april 26 - 28, 1988
amsterdam
the netherlands

traffic safety theory & research methods

Session 2: Models for evaluation

SWOV

This book was made possible through the support of 3M Europe

SESSION 2: MODELS FOR EVALUATION

Summaries of the papers presented by the additional speakers

Ekkehard BRÜHNING & Gabriele ERNST, Bundesanstalt für Straßenwesen, Bergisch Gladbach, Federal Republic of Germany
Recent developments in the methodology of effectiveness studies; New applications and statistical models for quasi-experimental designs

Heather WARD, R.E. ALLSOP, A.M. MACKIE & R.T. WALKER, University College, London, United Kingdom
Altering the pattern of traffic and accidents in urban areas; A methodology to detect change

Full papers of other contributors

Wolfgang SENK, Ruhr-University Bochum, Federal Republic of Germany
Area-wide traffic calming measures: Accident analysis

Talib ROTHENGATTER, University of Groningen, The Netherlands
A model for evaluating educational road safety measures

Alan J. NICHOLSON, University of Canterbury, Christchurch, New Zealand and Visiting Fellow, University of Leeds, United Kingdom
Accident count analysis: The classical and alternative approaches

D.F.JARRETT, C.R. ABBESS & C.C. WRIGHT, Middlesex Politechnic, London, United Kingdom
Emperical estimation of the regression-to-mean effect associated with road accident remedial treatment

Summaries of the papers presented by the additional speakers

Ekkehard BRÜHNING & Gabriele ERNST, Bundesanstalt für Straßenwesen, Bergisch Gladbach, Federal Republic of Germany
Recent developments in the methodology of effectiveness studies; New applications and statistical models for quasi-experimental designs

Heather WARD, R.E. ALLSOP, A.M. MACKIE & R.T. WALKER, University College, London, United Kingdom
Altering the pattern of traffic and accidents in urban areas; A methodology to detect change

RECENT DEVELOPMENTS IN THE METHODOLOGY OF EFFECTIVENESS STUDIES - NEW APPLICATIONS AND STATISTICAL MODELS FOR QUASI-EXPERIMENTAL DESIGNS

Ekkehard Brühning and Gabriele Ernst
German Federal Highway Research Institute

Summary

The aim of effectiveness studies is to describe the type, direction and extent of the effects of safety measures on the number of accidents, i.e. its particular objective is to **quantify** the effects. Because experiments under laboratory conditions are generally not possible, studies of traffic safety measures are carried out as quasi-experiments. In such quasi-experiments the experimental groups are studied one or several times, before and/or after application of the measure.

Quasi-experimental designs always raise the question of whether or not measured changes are due to effects of the particular measure or can be explained by other effective influences.

The control efforts required to take care of such possible interferences or distorting influences are connected with the particular experimental design. When there are no control groups it is necessary to carry out extensive and very costly control experiments to take into account all distorting influences. This additional control effort can be reduced considerably if it is possible to achieve design-immanent controls by choosing an adequate experimental design.

Effectiveness studies in accident research or on traffic safety measures are frequently carried out not only at one place but at several places at the same time. In principle, two different experimental plans can be associated with such studies:

a) Depending on the underlying design, all observations are **aggregated** to form experimental and control groups.

b) In simultaneous studies several experiments are planned and conducted at different places or different experimental groups (e.g. road sections) using one and the same experimental design. Each individual experiment should **by itself** furnish undistorted results on the effectiveness of a measure.

If simultaneous designs are used, findings on the success and the efficiency of a measure will be of a better quality than if the measure is studied only once (e.g. at one place). In principle, simultaneous comparisons reduce or even eliminate the threats to the validity of the results and increase their accuracy. Simultaneous designs are therefore increasingly employed for large-scale studies of traffic safety measures.

The adequate analysis of simultaneous comparisons for the evaluation of the effectiveness of a measure cannot, however, be carried out by using the conventional statistical methods.

To evaluate the effectiveness of a measure with the previously available statistical tools, it was necessary to aggregate the data of the individual experiments. The cell frequencies of the underlying design had to be summed up from all simultaneous comparisons. The statistical analysis was then based on a single contingency table.

Such an aggregation preserves the advantages of the simultaneous design concerning the avoidance of dangers and the increase of the number of cases. But the statistical analysis of simultaneous comparisons through this method (aggregation) cannot be satisfactory because the additional information provided by the simultaneous design is not exploited.

A simultaneous experimental design is based on the hypothesis that there is one true value of the effectiveness of a measure. The measure factors determined in the individual experiments then appear in random distribution around this true value.

Before the simultaneous measure factor is determined a test of whether the individual measure factors are homogeneous (jointly compatible) has to be made. Only when this condition is fulfilled can the simultaneous measure factor be regarded as consistent and therefore as an adequate solution.

New methods employing loglinear or Logit models have recently been developed for simultaneous experimental designs. The model formula in effectiveness studies depends directly on the underlying experimental design. Depending on the particular design, appropriate loglinear models with adequate test statistics have to be employed: Thus it is possible to formulate adequate models for simultaneous experimental designs. A simultaneous analysis can be carried out without aggregation of the data of the individual experiments. This method makes considerably better use of the available information.

The evaluation of a measure's effectiveness is based on different types of criterion variables. The frequency of events (accidents, possibly conflicts) is mostly used as the criterion variable. But other possible types of criterion variables are: interview results, measurements, percentage changes, monetary values and relativized quantities (e.g. the number of accidents related to kilometers driven = Accident Rate).

These different types of criterion variables are connected with different assumptions regarding their statistical distribution: e.g., numbers of accidents usually follow a Poisson distribution; vehicle speeds, percentage changes, interview results are often approximately in normal distribution.

Depending on the distribution type, adequate tests can be used to check whether an observed value is consistent with a hypothetical expected value or some other empirical value.

In the case of percentage changes it is, however, not permitted to check exclusively on the basis of assumptions regarding the distribution, i.e. without weighting, whether an observed value deviates significantly from a comparison value. Rather, the value of the reference variable which was used for determining the percentage change has to be taken into account.

In the case of relativized or risk quantities (e.g. Accident Rates) the exposure quantity is frequently taken as a deterministic (non-random) quantity with no error of measurement. In this case, loglinear models can be used whose parameters can be estimated by the classic maximum likelihood methods. This procedure is similar to that of weighted Poisson models.

If for the analysis of risk quantities the exposure quantity is not deterministic but stochastic (random), there are special problems because the type of the joint distribution of nominator and denominator of the risk quantity is unknown. But even for this case a solution can be found by using the recently developed theory of pseudo maximum likelihood estimation (Gourieroux, Monford, Trognon).

A manual about the statistical analysis of simultaneous effectiveness studies has recently been published. On the basis of loglinear and Logit models solutions are offered in this manual for a number of simultaneous quasi-experimental designs as well as for different types of criterion variables. A detailed description of the analysis methods is given on the basis of application examples employing standard software.

For reasons of transport policy and the efficient use of available funds effectiveness studies on traffic safety measures are carried out frequently and with considerable amounts of research money. It is a fact that better results are obtained at no extra cost if methodological knowledge is employed early in the planning phase of a study.

ALTERING THE PATTERN OF TRAFFIC AND ACCIDENTS IN URBAN AREAS A METHODOLOGY TO DETECT CHANGE

Heather Ward, R.E.Allsop, A.M.Mackie and R.T.Walker

1. INTRODUCTION

Since 1982 the Transport and Road Research Laboratory (TRRL) has been leading an Urban Safety Project which aims to demonstrate the effectiveness of introducing a package of low-cost accident countermeasures to improve the safety of residential areas of typical British free-standing towns. The Transport Studies Group (TSG) at University College London has been involved in the development of a methodology for evaluating urban safety schemes in conjunction with TRRL and Transport Operations Research Group (TORG) at the University of Newcastle upon Tyne.

The area-wide approach to road safety requires low-cost accident countermeasures to be combined to produce an area-wide effect, strengthening where possible the hierarchy of the street system and diverting traffic from primarily residential roads on to the local distributors and arterials whilst ensuring that suitable measures are taken on these roads to ease flow and improve safety.

The first stage in the planning and development of an urban safety scheme is the definition of the existing road hierarchy. This is followed by an appraisal of each route in turn to identify inadequacies in terms of safety or traffic management. The next stage is to define a new or improved road hierarchy based upon which safety objectives and strategies can be developed for each class of route and for the area as a whole. Each category of route in the hierarchy should be improved in line with its function, thus making them safer and making it practicable to discourage through traffic from residential areas. The individual measures should be chosen to support the newly defined hierarchy and to bring about accident reductions in accordance with the safety objectives. The measures required to achieve these objectives are, to a great extent, interactive and are not necessarily sited at locations which have an accident history.

2. MONITORING - THE DEVELOPMENT OF A METHODOLOGY TO DETECT CHANGE

A methodology has been developed which is capable of providing information to enable the assessment of

(i) the overall objective of reducing the total number of injury accidents,

(ii) the objectives defined for individual routes or residential areas in terms of the transfer of traffic to more suitable routes, the change in traffic flows entering and leaving the residential areas and reductions in injury accidents of particular kinds, and

(iii) areas of unforeseen difficulties of operation or inconvenience such as increased travel times on main routes, decreased accessibility to residential areas, or transfer of traffic and accidents to areas adjacent to the scheme area.

2.1 Scope of monitoring

In pursuit of their safety objectives, area-wide schemes are expected to affect the pattern of routeing and the speed of travel along main roads but would not be expected to affect the total number of trips made. By their nature, however, restrictive traffic measures are likely to result in increased journey distances for some residents. A method of estimating the order of magnitude of this effect has been developed. To establish these changes fully, it would be necessary to undertake origin-destination surveys on a before and after basis. Whilst this would provide a wealth of information, it has not been done because it is a very costly and intrusive exercise when compared with the budget for the total package of engineering measures of the type considered here.

2.2 Size of effect - size of sample

The Urban Safety Project schemes have been planned as projects with full experimental design. A pilot study was undertaken (Dalby and Ward, 1982) in which the variability of traffic and accident parameters was assessed to provide input into the statistical design of traffic surveys. Experimental design encompasses sample size, methods of data collection and subsequent statistical analysis to allow the monitoring team to draw conclusions about traffic and accident parameters before and after the introduction of the schemes with a given probability of establishing with a given level of confidence that a certain expected change has not occurred by chance. In order not to waste scarce resources, it is important to choose the correct sample size. If a sample is too small and insufficient data are collected, large real differences may not be established as statistically significant whilst if too much data is collected, real differences too small to be of practical importance appear statistically significant.

The size of areas used in the Urban Safety Project was determined by taking an expected reduction of 15 per cent in accidents as a starting point. A sample size of 500 injury accidents a year would be necessary to have about a 50 per cent chance of establishing a reduction of 15 per cent at the 5 per cent level of significance after one year of operation of the scheme. Five towns, Bradford, Bristol, Nelson, Reading and Sheffield, participated with study and comparison areas each with about 100 injury accidents per year, giving the required total sample size of 500 injury accidents in the study areas.

2.3 <u>Type of surveys and data collected</u>

The surveys undertaken by the monitoring teams were designed to allow the assessment of the effectiveness of the area-wide schemes in the five towns with relation to the stated safety objectives for each town. They also had to take into account the need to identify, in the short-term, areas of unforeseen difficulty of operation or inconvenience. The types of survey and data collected are described below.

Accident records were collected over a five year before period. The use of such an extended before period allows the detection of trends and seasonal variability in the accident pattern and enables the range to be established in which accident totals might be expected to fall in the after period. It also provides a basis for detection of effects on the number of road accidents, their severity and distribution over the road network among different groups of road users.

A small number of automatic traffic counters provided data about flows over extended periods which enabled the detection of trends and overall redistribution of traffic. Classified manual counts of flows and turning movements were carried out at about 50 key junctions throughout the area over a four-day period at four times a year before and after the changes in the study areas of each town. Sites were selected to enable changes to be detected in the points at which drivers choose to enter and leave the residential areas and to provide a measure of compliance at junctions where certain movements have been prohibited but not necessarily physically prevented.

The redistribution of traffic can affect journey times both within and adjacent to the area treated. Changes in layout and control at important junctions and in the type and number of points of access from residential areas to the main roads can have a substantial effect on both the duration and location of delays.

Journey time data were collected using the moving observer technique (Wardrop and Charlesworth, 1954) on a link-by-link basis on preselected routes in the study and comparison areas. The use of a portable in-car computer allowed accurate data to be collected at frequent intervals along the routes leading to a detailed assessment of delays incurred in approaching junctions and pedestrian crossings. By incorporating suitable loops into the journey time routes, key junctions could be approached from each arm and delays to side road traffic subsequently quantified. This addition is important in that extra distance travelled in the residential area may be set against gains made in time taken to exit from these areas when mini-roundabouts, or other changes in control, are introduced.

Pedestrian movement is difficult to monitor because of its complex and often diffuse patterns. Schemes of the type described here are unlikely to have an adverse effect on pedestrian movement within the residential areas but those crossing on the main roads are more likely to be affected by changes in traffic as well as in the provision and location of pedestrian facilities and local surveys may be appropriate to detect such effects.

3. EVALUATION OF EFFECTIVENESS - THE DETECTION OF CHANGE

3.1 Detection of changes in number and pattern of accidents

In order to test the effectiveness of schemes in reducing accidents, the accidents in each study and comparison area were divided into quarterly totals and log-linear models were fitted to these totals. This enables the effect of the scheme to be estimated after allowing for trends and seasonal effects, which may well differ between study and comparison areas. The accidents occurring in the implementation period should be analysed separately because the rate of occurrence may be atypical in this period as road users become accustomed to the changes, especially when required to find new routes.

Using the Reading data as an example, the log-linear model which gave the best fit was

$$y_{jklm} = \exp\{a+[b+(bc)_k]j+c_k+d_l+e_m+(ce)_{km}+(de)_{lm}$$

where j is time in quarters

k is area, $k=1$ comparison and $k=2$ study

l is season $l=1$ Nov-Jan ....$l=4$ Aug-Oct

m is period $m=1$ before, $m=2$ implementation, and $m=3$ and 4 two parts of after period

The term which provides information about the effectiveness of the scheme is $(ce)_{km}$, the inclusion of which means that there is a difference between before and after periods which is not the same in the study and comparison areas. Comparison of this parameter with its standard error of estimate shows how likely this effect is to have occurred by chance. The size of effect may be calculated by $\exp\{(ce)_{23}-(ce)_{21}-(ce)_{13}+(ce)_{11}\}-1$.

In the case of Reading, no third order interaction terms were statistically significant so are not included in the model. However, these terms should always be tested and included where necessary. The inclusion of third order terms in period and area complicates the estimation of size of effect as the second order term $(ce)_{km}$ on its own no longer does this. A method has, however, been developed which enables the calculation of both the size of effect and its standard error in such cases.

Log-linear models may be fitted in a similar way to disaggregate data, for example to pedestrian, motorcycle or pedal cycle accidents. The effect of the scheme on severity of injury may also be assessed, again by fitting linear models. The indicator of severity used has been the ratio of fatal plus serious accidents to total injury accidents. In this case one appropriate model is the logit model which is fitted to the proportion, p, of fatal plus serious accidents in the following way,

$$\ln[(P_{km}/(1-P_{km})] = a + c_k + e_m + (ce)_{km}$$

where the indicator of effect on severity is the term $(ce)_{km}$. If the reduction in deviance associated with adding this term to the model is statistically significant, this indicates that the change in severity between the before and after periods in the study area differs from that in the comparison area. It is important to consider severity of injury because it may be possible to reduce it in an area without necessarily significantly changing the total number of accidents occurring.

3.2 Detection of changes in pattern of traffic movement

The collection of traffic data was undertaken on a before and after basis extending over at least one year before the introduction of the schemes and for two years afterwards.

The junctions at which turning counts were made were divided into four groups which enabled a latin square analysis of variance to be carried out on the resulting data. This provided information regarding variations with respect to the time of day, day of week and week of survey, and whether these differed between before and after periods. The count data were not normally distributed and a standard square root transformation was used prior to analysis. Analysis of variance enabled junctions with statistically significant changes in mean flow to be confirmed as a first stage in the identification of new patterns of routeing. The analysis took account of differences at the three times of day surveyed in order to detect, for example, changes affecting routes into town but not the return journey.

The journey time data were analysed by fitting linear models to the reciprocals of the link travel times. The effect of flow was tested using an analysis of covariance to establish whether the journey time/flow relationship was significantly different in the before and after periods.

The final type of survey undertaken was of pedestrian movements. To accommodate the large fluctuation in flows sometimes found, proportions of pedestrians crossing in each sector of a site were analysed on a before and after basis. Traffic flows were observed at each site which allowed logit models of the form $\ln(p/1-p) = a + bq$ to be fitted, where p is the proportion of pedestrians using a crossing or crossing in a given section and q is the traffic flow during the corresponding survey periods. This allows the effect of traffic flow on where pedestrians choose to cross the road to be assessed.

4. THE ECONOMIC EVALUATION OF EFFECTIVENESS

When assessing the cost-effectiveness of a scheme there are on the benefit side net savings in accidents and there may well also be net savings in vehicle operating costs and time. On the disbenefits side are the cost of the scheme, its maintenance and monitoring, and possibly in some cases a net increase in vehicle operating costs and travel time and some extra accidents occurring during or just after implementation. Standard values of accident costs and value of time are provided by the Department of Transport (annual).

5. IMPLICATIONS FOR FUTURE PRACTICE

In this paper a methodology has been outlined which enables the monitoring of urban safety schemes. The monitoring programme allows an economic evaluation to be made of the main effects both in terms of changes in amounts and patterns of traffic and accidents. However, in routine applications of the resulting area-wide approach, it is not envisaged that local authorities will have the budget, staff resources or need to monitor on an equivalent scale to that undertaken in the Urban Safety Project. Even for routine monitoring purposes it is important to consider some aspects of experimental design described in this paper especially with respect to monitoring of accident patterns where changes in number of accidents are often the only input into an economic evaluation or justification for expenditure on such schemes.

The monitoring teams and TRRL will continue to learn from the experience gained in the Urban Safety Project and with further work, will be in a position to offer guidelines to the design implementation and monitoring of urban safety schemes by local authorities as part of their routine accident remedial programmes.

6. ACKNOWLEDGEMENTS

The work described in this paper forms part of the programme of the Transport and Road Research Laboratory and the paper is published by permission of the Director.

The authors are grateful for the help of colleagues, both past and present, in the TSG, especially to John Norrie for his work on journey distances, to David Lynam of TRRL for his continuing support and to colleagues in TORG.

7. REFERENCES

DALBY, E. and WARD, Heather (1981) Application of low-cost road accident countermeasures according to an area-wide strategy. Traffic Engineering & Control, 22(11), 567-575.

DEPARTMENT OF TRANSPORT (annual) Road accident costs. Highways Economics Note No. 1; Value of time and vehicle operating costs. Highways Economics Note No. 2. London.

WARDROP, J.G. and CHARLESWORTH, G. (1954) A method of estimating speed and flow of traffic from a moving vehicle. Proceedings Institute of Civil Engineers, Part II, 3, 158-171.

Full papers of other contributors

Wolfgang SENK, Ruhr-University Bochum, Federal Republic of Germany
Area-wide traffic calming measures: Accident analysis

Talib ROTHENGATTER, University of Groningen, The Netherlands
A model for evaluating educational road safety measures

Alan J. NICHOLSON, University of Canterbury, Christchurch, New Zealand and Visiting Fellow, University of Leeds, United Kingdom
Accident count analysis: The classical and alternative approaches

D.F.JARRETT, C.R. ABBESS & C.C. WRIGHT, Middlesex Politechnic, London, United Kingdom
Emperical estimation of the regression-to-mean effect associated with road accident remedial treatment

AREA-WIDE TRAFFIC CALMING MEASURES: ACCIDENT ANALYSIS

Wolfgang Senk, Ruhr-University, Bochum, West Germany

Introduction

The Federal Office of Environmental Protection (UBA),the Federal Institute for Regional Studies and Environmental Planning (BfLR) and the Federal Road Research Laboratory (BASt) are carrying out a common long-scale experiment in six German model cities to investigate the impact of area-wide traffic calming measures on urban areas and different traffic situations. The Chair of Traffic Engineering I at the Ruhr-University Bochum has been commissioned to analyse the accidents in the following six cities (see appendix 1):

- Berlin-Moabit, an area with residential and shopping streets in the midst of a large city;
- Borgentreich, a rural community in Eastern Westfalia with about 10.000 inhabitants;
- Buxtehude, a town near Hamburg with a population of 18.000 inhabitants;
- Esslingen, a middle order centre situated near the river Neckar;
- Ingolstadt, a town with a historical centre located in Bavaria;
- Mainz, a large city near the river Rhine.

First concluding reports about these six model cities will be published in May 1988.

The methodology of this accident analysis was developed in the course of a pilot study on an area in Berlin-Charlottenburg where traffic calming measures were carried out and investigated (see appendix 2). The objective of the study was to prove the practical applicability of statistical methods that are powerful enough to recognize changes in accident occurence even in cases of low accident rates. Furthermore, these statistical methods should also point at correlations between traffic calming measures and a decrease of accident rates.

All tests were before-and-after studies with control groups. For this, an area in Berlin similar to the one in Charlottenburg regarding its architectural structure and traffic situation had to be found. A part of Berlin-Moabit, the later area of the long-scale investigation mentioned above, proved to be suitable as a control group. During the pilot study, only sporadic traffic-calming measures were carried out in this area.

The tests were based on various figures (realizations of the random variables of the stochastic model) consisting both of absolute accident rates like

- total number of accidents
- accidents of a certain kind
- accidents with injuries to persons
- accidents in correlation with certain road users
- accidents of a certain severity

and relative accident rates, i.e. the quotients of absolute accident rates and suitable exposure values such as

- number of residents
- length of the road network
- kilometres travelled.

These accident rates were analysed as a whole, and they were furthermore differentiated according to

- accidents on road sections and
- accidents at intersections .

Overall accident occurrence

A first survey of the overall accident occurence was gained by temporally dividing the seven years of the investigated period into three parts, the time before the beginning of the traffic calming measures, the time of the architectural modifications and the adjustment of the residents, and the time after the end of the measures. Furthermore, the whole area was divided into 8 zones. These were a zone of architectural modifications, a zone with a speed limit of 30 km/h, a zone in which traffic had been calmed down as a side effect of measures carried out in areas next to it (zone of passive traffic calming measures), neighbouring arterials, limiting arterials, and a neighbouring area. In the control area, these zones were a residential area and neighbouring arterials.

A more detailed subdivision comprised 7 periods of one year and 11 local zones. In the course of this investigation, the access points to the urban motorway in the investigated area were considered additionally. In the control area, the "Turmstraße", a main road intersecting this zone, and a residential area where sporadic traffic calming measures had been carried out, were also analyzed (see appendix 3).

As a result of this, the accident rates were compiled in 8x3 or 11x7 contingency tables. These contingency tables were analysed using χ^2-tests. A χ^2-test is based on the hypothesis that the accident rates of the individual cells of a contingency table depend on accidental variations and are independent of each other. If the test variable is greater than the corresponding critical figure, the hypothesis has to be rejected, i.e. it may be considered as statistically proven that there are systematic divergences between the actual accident occurrence and the accident occurence expected according to the hypothesis. Since the test variable is calculated by regression of actual and expected values, one cannot deduce the cause of deviations from significant divergences from the expected accident occurrence when rejecting the hypothesis. This becomes evident to everyone calculating a χ^2-test of a 2x2 table with paper and pencil. An abstract of a χ^2-test is to be found in appendix 4.

Thus, more refined statistical methods were necessary to establish a causal connection between traffic calming measures and a substantial change in accident rates. For this, a log-linear approach and a Poisson-regression model were used.

The log-linear approach is based on the assumption that the accident rates of each cell of a contingency table result from various factors. These components comprise a universal factor, the influence of the area, the influence of time, and the interrelation between time and place. The name of this model is derived from the fact that this multiplicative approach is both logarithmitized and linearized in the course of the numerical evaluation . If the logarithms of the interrelation factor are negative in each cell of the investigated area in the time period after the traffic calming measures, one may conclude that the measures are responsible for this decrease. If they equal zero, the measures have no influence on accident occurrence. If the logarithms of the interrelation factor happen to be positive, this means that the measures cause an undesirable increase in accident rates. The values of the interrelation factors also allow for statements on the different effectiveness of the individual measures, if values in different cells correspond to different traffic calming measures. The log-linear analysis is described in appendix 5.

The Poisson-regression model is based on four plausible hypotheses:

- accident rates in different intervals are independent of each other;
- accident rates only depend on the length of the interval considered, but not on certain moments;
- the probability of the occurence of more than one accident during a very short interval almost equals zero;
- the probability of the occurence of exactly one accident during a very short interval is proportional to the length of the interval.

Based on these assumptions, one can mathematically derive that accident rates must be realizations of Poisson-distributed random variables.

The actual modelling approach, which is similar to the log-linear model, is founded on hypotheses about the number and kind of factors influencing accident occurence. Thus, the validity of the model depends decisively on the choice of the factors considered in the modelling approach.

For the accident analysis of the area in Berlin-Charlottenburg, an approach containing seven factors without interrelationships was chosen (see appendix 6). The model confirmed the results of the log-linear approach; however, it will be revised for further evaluations. Statements on the effectiveness of the measures were mainly based on the results of the log-linear evaluation.

In those cases in which the value almost reached the corresponding critical figure in the χ^2-test but did not exceed it, it could be supposed that changes in the number of accidents had occurred, but had not been classified as significant in the χ^2-test, which is not powerful enough for this. In these cases, the data were once again examined by means of a Bayes' method, which is a mere before-and-after comparison based on the assumption that accident rates are realizations of Poisson-distributed random variables. The parameter of the Poisson-distribution is assumed to be Γ-distributed and is calculated from the accident rates of the period before the traffic calming measures. This information is also considered when the confidence interval is calculated in the course of the evaluation of the accident rates occurring in the period after the traffic calming measures have been carried out, which increases the power of the test. Thus, a significant decrease in accidents on the road sections of the investigated area could be detected with the aid of this Bayes' method. (For further information on the Bayes' method see appendix 7.)

Individual aspects

Many aspects could not be investigated with the rather coarse 7x11 and 3x8 contingency tables. Although approximately 14.000 accidents were recorded, the sample sizes were too small to analyze certain aspects, e.g. how many pedestrians older than 65 years were killed. Several cells of the contingency tables would have contained insufficient values or no values at all. In these cases, the relevant accident rates of the investigated and the control area were comprised in a 2x2 contingency table for the time before and after the measures and were evaluated by means of a χ^2-test if there were enough values. If there were values in all cells, but in at least one cell insufficient values for a χ^2-test, Fisher's exact test was applied. Five actual accidents and five accidents expected due to the test hypothesis were considered as sufficient values for a χ^2-test. With the help of these tests it was possible to find out whether statistically valid changes in accident rates had occurred. As with the larger contingency tables, the log-linear approach was used again to analyze whether the measures were responsible for such changes. The results of the individual analyses were summed up in a large table (see appendix 8).

The evaluation of these aspects made it possible to analyse the effectivenes of the measures in detail. The measures were more effective in the zone calmed by architectural modifications (zone 1) than in the zone with the speed limit of 30 km/h (zone 2). However, the effects in this zone were still stronger than in the zone calmed as a result of the measures in the neighbouring area. It can therefore be concluded that architectural modifications are the most efficient means to reduce accident rates. On the other hand, the significant increase in accidents with oncoming traffic at narrowings of streets and staggered lanes indicates that the design of these architectural measures must still be improved.

Conclusion

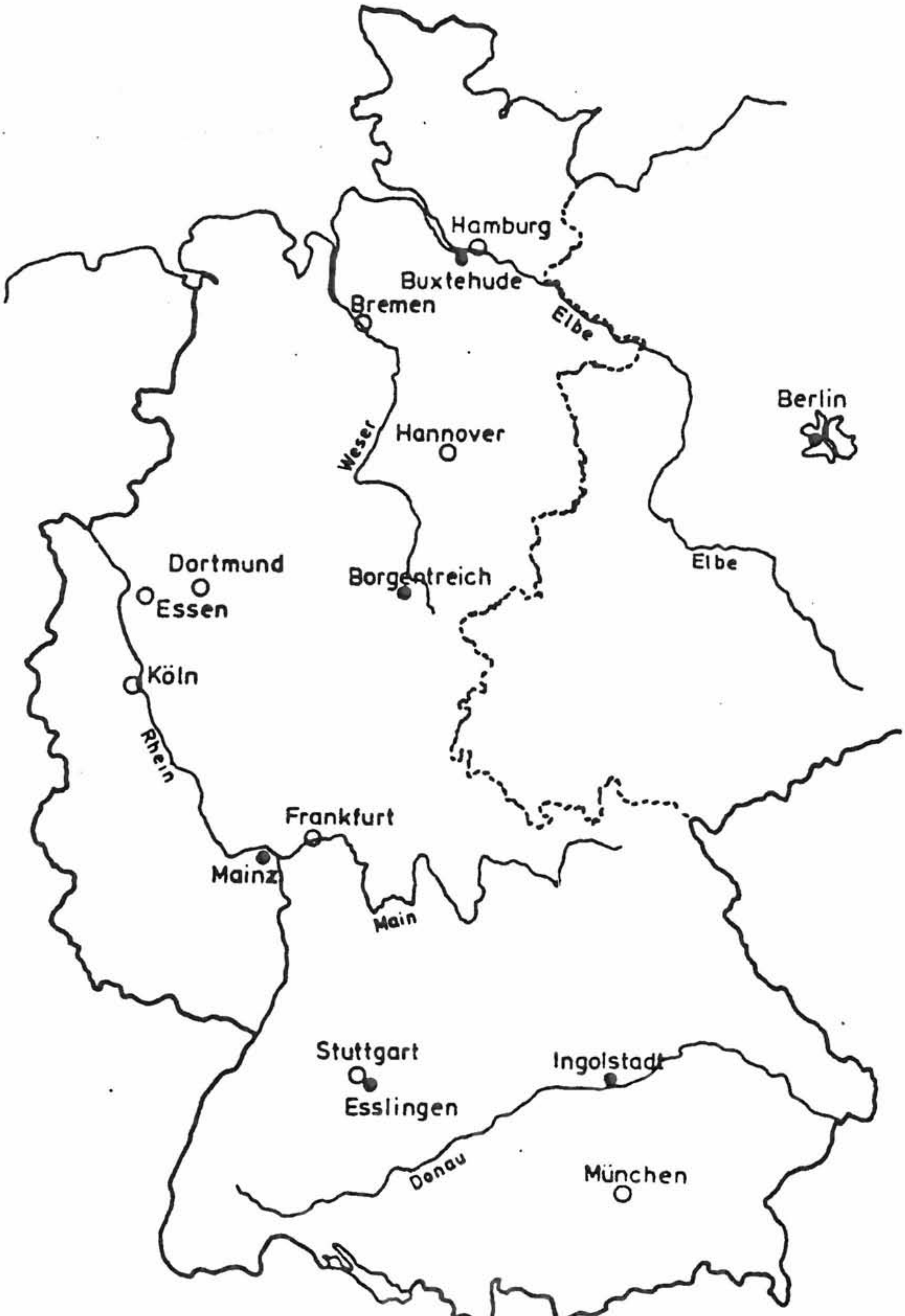
The pilot study showed that the χ^2-test is a simple means to determine changes in accident occurrence exceeding accidental values. However, causes for these changes cannot be deduced with a χ^2-test, they can only be determined in the course of the log-linear analysis. Moreover, Poisson-regression models are suitable means to establish causal connections between measures carried out and changes recorded afterwards.

Among other things, the study showed that architectural measures lead to a significant decrease in the following kinds of accidents:

- accidents with vehicles at intersections (-46%);
- accidents with pedestrians (-78%);
- accidents with children (-62%);
- accidents with motor cyclists (-39%).

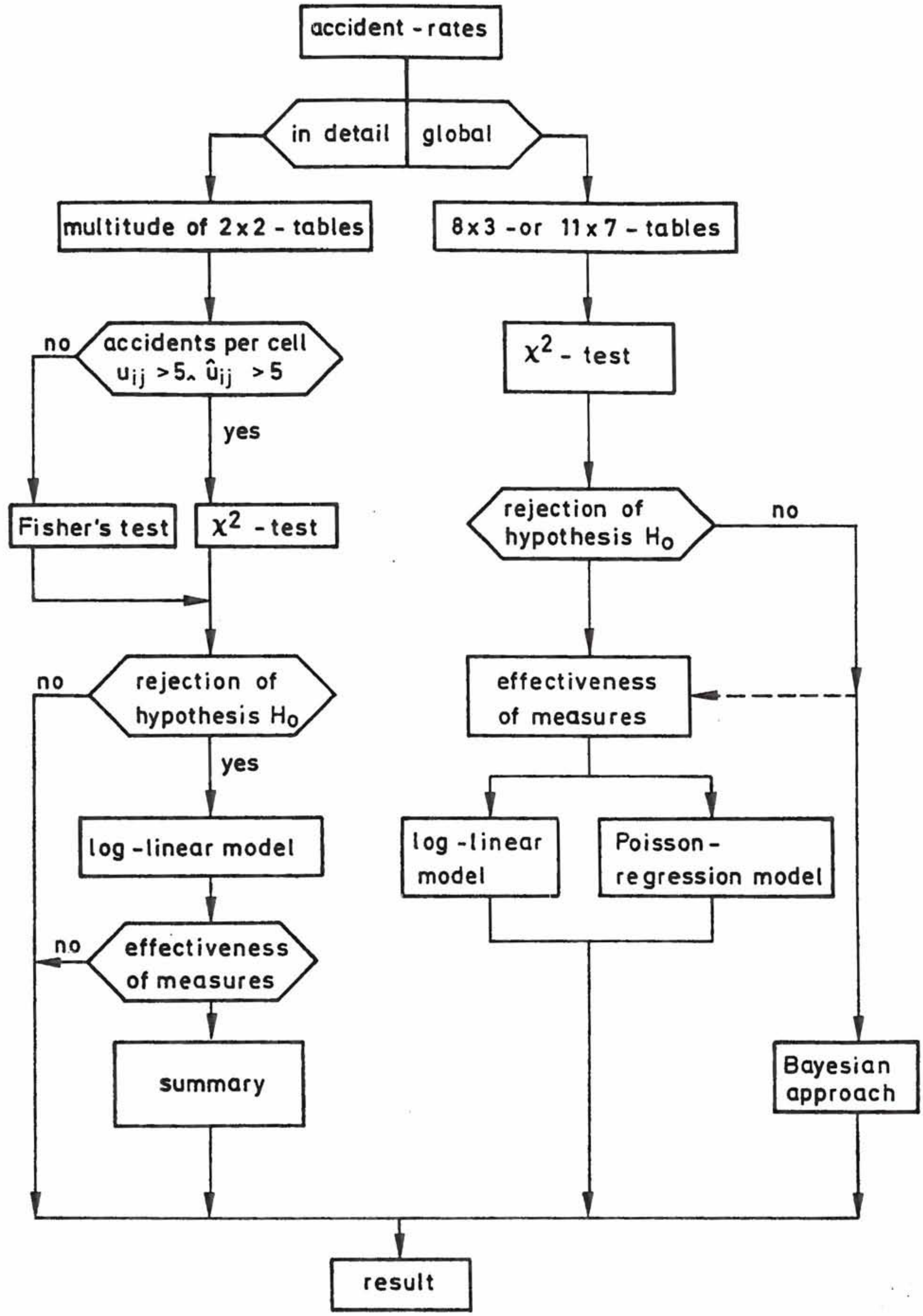
Literature

Brilon, W., Kahrmann, B., Senk, W., Thiel, R., Werner, H.: Area-wide traffic calming measures: Accident-analysis Berlin-Charlottenburg 1985. Research report of the BASt, No. 125. Bergisch Gladbach: 1985.



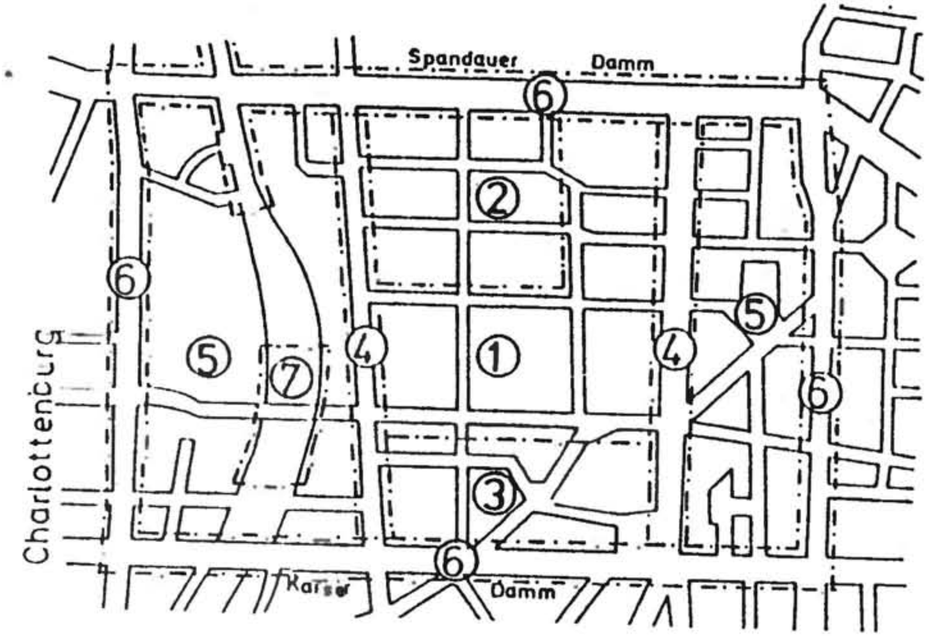
Appendix 1	Flächenhafte Verkehrsberuhigung Berlin - Charlottenburg	Lehrstuhl für Verkehrswesen I Ruhr-Universität-Bochum

plan for accident - analysis



Appendix 2	Flächenhafte Verkehrsberuhigung Berlin - Charlottenburg	Lehrstuhl für Verkehrswesen I Ruhr - Universität Bochum

INVESTIGATION AREA BERLIN-CHARLOTTENBURG



(1) modification of streets

(2) speed limit of 30 km/h

(3) traffic calming measures in neighbourhood

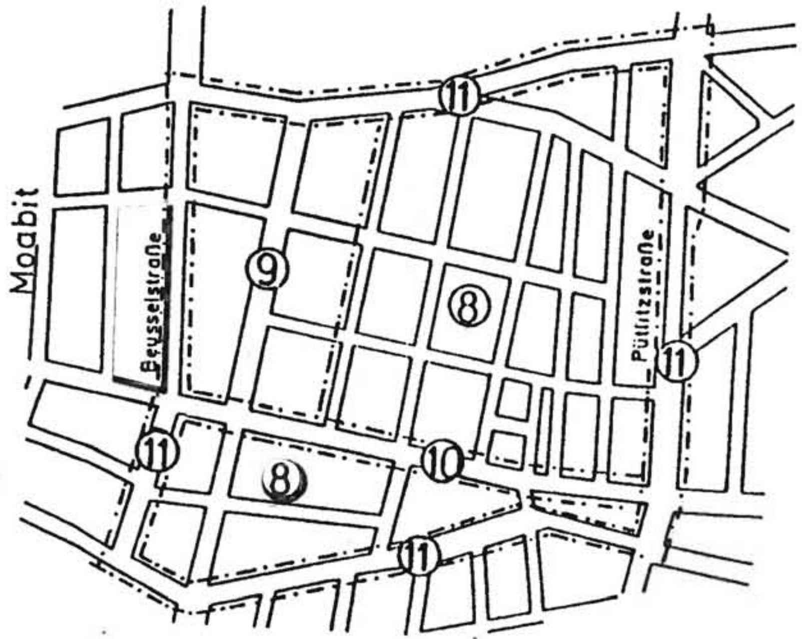
(4) neighbouring arterials

(5) neighbourhood areas

(6) bordering arterials in Charlottenburg

(7) access points to urban motorway

CONTROL AREA BERLIN-MOABIT



(8) residential area

(9) sporadic traffic calming measures

(10) the arterial "Turmstraße"

(11) bordering arterials in Moabit

Appendix 3	Flächenhafte Verkehrsberuhigung Berlin - Charlottenburg	Lehrstuhl für Verkehrswesen I Ruhr - Universität Bochum

X^2 - Test

Abbreviations ($1 \le i \le I$; $1 \le j \le J$):

u_{ij}: accident-rate in cell (i,j)
p_{ij}: probability that an accident will be observed in cell (i,j)

Expected Values of u_{ij}:

$$\hat{u}_{ij} = \frac{u_{i.} * u_{.j}}{u_{..}} \quad \text{with}$$

$$u_{i.} = \sum_{i=1}^{I} u_{ij} \qquad u_{.j} = \sum_{j=1}^{J} u_{ij} \qquad u_{..} = \sum_{i=1}^{I} \sum_{j=1}^{J} u_{ij}$$

Hypothesis H_0:

$p_{ij} = p_{i.} * p_{.j}$

i.e.: if hypothesis H_0 is true, then the rows and columns of the (I×J)-contingency-table are stochastically independent

X^2 - Statistic:

- for (I×J)-contingency tables: $$T = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(u_{ij} - \hat{u}_{ij})^2}{\hat{u}_{ij}}$$
- especially for (2×2) tables: $$T = \frac{u_{..} * (u_{11} u_{22} - u_{12} u_{21})^2}{u_{1.} u_{2.} u_{.1} u_{.2}}$$

Critical Region:

- if $T > X^2_{\alpha, (I-1)(J-1)}$, then hypothesis H_0 has to be rejected
- $X^2_{\alpha, (I-1)(J-1)}$ is the α-quantile of the X^2- distribution with (I-1)(J-1) degrees of freedom (the exact value can be found in most statistical books)

Result of the Test:

- if hypothesis H_0 has to be rejected, then the accident rates have significantly changed
- there is no hint due to which factors hypothesis H_0 has to be rejected

Log-linear Model

Abbreviations:

p_{ij}: probability, that an accident will happen in the i-th area during the j-th period of time $(1 \le i \le I;\ 1 \le j \le J)$

u_{ij}: accident-rate in cell (i,j)

u: number of all accidents observed

Log-Linear Model:

$$p_{ij} = e^{\lambda'}\, e^{\lambda_i^A}\, e^{\lambda_j^B}\, e^{\lambda_{ij}^{AB}}$$

λ' : general effect

λ_i^A : effect of the area

λ_j^B : effect of the time period

λ_{ij}^{AB} : interaction of area and time effects

Expected Accident Rates:

$$\hat{u}_{ij} = u \cdot p_{ij} = u \cdot e^{\lambda' + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}}$$

Linearisation:

$$\ln(\hat{u}_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, \quad \text{with} \quad \lambda := \ln(u) + \lambda'$$

Maximum-Likelihood Estimators for the Effects:

$$\hat{\lambda} = \frac{1}{I \cdot J} \sum_{i=1}^{I} \sum_{j=1}^{J} \ln(u_{ij})$$

$$\hat{\lambda}_i^A = \frac{1}{J} \sum_{j=1}^{J} \ln(u_{ij}) - \hat{\lambda}$$

$$\hat{\lambda}_j^B = \frac{1}{I} \sum_{i=1}^{I} \ln(u_{ij}) - \hat{\lambda}$$

$$\hat{\lambda}_{ij}^{AB} = \ln(u_{ij}) - \hat{\lambda}_i^A - \hat{\lambda}_j^B - \hat{\lambda}$$

Result:

If there are significant changes in the accident rates and there are traffic calming measures in the i-the area and the j-th period of time, then:

$$\lambda_{ij}^{AB} \begin{Bmatrix} < 0 \\ = 0 \\ > 0 \end{Bmatrix} \rightarrow \text{measures} \begin{Bmatrix} \text{cause a decrease in accident rates} \\ \text{are irrelevant} \\ \text{cause an increase in accident-rates} \end{Bmatrix}$$

Poisson-regression model

Abbreviations:

u_{ij}: Poisson-distributed accident-rate in the i-th local area and the j-th period of time $(1 \le i \le I;\ 1 \le j \le J)$

$\hat{u}_{ij}$: expected value of u_{ij}

the Model:

$$\hat{u}_{ij} = \exp\left\{ \sum_{k=0}^{6} \lambda_k \, x_k^{ij} \right\} \qquad \text{with}$$

$$x_k = \left\{ \begin{matrix} 1, \\ 0, \end{matrix} \text{ factor is } \begin{matrix} \text{present} \\ \text{absent} \end{matrix} \right\}$$

λ_k = (unknown) weight of factor x_k

Linearisation and Lexicographic Ordering of u_{ij}, x^k_{ij} and λ^k:

representation of the model by vectors $\underline{\hat{u}} = X \cdot \underline{\lambda}$

Least-square Estimator of $\underline{\lambda}$:

$\lambda = (X^T X)^+ X^T \cdot u$ with

X^T: transposed of matrix X

$(X^T X)^+$: Moore-Penrose-Inverse of matix $(X^T X)$

Baysian Approach

Abbreviations:

- u: accident-rate
- $\underline{u} := (u_1, \ldots, u_n)$ sample of accident-rates
- $\overline{u}$: sample mean
- s: sample deviation

A-priori Distribution:

- U is Poisson-distributed with parameter μ, this means

$$\Psi(u) = P[U = u] = \frac{\mu^u}{u!} e^{-\mu} \quad \text{with } E(U) = \mu \text{ and } Var(U) = \mu$$

- the parameter μ is Γ-distributed

$$\Psi(\mu) = \frac{n^p}{\Gamma(p)} \mu^{p-1} e^{-n\mu}$$

A-priori Estimators for n, p, μ:

$$n_o = \overline{u} / s^2 \qquad p_o = (\overline{u}/s)^2 \qquad \mu_o = p_o / n_o$$

A-posteriori Distribution:

$$\Psi(u) = \frac{\Gamma(p_o + n)}{\Gamma(p_o)\,\Gamma(u+1)} \left(\frac{n_o}{n_o + n}\right)^{p_o} \left(\frac{n}{n_o + n}\right)^u$$

A-posteriori Estimator for μ:

$$\mu_1 = \frac{n_o \mu_o + n \overline{u}}{n_o + n}$$

Confidence Interval for μ_1:

$$0 < \mu_1 < \frac{X^2_\alpha}{2(n_o + n)}$$

X^2_α : the α-Quantile of the X^2-distribution with $2(p_o + n\overline{u})$ degrees of freedom

Summary

		zone 1			zone 2			zone 3			zone 4			zone 5			zone 6		
		total	intersec	section	total	intersec.	section	total	intersec.	section	total	intersec.	section	total	intersec	section	total	intersec	section
	number of accidents	•	•	0	0	0	•	0	-	•	••	••	0	-	-	-	0	0	-
injured persons	seriously injured or dead	0	=	0	0	0	=	-	=	-	•	••	-	-	=	-	•	••	0
	slightly injured	••	••	0	0	0	-	-	-	-	••	••	0	0	0	=	0	0	0
	children	••	-	••	-	0	-	-	-	-	0	0	0	-	0	-	0	0	0
worst defect	seriously injured or dead	0	=	•	0	0	=	-	0	-	0	••	-	-	=	-	0	0	-
	slightly injured	0	•	0	0	0	-	0	0	=	••	=	0	0	•	=	0	0	-
	damage to property	0	•	0	0	-	•	0	-	•	••	••	0	-	-	-	0	0	-
road users	vehicles	••	••	•	•	-	••	0	-	•	••	••	0	-	-	-	0	0	-
	passenger car	••	••	••	•	-	••	0	-	•	••	••	0	-	-	-	0	0	-
	bike	•	•	0	=	0	-	-	-	-	0	-	••	-	0	-	-	--	-
	pedestrians	+	0	••	0	0	-	0	=	0	0	•	0	=	0	-	0	••	0
kind of accidents	turning and crossing	••	••	=	0	-	+	=	0	-	•	••	-	--	-	-	-	-	0
	pedestr. crossing the road	0	=	•	-	=	-	=	=	=	0	0	-	-	0	-	•	••	0
	parked vehicles	-	-	-	0	-	0	-	-	0	•	+	0	0	0	0	0	••	-
	parking vehicles	••	0	••	••	=	••	-	=	-	-	-	=	0	0	=	=	-	0
	following traffic	0	0	-	0	0	0	0	0	0	•	•	0	0	=	-	-	0	-
	oncoming traffic	-	--	0	=	0	-	0	0	0	=	=	=	0	0	=	-	-	-
	accidents per km and year	0			0			0			•			-			0		
	accidents per 1000 residents and year	0			0			0			0			-			0		

Symbols

- •• significant decrease in accident rates ($\alpha = 0.01$; X^2 - test)
- • significant decrease in accident rates ($\alpha = 0.05$; X^2 - test)
- \+ significant decrease in accident rates ($\alpha = 0.05$; Fisher's exact test)
- 0 decrease in accident rates, but not significant decrease
- \- increase in accident rates, but not significant increase
- -- significant increase of accident rates ($\alpha = 0.01$; X^2 - test)

A MODEL FOR EVALUATING EDUCATIONAL ROAD SAFETY MEASURES

Dr. J.A. Rothengatter,
Traffic Research Centre, University of Groningen

Abstract

The purpose of evaluating educational road safety measures is two fold. Firstly evaluation research serves as a method for obtaining information that can aid the development process of these measures. Secondly, evaluation research can provide information about the potential effects of educational road safety measures both in terms of behavioural changes and in terms of accident reduction. In practice, these two purposes are often confused, which leads to use of inappropriate evaluation methods and hence to incorrect conclusions regarding the development and implementation of road safety education programmes.
This paper presents a recently developed model for evaluating educational countermeasures. The model distinguishes process and product evaluation and outlines a sequential approach in terms of a number of discrete stages. For each of these stages the suitable research methodology is specified in terms of objectives, methods and conclusion validity. Examples of recent evaluation studies of educational programmes will be analysed to illustrate the use of the model, and it will be demonstrated how a stringent use of the model can improve both the development process and the decision making regarding the implementation of education measures.

ACCIDENT COUNT ANALYSIS:

THE CLASSICAL AND ALTERNATIVE APPROACHES

A J Nicholson
Senior Lecturer in Civil Engineering
University of Canterbury
Christchurch
New Zealand

Visiting Fellow, Institute for Transport Studies
University of Leeds
United Kingdom

ABSTRACT

The classical approach to estimating accident rates, and to testing the statistical significance of changes in accident rate, involves interpreting accident count data relating to a specific site over an extended period of time. An alternative approach, involving the analysis of accident data relating to groups of sites over a shorter period, has been proposed. This paper describes both approaches, discusses their strengths and weaknesses, and suggests avenues for further research.

INTRODUCTION

Much of the recent literature on accident analysis has been focussed on two problems:

(1) the identification of hazardous locations (or blackspots), and

(2) the estimation of the effectiveness of treatment.

Both involve estimation of what may be termed the "underlying true accident rate" (or UTAR); hazardous location identification requires estimation of the current UTAR only, while treatment effectiveness estimation requires estimation of the UTAR both before and after treatment.

It should be noted that the underlying true accident rate (UTAR) is not known with certainty, and is almost certainly not equal to the number of accidents observed per unit time (or per exposure). The observed number of accidents is merely an indication of the UTAR, which can only be estimated on the basis of observations.

Accidents are relatively rare, and are subject to both temporal and spatial variations; at a site where the UTAR is not changing, there is generally considerable variation in the annual accident counts about the UTAR, while it is generally accepted that at a point in time the UTAR varies from one location to another. In reality, it may well be that the UTAR for each specific location is varying with time.

When analysing accident count data for many sites over several years (see Figure 1), it must be remembered that a mixture of spatial and temporal variations underly the count data, and it is a difficult task to separately identify those variations, in order to identify hazardous locations and estimate treatment effectiveness accurately.

Years

		1	2	...	j	...	J
Sites	1	x_{11}	x_{12}	...	x_{1j}	...	x_{1J}
	2	x_{21}	x_{22}	...	x_{2j}	...	x_{2J}
	.	.	.		.		.
	.	.	.		.		.
	i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{iJ}
	.	.	.		.		.
	.	.	.		.		.
	I	x_{I1}	x_{I2}	...	x_{Ij}	...	x_{IJ}

Figure 1: Matrix of accident counts for I sites and J years.

The classical approach to the problem entails analysing the data for each site separately, in order to estimate the UTAR for each (or the pattern of variation of the UTAR, if it is not constant). One can then identify the sites with an unusually high UTAR (blackspots), or detect whether there has been a change in the UTAR since treatment. The longer the time period for which accident count data is available, the more precise the estimate of the UTAR (assuming it is constant). If the UTAR is changing, then the pattern of variation of the UTAR can be identified more accurately as the time period increases.

Road safety work is invariably undertaken in less than ideal circumstances, there being considerable pressure upon researchers and practitioners to adopt procedures which permit responses or results in a short time. For instance, a sudden spate of accidents at a site may lead to intense public pressure for immediate remedial treatment, and the practising traffic engineer will have difficult convincing the public (or their elected representatives) that any action should be deferred until it is known with a reasonable level of confidence that the spate of accidents did indeed indicate an increase in the UTAR, or is merely confirmation of the stochastic nature of accident occurrence. Similarly, there is often a demand for a quick assessment of the effect of some change to the road environment upon the accident rate. The development of the traffic conflicts technique is a reaction to this pressure, as is the development of statistical analysis procedures involving the analysis of accident data for groups of sites over a shorter time period.

THE CLASSICAL APPROACH

Estimating the Underlying True Accident Rate

Consider the case of x_{i1}, x_{i2}, ..., x_{in} accidents in n years at a single site, i. If it is assumed that the annual accident counts are governed by a stationary Poisson process, the mean of which is the UTAR α_i, then one can derive confidence limits for α_i.

If the accident counts are Poisson-distributed with mean α_i, then the sum of the counts is also Poisson-distributed (with mean $n\alpha_i$). Since the cumulative sum of the Poisson distribution is related to the cumulative Chi-square distribution, it follows that, with a level of confidence of (1-2k),

$$B_l < \alpha_i < B_u$$

where

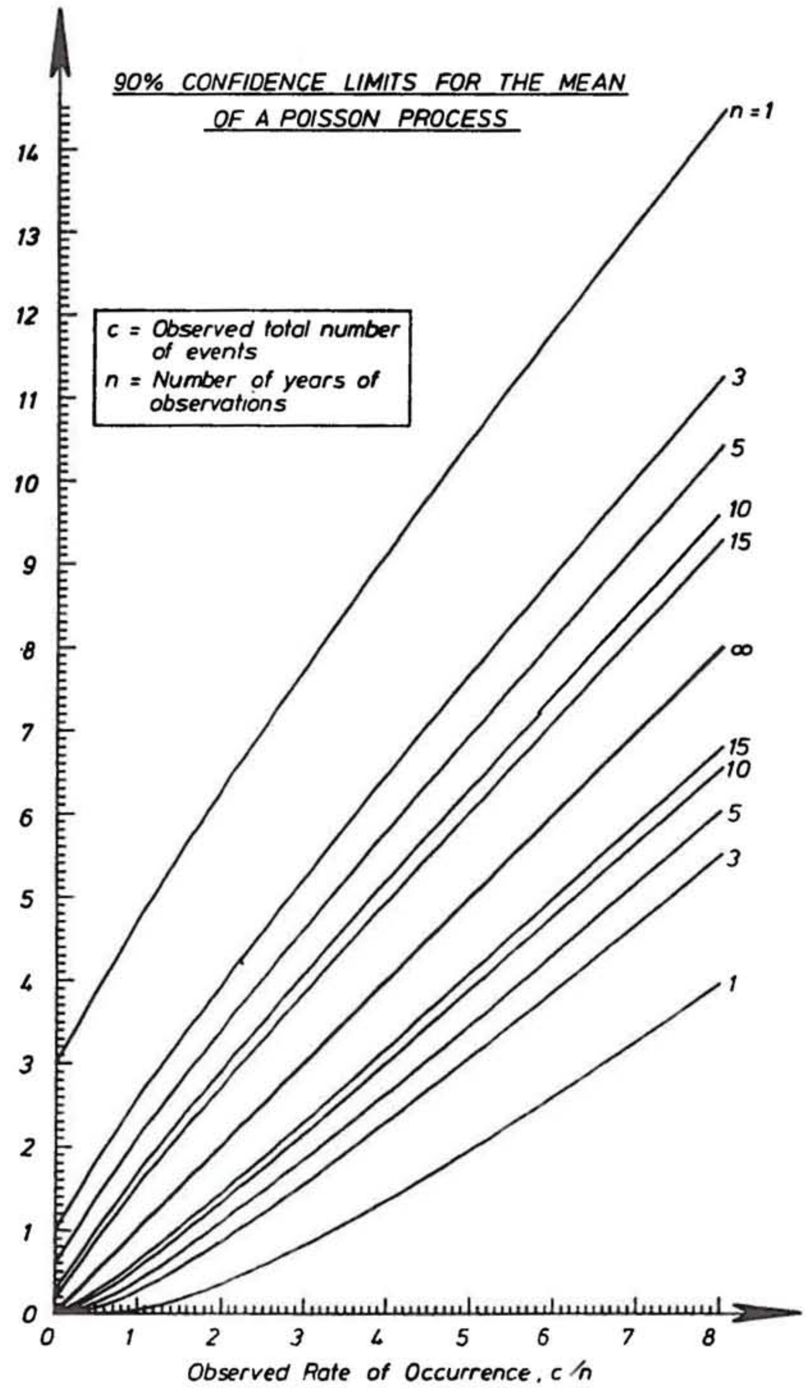
$$B_l = \chi^2 (k | \nu_i = 2c_i) \ / \ (2n)$$

$$B_u = \chi^2 (1-k | \nu_i = 2c_i + 2) \ / \ (2n)$$

and

$$c = \sum_{j=1}^{n} x_{ij}$$

Using these relationships and tables of the percentage points of the χ^2 distribution for integral and fractional degrees of freedom (Pearson and Hartley, 1976), graphs of confidence limits for the **UTAR** (α_i), for various values of the observed rate of accident occurrence (c_i/n) and time period (n), have been derived (Nicholson, 1987). An example is shown in Figure 2.



<u>Figure 2</u>: 90% Confidence Limits for the Underlying True Accident Rate

The width of the confidence interval for the estimate of the UTAR reduces as the number of years of observation increases, as shown in Table 1. Clearly, the rate of improvement in precision decreases as the period of observation increases. A graph of the width of the confidence interval (as a percentage of the observed accident rate) versus observation period (Nicholson, 1986) reveals that in the vicinity of n = 5, there is a marked decrease in the rate of improvement in precision as the observation period increases.

Total no. of accidents c	No. of years n	B_l	B_u	$\frac{B_l - B_u}{(c/n)}$
5	1	2.0	10.6	172 %
15	3	3.1	7.7	92 %
25	5	3.5	7.0	70 %
50	10	3.9	6.3	48 %
75	15	4.1	6.0	38 %

Table 1: Variation in width of 90% confidence interval with increasing observation period

It seems, from the viewpoint of statistical reliability, that five years is about the optimum time period for estimation of the UTAR. It might be argued that five years is too long a time period, in that it would prevent the quick detection of sudden changes in the UTAR, and many roading/highway authorities use a much shorter period (Zegeer, 1982; Silcock and Smyth, 1984). Such an argument implies that annual accident counts are governed by a non-stationary stochastic process. The procedure described above is based upon the assumption that the mean and variance of the accident process are constant and equal. Clearly, if non-stationarity is assumed, a greater observation period is required to identify the form of variation of the mean and/or variance of the accident process (and, hence, the UTAR at some point in time) than if non-stationarity is assumed.

Testing the Significance of Accident Rate Changes

Consider now the case of $x_{i1}, x_{i2}, \ldots, x_{in}$ accidents in n years before some change (remedial treatment, say) and $y_{i1}, y_{i2}, \ldots, y_{im}$ accidents in the m years afterwards. Assuming that the accident counts are Poisson-distributed, with means α_i and β_i "before" and "after" respectively, then the corresponding accident totals X and Y are also Poisson-distributed, with means $n\alpha_i$ and $m\beta_i$ respectively. According to Feller (1971), the probability distribution for the difference in accident totals is given by:

$$P[X-Y=d] = \exp(-n\alpha_i - m\beta_i)\,(n\alpha_i / m\beta_i)^{d/2}\, I(\alpha_i, \beta_i, m, n, d)$$

where

$$I(\alpha_i, \beta_i, m, n, d) = \sum_{\lambda=0}^{\infty} (\lambda!(\lambda + |d|)!)^{-1} (mn\alpha_i\beta_i)^{(2\lambda + |d|)/2}$$

is a modified Bessel function.

In this situation, one is interested in estimating the probability that the observed difference in the accident totals is due to chance, assuming that the UTAR "before" is equal to the UTAR "after" (i.e., $\alpha_i = \beta_i$). If it is also assumed that the observation periods are equal then the above expression can be simplified, to give:

$$P[X-Y=d] = \exp(-2c_i) \sum_{\lambda=0}^{\infty} (\lambda!(\lambda + |d|)!)^{-1} c_i^{(2\lambda + |d|)}$$

where

$$c_i = n\alpha_i = m\beta_i$$

Using this expression, the discrete density function of (X-Y) can be calculated, from which graphs of the critical change in accident rate, for various values of the observed rate of accident occurrence and time period, have been derived (Nicholson, 1987). An example is shown in Figure 3.

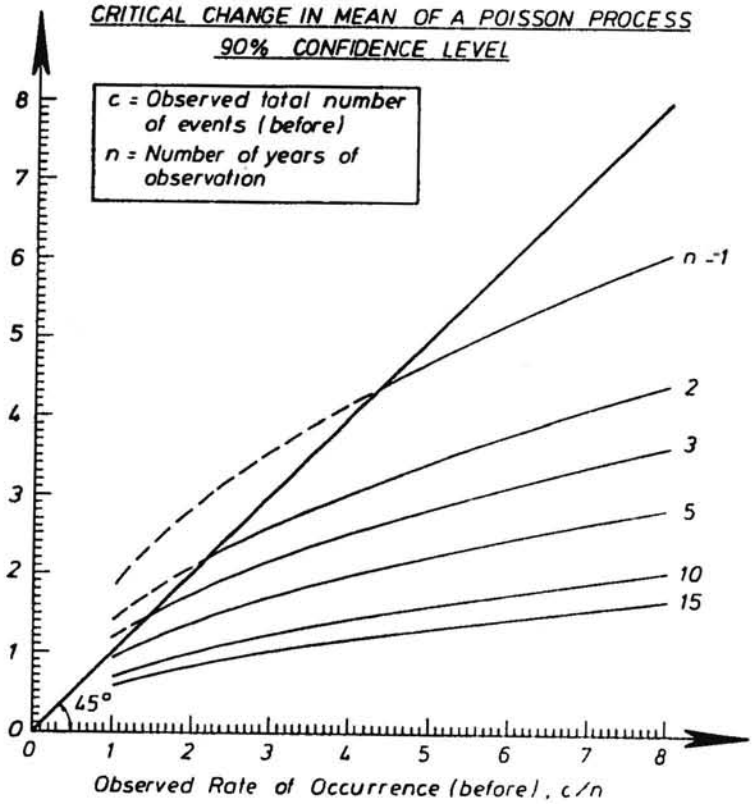


Figure 3: Critical Change in Accident Rate (90% Confidence Level)

The greater the number of years of observation, the smaller the required change in accident rate for statistical significance, as shown in Table 2. Clearly, the required change in accident rate decreases as the period of observation increases, with five years again appearing to be about optimum from the viewpoint of statistical reliability.

Total number of accidents "before" c	Number of years n	Critical change	Critical Change (c/n)
5	1	4.7	94 %
10	2	3.4	68 %
15	3	2.8	56 %
25	5	2.2	44 %
50	10	1.6	32 %
75	15	1.3	26 %

Table 2: Variation in critical change with increasing observation period (90% confidence level)

Depending upon the time period and the confidence level, the change in accident rate required for statistical significance may exceed the accident rate "before" (i.e. it may not be feasible to achieve statistical significance). The "zone of infeasibility" is shown in Figure 3.

When comparing the means of two stationary Poisson processes, it is easier to make inferences about the ratio of the means than the difference. If one is interested in testing the statistical significance of the difference in the UTAR's, Cox and Lewis (1966) suggest that one "shall almost always have to fall back on large-sample approximations". Such approximations are very often inappropriate for accident data analysis; the above-described test does not involve such approximations, and may be applied to the analysis of small numbers of accidents.

It may be thought satisfactory to estimate the confidence intervals for the UTAR's before and after a change, and if there is no overlap conclude that there has been a statistically significant change in the UTAR. Hence, if one observed 25 and 11 accidents in the five years both before and after a change, then the corresponding 90% confidence intervals would be 3.5 to 7.0 and 1.2 to 3.6 accidents per year (Figure 2), and given that they do overlap, it might be concluded that there has not been a statistically significant change in the UTAR. It can be seen from Figure 3, however, that the critical change in accident rate is 2.2, and since the observed change is 2.8 accidents per year, then it can be concluded that there has been a statistically significant change. Clearly, the rigorous test is more reliable; the use of the simple test gives a bias towards incorrectly concluding that a treatment has not had a statistically significant effect.

THE ALTERNATIVE APPROACH

The Non-Parametric and Empirical Bayesian Methods

Consider the case of x_{1j}, x_{2j}, ... , x_{nj} accidents in year j at n sites. The sites may be ranked according to their accident counts in year j, and those with a high ranking (i.e. a high accident count in year j) may be selected for treatment because they seem unusually hazardous in comparison with the other sites. Due to the stochastic nature of accident occurrence, the sites with a high ranking based on data for one year may not have a high ranking if data for another year is used. In fact, the sites with an above-average accident count in one year will tend to have a lower accident count in the next year. The nature and extent of the regression-to-the-mean effect is discussed in detail by Hauer (1980), who has clearly shown that if

(1) sites are selected for treatment because of a history of many accidents, and

(2) the regression-to-the-mean effect is ignored

then the effectiveness of the treatment will be exaggerated.

According to Persaud and Hauer (1984), there are two analytical methods which may be used to correct estimates of treatment effectiveness, in situations where control groups have not been established. They are

(1) the non-parametric (NP) method, and

(2) the empirical Bayesian (EB) method.

Both methods are aimed at providing an estimate of the total number of accidents that would have occurred at the group of sites selected for treatment had they not been treated; this estimate can be compared with the observed total number of accidents at the sites after treatment, in order to obtain an unbiased estimate of the overall effect of the treatment.

The non-parametric (NP) method is based upon the assumption that the number of accidents at each individual site is governed by a stationary Poisson process. It has been shown (Hauer, 1980; Hauer and Persaud, 1982) that if sites which during a period of time had k or more accidents are selected for treatment, then

$$A(k) = B(k + 1)$$

where

A(k) = the expected total number of accidents at the selected sites for an equivalent period after treatment, if the treatment has no effect and

$B(k+1)$ = the actual total number of accidents at those sites having (k+1) or more accidents during the before-treatment period.

Persaud and Hauer (1984) stated the NP method as follows:

$$a(k) = [(k+1)\ N_{k+1}] \ /\ N_k$$

where

$a(k)$ = the expected number of accidents during an equivalent after-treatment period at a site that had k accidents in the before-treatment period

N_k = the number of sites having k accidents in the before-treatment period.

They also stated that the empirical Bayesian (EB) method, first proposed by Abbess, Jarrett and Wright (1981), could be written in the similar form

$$a(k) = [(k+1)\ N^*_{k+1}\ ] \ /\ N^*_k$$

where

N^*_k = the number of sites expected to have k accidents in the before-treatment period.

The EB method, as proposed by Abbess, Jarrett and Wright, involved two assumptions:

(1) that the number of accidents at each individual site during a year, say, is governed by a stationary Poisson process, and,

(2) that the means of the Poisson processes varies between sites, according to a Gamma distribution.

Hence, the number of sites expected to have k accidents in a year, say, is given by the Negative Binomial distribution.

Abbess, Jarrett and Wright examined the actual distributions of annual accident numbers at blackspots and concluded that the Negative Binomial distribution gave "a reasonable fit". They also noted that "there tends to be more sites with zero accidents than one would expect" (about 33% had zero accidents). They therefore tried a truncated Negative Binomial, excluding sites with zero accidents, and claimed to have obtained a "good fit", although they did not give any goodness-of-fit statistics.

Andreassen and Hoque (1986) reported that a very high proportion (about 93%) of all intersections in Melbourne have zero accidents in a year, and they also chose to exclude the zero accident category. They concluded that the truncated Negative Binomial distribution was unsuitable, because the parameter estimation procedure gave a negative value for one parameter. They did conclude that the observed distribution of annual accident counts was well described by the logarithmic series distribution. Maher (1987a) has subsequently claimed that a negative parameter value is quite acceptable and that the truncated Negative Binomial gives a much better fit to the Melbourne data than does the logarithmic series distribution.

It is hard to imagine any p.d.f. providing a good fit to an observed accident count frequency distribution for all locations in a large network, as very many locations, most of which are low in the roading hierarchy (e.g. low volume roads/intersections in residential areas), experience zero accidents in any given year. Even if low-hierarchy locations are omitted, it must be remembered that a good fit of the Negative Binomial distribution to observed accident count data for the other locations does not mean that the Poisson and Gamma distributions are appropriate. It has traditionally been assumed that accident counts at a site are governed by a Poisson process, and the choice of the Gamma distribution is really one of mathematical convenience, as it is the natural conjugate of the Poisson distribution. Hauer and Persaud (1982) refer to the assumption of the Poisson distribution as being "empirically unproven", and there is evidence (Nicholson, 1985) that it is not generally valid, as some locations have either too much or too little variance in their accident counts for the Poisson distribution.

Persaud and Hauer compared the performance of the NP and EB methods for debiasing estimates of countermeasure effectiveness, using a large number and variety of data sets, and concluded that the EB method generally performed better and "should be used in assessing the safety effect of a treatment". They did, however, note that for sites having zero or one accident, the NP method gave slightly better results; this was probably due to the tendency to underestimate the number of sites having zero or one accidents when using the Negative Binomial distribution. It should be noted that Persaud and Hauer did not employ statistical tests in making the comparison, but relied upon graphical and numerical descriptive measures only.

One problem associated with the NP method (Hauer, 1980; Abbess, Jarrett and Wright, 1981) is that the estimate of the bias is unreliable when the number of systems treated is small. This is due to the small numbers of sites having x accidents in a year, and the large variation (from year to year) in the numbers of sites having x accidents, when x is large. Hence, the sequence of values for a(k), k = 0, 1, 2, ... , can exhibit considerable random noise. In a recent paper (Hauer, 1986), a procedure for reducing the random noise is described. It involves fitting a function to the calculated, unsmoothed a(k) in order to obtain a smoothed sequence of values for a(k).

A New Method

Another problem associated with the NP method noted by Abbess, Jarrett and Wright , is that the method can be applied to groups of sites only. Hauer (1986) has subsequently proposed a procedure for estimating the number of accidents expected to occur annually at a single site, given accident count data for several sites over several years, as follows:

$$E(x_i) = \bar{x}_{i.} + [\bar{x}/(J(s^2 - \bar{x}) + \bar{x})][\bar{x} - \bar{x}_{i.}]$$

where

$E(x_i)$ = expected annual number of accidents at site i

x_{ij} = the number of accidents at site i during year j

I = number of sites

J = number of years

$$\bar{x}_i = \sum_{j=1}^{J} x_{ij}/J$$

$$\bar{x} = \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}/(IJ)$$

$$s^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \bar{x})^2/ (IJ)$$

Like Abbess et al, Hauer assumed that UTAR's are Gamma distributed and annual accident counts for each site are Poisson distributed.

The form of the expression for $E(x_i)$ is such that

(1) the first term is the mean accident count for site i

(2) the second term is an adjustment, which depends upon the spatial and temporal distribution of accidents (over the I sites and J years) and the duration of the accident history.

It is helpful to consider the accident count matrix (Figure 1), and think in analysis of variance terms. It can be seen that

(1) $\bar{x}$ is simply the overall mean accident count.

(2) s^2 is simply the total sum of squares, divided by the total number of accident counts.

(3) the second part of the second term is the difference between the overall mean accident count and the mean accident count for site i.

Hauer suggests that

(1) the first term converges on the UTAR as J increases;

(2) the adjustment term tends towards zero as J gets large, so that the more information one has about a particular site, the less is the effect of using data for other sites.

An analysis of the expression for the expected annual number of accidents at a particular site (see the Appendix) reveals that the adjustment term may not reduce to zero as J increases.

Hauer also investigated the error associated with the above procedure for estimating $E(x_i)$. He suggested that the mean-square-error is comprised of two parts:

(1) the difference between the underlying true accident rate (UTAR) at site i from the mean of the UTAR's for the group of sites having the same total number of accidents;
(2) the difference between the estimated accident rate at site i and the UTAR at site i.

Using his data and estimation procedure (as described above), Hauer found that the first component of the mean-square-error was very much larger (at least one order of magnitude) than the second component. The first component arises from the grouping together of sites and Hauer suggests that the only way to reduce this component is by judiciously changing the criteria for deciding whether sites are sufficiently similar to be grouped together.

APPLICATION OF BOTH APPROACHES

Consider the case of ten sites for which there is five years of accident count data, the accident count matrix being as shown in Figure 4. Row and column totals and means are also shown.

		Years						
		1	2	3	4	5	Row Total	Row Mean
	1	5	2	1	4	3	15	3
	2	5	7	4	6	3	25	5
	3	3	1	4	0	2	10	2
	4	3	5	1	7	4	20	4
Sites	5	0	6	2	4	3	15	3
	6	0	4	1	2	3	10	2
	7	1	2	0	1	1	5	1
	8	4	3	1	5	2	15	3
	9	3	5	4	2	6	20	4
	10	1	5	2	4	3	15	3
Column Total		25	40	20	35	30	150	
Column Mean		2.5	4.0	2.0	3.5	3.0		

Figure 4: Example accident count matrix

For this case,

(1) the total sum of squares = 172

(2) the overall mean accident count $\bar{x}$ = 3.0

(3) the variance s^2 = 3.44

so the correction term will reduce as the number of years is increased.

Table 3 shows the results of applying Hauer's estimation procedure to the year 1 data, the years 1 and 2 data, and so on. The values of $\bar{x}_i$ and $E(x_i)$ are both shown, and it can be seen that the absolute magnitude of the adjustment term is quite substantial even after five years of data; for site 7, after five years, it is still greater than $\bar{x}_i$. It is also evident that the effect of the adjustment term is to give much less variance in the values of $E(x_i)$, i = 1, ... , 10, than in the values of $\bar{x}_i$.

	Number of years of observation				
Site	1	2	3	4	5
1	5, 3.08	3.5, 3.32	2.67, 2.76	3.00, 3.00	3.00, 3.00
2	5, 3.08	6.0, 4.03	5.33, 3.96	5.50, 4.36	5.00, 3.85
3	3, 2.62	2.0, 2.90	2.67, 2.76	2.00, 2.45	2.00, 2.58
4	3, 2.62	4.0, 3.46	3.00, 2.91	4.00, 3.55	4.00, 3.42
5	0, 1.92	3.0, 3.18	2.67, 2.76	3.00, 3.00	3.00, 3.00
6	0, 1.92	2.0, 2.90	1.67, 2.31	1.75, 2.32	2.00, 2.58
7	1, 2.15	1.5, 2.76	1.00, 2.01	1.00, 1.91	1.00, 2.15
8	4, 2.85	3.5, 3.32	2.67, 2.76	3.25, 3.14	3.00, 3.00
9	3, 2.62	4.0, 3.46	4.00, 3.36	3.50, 3.27	4.00, 3.42
10	1, 2.15	3.0, 3.18	2.67, 2.76	3.00, 3.00	3.00, 3.00

Table 3: Values of $\bar{x}_i$ and $E(x_i)$ for J = 1, 2, ... , 5

Table 4 shows the results of using the classical approach, embodied in Figure 2, to estimate the 90% confidence intervals for the UTAR's after 1, 3 and 5 years. As expected, those confidence intervals are generally reduced considerably by the use of data for a longer time period.

Site	Number of years of observation 1	3	5
1	5, (2.0-10.6)	2.67, (1.3-4.8)	3.0, (1.9-4.7)
2	5, (2.0-10.6)	5.33, (3.4-8.1)	5.0, (3.5-7.0)
3	3, (0.8- 7.8)	2.67, (1.3-4.8)	2.0, (1.1-3.4)
4	3, (0.8- 7.8)	3.00, (1.6-5.3)	4.0, (2.7-5.9)
5	0, (0.0- 3.0)	2.67, (1.3-4.8)	3.0, (1.9-4.7)
6	0, (0.0- 3.0)	1.67, (0.7-3.5)	2.0, (1.1-3.4)
7	1, (0.1- 4.8)	1.00, (0.3-2.7)	1.0, (0.4-2.1)
8	4, (1.4- 9.2)	2.67, (1.3-4.8)	3.0, (1.9-4.7)
9	3, (0.8- 7.8)	4.00, (2.3-6.5)	4.0, (2.7-5.9)
10	1, (0.1- 4.8)	2.67, (1.3-4.8)	3.0, (1.9-4.7)

Table 4: Best estimates (with 90% confidence lower and upper bounds) for J = 1, 3 and 5.

It is interesting to note that for site 7, the value of $E(x_i)$ is 2.15 (Table 3), and this is outside the 90% confidence limit (0.4 to 2.1).

REGRESSION-TO-THE-MEAN

It is well known that regression-to-the-mean, in combination with the selection of sites for treatment on the basis of high observed accident counts over a short period, can lead to biased estimates of the effect of treatment (Hauer, 1980; Abbess, Jarrett and Wright, 1981).

When considering whether to treat a site, any one of six possible conditions may exist:

(1) $k < \alpha < \hat{\alpha}$

(2) $k < \hat{\alpha} < \alpha$

(3) $\hat{\alpha} < k < \alpha$

(4) $\alpha < k < \hat{\alpha}$

(5) $\alpha < \hat{\alpha} < k$

(6) $\hat{\alpha} < \alpha < k$

where

α = UTAR for the site

$\hat{\alpha}$ = observed accident rate for the site

k = critical accident rate

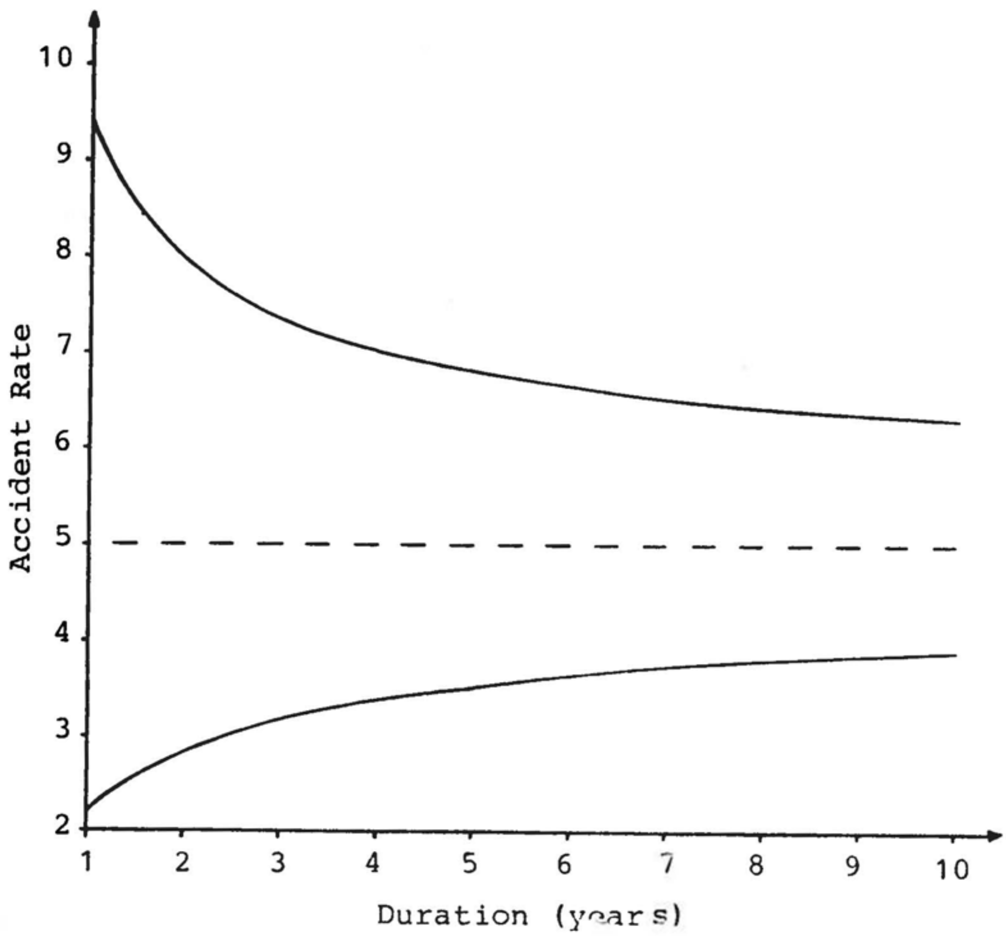
Ideally, the site should be treated if $\alpha > k$ and should not be treated if $\alpha < k$, but in reality α is unknown and is estimated by $\hat{\alpha}$, so that treatment will occur if $\hat{\alpha} > k$ and will not occur if $\hat{\alpha} < k$.

In virtually all discussion of regression-to-the-mean, attention is focussed upon cases 1, 4 and 5, where $\alpha < \hat{\alpha}$ and the regression will be downwards. Abbess et al do mention the possibility of the regression-to-the-mean effect being "completely reversed". For cases 2, 3 and 6, where $\alpha > \hat{\alpha}$, the regression-to-the-mean will be upwards.

If one has a large number of sites under consideration for treatment, then one would expect to have the same number of cases 1 and 2, and of cases 5 and 6. Hence, the regression effects of cases 1 and 2 would be expected to be equal and opposite, and to thus cancel. Likewise, the effects of cases 5 and 6 would be expected to cancel. Case 3 sites should not be selected for treatment and should have no effect, except in certain circumstances as discussed below. Case 4 sites should be selected, with the consequence that there is an expected nett downwards regression effect.

If case 3 sites should happen to be included in the set of control sites, there will be an expected nett upwards regression effect in those sites, and this might be taken as evidence of an accident migration effect. The existence of an accident migration effect has been a matter of considerable debate since it was raised by Boyle and Wright (1984), and Maher (1987b) has suggested that there is a statistical explanation, which seems essentially the same as that given here.

As the observation period increases, one would naturally expect the observed accident rate, $\hat{\alpha}$, to more closely approximate the underlying true accident rate, α , and Figure 2 shows how the confidence interval for α decreases in width as the observation period increases. If one assumes that accident counts are Poisson distributed about a constant UTAR, then one can (from Poisson probability tables or charts) readily derive confidence intervals for $\hat{\alpha}$, for varying observation period length, as shown in Figure 5. The 90% confidence interval for $\hat{\alpha}$, given α = 5, narrows quickly from 2.2 to 9.4 for one year, to 3.5 to 6.8 for five years. Thereafter the rate of narrowing is much less, and after 10 years the 90% confidence interval for $\hat{\alpha}$ is 3.9 to 6.3, or 22% below to 26% above the UTAR.



<u>Figure 5</u>: 90% Confidence Interval for $\hat{\alpha}$ given α = 5.0

Hauer (1980) investigated the effect of both the observation period and the fraction of sites selected for treatment (the higher the critical accident rate, k, the lower that fraction) on the magnitude of the bias. It was assumed that all sites had the same UTAR, and the selection-fraction was assumed to vary from 1% to 50%. It was found that the larger the selection-fraction and the observation period, the smaller was the bias. The assumption of the same UTAR for all sites may well have lead to overestimation of the bias, however, for Abbess, Jarrett and Wright (1981) assumed the UTAR varied between sites according to a Gamma distribution, and obtained substantially smaller estimates of the bias. In addition, whereas Hauer's results suggest that the bias using three and five years is 50%-55% and about 40% (respectively) of the bias using only one year, the results of Abbess et al indicate a much more rapid reduction in bias as the observation period is extended (55%-60% and about 40% of the bias using only one year, for two and three years, respectively).

In a later paper, Hauer and Persaud (1982) again examined the relationship between the magnitude of the regression-to-the-mean and the duration of the observation period. An empirical approach was employed, with a seven year accident history for rural roads in Israel being treated as follows:

(1) the 7th year was regarded as the "after" period,
(2) years 6, 5 and 6, etc. were regarded as "before" periods, of duration 1, 2, etc. years.

Comparison of the accident rates (per year) for years 6, 5 and 6, etc. with that for year 7 suggested that the regression-to-the-mean effect did reduce as the duration of the "before" period was increased, but not as quickly as the results of Abbess et al indicate.

If the accident rates for the "before" period are compared with the overall accident rates for the "before" and "after" periods, it can be seen that the effect of increasing the duration of the "before" period is very marked (see Table 5).

Number of Sections	Before Period (yrs)	Accidents in 7th Year	Mean Accident Rate "Before" A	Overall Mean Accident Rate B	$\frac{A-B}{B}$
337	1	317	530	424	25.1%
258	2	277	393	354	10.9%
231	3	250	321	303	5.9%
191	4	230	292	280	4.4%
178	5	224	272	264	3.0%
170	6	222	258	253	2.0%

Table 5: Regression-to-the-Mean and the Duration Effect.

DISCUSSION

In the frequent references to Sir Francis Galton's observations of the height of offspring relative to that of their progenitors, no mention is made of the ethnic group to which the people belonged. Clearly, were one to group together the pygmies of equatorial Africa with an ethnic group noted for their considerable height, then evidence of regression-to-the-mean would be difficult to find. In such a case, there would not be regression to the overall mean, although there may be regression to the ethnic group means.

Previous discussions of regression-to-the-mean in road safety literature have been in the context of accidents at groups of sites, and the grouping together of sites seems to imply that there is a relationship between the accident processes at those sites. The grouping of sites is often done on an arbitrary basis, and there is no basis for supposing that there is a relationship between the accident process at the sites so grouped, such that there is regression to the group mean.

The classical approach entails looking at data for individual sites, with the accident counts varying about the underlying true accident rate. In this context, an accident count well above or below the underlying true accident (or UTAR) is likely to be followed by a count that is closer to that UTAR. It follows that regression can be upwards or downwards, and the inclusion of sites where upwards regression is likely within a control group may lead to the appearance of accident migration.

The consideration of individual sites shows clearly how one may improve the precision of the estimate of the UTAR by extending the duration of the observation period, thereby reducing the impact of the regression-to-the-mean effect on the estimates of the UTAR's before and after treatment. The alternative approach, involving consideration of groups of sites, also indicates that the regression-to-the-mean effect dimishes as the duration of the observation period increases. There is, however, some discrepancies between the estimates of the effect of increasing the observation period, and this matter needs to be resolved.

Practising traffic safety engineers now seem aware of the regression-to-the-mean effect, and the need now is for advice on how to take proper account of the effect. If they are evaluating a remedial treatment at a site, one option is to simply increase the observation period. Alternatively, they may opt for using data for a large number of sites. The disadvantage of extending the observation period is obvious; the time to detect a change in the UTAR is increased, and the corresponding increase in statistical reliability seems to count for very little. The disadvantage of considering groups of sites is less obvious. The results of Hauer (1986) suggest that the error associated with the grouping together of sites is the dominant one, and it is not one which can be readily quantified by practitioners, who might be better off simply extending the observation period and using the classical approach.

The matter of grouping sites on the basis of their having similar characteristics (including similar underlying true accident rates) needs attention. There seems to be scope for trying to improve the statistical efficiency of the stratification, so that the between-group variance of the UTAR is maximised and the within-group variance of the UTAR is minimised. The greater the ratio of the between-group variance to the within-group variance, the greater the statistical efficiency of the stratification. If the total number of sites is N and the number of groups is k, then the quantity

$$[(\text{between-group variance})/(\text{within-group variance})][(N-k)/(k-1)]$$

is F-distributed, and one can test whether there is a statistically significant relationship between the UTAR and the criteria for grouping sites. Anyway, such an approach should give a reduced standard error for the estimate of the expected annual number of accidents.

Statistical efficiency is, of course, not the only goal. The criteria for grouping sites should be readily applicable by practitioners (e.g. the form of intersection control, the approximate traffic flow), and a compromise is likely to be required.

It has been noted above that the procedure proposed by Hauer (1986), for estimating the expected annual number of accidents, involves terms that have their equivalents in the standard analysis of variance procedures. Given a matrix of accident counts, there seems to be potential for using analysis of variance in order to separately identify spatial and temporal variations in accident occurrence. If the mean accident counts and the variance of those counts for individual sites are proportional (if the accident counts are Poisson distributed, then the mean and variance will be approximately equal), then it is necessary to transform each count (by taking the square root of the count-plus-one-half), in order to satisfy the assumptions upon which standard analysis of variance procedures are based.

If one is treating an area containing many individual sites (e.g. the area studies being undertaken in the UK), then one may be content with an estimate of the overall effect, in which case one can use either the non-parametric (NP) or empirical Bayes (EB) procedure (Persaud and Hauer, 1984). These procedures do not provide information about the effect of treatment at individual locations. It may well be that a treatment is not uniformly effective; its effect may well vary from site to site. If so, then it is important that this be known, so that the reasons for the variation can be investigated and understood, and the knowledge incorporated into the detailed design of future applications of the treatment.

By focussing on the accident counts for individual sites (as the classical approach entails), it may be found that a site exhibits unusually low or high variance in the annual accident counts, in which case use of the Poisson distribution will increase the probability of mistaking a change in the UTAR for a simple fluctuation in the accident counts about a constant UTAR, or vice versa. There is some evidence (Nicholson, 1985) that a substantial proportion of sites may have accident counts not well-described by the Poisson distribution.

Finally, both the classical and alternative approaches have strengths and weaknesses. There is further work to do before one can clearly identify the circumstances in which one approach will be better than the other, and it seems unlikely that one approach will be better in all circumstances.

REFERENCES

Abbess, C., Jarrett, D. and Wright, C.C., "Accidents at Blackspots: Estimating the Effectiveness of Remedial Treatment, with Special Reference to the 'Regression-to-Mean' Effect", Traff. Eng. & Control, 22(10), 1981.

Andreassen, D.C. and Hoque, M.M., "Intersection Accident Frequencies", Traff. Eng. & Control, 27(10), 1986.

Boyle, A.J. and Wright, C.C., "Accident 'Migration' After Remedial Treatment at Accident Blackspots", Traff. Eng. & Control, 25(5), 1984.

Cox, D.R. and Lewis, P.A.W., "The Statistical Analysis of Series of Events", Methuen, 1966.

Feller, W., "An Introduction to Probability Theory and Its Applications: Vol. 2 (2nd Ed.)", Wiley, 1971.

Hauer, E., "Bias-by-Selection: Overestimation of the Effectiveness of Safety Countermeasures Caused by the Process of Selection for Treatment", Accid. Anal. & Prev., 12(2), 1980.

Hauer, E. and Persaud, B., "Common Bias in Before-and-After Accident Comparisons and Its Elimination", Transportation Research Record 905, Transportation Research Board, 1982.

Hauer, E., "On the Estimation of the Expected Number of Accidents", Accid. Anal. & Prev., 18(1), 1986.

Maher, M.J., "Fitting Probability Distributions to Accident Frequency Data", Traff. Eng. & Control, 28(6), 1987(a).

Maher, M.J., "Accident Migration: A Statistical Explanation", Traff. Eng. & Control, 28(9), 1987(b).

Nicholson, A.J., "The Variability of Accident Counts", Accid. Anal. & Prev., 17(1), 1985.

Nicholson, A.J., "Estimation of the Underlying True Accident Rate: A New Procedure", Proc. 13th ARRB-5th REAAA Conference, 13(9), Australian Road Research Board, 1986.

Nicholson, A.J., "The Estimation of Accident Rates and Countermeasure Effectiveness", Traff. Eng. & Control, 28(10), 1987.

Persaud, B. and Hauer, E., "Comparison of Two Methods for Debiasing Before- and- After Accident Studies", Transportation Research Record 975, Transportation Research Board, 1984.

Pearson, E.S. and Hartley, H.O., "Biometrika Tables for Statisticians: Vol. 2", Charles Griffen, 1976.

Silcock, D.T. and Smyth, A.W., "The Methods Used by British Highway Authorities to Identify Accident Blackspots", Traff. Eng. & Control, 25(11), 1984.

Zegeer, C.V., "Highway Accident Analysis Systems", Synthesis of Highway Practice 91, National Cooperative Highway Research Program, Transportation Research Board, 1982.

APPENDIX

The expected annual number of accidents at site i is

$$E(x_i) = \bar{x}_i + F$$

where

$$\bar{x}_i = \sum_{j=1}^{J} x_{ij}/J$$

$$F = \bar{x}(\bar{x} - \bar{x}_i)/[J(s^2-\bar{x}) + \bar{x}]$$

For J = 1,

$$F = \bar{x}\ (\bar{x}-\bar{x}_i)\ /\ s^2$$

and for $\bar{x} > 0$ and $s^2 > 0$,

$$F > 0 \quad \text{if} \quad \bar{x} > \bar{x}_i$$
$$F < 0 \quad \text{if} \quad \bar{x} < \bar{x}_i$$

That is, if the average accident count over J years for site i is less than the average for all sites, then $E(x_i)$ will be greater than $\bar{x}_i$ (and vice versa). The adjustment is zero if $\bar{x} = \bar{x}_i$.

The absolute magnitude of the adjustment F will decrease with increasing J if and only if F and $\partial F/\partial J$ are of opposite sign. Since

$$(\partial F/\partial J) = -\bar{x}\ (\bar{x}-\bar{x}_i)\ (s^2-\bar{x})\ /\ [J(s^2-\bar{x}) + \bar{x}]^2$$

then it follows that $\partial F/\partial J$ is

(1) < 0 if $\bar{x} > \bar{x}_i$ and $s^2 > \bar{x}$, or $\bar{x} < \bar{x}_i$ and $s^2 < \bar{x}$
(2) > 0 if $\bar{x} > \bar{x}_i$ and $s^2 < \bar{x}$, or $\bar{x} < \bar{x}_i$ and $s^2 > \bar{x}$.

Hence, the absolute magnitude of F will decrease with increasing J if and only if $s^2 > \bar{x}$.

If one has a group of sites with very similar UTAR's and little variation in the accident counts about those UTAR's, then it may well be that $s^2 < \bar{x}$, and the adjustment F will not tend to zero as J gets large.

Now,

$$(\partial F/\partial s^2) = - J \bar{x} (\bar{x}-\bar{x}_i) / [J (s^2-\bar{x}) + \bar{x}]^2$$

and this is, for J = 1, 2, ... ,

(1) < 0 if $\bar{x} > \bar{x}_i$
(2) > 0 if $\bar{x} < \bar{x}_i$

Since F and $(\partial F/\partial s^2)$ are of opposite sign for both $\bar{x} > \bar{x}_i$ and $\bar{x} < \bar{x}_i$, then the incorporation of data from another site, such that the variance s^2 is increased will, all other things being equal, lead to a decrease in the adjustment F. This is consistent with the result that a lack of variance will lead to F increasing as J increases.

In order that the adjustment F be

(1) > 0 for $\bar{x} > \bar{x}_i$
(2) < 0 for $\bar{x} < \bar{x}_i$

it is necessary that

$$\bar{x} / [J (s^2-\bar{x}) + \bar{x}] > 0$$

and for $s^2 < \bar{x}$ this may not hold, especially as J becomes larger.

It appears that so long as $s^2 < \bar{x}$, then the estimator of the annual number of accidents is well-behaved.

EMPIRICAL ESTIMATION OF THE REGRESSION-TO-MEAN EFFECT ASSOCIATED WITH ROAD ACCIDENT REMEDIAL TREATMENT

D.F. Jarrett
C.R. Abbess
C.C. Wright

Road Traffic Research Centre
Middlesex Polytechnic
England

INTRODUCTION

It is well known that the true or long-term mean accident rate at a blackspot tends to be less than the observed frequency over any given period. This apparently contradictory statement is explained by the fact that, by definition, a 'blackspot' is a site with a high observed accident frequency relative to the rest of the population; the distribution of observed accident frequencies for sites as a whole will be more widely dispersed than the underlying distribution of true means, so, for sites with observed frequencies which are large compared with the rest of the population, the observed frequency is likely to be somewhat greater than the long-term mean.

It is now known that an appreciable proportion of the apparent reduction in accidents at a black spot following remedial treatment is attributable to the regression-to-mean effect. Several authors have suggested methods for predicting the effect so that it can be corrected for during the evaluation process: for example Gipps (1980), Hauer (1981), and Abbess et al. (1981). Abbess (1984) has developed a computer program PRAYERMATS which processes before and after data and provides estimates of the effectiveness of treatment for individual blackspots and collectively for a whole group. At the same time, the results are adjusted automatically for trend using control data supplied by the user.

Most of the existing methods rely on one or more assumptions about the nature of the distribution of the long-term mean accident rates between sites in the population. For a brief review and comparison of the methods, the reader is referred to Wright et al. (1985, 1988). One particular method, which was tentatively suggested by Jarrett et al. (1982), is very close in spirit to the original concept of 'regression-to-mean', which was discussed and named as such by Francis Galton (1889). This method is based on the simple idea of fitting a regression model to accident frequencies for two separate time periods. The approach calls for little by way of assumptions about the data, and in principle can be made to yield acceptable results for cases where other methods would fail. Because of this, and because it involves the fitting of a line or curve to a scatter plot of the data, the approach can be called an <u>empirical</u> method of estimating the regression effect.

For many underlying distributions of long-term mean accident frequencies, it turns out that the appropriate regression model is a straight line (see Wright et al., 1985). However, the usual least-squares procedure for estimating the parameters of the model is not appropriate for accident data since the usual assumptions of regression theory are violated. The aim of this paper is to suggest a technique for fitting the regression function and to test it using both simulated data and accident data from the London area. The suggested method will be compared with existing methods of estimating the regression effect.

THEORETICAL BACKGROUND

The approach to estimating the regression-to-mean effect suggested by Gipps (1980) and developed by Abbess et al. (1981) is based on the following three assumptions:

(i) in a time period of fixed duration (say one year), the number of accidents x at a given site has a Poisson distribution with mean m, independently of other sites;

(ii) the value of the true mean accident frequency m varies between sites according to a gamma distribution;

(iii) the mean accident frequencies for different sites are the values of independent random variables.

Under these assumptions, the conditional expectation of m given x, $E(m|x)$, is a linear function of x: if the shape parameter of the gamma distribution is denoted by k, and the scale parameter by c, then

$$E(m|x) = \frac{k}{c+1} + \frac{1}{c+1}x$$

(see Jarrett et al., 1982).

This is the <u>regression function</u> of m on x, and the magnitude of the regression-to-mean effect is the difference between x and $E(m|x)$. The parameters k and c can be estimated from accident data for a sample of sites using the fact that the distribution of accidents over all sites has a negative binomial distribution, which can be fitted to the data by standard methods. Abbess et al. (1981) gives full details of the estimation procedure; this general approach to estimating the parameters of an underlying 'prior' distribution is known as an <u>empirical Bayes method</u> (Maritz, 1970).

Now suppose that data for two time periods are available; for convenience it will be assumed that the periods are of equal duration, although the methods proposed can easily be generalised to periods of different duration. For a particular site, let x denote the accident frequency in the first ('before') period, and y the frequency in the second ('after') period. Assume that, at this site, x and y have independent Poisson distributions with mean m. Then, as is shown in Jarrett et al. (1982), assumptions (ii) and (iii) above imply the following:

(iv) the joint distribution of x and y is a bivariate negative binomial distribution;

(v) the conditional distribution of y given x is negative binomial with mean $E(m|x)$ and variance proportional to $E(m|x)$; in addition, the constant of proportionality is equal to $1 + 1/(c+1)$.

Since the regression function is linear, this result suggests that the regression function can be estimated from the bivariate data using a model of the form

$$E(y|x) = A + Bx, \qquad var(y|x) = constant*E(y|x);$$

the constant is known as the <u>scale factor</u>. The estimation of such a model can be carried out easily using a statistical package such as GLIM (Payne, 1985).

There are some problems in adopting such an approach but also a number of advantages. First it should be realised that ordinary least squares is not an appropriate method for the estimation of the regression coefficients A and B; since the variance of y is not constant, the least-squares estimators of A and B will be unbiased but not efficient. Secondly, if the bivariate negative binomial model is appropriate, then fully efficient estimates of A and B can be obtained using the fact that $x + y$ (the total number of accidents over both periods) has a (univariate) negative binomial distribution, while direct estimation of the regression function (even if taking account of the non-constant variance) will give less efficient estimates of the regression coefficients.

However, the main advantage of using this regression model is that it can be derived under much weaker assumptions than those made above. Most importantly, it is no longer necessary to assume that the mean accident frequencies m have been drawn independently from a gamma distribution since the 'before' data are regarded as fixed; so long as the sites are selected purely on the basis of the 'before' data (or more generally on features of the sites in the 'before' period), unbiased estimates of A and B will be obtained. In addition, it is no longer necessary to assume that the accident frequency at a particular site has a Poisson distribution; it is sufficient to assume that the frequency has a distribution with variance proportional to the mean.

More precisely, we assume the following:

(a) At a particular site, with true mean accident frequency m in the before period, the before frequency x and the after frequency y are independent random variables with

$$E(x|m) = m, \quad var(x|m) = vm$$

$$E(y|m) = rm, \quad var(y|m) = vrm.$$

Here v is a scale factor (equal to 1 for a Poisson distribution), and r represents a multiplicative trend in the underlying mean.

(b) the distribution of m over sites is such that

$$E(m|x) = A + Bx, \quad var(m|x) = h(A + Bx).$$

It is straightforward to prove from these assumptions that:

$$E(y|x) = rA + rBx, \quad var(y|x) = v(1+rB)E(y|x).$$

(The parameter h does not appear in this model since it can be shown that the assumptions imply that $h = vB$.)

This is just the regression model above with scale factor depending on v, r and B. It is important to realise that this model is implied by, but does not imply, the bivariate negative binomial model; moreover, as indicated above, the model is a conditional one - it does not matter how the x (before) values were selected. The method can therefore be applied to data for sites where the method of selection suggests that the negative binomial model will not be appropriate. This aspect of the method is illustrated in some of the examples below.

APPLICATION OF THE METHOD

In this section the regression model is fitted to four different sets of data. The first two examples use simulated data so that the estimates obtained for the regression model can be compared with those obtained from the fitting of a bivariate negative binomial distribution. The remaining examples use real data. One data set is that for the City of Westminster previously analysed in Jarrett et al. (1982), and for which the negative binomial distribution was found to give a good fit. The other data set is unusual in that it relates to a group of 'candidate' sites which at one time were selected for remedial treatment, but the treatment was subsequently abandoned or deferred in each case.

In each example, the regression is fitted to the data using GLIM. To take account of the fact that the regression is linear, with the variance of the response (dependent) variable being proportional to its expectation, one defines the model to have an identity link and a Poisson error term and uses the $SCALE directive; Appendix 1 gives an example of a GLIM analysis. (This method of fitting the model can be justified by the idea of quasi-likelihood - see McCullagh and Nelder, 1983.) As well as the parameter estimates, GLIM gives estimates of the standard errors of the coefficients; these can only be regarded as approximations for small sample sizes. Note that it is impossible to obtain an estimate of the trend factor r from the fit; this is not important if the estimate of the regression effect is required for sites subject to the same trend as those sites used to fit the model. Alternatively, if an independent estimate of r is available (e.g. as a 'control factor') then this can be used to obtain estimates of the regression coefficients A and B which would apply in the absence of trend.

Examples using simulated data

Two simulated data sets were used, the first representing a sample of 20 sites, the second a sample of 200. In both cases it is assumed that the before and after frequencies x and y were Poisson-distributed with mean m, where m varies from site to site according to a gamma distribution with $k = 1.5$ and $c = 0.5$. The theoretical regression function is therefore

$$E(y|x) = k/(c+1) + (1/(c+1))\,x = 1 + 0.667\,x,$$

with scale parameter $1 + 0.667 = 1.667$.

Plots of y against x are shown in Figures 1 and 2, and the estimation results are summarised in Table 1. In addition to the fitted empirical regression model, the table shows the efficient estimates obtained from fitting the bivariate negative binomial model. For the larger set of data, the model is also fitted to two subsets obtained by restricting to sites with a limited range of values of x; this is to illustrate the point made above that no bias is introduced by selecting sites on the basis of the 'before' values, whereas the bivariate negative binomial model would certainly <u>not</u> be applicable to these restriced data sets. It will be seen that in all cases the estimated coefficients are well within two standard errors of both the true values and the efficient estimates. The standard errors are, however, relatively large for small sample sizes.

FIGURE 1: 20 sites (simulated data)

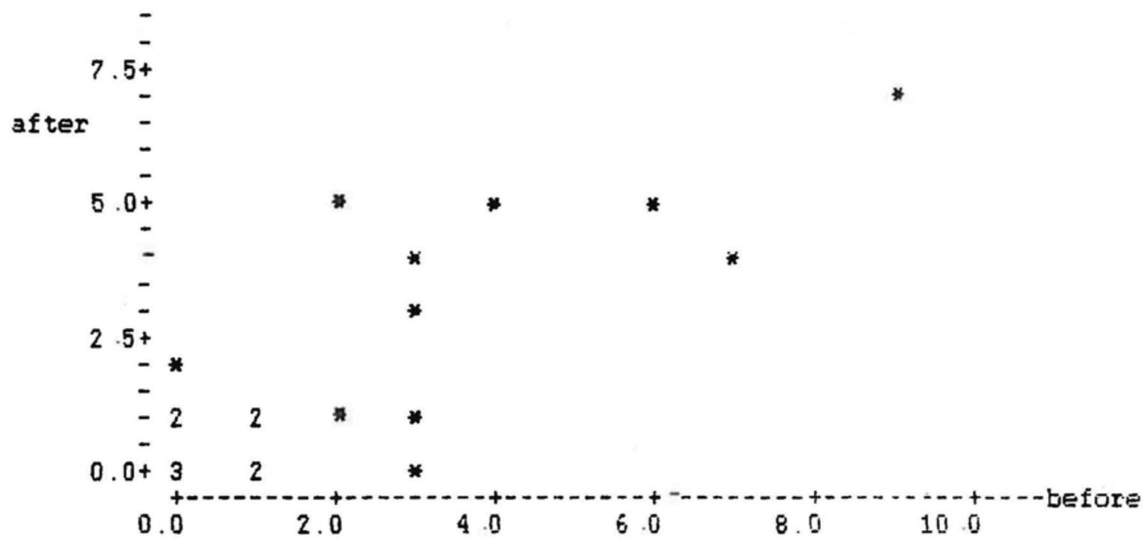


FIGURE 2: 200 sites (simulated data)

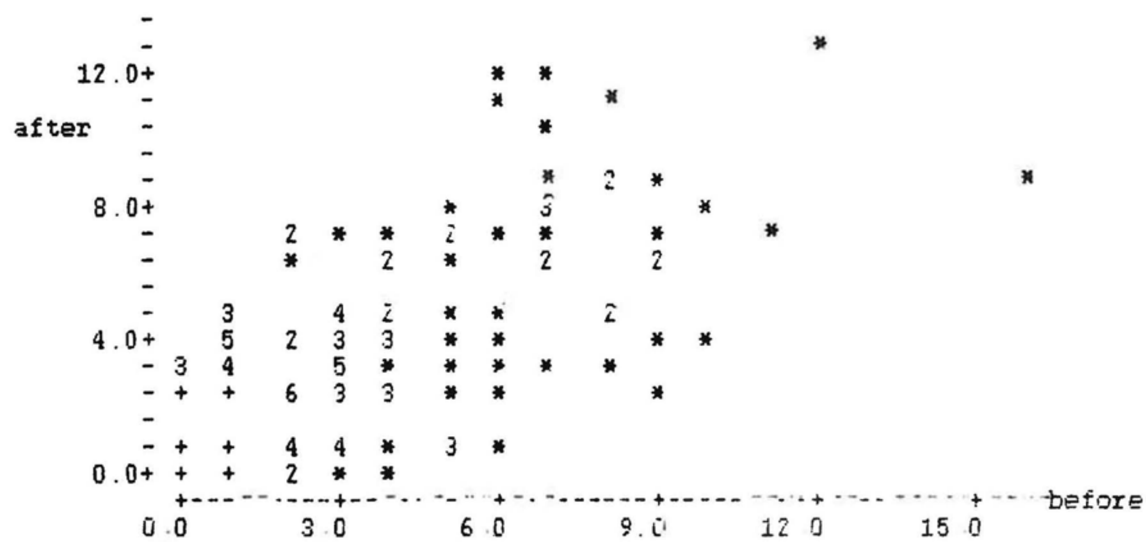


TABLE 1: Estimation from simulated data

Samples of size (i) 20 (ii) 200 with means independently drawn from a gamma distribution with shape parameter k=1.5 and scale parameter c=0.5

	Regression function			Scale factor
True values	1	+	0.667 x	1.667
(i) n = 20				
Efficient estimate from total	0.734	+	0.662 x	1.662
Empirical estimate	0.544	+	0.654 x	1.303
(SEs)	(0.311)		(0.173)	
(ii) n = 200				
Efficient estimate from total	0.873	+	0.686 x	1.686
Empirical estimate (whole sample)	0.916	+	0.717 x	1.452
(SEs)	(0.136)		(0.0596)	
Subsample 1: 2⩽x⩽12				
Empirical estimate	1.052	+	0.705 x	1.537
(SEs)	(0.532)		(0.116)	
Subsample 2: 2⩽x⩽7				
Empirical estimate	0.613	+	0.848 x	1.584
(SEs)	(0.666)		(0.175)	

Westminster data

This set of data consists of accident frequencies at 'nodes' (major junctions on the Greater London road network) in the City of Westminster, for the years 1976 and 1977. It was investigated in Jarrett et al. (1982), where the negative binomial distribution was found to give a reasonably good fit. A plot of the data is shown in Figure 3, and the results of the estimation procedure in Table 1; the empirical method is again compared with the efficient estimates obtained from the negative binomial fit. The fitted negative binomial distribution was truncated at zero, thus ignoring those sites where there were no accidents in either year, since the total number of zeros in the data is larger than would be expected from the negative binomial. Similarly, those sites for which the 'before' frequency x was zero were excluded from the fit of the empirical regression function; as explained above, this does not invalidate the method. Again it should be noted that the empirical estimates are fairly close to the efficient estimates.

FIGURE 3: Westminster data

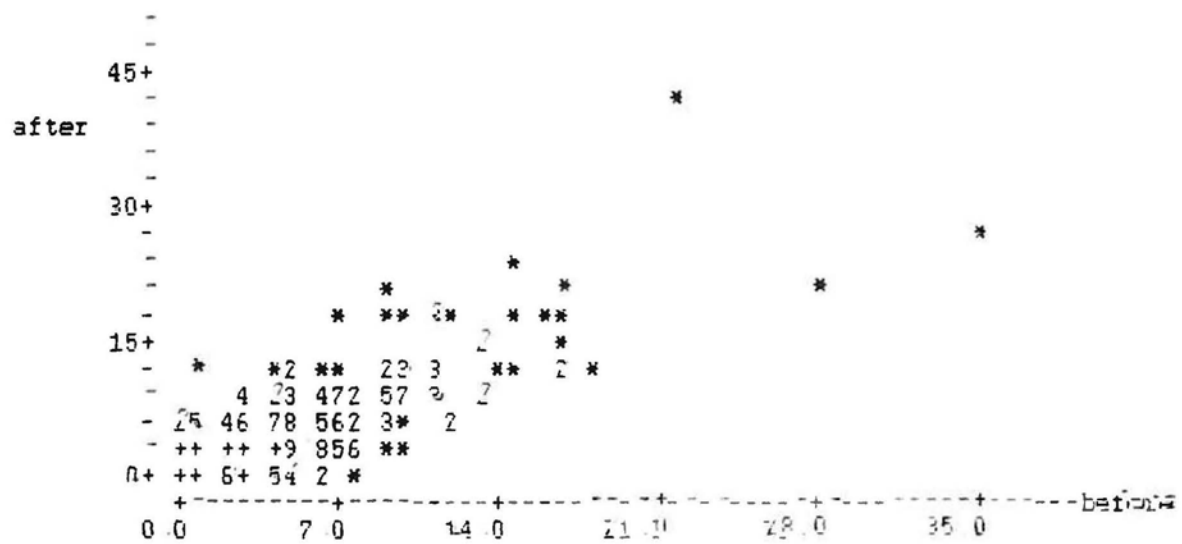


TABLE 2: Westminster data (n = 470)

	Regression function	Scale factor
Efficient estimate from total (omitting sites where x+y = 0; n = 307)	1.015 + 0.788 x	1.788
Empirical estimate (omitting sites where x = 0; n = 281)	0.959 + 0.856 x	2.020
(SEs)	(0.249) (0.0544)	

Candidate sites

The final example concerns a set of data collected by Viola and Wright (1983) for the specific purpose of investigating the regression-to-mean phenomenon for sites which had been selected for remedial treatment but where treatment had subsequently been deferred or abandoned. For the purposes of their study it was decided to limit coverage to a random sample of 16 of the 32 London boroughs; the data consist of accident frequencies for 167 sites, for a 'before' period 1975-77 and an 'after' period 1978-80. The magnitude of the regression effect observed for these sites is of special interest, since it can be argued that sites which are selected for treatment will display a quite different statistical behaviour from the population as a whole, because their selection implies not only a relatively high accident frequency, but a consistent pattern in the <u>type</u> of accident observed. In other words, the engineer will have taken into account some additional information which can be regarded as evidence that the high accident frequency observed at the site in question is not merely a random fluctuation. The distribution of accident frequencies can therefore be expected to be different from that for the population as a whole, and very possibly not of a negative binomial form. Moreover the regression-to-mean effect should be reduced, and provide a more satisfactory estimate of the effect for those sites which are treated.

A scatter plot of the data is shown in Figure 4. It will be noted that, in contrast to the earlier examples, there are comparatively few sites with small accident frequencies; one reason for this is that x and y are now accident frequencies over 3-year periods. The empirical method is particularly appropriate here, because of the way in which the sites were selected, and the results of the estimation are shown in Table 3. Also shown are the coefficients after correction for trend: the estimate of the trend term r was obtained as a control ratio obtained from the numbers of accidents in the two periods at all the untreated sites in the 16 boroughs. The constant term in the regression function is considerably larger than in the other examples, reflecting the absence of low accident frequencies; however, when account is taken of the fact that the accident frequencies are 3-year totals, the results do not appear very different from those for the Westminster data. Thus, perhaps surprisingly, the data do not reveal the different behaviour predicted above.

FIGURE 4: London candidate sites

```
         -
         -                                             *
      90+
         -
 after   -                                                          *
         -
         -
      60+
         -                                  *
         -                                     *
         -                   *  *   *
         -             * **    *   *
      30+                 2  *2*3     *
         -           *33*223 *4    *
         -         *22+3363 ** *
         -       2396949 32*
         -  **555474422
       0+  ** *
          +---------+---------+---------+---------+---------+---------+------before
          0        20        40        60        80       100
```

TABLE 3: London candidate sites (n = 167)

	Regression function	Scale factor
Empirical estimate (SEs)	2.827 + 0.815 x (0.723) (0.0496)	2.322
Corrected for trend (based on r = 0.9792)	2.887 + 0.832 x	

CONCLUSION

In this paper, a new method of estimating the regression-to-mean effect has been proposed. Unlike other approaches to this problem, the method is based on the simple idea of fitting a straight line to a scatter plot of 'before' and 'after' accident frequencies. Although the method gives less efficient estimates of the regression coefficients than approaches which require a negative binomial distribution to be fitted, it is valid under less restrictive assumptions and can therefore be regarded as a more 'robust' estimation method. Furthermore, the method gives similar results to the negative binomial approach in cases where the latter is valid. However, the standard errors of the regression coefficients produced by this method are relatively large, even for moderate sample sizes; thus considerable uncertainty will remain about the size of the regression effect. The relative efficiency of different methods of estimating the regression-to-mean effect requires further investigation, as does the development of techniques for determining the validity of the underlying model. Some progress has been made in each of these areas and it is hoped to report on this at a later date.

REFERENCES

ABBESS C. (1984). 'Assessment of remedial treatment', Proceedings of the Seminar on Traffic Operation and Management, PTRC Summer Annual Meeting, University of Sussex, Brighton, 71-82

ABBESS C., D.F. JARRETT & 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

GALTON F. (1889). Natural Inheritance, London: Macmillan

GIPPS P.G. (1980). 'Examining the safety contributions of traffic control devices', World Conference on Urban Transport Research, London, paper E26

HAUER E. (1980). 'Bias by selection: overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment', Accid. Anal. Prev., 15(5), 323-328

JARRETT D.F., C. ABBESS & C.C. WRIGHT (1982). 'Bayesian methods applied to recent road accident blackspot studies: some recent progress', Seminar on Short-term and Area-wide Evaluation of Safety Measures, Institute for Road Safety Research SWOV, Amsterdam

MARITZ J.S. (1970). Empirical Bayes Methods, London: Methuen

McCULLAGH P. & J.A. NELDER (1983). Generalised Linear Models, London: Chapman and Hall

PAYNE C.D. (ed.) (1985). The GLIM System Release 3.77 Manual, Oxford: Numerical Algorithms Group

VIOLA A. & C.C. WRIGHT (1983). 'Empirical measurements of the regression-to-mean effect for road accident blackspots', report prepared for the Nuffield Foundation (unpublished)

WRIGHT C.C., C. ABBESS & D.F. JARRETT (1985). 'Evaluating the effectiveness of road accident remedial measures: towards a more realistic approach', Proceedings of EVALUATION 85, Paris

WRIGHT C.C., C. ABBESS & D.F. JARRETT (1988). 'Estimating the regression-to-mean effect associated with road accident black spot treatment: towards a more realistic approach', Accid. Anal. Prev. (to appear)

APPENDIX 1: EXAMPLE OF A GLIM ANALYSIS

```
[o] GLIM 3.77 update 0 (copyright)1985 Royal Statistical Society, London
[o]
[i] ? $INPUT 7$
[i] $echo
[i] $output 6 80
[i] $C Empirical estimation of the regression-to-mean effect :
[i]    simulated data for 20 sites
[i]
[i] $units 20
[i] $data bef $read
[i] 0 9 3 1 1 0 1 6 3 1 0 0 0 7 0 2 4 2 3 3
[i] $data aft $read
[i] 0 7 4 1 1 1 0 5 1 0 0 0 2 4 1 5 5 1 3 0
[i]
[i] $plot aft bef $
[o]     7.600 |
[o]     7.200 |                                            A
[o]     6.800 |
[o]     6.400 |
[o]     6.000 |
[o]     5.600 |
[o]     5.200 |         A         A         A
[o]     4.800 |
[o]     4.400 |
[o]     4.000 |              A                   A
[o]     3.600 |
[o]     3.200 |              A
[o]     2.800 |
[o]     2.400 |
[o]     2.000 A
[o]     1.600 |
[o]     1.200 2    2    A    A
[o]     0.800 |
[o]     0.400 |
[o]     0.000 3    2         A
[o] ----------:---------:---------:---------:---------:---------:---------
[o]         0.00      2.00      4.00      6.00      8.00     10.00
[i]
[i] $error p $link i $scale
[i]
[i] $yvar aft $fit $
[o] deviance = 48.208 at cycle  3
[o]     d.f. = 19
[o]
```

```
[i]
[i] $C Use RECYCLE to guard against the possibility of negative
[i]    fitted values
[i] $recycle $fit + bef $dis me $
[o] deviance = 23.455 (change =  -24.75) at cycle  2
[o]     d.f. = 18     (change =   -1   )
[o]
[o] Current model:
[o]
[o]   number of units is  20
[o]
[o]   y-variate  AFT
[o]   weight     *
[o]   offset     *
[o]
[o]   probability distribution is POISSON
[o]               link function is IDENTITY
[o]               scale parameter is to be estimated by the mean deviance
[o]
[o]   terms =  1 + BEF
[o]
[o]           estimate        s.e.      parameter
[o]     1      0.5440      0.3112      1
[o]     2      0.6548      0.1730      BEF
[o]     scale parameter taken as  1.303
[o]
[i]
[i] $stop
```