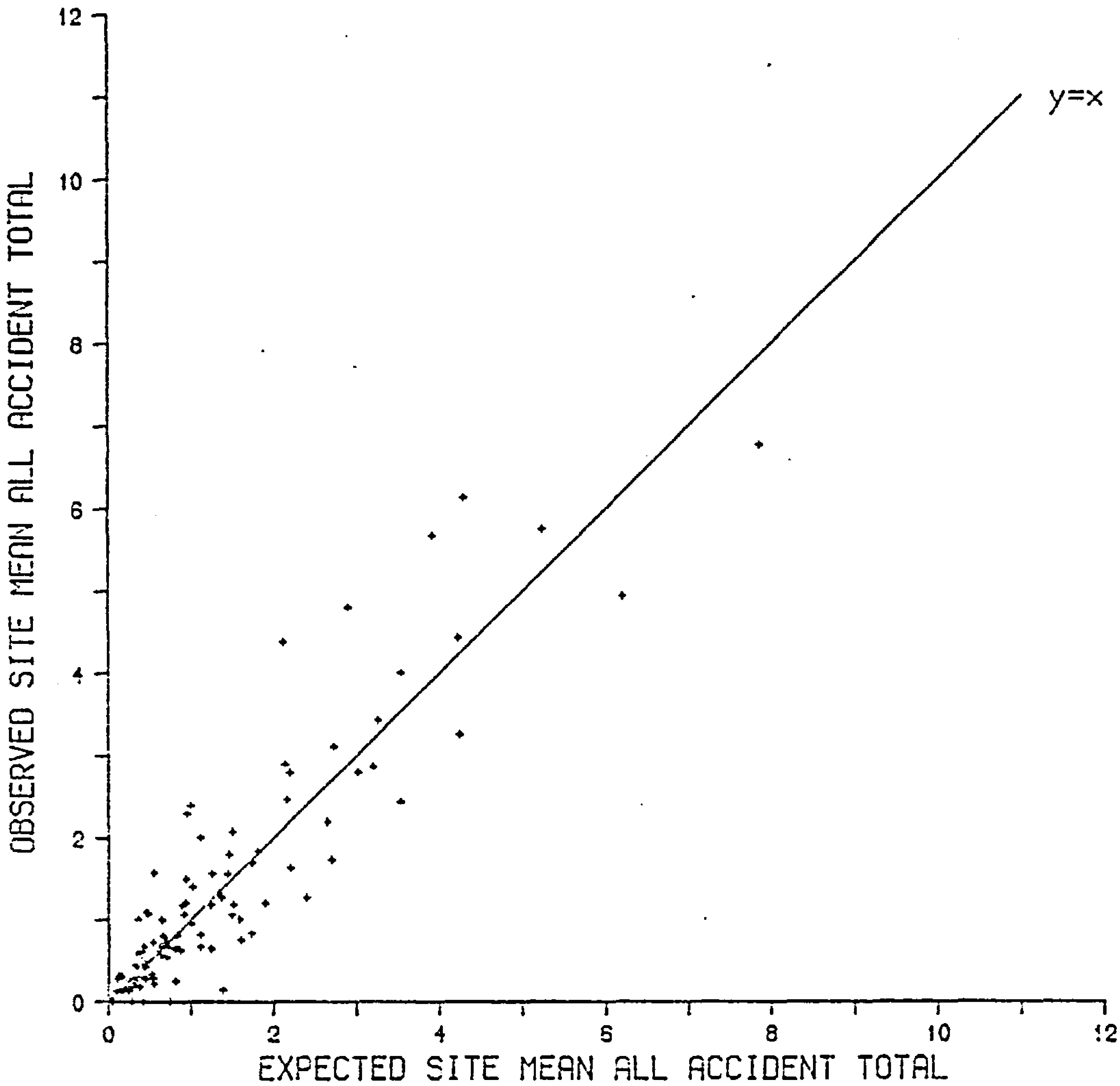


FIGURE 11.2: SCATTERPLOT OF EXPECTED v. OBSERVED SITE MEAN ACCIDENT TOTALS FOR LINK ACCIDENT CONTINGENCY TABLE CELLS (WHERE NUMBER OF SITES/CELL > 5).

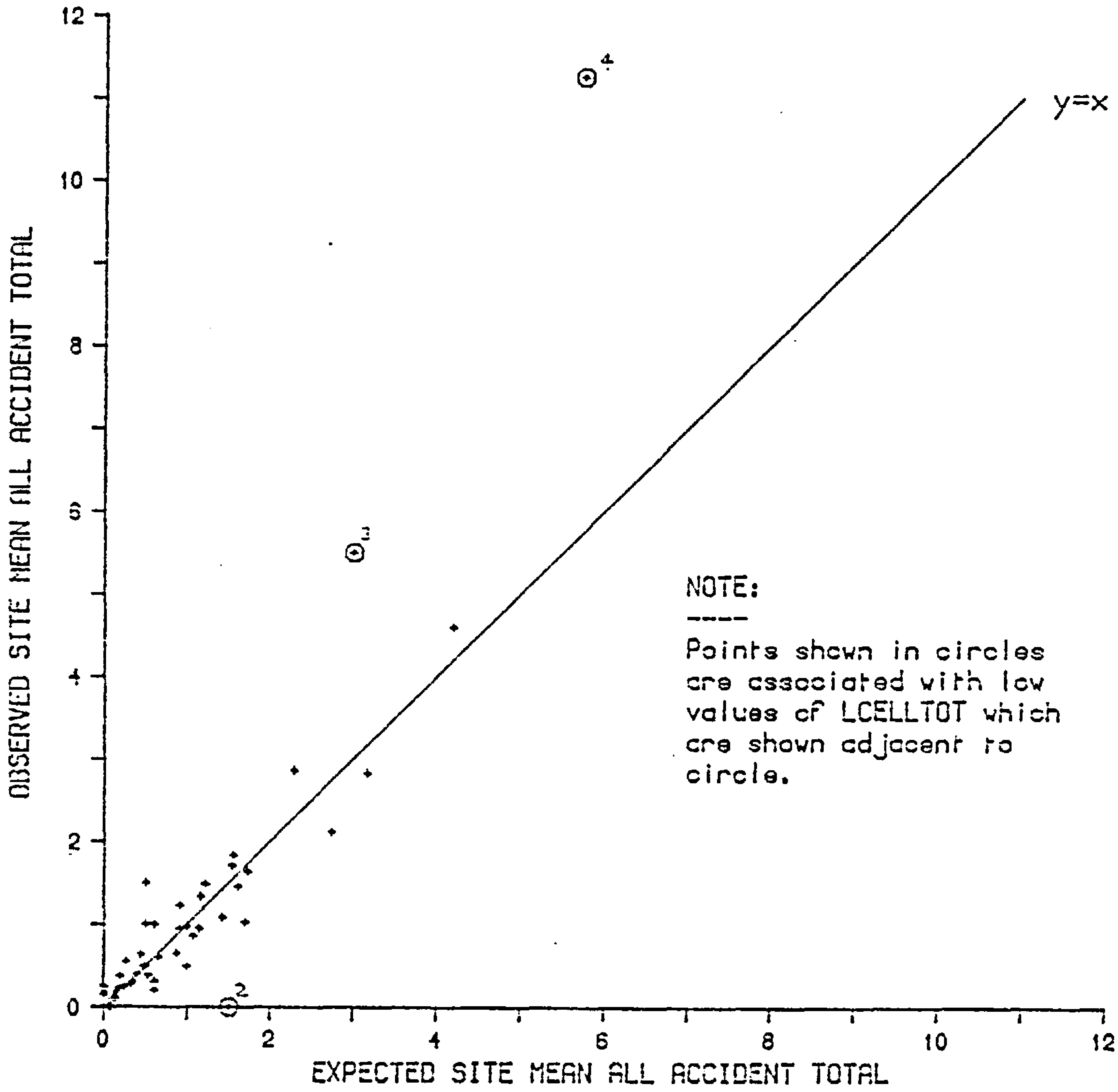


TABLE 11.1: COMPARISON OF BLACKSPOT RANKING CRITERIA FOR THE 41 TOP JUNCTIONS RANKED BY JAAT

JREF-RNCE	JUNC-TYPE	JCON-TROL	JLOC-ATN	JACC-TOT12	JEACC-TOT12	JEB1ACC-TOT12	JPAR	JAAT-RANK	JPAR-RANK	JACC-TOT34
8782	4	2	3	7.5	3.13	5.32	2.18	1	3	7.0
5394	4	2	4	7.0	4.30	5.65	1.35	2	26	5.0
5273	2	1	4	6.5	1.82	4.16	2.34	3	1	3.0
4817	4	2	4	6.5	4.31	5.41	1.09	4	41	4.0
13762	4	2	4	6.5	4.04	5.27	1.23	5	32	2.0
1371	5	2	4	6.5	3.30	4.90	1.60	6	16	6.0
13377	4	2	2	6.0	2.54	4.27	1.73	7	11	8.5
5696	4	2	4	5.5	4.77	5.14	0.36	8	324	8.5
5866	4	1	4	5.5	3.12	4.31	1.19	9	36	4.0
5856	5	1	4	5.5	1.76	3.63	1.87	10	7	3.0
6192	2	1	2	5.0	0.77	2.89	2.11	11	4	2.0
11404	4	2	3	5.0	2.11	3.55	1.45	12	21	3.5
4882	3	1	4	5.0	0.56	2.78	2.22	13	2	1.5
5694	5	2	4	5.0	1.77	3.38	1.62	14	15	3.0
5693	5	2	4	5.0	2.46	3.73	1.27	15	31	2.5
6851	5	2	4	5.0	3.11	4.05	0.95	16	62	3.5
10033	2	1	3	4.5	0.66	2.58	1.92	17	6	0.0
6744	4	2	2	4.5	1.14	2.82	1.68	18	12	1.0
9844	4	2	2	4.5	2.13	3.31	1.19	19	37	7.0
3907	4	1	2	4.5	1.15	2.83	1.67	20	13	1.0
3565	4	2	3	4.5	2.59	3.54	0.96	21	58	2.5
6464	4	2	4	4.5	3.84	4.17	0.33	22	362	2.5
4836	4	2	4	4.5	4.31	4.40	0.10	23	788	4.0
5357	4	1	4	4.5	1.95	3.22	1.28	24	30	1.5
5545	3	1	2	4.5	0.37	2.43	2.07	25	5	3.0
5254	5	2	4	4.5	2.11	3.30	1.20	26	35	5.0
5251	5	2	4	4.5	2.51	3.50	1.00	27	51	2.0
6081	2	1	2	4.0	0.43	2.22	1.78	28	9	1.5
5579	2	2	3	4.0	1.58	2.79	1.21	29	34	0.5
6098	2	1	3	4.0	0.49	2.24	1.76	30	10	2.0
6012	2	1	3	4.0	1.27	2.64	1.36	31	24	1.0
6982	2	1	4	4.0	1.39	2.70	1.30	32	28	1.5
14518	4	1	2	4.0	0.68	2.34	1.66	33	14	1.0
6467	4	1	2	4.0	1.05	2.53	1.47	34	19	0.0
5847	4	2	4	4.0	3.87	3.93	0.07	35	867	4.0
14506	4	2	4	4.0	4.29	4.15	-0.15	36	2296	5.0
5757	4	1	4	4.0	2.75	3.37	0.63	37	133	1.5
5546	3	1	2	4.0	0.36	2.18	1.82	38	8	1.0
5332	3	1	4	4.0	1.07	2.53	1.47	39	20	1.0
5318	5	1	4	4.0	0.84	2.42	1.58	40	18	2.0
4808	5	1	4	4.0	1.26	2.63	1.37	41	23	1.5
AVERAGE				4.841	2.145	3.493	1.348	21	135	2.927

TABLE 11.2: COMPARISON OF BLACKSPOT RANKING CRITERIA FOR THE 41 TOP JUNCTIONS RANKED BY JPAR

JREF-RNCE	JUNC-TYPE	JCON-TROL	JLOC-ATN	JACC-TOT12	JEACC-TOT12	JEB1ACC-TOT12	JPAR	JAAT-RANK	JPAR-RANK	JACC-TOT34
5273	2	1	4	6.5	1.82	4.16	2.34	3	1	3.0
4882	3	1	4	5.0	0.56	2.78	2.22	13	2	1.5
8782	4	2	3	7.5	3.13	5.32	2.18	1	3	7.0
6192	2	1	2	5.0	0.77	2.89	2.11	11	4	2.0
5545	3	1	2	4.5	0.37	2.43	2.07	25	5	3.0
10033	2	1	3	4.5	0.66	2.58	1.92	17	6	0.0
5856	5	1	4	5.5	1.76	3.63	1.87	10	7	3.0
5546	3	1	2	4.0	0.36	2.18	1.82	38	8	1.0
6081	2	1	2	4.0	0.43	2.22	1.78	28	9	1.5
6098	2	1	3	4.0	0.49	2.24	1.76	30	10	2.0
13377	4	2	2	6.0	2.54	4.27	1.73	7	11	8.5
6744	4	2	2	4.5	1.14	2.82	1.68	18	12	1.0
3907	4	1	2	4.5	1.15	2.83	1.67	20	13	1.0
14518	4	1	2	4.0	0.68	2.34	1.66	33	14	1.0
5694	5	2	4	5.0	1.77	3.38	1.62	14	15	3.0
1371	5	2	4	6.5	3.30	4.90	1.60	6	16	6.0
13214	2	1	1	3.5	0.33	1.92	1.58	46	17	2.0
5318	5	1	4	4.0	0.84	2.42	1.58	40	18	2.0
6467	4	1	2	4.0	1.05	2.53	1.47	34	19	0.0
5332	3	1	4	4.0	1.07	2.53	1.47	39	20	1.0
11404	4	2	3	5.0	2.11	3.55	1.45	12	21	3.5
4498	5	1	3	3.5	0.61	2.06	1.44	53	22	4.0
4808	5	1	4	4.0	1.26	2.63	1.37	41	23	1.5
6012	2	1	3	4.0	1.27	2.64	1.36	31	24	1.0
2588	2	1	2	3.5	0.78	2.14	1.36	43	25	1.5
5394	4	2	4	7.0	4.30	5.65	1.35	2	26	5.0
6137	5	1	2	3.5	0.81	2.16	1.34	52	27	1.0
6982	2	1	4	4.0	1.39	2.70	1.30	32	28	1.5
3846	2	1	3	3.0	0.43	1.71	1.29	61	29	1.0
5357	4	1	4	4.5	1.95	3.22	1.28	24	30	1.5
5693	5	2	4	5.0	2.46	3.73	1.27	15	31	2.5
13762	4	2	4	6.5	4.04	5.27	1.23	5	32	2.0
6040	2	1	2	3.5	1.07	2.28	1.22	44	33	3.0
5579	2	2	3	4.0	1.58	2.79	1.21	29	34	0.5
5254	5	2	4	4.5	2.11	3.30	1.20	26	35	5.0
5866	4	1	4	5.5	3.12	4.31	1.19	9	36	4.0
9844	4	2	2	4.5	2.13	3.31	1.19	19	37	7.0
4132	2	2	3	3.5	1.25	2.38	1.12	45	38	2.5
5699	3	2	4	3.5	1.26	2.38	1.12	51	39	3.5
4659	2	1	2	3.0	0.79	1.89	1.11	58	40	0.5
4817	4	2	4	6.5	4.31	5.41	1.09	4	41	4.0
AVERAGE				4.598	1.542	3.070	1.528	26.6	21	2.573

TABLE 11.3: COMPARISON OF BLACKSPOT RANKING CRITERIA FOR THE 32 TOP RANKED LINKS BY LAAT

LREFRNCE ROUTE	/ANODE	LDEV-TYPEA	LCWY-TYPE	LACC-TOT12	LEACC-TOT12	LEB1ACC-TOT12	LPAR	LAAT-RANK	LPAR-RANK	LACC-TOT34
6500	5394	3	1	16.5	1.40	8.95	7.55	1	1	3.5
2229	1572	1	2	7.0	4.66	5.83	1.17	2	22	3.0
298	5209	3	1	6.0	1.07	3.53	2.47	3	3	4.0
819	11762	1	1	6.0	0.81	3.40	2.60	4	2	1.5
1209	1042	1	1	6.0	2.49	4.25	1.75	5	7	4.0
395	5254	3	2	5.5	2.12	3.81	1.69	6	9	10.0
819	9492	2	1	5.5	1.12	3.31	2.19	7	4	4.0
2229	963	1	2	5.5	4.54	5.02	0.48	8	129	3.5
6500	5840	3	1	5.0	0.75	2.88	2.12	9	5	1.0
118	5696	3	1	5.0	2.12	3.56	1.44	10	12	3.0
395	5242	3	2	5.0	2.37	3.68	1.32	11	17	9.5
6500	5860	3	1	4.5	0.56	2.53	1.97	12	6	1.0
395	5254	3	2	4.5	2.48	3.49	1.01	13	31	2.0
2230	1575	1	2	4.5	5.22	4.86	-0.36	14	2328	5.0
2156	1856	1	1	4.5	1.29	2.89	1.61	15	11	3.0
118	14499	3	1	4.0	0.62	2.31	1.69	16	10	1.0
1515	4496	3	1	4.0	1.86	2.93	1.07	17	27	1.5
298	5953	3	1	4.0	1.98	2.99	1.01	18	32	2.5
500	4929	3	1	4.0	3.62	3.81	0.19	19	405	2.5
4035	9777	2	1	4.0	2.53	3.26	0.74	20	64	2.0
483	9014	2	1	4.0	1.54	2.77	1.23	21	20	1.5
2928	5339	2	1	4.0	0.56	2.28	1.72	22	8	2.0
500	5023	3	1	3.5	1.53	2.52	0.98	23	35	0.5
1209	13301	3	1	3.5	1.50	2.50	1.00	24	33	3.5
2928	5411	3	1	3.5	1.90	2.70	0.80	25	54	1.0
1068	4209	2	1	3.5	0.74	2.12	1.38	26	14	0.0
9000	9467	2	1	3.5	0.69	2.10	1.40	27	13	1.0
793	8645	2	1	3.5	2.44	2.97	0.53	28	111	4.0
2336	8043	2	1	3.5	2.56	3.03	0.47	29	135	2.0
2387	3747	2	1	3.5	1.77	2.63	0.87	30	47	4.0
1068	14414	1	2	3.5	2.80	3.15	0.35	31	216	5.0
1067	14413	1	2	3.5	2.94	3.22	0.28	32	260	2.0
AVERAGE				4.813	2.019	3.416	1.397	16.5	127	2.938

TABLE 11.4: COMPARISON OF BLACKSPOT RANKING CRITERIA FOR THE 32 TOP RANKED LINKS BY LPAR

LREFRNCE ROUTE	LREFRNCE ANODE	LDEV-TYPEA	LCWY-TYPE	LACC-TOT12	LEACC-TOT12	LEB1ACC-TOT12	LPAR	LAAT-RANK	LPAR-RANK	LACC-TOT34
6500	5394	3	1	16.5	1.40	8.95	7.55	1	1	3.5
819	11762	1	1	6.0	0.81	3.40	2.60	4	2	1.5
298	5209	3	1	6.0	1.07	3.53	2.47	3	3	4.0
819	9492	2	1	5.5	1.12	3.31	2.19	7	4	4.0
6500	5840	3	1	5.0	0.75	2.88	2.12	9	5	1.0
6500	5860	3	1	4.5	0.56	2.53	1.97	12	6	1.0
1209	1042	1	1	6.0	2.49	4.25	1.75	5	7	4.0
2928	5339	2	1	4.0	0.56	2.28	1.72	22	8	2.0
395	5254	3	2	5.5	2.12	3.81	1.69	6	9	10.0
118	14499	3	1	4.0	0.62	2.31	1.69	16	10	1.0
2156	1856	1	1	4.5	1.29	2.89	1.61	15	11	3.0
118	5696	3	1	5.0	2.12	3.56	1.44	10	12	3.0
9000	9467	2	1	3.5	0.69	2.10	1.40	27	13	1.0
1068	4209	2	1	3.5	0.74	2.12	1.38	26	14	0.0
1520	4896	2	2	3.0	0.25	1.62	1.38	40	15	0.0
819	8672	2	1	3.0	0.36	1.68	1.32	39	16	2.0
395	5242	3	2	5.0	2.37	3.68	1.32	11	17	9.5
6500	12423	1	1	3.0	0.48	1.74	1.26	48	18	0.5
1646	6291	2	1	3.0	0.52	1.76	1.24	37	19	0.5
483	9014	2	1	4.0	1.54	2.77	1.23	21	20	1.5
1068	3962	2	1	2.5	0.13	1.31	1.19	60	21	0.0
2229	1572	1	2	7.0	4.66	5.83	1.17	2	22	3.0
1560	7005	3	1	3.0	0.68	1.84	1.16	33	23	4.5
1336	3427	2	2	2.5	0.19	1.34	1.16	59	24	0.5
197	6604	2	1	3.0	0.72	1.86	1.14	49	25	1.0
793	9628	1	1	3.0	0.76	1.88	1.12	46	26	3.5
1515	4497	3	1	4.0	1.86	2.93	1.07	17	27	1.5
1560	3349	2	1	2.5	0.41	1.46	1.04	76	28	1.5
6500	13451	3	1	2.5	0.42	1.46	1.04	51	29	2.0
118	5700	3	1	3.0	0.96	1.98	1.02	34	30	0.0
395	5254	3	2	4.5	2.48	3.49	1.01	13	31	2.0
298	5953	3	1	4.0	1.98	2.99	1.01	18	32	2.5
AVERAGE				4.438	1.160	2.799	1.639	25.5	16.5	2.344

CHAPTER 12 CONCLUDING REMARKS

This study was based on all accidents recorded on Lothian Region's computerised Accident Network with the exception of those accidents occurring on junctions and links which were the subject of remedial works or significant changes in traffic flow during the period of the study (ie 1979–1982). The data comprised 6,037 accidents on 3,112 junctions and 3,974 accidents on 2,492 links.

The within-site or temporal distribution of the data was shown not to be significantly at odds with a Poisson assumption. On the other hand, the between-site or spatial distribution of the data was found to be over-dispersed and, as such, capable of being modelled by the negative binomial distribution. A number of possible hypotheses to explain the over-dispersion were examined and only that of apparent contagion was found to be justifiable. This finding is fundamentally important because it suggests that the differences between the rates of accident occurrence at sites may vary as a result of inter alia real physical differences between the sites which may be treatable by appropriate engineering works. Had this hypothesis not been justified by the data, the conclusion would have suggested that the traditional treatment of sites identified as blackspots – as carried out by local authorities – may have had no significant effect on the accident frequency.

With an understanding of the distribution of the accident data, statistical models for junction and link accidents were determined for a number of individual accident types. Although these models for individual accident types proved to be of considerable descriptive interest, they did not, in aggregate, produce better models for all accidents than a simple all accident model.

For junction accidents, significant differences were observed between the type of junction, form of traffic control, and location. Interestingly, the accident rate was found to decrease with increasing traffic intensity. Similar results were found for the link accidents where significant differences were observed between the adjacent roadside development type and carriageway type. As with the junction accidents, the accident rate was found to be inversely related to the traffic intensity and more markedly so than for the junction accidents.

The models were applied to both the junction and link data for two non-overlapping time periods at an individual site level and were found to provide a better basis for predicting future accident levels than the observed (prior) accident levels, particularly where empirical Bayesian estimates were used. The importance of the regression-to-mean effect which is modelled by the empirical Bayesian approach is, therefore, evident.

Finally it is suggested that the use of PAR as a blackspot identification criterion would, for Lothian Region data, have offered the potential to determine a more cost-effective road safety works programme than that offered by the AAT criterion and, as it is unlikely that the Lothian Region data are significantly at odds with those for Great Britain (Chapter 3 refers), this conclusion will, in all probability, have a wider application.

REFERENCES

Those references not examined are marked with an asterisk (*) and are included only where they are specifically mentioned in another quoted text.

Abbess C., D. Jarrett and C.C. Wright [1981]. Accidents at blackspots: estimating the effectiveness of remedial treatment, with special reference to the 'regression-to-mean' effect. Traff. Engng. Control, 22(10), 535-542, Printerhall, London.

Adams J.G.U. [1985]. Risk and Freedom: the record of road safety regulation. Transport Publishing Projects, Cardiff.

Andreassen D.C. [1987]. Reply to an article by Maher M. [1987]. Traff. Engng. Control, 28(6), 357-358, Printerhall, London.

Andreassen D.C. and M.M. Hoque [1986]. Intersection accident frequencies. Traff. Engng. Control, 27(10), 514-517, Printerhall.

Arbous A.G. and J.E.Kerrich [1951]. Accident statistics and the concept of accident proneness. Biometrics, 7, 340-432.

Baker R.J. and J.A. Nelder [1978]. Generalised Linear Interactive Modelling Manual. Numerical Algorithms Group, Oxford, UK.

*Bates G.E. and J. Neyman [1952]. Contributions to the theory of accident proneness: I An optimistic model of correlation between light and severe accidents; II True or false contagion. Univ. Calif. Publ. Statist. I., 215-275.

Brilon W. [1972]. Relationship between accident rates and hourly traffic volumes on German highways. Proceedings of a Road Accidents Seminar, PTRC, London.

Brodsky H. and A.S. Hakkert [1983]. Highway accident rates and rural travel densities. Accid. Anal. & Prev., 15(1), 73-84, Pergammon Press, London.

Buckinghamshire County Council [1981]. Road safety statistics and traffic flows 1981, County Engineer's Department, Buckinghamshire County Council.

Chapman R.A. [1969]. Five studies of accident rates. Traff. Engng. Control, 11(5), 224-227, Printerhall, London.

Chapman R.A. [1970]. Numbers of accidents per day and times between accidents, Traff. Engng. Control, 12(2), 82-84, Printerhall, London.

Chapman R.A. [1973]. The concept of exposure. Accid. Anal. & Prev., 5, 95-110, Pergammon Press, London.

Chapman R.G. [1978]. Accidents on urban arterial roads. Laboratory Report 838, TRRL, Crowthorne, UK.

Colgate M.G. and J.C. Tanner [1967]. Accidents at rural three-way junctions. Laboratory Report 87, TRRL, Crowthorne, UK.

Cox D.R. and H.D. Millar [1965]. The Theory of Stochastic Processes. Methuen and Co., London.

Cresswell W.L. and Froggatt P. [1963]. The causation of bus driver accidents. Oxford University Press, London.

Elliott J.M. [1983]. Some Methods for the Statistical Analysis of Examples of Benthic Invertebrates. Scientific Publication No 25, The Freshwater Biological Association, Ambleside, UK.

*Farmer E and E.G. Chambers [1926]. A psychological study of individual differences in accident rate. Rep. Industr. Fatigue Res. Bd. No 38, London.

*Farmer E and E.G. Chambers [1939]. A study of accident proneness amongst motor drivers. Rep. Industr. Fatigue Res. Bd. No 84, London.

Gipps P.G. [1980]. Examining the safety contribution of traffic control devices. Proceedings of the World Conference on Transport Research, Volume 4, 2604-2609, London,

Greenwood M. and H.M. Woods [1919] (Reproduced in Haddon et al [1964]). Incidence of industrial accidents upon individuals with special reference to multiple accidents. Rep. Industr. Fatigue Res. Bd. No 4, London.

Greenwood M. and G.U. Yule [1920]. An Inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. J. Roy. Stat. Soc., 83, 225-279.

Gwynn D.W. [1967]. Relationship of accident rates and accident involvement with hourly traffic volumes. Traffic Quarterly, 21(3), 407-418, Eno Foundation for Highway Traffic Control.

Gwynn D.W. and W.T. Baker [1970] Relationship of accident rates with hourly traffic volumes. Traffic Engineering, 40(5), 42-47.

Haddon W., E.A. Suchman and D. Klein [1964]. Accident Research: methodology and approachs. Harper and Row, New York.

Hakkert A.S. and D. Mahalel [1978]. Estimating the number of accidents at intersections from a knowledge of the traffic flows on the approaches. Accid. Anal. & Prev., 10(1), 69-79, Pergammon Press, London.

Hauer E. [1978]. Traffic conflict surveys: some study design considerations. Supplementary Report 352, TRRL, Crowthorne, UK.

Hauer E. [1980]. Selection for treatment as a source of bias in 'before-and-after' studies. Traff. Engng. Control, 21(8/9), 419-421, Printerhall, London.

Hauer E. [1982]. Traffic conflicts and exposure. Accid. Anal. & Prev., 14(5), 359-364, Pergammon Press, London.

Hauer E. and B. Persaud [1983a]. Common bias in before-and-after accident comparisons and its elimination. Transportation Research Record 905, 164-174, Transportation Research Board, USA.

Hauer E. and B. Persaud [1983b]. Theoretical foundations of a rational procedure for blackspot identification and its application. Proceedings of Seminar K, PTRC 11th Summer Annual Meeting, P240, 231-243, PTRC, London.

Hutchinson T.P. and A.J. Mayne [1977]. The year-to-year variation in the numbers of road accidents. Remedial Measures), University of Salford, UK.

Jorgensen N.O. [1972]. The statistical detection of accident "blackspots". 11th International Study Week in Traffic Engineering and Safety, Montreal, Canada.

Kemp C.D. [1970]. "Accident proneness" and discrete distribution theory. Random Counts in Scientific Work, Volume 2, (Ed Patil G.P.), 42-65, The Pennsyvania State University Press, USA.

Lawson S.D. [1986]. Descriptions and predictions of accidents on urban arterial roads. Traff. Engng. Control, 27(6), 310-319, Printerhall, London.

Leong H.J.W. [1973]. Relationship between accidents and traffic volumes at urban intersections. Australian Road Research Bulletin, 5(3), 72-90.

Lothian Regional Council [1979]. Lothian Region Accident Statistics System Manual. Lothian Regional Council, Edinburgh.

McDonald J.W. [1953]. Relation between number of accidents and traffic volume at divided-highway intersections. Highway Research Board Bulletin, Part 74, 7-17, Highway Research Board, USA.

McGuigan D.R.D. [1980]. The relationship between accidents and traffic at road junctions. M. Sc. Thesis, University of Strathclyde, Glasgow, UK.

McGuigan D.R.D. [1982]. Non-junction accident rates and their use in 'blackspot' identification. Traff. Engng. Control, 23(2), 60-65, Printerhall, London.

McGuigan D.R.D. [1985a]. Accident 'migration': or a flight of fancy? Traff. Engng. Control, 26(4), 229-233, Printerhall, London.

McKenna F.P. [1983]. Accident proneness: a conceptual analysis. Accid. Anal. & Prev., 15(1), 65-71, Pergammon Press, London.

Maher M.J. [1987]. Fitting probability distributions to accident frequency data. Traff. Engng. Control, 28(6), 356-357, Printerhall, London.

Maher M.J. and L.J. Mountain [1988]. The identification of accident blackspots: a comparison of current methods. Accid. Anal & Prev., 20(2), 143-151, Pergammon Press, London.

Maltby D. and R.W.G. Bennett [1986]. Variations in accident frequencies on rural single carriageway roads in England. Proceedings of Seminar R, PTRC 14th Summer Meeting, PTRC, London.

Maritz J.S. [1950]. On the validity of inferences from the fitting of Poisson and negative binomial distributions to observed accident data. Psychology Bulletin, 47, 434-443.

Maycock G. and R.D. Hall [1984]. Accidents at 4-arm roundabouts. Laboratory Report 1120, TRRL, Crowthorne, UK.

Moran P.A.P. [1968]. An Introduction to Probability Theory. Clarendon Press, Oxford, UK.

Nie N.H. [1983]. SPSSX User's Guide. SPSS Inc., Chicago.

Nicholson A.J. [1985]. The variability of accident counts. Accid. Anal. & Prev., 17(1), 47-56, Pergammon Press, London.

Palmer, P.O. [1979]. The effect of biorythms on road accidents. Supplementary Report 535, TRRL, Crowthorne, UK.

Pfundt, K. [1969]. Three difficulties in the comparison of accident rates. Accid. Anal. & Prev., 1, 253-259, Pergammon Press, London.

Phillips, G. [1979]. Accuracy of annual traffic flow estimates from short period counts. Supplementary Report 514, TRRL, Crowthorne, UK.

Pickering D., R.D. Hall and M. Grimmer [1986]. Accidents at rural T-junctions. Research Report 65, TRRL, Crowthorne, UK.

Proctor S. [1985]. Accidents at central reservation gaps in the West Midlands. Paper presented to the UTSG Conference, Birmingham University, UK.

Quimby A.R., G. Maycock, I.D. Carter, R. Dixon and J.G. Wall [1986]. Perceptual abilities of accident involved drivers. Research Report 27, TRRL, Crowthorne, UK.

Robinson B.N., G.D. Anderson, E.Cohen, W.F. Gazdzik, L.C. Karpel, A.H. Miller and J.R. Stein [1980]. Scientific Information Retrieval User's Manual Version 2. Sir Inc, Evanston, IL, USA.

Ross G.J.S. [1980]. Maximum Likelihood Program. Rothamsted Experimental Station, Harpenden, UK.

Satterthwaite S.P. [1976]. Monthly variations in road accident frequency. Traff. Engng. Control, 17(1), 18-20, Printerhall, London.

Satterthwaite S.P. [1981]. A survey of research into relationships between traffic accidents and traffic volumes. Supplementary Report 692, TRRL, Crowthorne, UK.

Scott P.P. [1980]. The relationship between road lighting quality and accident frequency. Laboratory Report 929, TRRL, Crowthorne, UK.

Silcock D.T. and A.W. Smyth [1984]. The methods used by British highway authorities to identify accident blackspots. Traff. Engng. Control, 25(11), 542-545, Printerhall, London.

Smeed R.J. [1949]. Some statistical aspects of road safety research. J Roy Statist Soc A 112, 1-34.

Sokal R.R and F.J. Rohlf [1969]. Biometry - The Principles and Practice of Statistics in Biological Research. W.H. Freeman and Co., San Francisco.

Tanner J.C. [1953]. Accidents at rural three-way junctions. The Journal of The Institution of Highway Engineers, 2(11), 56-67, The Institution of Highway Engineers, London.

Transport, Department of [1986]. Accident Investigation Manual. Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1977a]. Stats 20 Instructions for the Completion of Road Accident Reports, Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1977b]. Stats 21 Procedures for Submitting Road Accident Data to the Department of Transport, Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1980]. Transport Statistics Great Britain 1969-1979. Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1981a]. Transport Statistics Great Britain 1970-1980. Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1981b]. Road Accidents Great Britain 1979. Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1981c]. Road Accidents Great Britain 1980. Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1982a]. Transport Statistics Great Britain 1971-1981. Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1982b]. Road Accidents Great Britain 1981. Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1983a]. Transport Statistics Great Britain 1972-1982. Department of Transport.

Transport, Department of, Scottish Development Department, Welsh Office [1983b]. Road Accidents Great Britain 1982. Department of Transport.

Turner D.J. and R. Thomas [1986]. Motorway accidents: an examination of accident totals, rates and severity and their relationships with traffic flow. Traff. Engng. Control, 27(7/8), 377-383 and 387, Printerhall, London.

Van Maren P.A. [1980]. Correlation of design and control characteristics with accidents at rural multi-lane highway intersections. Report FHWA/IN/JHRP-77/22, Joint Highway Research Project, Purdue University, USA.

Yu J.C. [1972]. Establishing relationship of level of service and highway safety. Traffic Engineering, 43(1), 50-52.

Zahavi Y., The problem of accident distribution. Traffic Quarterly, 16, 540-548, Eno Foundation for Transportation.

APPENDICES

I. GLOSSARY OF THE VARIABLE NAMES USED

I.I. Junction Data

JAAT	Annual average accident total [Section 11.3].
JAATRANK	Rank value for JAAT [11.3].
JACCTOT	All accident total for the whole study period 1979-1982 [2.6.2].
JACCTOT1	All accident total for the year 1979 [8.6].
JACCTOT2	All accident total for the year 1980 [8.6].
JACCTOT3	All accident total for the year 1981 [8.6].
JACCTOT4	All accident total for the year 1982 [8.6].
JACCTOT12	All accident total for the two-year period 1979-1980 [6.3].
JACCTOT34	All accident total for the two-year period 1981-1982 [6.3].
JCELLTOT	Number of junctions per cell in contingency table [8.2].
JCONTROL	Form of junction control [2.6.2].
JEACCTOT	Expected (modelled) accident total for the whole study period based on JACCTOT data [8.5].
JEACCTOT12	Expected (modelled) two-year accident totals based on JACCTOT12 data [10.2].
JEB1ACCTOT12	Empirical Bayesian expected two-year accident total based on JACCTOT12 data (Type 1) [10.2].
JEB2ACCTOT12	Empirical Bayesian expected two-year accident total based on JACCTOT12 data (Type 2) [10.2].
JEB3ACCTOT12	Empirical Bayesian expected two-year accident total based on JACCTOT12 data (Type 3) [10.2].
JERACCTOT	Expected (modelled) re-aggregated accident total based on individual accident type data for the whole study period 1979-1982 [8.4.2].
JEXPOS72	Exposure function $JMAINFLOW^{0.7}.JSIDEFLOW^{0.2}$ [8.2].
JEXPOS535	Exposure function $JMAINFOLW^{0.5}.JSIDEFLOW^{0.35}$ [8.2].
JEXPOS815	Exposure function $JMAINFLOW^{0.8}.JSIDEFLOW^{0.15}$ [8.2].
JFLOWRATIO	Ratio of JMAINFLOW to JSIDEFLOW [8.4.1].

JLOCATION	Junction location [2.6.2].
JMAINBAND	Band value for JMAINFLOW [2.6.4].
JMAINFLOW	Main road flow expressed in millions of vehivles per year [2.6.4].
JMNVP	Multi-vehicle non-pedestrian accident total for the whole study period 1979-1982 [2.6.2].
JPAR	Potential accident reduction [11.3].
JPARRANK	Rank value for JPAR [11.3].
JPED	Pedestrian accident total for the whole study period 1979-1982 [2.6.2].
JPRODPERVEH	Ratio of JPRODUCT to JTHROUGHPUT [8.4.1].
JPRODUCT	Product of JMAINFLOW and JSIDEFLOW [2.6.4].
JPSV	Public Service Vehicle passenger accident total for the whole study period 1979-1982 [2.6.2].
JREFRNCE	Node number of junction [2.6.2].
JRES1	Residual value: JACCTOT34-JACCTOT12 [10.2].
JRES2	Residual value: JACCTOT34-JEACCTOT12 [10.2].
JRESB1	Residual value: JACCTOT34-JEB1ACCTOT12 [10.2].
JRESB2	Residual value: JACCTOT34-JEB2ACCTOT12 [10.2].
JRESB3	Residual value: JACCTOT34-JEB3ACCTOT12 [10.2].
JROOTBAND	Band value for JROOTPROD [2.6.4].
JROOTPROD	Square root of JPRODUCT [2.6.4].
JSIDEBAND	Band value for JSIDEFLOW [2.6.4].
JSIDEFLOW	Side road flow expressed in millions of vehicles per year [2.6.4].
JSVNP	Single vehicle non-pedestrian accident total for the whole study period 1979-1982 [2.6.2].
JTHROUGHPUT	Number of vehicles negotiating the junction expressed as millions of vehicles per year (ie JMAINFLOW+JSIDEFLOW) [8.4.1].
JTINT	Traffic intensity index [8.4.1].
JTINTSQ	The square of the traffic intensity index (ie $JTINT^2$) [8.4.1].

JTVNP1	Two vehicle non-pedestrian accident total (Type 1) for the whole study period 1979-1982 [2.6.2].
JTVNP2	Two vehicle non-pedestrian accident total (Type 2) for the whole study period 1979-1982 [2.6.2].
JTVNP3	Two vehicle non-pedestrian accident total (Type 3) for the whole study period 1979-1982 [2.6.2].
JTVNP4	Two vehicle non-pedestrian accident total (Type 4) for the whole study period 1979-1982 [2.6.2].
JTVNP5	Two vehicle non-pedestrian accident total (Type 5) for the whole study period 1979-1982 [2.6.2].
JTVNP6	Two vehicle non-pedestrian accident total (Type 6) for the whole study period 1979-1982 [2.6.2].
JTVNP7	Two vehicle non-pedestrian accident total (Type 7) for the whole study period 1979-1982 [2.6.2].
JTVNP8	Two vehicle non-pedestrian accident total (Type 8) for the whole study period 1979-1982 [2.6.2].
JTVNP9	Two-vehicle non-pedestrian accident total (Type 9) for the whole study period 1979-1982 [2.6.2].
JTVNPA1	Aggregated two vehicle non-pedestrian accident total (Type 1) for the whole study period 1979-1982 (=JTVNP1+JTVNP2+JTVNP3+JTVNP4+JTVNP8) [8.2].
JTVNPA2	Aggregated two vehicle non-pedestrian accident totals (Type 2) for the whole study period 1979-1982 (=JTVNP5+JTVNP6+JTVNP7+JTVNP9) [8.2].
JUNCTYPE	Type of junction [2.6.2].
JUNCTYPEA	Aggregated type of junction [8.2].

I.II. Link Data

LAAT	Annual average accident total [Section 11.3].
LAATRANK	Rank value for LAAT [11.3].
LACCTOT	All accident total for the whole study period 1979-1982 [2.6.3].
LACCTOT1	All accident total for the year 1979 [9.5].
LACCTOT2	All accident total for the year 1980 [9.5].
LACCTOT3	All accident total for the year 1981 [9.5].
LACCTOT4	All accident total for the year 1982 [9.5].
LACCTOT12	All accident total for the two-year period 1979-1980 [6.3].
LACCTOT34	All accident total for the two-year period 1981-1982 [6.3].
LCELLTOT	Number of links per cell in contingency table [9.2].
LCWYTYPE	Carriageway type [2.6.3].
LDEVTYPE	Adjacent roadside development type [2.6.3].
LDEVTYPEA	Aggregated adjacent roadside development type [2.6.4].
LEACCTOT	Expected (modelled) accident total for the whole study period based on LACCTOT data [9.4].
LEACCTOT12	Expected (modelled) two-year accident totals based on LACCTOT12 data [10.2].
LEB1ACCTOT12	Empirical Bayesian expected two-year accident total based on LACCTOT12 data (Type 1) [10.2].
LEB2ACCTOT12	Empirical Bayesian expected two-year accident total based on LACCTOT12 data (Type 2) [10.2].
LEB3ACCTOT12	Empirical Bayesian expected two-year accident total based on LACCTOT12 data (Type 3) [10.2].
LERACCTOT	Expected (modelled) re-aggregated accident total based on individual accident type data for the whole study period 1979-1982 [9.4].
LENGTHBAND	Band value for LINKLENGTH [2.6.4].
LFLOWBAND	Band value for LINKFLOW [2.6.4].
LINKFLOW	Annual average traffic flow expressed in millions of vehicles per year [2.6.4].

LINKLENGTH — Length of link in metres [2.6.3].

LMVKM — Travel distance on link expressed in million vehicle kilometres (ie LINKFLOW x LINKLENGTH/1000) [2.6.4].

LMVKMBAND — Band value for LMVKM [2.6.4].

LMVNP — Multi-vehicle non-pedestrian accident total for the whole study period 1979-1982 [2.6.3].

LPAR — Potential accident reduction [11.3].

LPARRANK — Rank value for LPAR [11.3].

LPED — Pedestrian accident total for the whole study period 1979-1982 [2.6.3].

LPSV — Public Service Vehicle passenger accident total for the whole study period 1979-1982 [2.6.3].

LREFRNCE — ROUTE and ANODE numbers for link [2.6.3].

LRES1 — Residual value: LACCTOT34-LACCTOT12 [10.2].

LRES2 — Residual value: LACCTOT34-LEACCTOT12 [10.2].

LRESB1 — Residual value: LACCTOT34-LEB1ACCTOT12 [10.2].

LRESB2 — Residual value: LACCTOT34-LEB2ACCTOT12 [10.2].

LRESB3 — Residual value: LACCTOT34-LEB3ACCTOT12 [10.2].

LSVNP — Single vehicle non-pedestrian accident total for the whole study period 1979-1982 [2.6.3].

LTINT — Traffic intensity index [9.3].

LTINTSQ — The square of the traffic intensity index (ie $LTINT^2$) [9.3].

LTVNP1 — Two vehicle non-pedestrian accident total (Type 1) for the whole study period 1979-1982 [2.6.3].

LTVNP2 — Two vehicle non-pedestrian accident total (Type 2) for the whole study period 1979-1982 [2.6.3].

LTVNP3 — Two vehicle non-pedestrian accident total (Type 3) for the whole study period 1979-1982 [2.6.3].

LTVNP4 — Two vehicle non-pedestrian accident total (Type 4) for the whole study period 1979-1982 [2.6.3].

LTVNPA1 — Aggregated two vehicle non-pedestrian accident total (Type 1) for the whole study period 1979-1982 (=LTVNP1) [9.2].

LTVNPA2	Aggregated two vehicle non-pedestrian accident total (Type 2) for the whole study period 1979-1982 (=LTVNP2) [9.2].
LTVNPA3	Aggregated two vehicle non-pedestrian accident total (Type 3) for the whole study period 1979-1982 (=LTVNP3+LTVNP4) [9.2].

II. JUNCTION CONTINGENCY TABLE DATA IN GLIM FORMAT

```
$C FILENAME IS (BACKUPDATA_)JC26
$UNITS 184
$DATA J C L MB SB N S T1 T2 M PED PSV Y1 Y2 Y3 Y4 FM FS PROD ROOT
      E72 E535 E815 IPPV
$FACTOR J 5 L 4 C 2 MB 3 SB 3
$C JUNCTION DATA FOR CONTINGENCY TABLE ANALYSIS:
```

TRUNCATED VARIABLE NAME	FULL VARIABLE NAME	DESCRIPTION
J	JUNCTYPE	junction type (5 factor levels)
C	JCONTROL	junction control (2)
L	JLOCATION	junction location (4)
MB	JMAINBAND	main road flow band (3)
SB	JSIDEBAND	side road flow band (3)
N	JCELLTOT	number of junctions per cell
S	JSVNP	single veh non-ped accidents
T1	JTVNPA1	two veh non-ped accidents (type 1)
T2	JTVNPA2	two veh non-ped accidents (type 2)
M	JMVNP	multi veh non-ped accidents
PED	JPED	pedestrian accidents
PSV	JPSV	psv accidents
Y1	JACCTOT1	total accidents for 1979
Y2	JACCTOT2	total accidents for 1980
Y3	JACCTOT3	total accidents for 1981
Y4	JACCTOT4	total accidents for 1982
FM	JMAINFLOW	main road flow (MV/year) per cell
FS	JSIDEFLOW	main road flow (MV/year) per cell
PROD	JPRODUCT	JMAINFLOW.JSIDEFLOW per cell
ROOT	JROOTPROD	square root of JPRODUCT per cell
E72	JEXPOS72	exposure JMAINFLOW**.7*JSIDEFLOW**.2
E535	JEXPOS535	exposure JMAINFLOW**.5*JSIDEFLOW**.35
E815	JEXPOS815	exposure JMAINFLOW**.8*JSIDEFLOW**.15
IPPV	JPRODPERVEH	intensity measure=JROOTPROD/JTHROUGHPUT

FACTOR LEVELS ARE:

JUNCTYPE

1=ROUNDABOUT
2=T- and Y- JUNCTIONS
3=OTHER 3-WAY JUNCTIONS
4=CROSSROADS
5=OTHER 4-WAY JUNCTIONS

JCONTROL

1=PRIORITY CONTROL
2=SIGNAL CONTROL

JLOCATION

1=RURAL
2=SUBURBAN
3=LOCAL CENTRE
4=CITY CENTRE

JMAINBAND

1=LOW MAIN ROAD FLOW BAND
2=MEDIUM MAIN ROAD FLOW BAND
3=HIGH MAIN ROAD FLOW BAND

JSIDEBAND

1=LOW SIDE ROAD FLOW BAND
2=MEDIUM SIDE ROAD FLOW BAND
3=HIGH SIDE ROAD FLOW BAND

$READ

1 1 1 1 1	6	0	1	0	0	0	0	1	0	0	0	0.776	0.017	0.013	0.114	0.370	0.211	0.443	0.017
1 1 1 1 2	4	0	0	0	0	0	0	0	0	0	0	1.417	0.083	0.124	0.341	0.771	0.494	0.906	0.078
1 1 1 1 3	47	10	3	1	0	0	0	4	1	4	5	1.226	0.506	0.636	0.762	0.983	0.845	1.044	0.340
1 1 1 2 3	18	3	3	1	1	1	0	4	2	1	2	2.454	0.667	1.690	1.202	1.656	1.280	1.865	0.488
1 1 1 3 2	1	0	0	0	0	0	0	0	0	0	0	7.575	0.084	0.636	0.797	2.514	1.157	3.485	0.083
1 1 1 3 3	7	5	10	2	0	0	1	3	4	5	6	5.958	2.004	11.78	3.401	3.967	3.070	4.589	1.468
1 1 2 1 1	2	0	0	0	0	0	0	0	0	0	0	1.386	0.008	0.011	0.103	0.475	0.215	0.626	0.008
1 1 2 1 2	18	4	1	0	1	2	2	3	2	1	4	0.867	0.054	0.047	0.200	0.477	0.312	0.552	0.048
1 1 2 1 3	38	11	4	4	0	2	2	8	5	6	4	1.227	0.594	0.809	0.817	1.005	0.882	1.061	0.374
1 1 2 2 2	2	1	0	0	0	0	0	0	0	0	1	2.176	0.084	0.183	0.427	1.050	0.620	1.284	0.081
1 1 2 2 3	33	8	3	4	3	3	5	4	8	8	6	3.020	0.883	2.635	1.496	1.993	1.534	2.266	0.611
1 1 2 3 3	59	33	26	14	5	8	4	26	16	24	24	8.317	2.302	21.92	4.049	4.886	3.553	5.877	1.643
1 1 3 1 2	4	1	2	5	0	2	1	5	3	2	1	1.679	0.122	0.202	0.448	0.939	0.616	1.100	0.113
1 1 3 1 3	7	2	1	1	0	2	4	2	1	3	4	1.373	0.537	0.715	0.818	1.067	0.903	1.143	0.360
1 1 3 2 3	3	0	0	0	0	0	1	0	1	0	0	2.192	0.602	1.322	1.147	1.564	1.238	1.735	0.472
1 1 3 3 3	2	1	0	0	0	1	0	0	1	0	1	4.498	3.666	16.51	4.060	3.715	3.341	4.046	2.019
1 1 4 2 3	2	1	5	0	0	1	0	0	2	3	2	3.947	2.446	9.657	3.106	3.127	2.717	3.430	1.510
2 1 1 1 1	127	3	14	9	9	0	2	10	9	8	10	1.394	0.010	0.014	0.112	0.487	0.226	0.639	0.010
2 1 1 1 2	30	1	5	4	0	2	0	7	1	2	2	1.270	0.057	0.075	0.256	0.643	0.394	0.766	0.054
2 1 1 1 3	28	9	11	9	4	0	0	4	12	9	8	1.405	0.382	0.577	0.690	1.005	0.799	1.101	0.276
2 1 1 2 1	38	2	7	3	0	1	1	2	5	2	5	2.701	0.010	0.028	0.163	0.792	0.325	1.102	0.010
2 1 1 2 2	12	1	3	0	1	4	1	1	1	4	4	2.586	0.060	0.154	0.382	1.086	0.585	1.380	0.059
2 1 1 2 3	9	0	4	6	2	1	0	1	3	3	6	2.654	0.709	1.785	1.236	1.726	1.315	1.962	0.493
2 1 1 3 1	22	5	5	3	3	0	0	4	1	3	8	4.527	0.010	0.046	0.211	1.137	0.421	1.667	0.010
2 1 1 3 2	7	0	4	2	0	1	0	1	2	1	3	4.803	0.048	0.224	0.465	1.602	0.737	2.190	0.048
2 1 1 3 3	10	6	11	15	4	1	0	8	6	12	11	4.582	1.300	5.715	2.162	2.811	2.096	3.282	0.884
2 1 2 1 1	116	5	8	8	5	12	1	6	13	11	9	1.148	0.014	0.015	0.119	0.452	0.229	0.572	0.013
2 1 2 1 2	157	19	19	22	5	35	2	25	36	25	16	1.261	0.053	0.067	0.245	0.630	0.382	0.753	0.049
2 1 2 1 3	57	7	18	14	2	24	5	10	19	23	18	1.459	0.318	0.479	0.645	0.999	0.768	1.107	0.243
2 1 2 2 1	155	9	24	12	10	38	10	26	30	26	21	2.857	0.012	0.034	0.180	0.846	0.351	1.177	0.012
2 1 2 2 2	249	42	73	48	16	81	17	56	67	77	77	2.851	0.054	0.153	0.379	1.136	0.590	1.466	0.053
2 1 2 2 3	103	32	40	75	15	66	13	60	71	59	51	2.931	0.648	1.875	1.210	1.791	1.307	2.073	0.451
2 1 2 3 1	114	31	25	24	19	45	8	35	42	33	42	6.070	0.012	0.074	0.264	1.431	0.512	2.146	0.012
2 1 2 3 2	190	45	67	75	30	122	25	90	93	93	88	5.699	0.055	0.323	0.544	1.854	0.842	2.563	0.055
2 1 2 3 3	65	21	45	57	15	39	6	39	43	49	52	6.430	0.642	3.970	1.696	3.010	1.850	3.795	0.498
2 1 3 1 1	6	1	0	0	0	1	1	0	0	1	2	1.543	0.016	0.026	0.157	0.591	0.292	0.760	0.016
2 1 3 1 2	11	2	1	0	0	16	1	3	5	7	5	1.281	0.052	0.071	0.247	0.636	0.384	0.763	0.049
2 1 3 1 3	12	0	2	7	1	2	0	3	6	3	0	1.569	0.431	0.662	0.743	1.086	0.853	1.201	0.294

2 1 3 2 1	14	0	1	2	0	3	0	1	2	2	1	2.841	0.012	0.033	0.178	0.840	0.348	1.169	0.012
2 1 3 2 2	28	6	6	6	10	30	3	14	14	10	23	2.895	0.066	0.195	0.429	1.207	0.646	1.542	0.064
2 1 3 2 3	11	0	3	1	0	17	1	5	4	5	8	2.901	0.406	1.208	1.022	1.695	1.176	1.988	0.334
2 1 3 3 1	48	5	10	7	7	62	24	31	26	27	31	5.127	0.014	0.074	0.267	1.327	0.505	1.938	0.014
2 1 3 3 2	63	16	43	25	13	101	16	48	53	46	67	6.226	0.056	0.343	0.561	1.947	0.868	2.721	0.055
2 1 3 3 3	35	5	25	22	13	47	15	33	37	29	28	7.192	0.540	3.582	1.745	3.246	1.940	4.144	0.469
2 1 4 1 2	1	1	2	1	0	1	0	2	1	2	0	1.241	0.084	0.104	0.322	0.709	0.468	0.820	0.079
2 1 4 1 3	1	1	0	0	0	0	0	0	1	0	0	1.241	0.168	0.208	0.456	0.814	0.597	0.910	0.148
2 1 4 2 1	4	0	2	0	0	4	0	0	1	5	0	3.107	0.008	0.025	0.156	0.837	0.323	1.195	0.008
2 1 4 2 2	6	0	3	0	0	4	0	2	1	1	3	2.730	0.076	0.210	0.449	1.195	0.662	1.506	0.073
2 1 4 2 3	11	2	4	4	1	9	2	3	6	6	7	2.993	0.290	0.899	0.896	1.638	1.081	1.956	0.258
2 1 4 3 1	14	5	1	1	3	9	0	3	7	4	5	5.690	0.014	0.077	0.272	1.408	0.519	2.083	0.014
2 1 4 3 2	18	9	13	8	7	22	4	16	14	13	20	5.554	0.071	0.403	0.616	1.928	0.916	2.621	0.070
2 1 4 3 3	41	17	29	39	9	68	32	49	55	47	43	6.316	0.880	5.609	2.099	3.278	2.157	4.030	0.697
2 2 2 1 2	1	0	2	0	0	1	0	1	2	0	0	1.993	0.084	0.167	0.409	0.987	0.593	1.197	0.081
2 2 2 2 2	1	0	0	0	0	0	0	0	0	0	0	3.161	0.084	0.266	0.515	1.364	0.747	1.732	0.082
2 2 2 2 3	2	1	0	0	0	0	0	0	1	0	0	3.120	0.444	1.383	1.167	1.876	1.320	2.192	0.386
2 2 2 3 3	8	5	11	25	9	13	2	8	22	13	22	7.975	2.404	21.19	4.163	4.896	3.629	5.824	1.729
2 2 3 2 3	1	0	0	1	0	2	1	1	0	1	2	3.461	2.566	8.881	2.980	2.879	2.587	3.110	1.474
2 2 3 3 3	10	4	15	13	3	31	14	29	20	16	15	6.576	2.626	18.25	3.996	4.403	3.452	5.099	1.770
2 2 4 2 3	1	1	0	0	0	7	0	1	4	2	1	3.619	0.168	0.608	0.779	1.722	1.019	2.141	0.161
2 2 4 3 3	7	3	7	9	1	12	3	11	7	7	10	6.614	3.282	21.91	4.612	4.726	3.860	5.388	2.156
3 1 1 1 1	16	1	3	2	0	0	0	1	3	1	1	1.105	0.012	0.013	0.107	0.422	0.210	0.538	0.012
3 1 1 1 2	27	3	0	4	0	0	0	1	2	3	1	1.364	0.066	0.087	0.285	0.697	0.431	0.830	0.062
3 1 1 1 3	54	15	11	2	2	4	0	5	15	9	5	1.526	0.497	0.786	0.830	1.131	0.924	1.230	0.350
3 1 1 2 1	6	2	0	0	0	0	0	0	1	0	1	2.613	0.011	0.028	0.166	0.783	0.328	1.083	0.011
3 1 1 2 2	9	1	0	2	2	0	0	0	1	2	2	2.761	0.073	0.201	0.440	1.190	0.653	1.505	0.071
3 1 1 2 3	29	7	7	4	0	2	0	6	6	3	5	2.519	0.757	1.999	1.325	1.755	1.383	1.963	0.552
3 1 1 3 1	5	0	2	0	0	0	0	0	0	2	0	6.135	0.016	0.095	0.304	1.534	0.570	2.269	0.016
3 1 1 3 2	6	1	1	0	0	0	0	0	0	2	0	5.746	0.067	0.371	0.602	1.941	0.907	2.656	0.066
3 1 1 3 3	14	5	6	3	2	0	0	4	2	3	7	5.241	0.613	3.111	1.711	2.806	1.852	3.414	0.529
3 1 2 1 1	16	4	1	2	0	0	0	2	0	2	3	1.013	0.013	0.014	0.110	0.410	0.212	0.514	0.013
3 1 2 1 2	40	5	5	3	0	8	6	4	11	6	6	0.843	0.061	0.053	0.210	0.478	0.319	0.548	0.054
3 1 2 1 3	24	2	3	10	0	4	3	5	3	9	5	1.221	0.379	0.493	0.644	0.910	0.745	0.983	0.269
3 1 2 2 1	10	0	1	1	1	0	0	0	3	0	0	2.725	0.015	0.041	0.198	0.860	0.373	1.177	0.015
3 1 2 2 2	28	5	6	1	1	3	0	5	1	5	5	2.602	0.054	0.144	0.364	1.069	0.566	1.367	0.053
3 1 2 2 3	24	3	3	17	3	2	2	9	8	8	5	3.264	0.681	2.261	1.377	2.010	1.465	2.332	0.516
3 1 2 3 1	14	5	12	9	2	7	1	10	11	11	4	5.853	0.012	0.072	0.260	1.402	0.505	2.094	0.012
3 1 2 3 2	21	0	10	6	3	10	7	10	12	9	5	7.414	0.062	0.467	0.646	2.253	0.983	3.187	0.061
3 1 2 3 3	27	5	8	16	6	7	1	10	11	7	15	7.023	1.052	7.236	2.378	3.608	2.382	4.455	0.811
3 1 3 1 3	2	0	0	0	0	0	0	0	0	0	0	1.260	0.334	0.364	0.598	0.897	0.715	0.979	0.238
3 1 3 2 1	2	0	0	0	0	2	1	0	0	1	2	3.170	0.013	0.040	0.197	0.924	0.380	1.293	0.013
3 1 3 2 2	9	1	1	3	1	4	4	4	4	3	3	3.120	0.056	0.170	0.403	1.218	0.625	1.583	0.055
3 1 3 2 3	10	3	5	3	1	12	7	8	8	4	11	3.112	0.740	2.307	1.354	1.937	1.431	2.237	0.516
3 1 3 3 1	2	0	0	0	1	0	0	0	0	0	1	5.966	0.013	0.081	0.275	1.454	0.528	2.164	0.013
3 1 3 3 2	4	2	2	2	0	1	0	2	3	1	1	5.799	0.071	0.404	0.628	1.987	0.936	2.711	0.071
3 1 3 3 3	4	1	4	5	1	2	2	4	7	3	1	4.625	0.503	2.350	1.373	2.382	1.537	2.913	0.419
3 1 4 1 2	1	0	0	0	0	0	0	0	0	0	0	1.408	0.105	0.148	0.384	0.810	0.539	0.938	0.098
3 1 4 1 3	1	0	0	0	0	0	0	0	0	0	0	1.697	0.867	1.471	1.212	1.407	1.239	1.494	0.574
3 1 4 2 2	4	2	1	2	0	6	1	6	1	1	4	3.581	0.089	0.318	0.563	1.503	0.810	1.928	0.087
3 1 4 2 3	16	6	8	7	3	13	8	10	10	16	9	3.130	0.972	2.967	1.632	2.114	1.649	2.393	0.682
3 1 4 3 1	4	0	0	1	0	1	0	0	0	0	2	7.043	0.013	0.097	0.299	1.630	0.573	2.468	0.013
3 1 4 3 2	3	0	3	3	0	2	0	1	1	2	4	5.168	0.051	0.283	0.506	1.724	0.790	2.365	0.050
3 1 4 3 3	15	13	7	3	2	13	7	12	15	6	12	6.318	1.462	8.888	2.714	3.620	2.583	4.334	1.067
3 2 2 2 1	2	0	0	0	0	0	0	0	0	0	0	3.161	0.017	0.054	0.231	0.991	0.427	1.363	0.017
3 2 2 2 2	2	0	1	0	0	0	0	0	0	0	1	2.409	0.059	0.139	0.365	1.031	0.561	1.301	0.057
3 2 2 2 3	8	0	0	1	0	0	0	0	1	0	0	2.906	0.789	2.554	1.406	1.913	1.457	2.178	0.564
3 2 2 3 2	5	0	3	1	0	1	1	2	0	1	3	4.768	0.045	0.210	0.456	1.590	0.728	2.175	0.044
3 2 2 3 3	11	8	5	7	2	5	1	5	9	6	8	6.724	2.457	16.76	3.942	4.443	3.447	5.165	1.721
3 2 3 1 3	1	0	0	1	0	0	0	1	0	0	0	1.681	0.487	0.819	0.904	1.246	1.008	1.360	0.378
3 2 3 2 2	1	0	0	0	0	0	0	0	0	0	0	2.373	0.142	0.337	0.580	1.239	0.778	1.490	0.134
3 2 3 2 3	4	1	1	1	0	2	3	4	1	1	2	2.710	0.852	2.547	1.375	1.819	1.413	2.055	0.565
3 2 3 3 3	8	1	3	1	0	6	1	4	2	4	2	5.844	0.899	5.067	2.083	3.161	2.134	3.840	0.718
3 2 4 1 3	1	0	0	0	0	1	0	0	0	1	0	1.575	0.867	1.366	1.168	1.336	1.194	1.408	0.559
3 2 4 2 3	14	2	4	0	2	30	9	11	7	19	10	3.445	0.967	3.403	1.730	2.277	1.744	2.600	0.708
3 2 4 3 3	24	13	12	14	4	55	24	20	35	34	33	5.931	2.189	13.39	3.456	3.941	3.072	4.556	1.516
4 1 1 1 1	14	0	7	6	1	1	0	3	6	2	4	1.322	0.011	0.014	0.113	0.475	0.225	0.617	0.011
4 1 1 1 2	7	1	2	3	0	0	0	3	1	0	2	1.072	0.075	0.086	0.272	0.606	0.400	0.700	0.068
4 1 1 1 3	5	0	2	18	3	0	0	6	5	5	7	1.314	0.588	0.888	0.868	1.072	0.932	1.137	0.398
4 1 1 2 1	2	0	0	0	0	0	0	0	0	0	0	2.746	0.010	0.029	0.168	0.813	0.335	1.130	0.010
4 1 1 2 2	5	0	7	8	3	0	0	4	4	7	3	2.694	0.080	0.215	0.455	1.193	0.667	1.498	0.078

41123	3	2	4	5	0	2	0	4	5	1	3	2.780	0.503	1.522	1.147	1.742	1.268	2.008	0.416
41131	2	1	0	0	0	0	0	0	1	0	0	4.499	0.008	0.036	0.189	1.091	0.391	1.614	0.008
41133	1	0	1	3	0	0	0	2	0	2	0	4.499	0.166	0.747	0.864	2.001	1.131	2.544	0.160
41211	5	0	0	1	0	0	0	0	1	0	0	0.973	0.016	0.016	0.117	0.408	0.217	0.507	0.015
41212	22	6	5	10	3	14	1	13	12	6	8	1.377	0.054	0.072	0.259	0.676	0.405	0.813	0.051
41213	9	4	1	16	1	9	3	12	10	7	5	1.423	0.514	0.790	0.786	1.054	0.870	1.144	0.336
41221	2	2	1	0	0	1	0	2	1	1	0	2.395	0.013	0.031	0.175	0.771	0.337	1.047	0.013
41222	33	5	12	22	7	22	1	15	19	24	11	2.954	0.060	0.182	0.407	1.189	0.622	1.532	0.059
41223	13	6	12	31	8	11	3	18	21	19	13	2.658	0.699	1.806	1.272	1.762	1.350	1.996	0.510
41231	6	2	1	2	1	10	0	4	7	2	3	5.193	0.018	0.097	0.305	1.413	0.556	2.039	0.018
41232	30	6	14	39	5	22	3	28	23	21	17	5.767	0.055	0.320	0.548	1.875	0.849	2.593	0.054
41233	7	3	6	26	6	4	2	13	10	8	16	5.915	0.289	1.689	1.269	2.653	1.533	3.386	0.272
41311	2	0	0	1	1	2	0	0	1	3	0	1.219	0.017	0.020	0.141	0.504	0.262	0.631	0.017
41312	2	0	0	1	1	0	0	2	0	0	0	1.744	0.046	0.078	0.278	0.788	0.443	0.974	0.045
41313	4	0	0	5	0	1	0	2	1	3	0	1.955	0.464	0.900	0.904	1.328	1.021	1.486	0.350
41321	2	1	0	0	0	1	0	1	0	1	0	3.794	0.014	0.056	0.228	1.074	0.433	1.521	0.014
41322	3	0	1	2	1	1	0	0	2	1	2	3.250	0.095	0.309	0.549	1.414	0.782	1.792	0.092
41323	1	1	0	1	1	3	0	3	1	2	0	3.553	1.108	3.937	1.984	2.479	1.954	2.800	0.845
41331	2	0	1	0	0	10	5	4	2	7	3	4.762	0.018	0.086	0.292	1.335	0.535	1.908	0.018
41332	3	0	1	1	0	5	1	3	1	4	0	5.780	0.065	0.385	0.610	1.969	0.919	2.694	0.064
41413	1	1	0	0	0	1	0	1	0	1	0	1.687	0.412	0.695	0.833	1.208	0.952	1.330	0.331
41422	2	0	0	1	0	2	0	0	1	1	1	2.512	0.088	0.221	0.469	1.172	0.677	1.451	0.085
41423	2	0	2	6	0	8	3	7	6	3	3	3.740	0.168	0.628	0.792	1.762	1.036	2.199	0.161
41432	1	0	1	2	0	6	0	2	4	3	0	4.245	0.092	0.391	0.624	1.707	0.894	2.223	0.090
41433	4	3	5	12	2	10	6	11	13	7	7	5.052	0.693	3.578	1.789	2.800	1.893	3.375	0.589
42113	1	0	1	5	0	0	0	2	1	3	0	1.362	0.746	1.016	1.007	1.171	1.053	1.225	0.482
42123	1	3	2	2	0	1	0	0	3	1	4	2.590	1.716	4.444	2.108	2.169	1.944	2.322	1.032
42212	1	0	0	3	0	0	0	1	0	2	0	1.218	0.084	0.102	0.319	0.700	0.464	0.808	0.079
42213	2	0	1	13	0	1	2	5	5	1	6	1.779	0.938	1.684	1.291	1.477	1.303	1.570	0.614
42222	1	0	0	4	0	0	0	2	0	1	1	3.401	0.034	0.116	0.340	1.198	0.565	1.603	0.034
42223	10	2	6	36	5	14	5	19	21	13	15	3.080	1.548	4.759	2.142	2.368	2.010	2.600	0.997
42232	1	1	0	2	1	0	0	0	2	2	0	4.012	0.144	0.578	0.760	1.795	1.016	2.272	0.139
42233	15	10	18	93	19	20	7	40	35	46	46	5.452	2.571	14.15	3.662	3.898	3.181	4.420	1.687
42313	1	0	0	1	0	2	0	1	1	0	1	1.793	0.814	1.460	1.208	1.444	1.246	1.547	0.560
42323	6	4	6	20	1	29	7	17	16	15	19	2.805	1.328	3.732	1.894	2.151	1.817	2.357	0.873
42332	1	1	0	0	0	0	0	0	0	0	1	4.362	0.059	0.257	0.507	1.592	0.776	2.125	0.058
42333	15	10	21	59	6	33	9	31	33	29	45	6.300	1.623	10.56	3.036	3.849	2.826	4.553	1.208
42423	2	2	0	9	1	3	2	3	6	6	2	3.452	1.863	6.849	2.511	2.674	2.281	2.941	1.189
42433	17	16	23	77	15	83	29	56	72	67	48	6.231	2.498	16.57	3.777	4.184	3.288	4.832	1.682
51111	1	0	0	0	0	0	0	0	0	0	0	1.726	0.008	0.014	0.117	0.558	0.242	0.750	0.008
51112	5	0	0	0	1	1	0	0	1	0	1	0.932	0.044	0.051	0.198	0.495	0.312	0.580	0.042
51113	9	4	2	6	1	2	0	4	4	3	4	1.346	0.572	0.868	0.847	1.071	0.917	1.142	0.380
51121	1	0	0	0	0	0	0	0	0	0	0	2.238	0.008	0.018	0.133	0.669	0.276	0.923	0.008
51122	2	3	0	2	0	0	0	3	1	0	1	2.294	0.071	0.165	0.404	1.054	0.600	1.307	0.069
51123	3	1	1	2	0	0	0	0	1	1	2	2.699	0.584	1.826	1.209	1.748	1.303	1.998	0.461
51131	1	0	0	3	0	0	0	0	1	0	2	6.730	0.008	0.054	0.232	1.446	0.479	2.228	0.008
51132	7	0	0	1	3	0	0	2	1	0	1	6.754	0.072	0.488	0.671	2.193	0.998	3.043	0.071
51211	5	1	1	0	0	0	0	1	0	1	0	1.048	0.015	0.016	0.123	0.441	0.231	0.548	0.015
51212	11	0	0	4	1	4	0	3	4	1	1	0.850	0.058	0.052	0.204	0.475	0.314	0.547	0.051
51213	16	2	4	6	0	7	0	5	4	5	5	1.335	0.404	0.580	0.727	1.011	0.829	1.092	0.306
51221	4	1	2	0	0	0	0	1	0	0	2	2.802	0.015	0.043	0.202	0.880	0.380	1.206	0.015
51222	10	3	7	6	1	10	1	7	7	8	6	2.590	0.088	0.234	0.472	1.188	0.680	1.478	0.085
51223	9	3	4	5	3	5	4	7	8	2	7	2.931	0.764	2.262	1.443	1.959	1.505	2.223	0.583
51231	2	0	0	0	0	0	0	0	0	0	0	6.709	0.013	0.087	0.294	1.590	0.567	2.390	0.013
51232	10	3	2	13	5	10	3	7	12	7	10	5.959	0.079	0.453	0.647	2.017	0.953	2.758	0.077
51233	11	3	11	8	4	8	2	8	9	8	11	6.891	1.273	8.052	2.679	3.793	2.613	4.603	0.969
51312	1	1	0	0	0	0	0	0	1	0	0	1.310	0.135	0.177	0.420	0.809	0.568	0.919	0.122
51322	2	0	3	3	1	10	4	6	4	5	6	3.122	0.040	0.124	0.350	1.161	0.570	1.530	0.039
51323	4	0	1	0	0	6	2	1	2	5	1	3.132	0.703	2.284	1.421	2.008	1.499	2.305	0.554
51331	1	0	1	3	1	5	1	0	2	3	6	5.350	0.021	0.112	0.335	1.494	0.598	2.143	0.021
51332	3	1	4	4	2	8	2	3	6	4	8	7.298	0.079	0.550	0.716	2.331	1.056	3.252	0.078
51333	4	2	1	3	0	11	0	6	2	5	4	5.559	1.848	10.84	3.152	3.708	2.868	4.283	1.358
51413	9	1	2	31	1	16	9	20	14	14	12	1.829	0.967	1.787	1.285	1.480	1.294	1.582	0.601
51422	1	0	0	0	0	0	0	0	0	0	0	3.894	0.084	0.327	0.571	1.578	0.829	2.046	0.082
51423	5	2	6	17	3	5	14	12	11	13	11	2.626	1.615	4.235	2.054	2.159	1.912	2.323	0.995
51431	1	1	1	0	0	1	0	1	1	1	0	5.511	0.018	0.099	0.314	1.479	0.575	2.144	0.018
51432	3	1	0	4	0	5	1	3	4	2	2	5.845	0.090	0.561	0.720	2.111	1.030	2.848	0.088
51433	8	6	8	14	4	18	10	22	12	12	14	5.965	0.993	5.340	2.190	3.243	2.218	3.930	0.786
52223	4	0	0	4	0	5	2	0	3	4	4	2.938	1.018	2.961	1.689	2.101	1.688	2.345	0.730
52233	7	3	5	7	3	9	3	6	4	14	6	6.973	1.519	10.26	3.124	4.108	2.943	4.915	1.198
52323	3	1	0	2	2	2	3	5	0	2	3	3.258	0.474	1.455	1.154	1.881	1.304	2.217	0.386

5 2 3 3 3	7	5	7	12	2	18	3	8	8	13	18	5.877	2.326	13.16	3.613	4.030	3.195	4.625	1.605
5 2 4 1 3	1	0	0	0	0	3	0	0	1	1	1	1.370	0.560	0.767	0.875	1.110	0.955	1.179	0.398
5 2 4 2 3	12	0	16	25	3	30	19	23	22	25	23	3.024	1.769	5.418	2.271	2.399	2.085	2.612	1.084
5 2 4 3 3	12	5	21	24	5	71	20	42	35	37	32	5.847	2.584	15.68	3.825	4.110	3.314	4.691	1.745

$RETURN

III. LINK ACCIDENT CONTINGENCY TABLE DATA IN GLIM FORMAT

```
$C FILENAME IS (BACKUPDATA_)LC40
$UNITS 52
$DATA D C FB LB N S T1 T2 T3 M PED PSV Y1 Y2 Y3 Y4 F L Q
$FACTOR D 3 C 2 FB 3 LB 3
$C LINK DATA FOR CONTINGENCY TABLE ANALYSIS.
```

TRUNCATED VARIABLE NAME	FULL VARIABLE NAME	DESCRIPTION
D	LINKTYPE	link type (3 factor levels)
C	LCWYTYPE	carriageway type (2)
FB	LFLOWBAND	flow band
LB	LENGTHBAND	length band
N	LCELLTOT	number of links per cell
S	LSVNP	single veh non-ped accidents
T1	LTVNP1	two veh non-ped accidents (type 1)
T2	LTVNP2	two veh non-ped accidents (type 2)
T3	LTVNP3	two veh non-ped accidents (type 3)
T4	LTVNP4	two veh non-ped accidents (type 4)
T5	LTVNP5	two veh non-ped accidents (type 5)
M	LMVNP	multi veh non-ped accidents
PED	LPED	pedestrian accidents
PSV	LPSV	PSV accidents
Y1	LACCTOT1	total accidents for year 1979
Y2	LACCTOT2	total accidents for year 1980
Y3	LACCTOT3	total accidents for year 1981
Y4	LACCTOT4	total accidents for year 1982
F	LINKFLOW	average traffic flow
L	LINKLENGTH	average link length
Q	LMVKM	average MVkm

FACTOR LEVELS ARE:

LDEVTYPE

3=SHOPPING OR COMMERCIAL ON AT LEAST ON SIDE OF ROAD)
2=OTHER URBAN (IE WITH NO SHOPPING OR COMMERCIAL DEVELOPMENT
1=RURAL

LCWYTYPE

1=SINGLE CARRIAGEWAY
2=DUAL CARRIAGEWAY

LFLOWBAND

1=LOW FLOW BAND
2=MEDIUM FLOW BAND
3=HIGH FLOW BAND

LENGTHBAND

1=SHORT LENGTH BAND
2=MEDIUM LENGTH BAND
3=LONG LENGTH BAND

$READ

3	1	1	1	29	2	3	0	1	4	7	0	3	3	3	8	1.489	24.24	0.036
3	1	1	2	13	2	1	2	2	0	4	1	3	4	2	3	1.441	72.54	0.103
3	1	1	3	13	3	2	2	1	2	16	2	7	5	8	8	1.335	175.8	0.243
3	1	2	1	68	4	9	2	1	1	38	18	17	13	18	25	3.083	24.40	0.076
3	1	2	2	39	11	3	2	11	3	33	14	20	19	16	22	3.079	69.28	0.213
3	1	2	3	41	21	13	7	11	12	59	15	29	42	26	41	2.983	189.5	0.565
3	1	3	1	118	14	11	2	14	7	96	36	45	58	53	24	5.708	24.75	0.140
3	1	3	2	52	19	28	2	16	20	126	42	80	63	59	51	6.371	69.00	0.437
3	1	3	3	35	19	20	2	31	18	97	23	44	67	54	45	6.941	151.9	1.031
3	2	1	3	7	1	8	0	0	4	14	9	9	7	10	10	1.293	177.1	0.189
3	2	2	1	5	0	0	0	0	1	0	3	2	0	1	1	2.871	20.80	0.062
3	2	2	2	2	0	1	0	0	0	3	0	0	1	2	1	3.262	80.00	0.259
3	2	2	3	4	1	5	1	0	1	39	21	10	13	16	29	3.212	150.8	0.486
3	2	3	1	5	0	1	0	0	0	2	5	1	2	1	4	6.096	20.20	0.127
3	2	3	2	2	0	2	0	0	0	14	1	3	3	5	6	4.088	89.00	0.364
3	2	3	3	5	2	6	0	2	1	27	6	12	9	13	10	4.481	148.2	0.664
2	1	1	1	115	1	1	0	2	3	19	1	7	7	7	6	1.212	26.63	0.033
2	1	1	2	95	12	1	1	3	2	18	2	11	7	12	9	1.228	70.21	0.086
2	1	1	3	140	53	13	11	17	10	63	9	53	38	39	46	1.280	231.9	0.299
2	1	2	1	166	16	1	4	2	8	17	8	16	9	14	17	2.821	28.05	0.080
2	1	2	2	122	11	2	6	5	7	36	9	22	19	16	19	2.837	69.36	0.197
2	1	2	3	210	136	37	29	51	24	134	28	115	125	105	94	2.869	243.2	0.696
2	1	3	1	143	15	8	1	5	8	37	19	20	29	20	24	5.885	24.97	0.147
2	1	3	2	114	17	15	4	15	11	40	8	26	28	27	29	5.893	69.89	0.413
2	1	3	3	176	133	74	28	75	49	147	34	142	142	128	128	6.570	228.7	1.491
2	2	1	1	18	0	0	0	0	0	1	0	1	0	0	0	0.694	15.67	0.012
2	2	1	2	8	3	0	0	0	1	8	0	1	3	5	3	1.008	83.63	0.085
2	2	1	3	14	2	7	0	2	1	10	5	9	6	3	9	0.678	196.4	0.120
2	2	2	1	6	0	0	0	0	0	0	0	0	0	0	0	2.645	21.33	0.057
2	2	2	2	12	4	3	0	0	1	15	7	9	8	4	9	2.795	80.92	0.227
2	2	2	3	23	7	4	1	2	1	6	0	7	7	1	6	2.787	251.6	0.732
2	2	3	1	11	0	2	0	0	0	7	0	2	1	4	2	5.657	30.09	0.167
2	2	3	2	5	0	2	0	0	2	0	0	1	2	1	0	5.499	64.00	0.342
2	2	3	3	12	7	3	0	0	4	13	3	7	7	6	10	5.420	302.3	1.569
1	1	1	1	18	0	0	0	0	0	0	0	0	0	0	0	1.462	24.72	0.038
1	1	1	2	11	1	0	0	0	0	1	0	1	0	0	1	1.513	73.36	0.113
1	1	1	3	253	248	26	89	46	26	32	3	123	107	116	124	1.359	818.6	1.094
1	1	2	1	8	0	0	0	1	0	1	0	0	0	1	1	2.551	29.63	0.074
1	1	2	2	4	0	1	0	0	0	0	0	0	0	1	0	2.920	87.50	0.254
1	1	2	3	81	106	12	37	24	12	22	6	48	51	64	56	2.720	804.2	2.201
1	1	3	1	7	0	0	0	1	0	0	0	0	0	1	0	4.696	22.00	0.100
1	1	3	2	2	0	0	0	1	0	1	0	0	1	0	1	5.911	66.50	0.393
1	1	3	3	54	84	17	33	14	17	16	2	41	43	44	55	4.613	590.1	2.692
1	2	1	1	19	0	0	0	0	0	0	0	0	0	0	0	1.201	26.42	0.030
1	2	1	2	11	1	0	0	0	0	0	0	1	0	0	0	1.073	67.64	0.073
1	2	1	3	81	15	3	1	0	2	0	0	5	7	5	4	0.964	475.2	0.575
1	2	2	1	11	2	0	0	0	0	0	0	0	0	0	2	2.770	19.00	0.053

1	2	2	2	8	2	1	0	0	0	1	0	2	0	0	2	2.297	80.88	0.186
1	2	2	3	36	59	17	7	1	7	6	1	38	23	13	24	2.645	1339	3.963
1	2	3	1	2	0	0	0	2	1	0	0	1	2	0	0	6.759	4.500	0.030
1	2	3	2	2	1	0	0	0	2	0	0	2	0	1	0	6.022	70.50	0.429
1	2	3	3	24	24	18	3	3	23	6	1	18	19	20	21	6.048	736.8	4.276

$RETURN