PROFESSOR MIKE MAHER

Napier University

Mike Maher has been Professor of Transportation at Napier University's Transport Research Institute for 12 years. Throughout his career, he has specialised in the mathematical and statistical modelling of transport problems, researching particularly in the areas of network modelling and traffic safety. After leaving Cambridge University in 1970 with a BA in Mechanical Sciences and a PhD in Operations Research, he held lectureships at the Institute for Transport Studies, Leeds University and the Department of Probability & Statistics, Sheffield University, before moving to work at TRL for seven years and thence, in 1994, to Napier. He is on the Editorial Advisory Board for the international journals Transportation Research B and Networks and Spatial Theory, and is a Fellow of the Institution of Highways & Transportation and a Chartered Statistician.

Abstract

Road accidents are, fortunately, rare events. Individual accidents are effectively unpredictable and occur generally because of road user error. Accident frequencies (at a site or by a driver) over a period of time are subject to considerable random variation but nevertheless display systematic features that can be analysed and, to some degree, explained. The talk starts with the obvious need for exposure measures in order to make meaningful and fair comparisons between accident numbers for different entities, whether sites, driver age groups, countries or time periods. It then moves on to look at the relationships between accidents, traffic flow and road design, as exemplified in the TRL series of junction accident studies. Whilst such predictive accident models are far from perfect, they do provide the potential for an intelligent assessment of how poorly a site is performing and hence of the possible need for remedial treatment. They also provide the means of carrying out, *post hoc,* an assessment of the effectiveness of a remedial treatment, by allowing for the 'regression to mean' effect in before and after comparisons, as applied in recent research on the effectiveness of speed cameras.

Accidents, Exposure and Cameras: the Role of Predictive Accident Models in Traffic Safety Analysis

Professor Mike Maher

Napier University

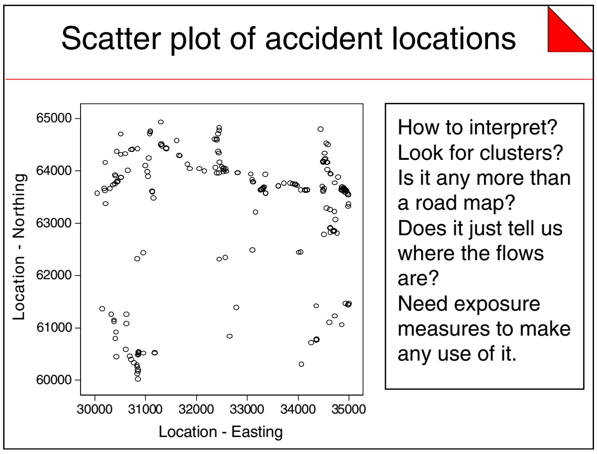
Introduction

I am hoping that I will be able to shed some light on predictive accident models for those of you who are not specialists in statistical modelling or traffic safety and that I can explain some of the interesting issues and slightly difficult concepts that arise. I hope I can explain them to a general audience of transport people. It's always a bit difficult giving a talk like this because one is never quite sure how the audience is going to be comprised, what they have come for and what they are hoping to get out of it. I hope I can hit the mark.

Theme

> To make meaningful comparisons of accident frequencies we need to have exposure measures, and we need to establish relationships between accidents and flow

We have a very good accident database in the UK. We are very fortunate to have STATS19 which records the location and circumstances of every accident on the road that involves a personal injury. It is a wonderful resource from which we can pull out the number of accidents at a particular location, along any stretch of road or in any region. But it always seems to me that to make any sense of it, to make any meaningful comparisons, to be able to say anything useful about the figures, whether they are good or bad, then we need exposure measures, whether in the form of traffic flows or vehicle kilometres. As soon as one introduces exposures and one has accident data, then one becomes interested in the relationship between them; and that is the theme of what I want to tackle today.



Any particular entry of the STATS19 database contains the location of each accident. This example just happens to be a 50 km by 50 km region of the Scottish Borders. There is nothing special about it at all really. So one can produce a scatter plot like this one showing the accidents. This type of geographical plot is often displayed using other techniques such as GIS (Geographical Information Systems). With this accident location information one can then look for clusters of accidents on a stretch of road. But does that tell us anything on its own or do the accident locations just inform one of where the traffic flows are? In this example, there are areas of moorland and there are roads. One would not expect road accidents on moorland, and that is broadly the case; so at face value we simply get a trace of the roads, outlined by the accidents occurring on them. Now, my contention is that to make any sense of this accident location data, we need information about the flows that are passing through the junctions and along those roads.

Outline of the talk

From STATS19 (or its like) we can find out how many accidents there have been over any preceding number of years either at any junction or along any stretch of road. Typically we want to be able to compare things such as accident numbers and types at different sites. Or maybe we want to identify potential problem sites, 'cluster sites'. From this one can possibly then decide on some remedial treatment if these really are problem sites. When we have applied remedial treatment, Outline of talk (1)

- Can find out how many accidents
  - at any junction
  - along any stretch of road
- How should we compare different sites?
- How should we identify “problem” sites?
  - and select for remedial treatment
- How should we evaluate effectiveness of remedial treatment?

hopefully to reduce the number of accidents at those problem sites, we will want to go back and do a *before* and *after* type of comparison to evaluate the effectiveness of the treatment.

Outline of talk (2)

- How many accidents “should” there be?
- Predictive accident models
  - expected number, given information about the site (type, flows, design…)
- Properties and uses
  - how good are they?
  - blackspot identification
  - estimation of treatment effectiveness
  - regression to mean, safety cameras
- Future needs

These are some of the things that I want to talk about. The first item I will discuss will be: at any site, whether it is a length of road or a junction, how many accidents ‘should’ there be? At first that may sound a bit odd: how many should there be? Perhaps it would be better to say ‘how many can we expect there to be for such a site?’ That will lead us into what we call predictive accident models; that is, given information about the site (its type, perhaps the flow that it carries, perhaps the design of the site) and using models to help us estimate, how many accidents would we expect there to be? Then I will be going on to look at the properties of these predictive accident models. How good are they? How well do they predict, how can we use them for blackspot identification and how can we use such models in the estimation of treatment effectiveness? This last topic will be a substantial part of what I will be talking about tonight.

The main concept that I will spend a little time trying to explain; that some of you may have heard about and wondered about—is called *‘regression to the mean’*. It is a problem that afflicts this type of comparison. Then I will go on to look at some recent work on applying that to the effectiveness of safety cameras, and then finally I will try to summarise things and look ahead a little bit to what we need for future types of models.

A very brief history of accident models

In a short talk, I can only give a very brief glimpse of accident models. It seems to me that Ruben Smeed was really the one who started looking at traffic safety as a science. He argued, way back in 1949[1], that one might expect that the number of single vehicle collisions in a region or in a country will be proportional to the number of vehicles on the road, and the number of two-vehicle collisions should be proportional to that number squared.

A very brief history of accident models

- Smeed argued (1949):
  - single vehicle collisions $\propto N$
  - two-vehicle collisions $\propto N^2$
- Later (1972), he argued:
  - collisions per year between vehicles of particular types should be proportional to product of distances travelled
  - but not borne out by time-series data

Later on, he modified that model, recognising that it did not include any sort of measures of exposure at all—about just what mileage was being driven. He did this in 1972, in a paper, and also at a seminar he gave when I was a very raw new lecturer at Leeds University[2]. The seminar really made an impact on me at the time. He modified what he had said before and argued that collisions per year between vehicles of particular types should be proportional to the product of the distances travelled by these types of vehicles. But he also showed that that hypothesis was not supported by time series data where there were measures of exposure—information about the mileage covered by different types of vehicles.

So there was something wrong. If we look at the same sort of idea but just at the junction level (Smeed was looking at national figures), at, say, a T-junction where you have got a minor road flow and a major road flow, and you look at the collisions between minor road and major road vehicles, then, on a similar sort

---

[1] **Smeed R J (1949).** Some statistical aspects of road safety research. *Journal of the Royal Statistical Society*, Series A (1949), 1-34

[2] **Smeed R J (1972).** The usefulness of formulae in traffic engineering and road safety. *Accident Analysis and Prevention*, 4 (4) 303-312

At junction level

Q_1 → Q_2 ↑ Right-angle accidents $Y = k\,Q_1 Q_2$?

- Tanner (1953)
  - empirical study of rural T junctions
  - found that $Y = k\,(Q_1 Q_2)^{1/2}$
- Brownian Motion model not correct!

of reasoning you might expect that the number of such collisions be proportional to the flow of the minor road vehicles, Q_1, multiplied by the flow of the major road vehicles, Q_2. In fact John Tanner in 1953 had thought about that, gathered some data and had done an empirical study of the numbers of collisions at rural T-junctions. He found that instead of being proportional to the traffic flows into the junction (or their product for example) there was something like a square root law. Instead of being proportional to Q_1Q_2, the numbers of collisions were proportional to something like the square root of Q_1Q_2. So this idea that vehicle collisions are something like elementary particles in a cloud chamber bumping into each other, fortunately does not seem to be correct. There seems to be evidence that there is some modification of behaviour in the presence of other vehicles.

Rate models

The very simplest form of predictive accident model is what I refer to as a rate model. This works from the total numbers of accidents on different types of roads and takes the estimates of the vehicle mileage on those different types of roads. In *Road Casualties of Great Britain*[3] you can see very easily that motorways are far and away the safest roads, with nine accidents per 100,000,000 vehicle kilometres. This rate goes up almost eight or nine times that for other types of road.

Now the rate model carries with it an implicit assumption that the number of accidents is proportional to *L*, the length of any section of road; and that seems fair enough—if you have 20 km of road you would expect twice as many accidents as if you had 10 km of road. The rate is also proportional to time—if you count up accidents over two years rather than one year, you would expect to double the number of accidents. It also assumes that the number

Rate models

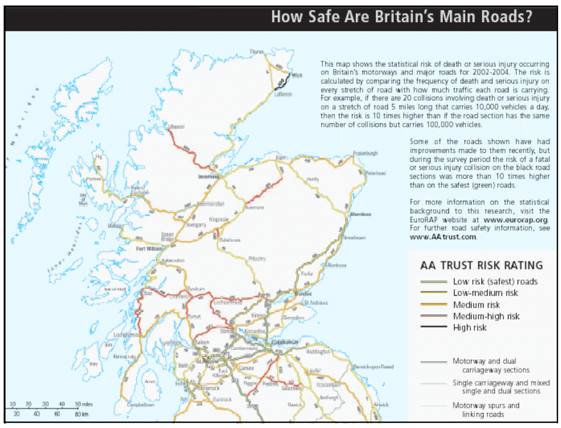
- Accidents per 100 million veh-kms
  - motorways 9
  - rural A roads 25
  - other rural roads 46
  - urban roads 67
- Implicit assumption that $Y = k\,L\,T\,Q$
  - is this correct? Power of Q = 1?
- Basis of Euro RAP

of accidents is proportional to *Q*, the flow rate on the road. That is not quite as obvious—that if you double the flow you will double the accidents—and, the previous work by Tanner and Smeed casts doubt on it.

Euro RAP

- European Road Assessment Programme
- Risk rates and maps
  - accidents per km of road
  - accidents per veh-km
  - accident rate in relation to similar roads
  - potential for accident reduction
- In Great Britain
  - 871 sections of motorways, trunk/major roads
  - total network length 27,500 km
  - FSAs (2002-2004) per billion veh-km

Thus, useful and interesting as these rate models are, they do not really seem to tell the whole story; the power of Q is not necessarily unity. But such rate models, measuring statistics like k, the accident risk, are the basis of EuroRAP—the European Road Assessments Programme showing maps of where accident rates are by road[4].



[3] **Transport Statistics: Department for Transport (2005)**. *Road Casualties Great Britain 2005*. The Stationery Office: London. Retrieved: 2006, from www.dft.gov.uk/transtat

[4] **AA Motoring Trust (2006).** *How safe are Britain's main roads? EuroRAP 2006: British results*. The AA Motoring Trust. Retrieved: 2006, from www.eurorap.org

One can get either accidents per km of road, or accidents per vehicle-km, which from the driver's point of view are the more relevant, or the accident rate in relation to similar roads. At present in Great Britain there is a total network length of 380,000 km. An example of one of the kinds of map you can get is shown here for the Scottish end of the network. One gets the risk rating—colour coded—for the roads indicating the numbers of fatal and serious accidents, per 1,000,000,000 vehicle-km.

Non-rate models for links

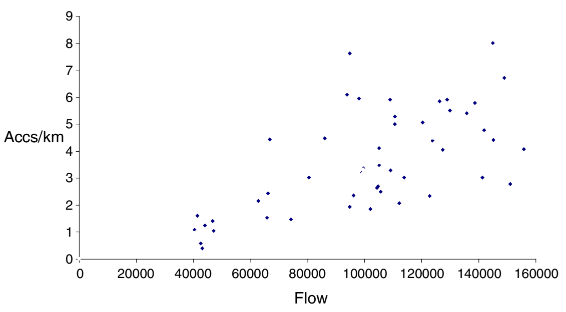
- Not necessarily proportional to flow
  - mix of various accident types
  - high flow linked to lower speed
  - so, analyse data to see what fits best
- Example
  - 50 motorway sections (between junctions)
  - accidents in one year (2002) on each section
  - section length *L*
  - AADT from TRADS2 database

But what happens if it is not necessarily the case that the number of accidents is proportional to flow? We have seen that there is a mix of different accident types, and from the way that Smeed was arguing before one might expect a different sort of relationship for single vehicle accidents from that for two vehicle accidents, and so on. We know also that high flow engenders lower speed which allows greater scope for the driver to avoid an accident, or lower the severity of an accident if it does occur. It is not really possible to think one's way through all of this, and really the only satisfactory approach is just to collect data and analyse it to see what fits best.

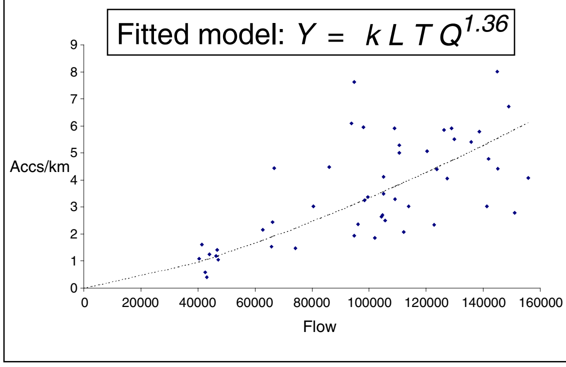
Motorway accident data

As a very simple illustration, I have taken 50 motorway sections—that is, sections between junctions on motorways (with known section length). The numbers of accidents in any one year is easy to get from STATS19 and one can get the flow rates—the AADT (Annual Average Daily Traffic)—from the TRADS2 database, also very easily. So this sort of diagram is straightforward to construct. For those 50 motorway sections we can plot accidents per km against flow. Now, accidents are random events, following some sort of Poisson process, so we would expect there to be scatter and there certainly is—but clearly there is some sort of increasing relationship.

Motorway sections: accs/km v AADT



Motorway sections: accs/km v AADT



If we fit a model to the data, it turns out that in this case the best fitting model has a power of the flow rate—the AADT—of something of the order of 1.36—this is not terribly precise, but it is somewhere around that.

Time of day effect

- Disaggregate accidents and flow by:
  - time of day (night, am peak, off-peak, pm peak)
  - weekday / weekend
  - data: 7 periods x 50 sections
- Fitted model: = $k\,L\,T\,Q^{1.30}\,F_t$
  - with F = 1 for "daytime"
  - and F = 2 for "nighttime"

Whilst it is not really the main thrust of what I want to say here I will just mention in passing that if one breaks the data down a little further and looks at different time periods—at weekday and weekends and different periods within the day, such as AM peak, night time, off peak, PM peak and so on—and then fits a relationship to each section, then one finds that a model of the type shown at the bottom of the figure can be fitted. It turns out that most of the time periods do not differ, except that broadly speaking one gets a factor 'F' of 1 for day time and a factor of 2 for night time. So just from this simple data set, one gets a very clear suggestion that numbers of accidents are doubled in the night time as compared with day time.

How good is the model?

- Y_k is Poisson, with mean $\mu_k = f(Q_k)$
- We can predict μ_k but not Y_k
- On motorways, design fairly uniform
  - so few other explanatory variables?
- Are the residuals pure Poisson?
  - no; other variables at work
  - "over-dispersion"
- Negative Binomial error structure
  - gamma distribution of mean μ_k, $C_v = 0.25$

Returning to the original model and the scatterplot one can see that there is quite a lot of scatter, a lot of variability. One of the basic assumptions in this modelling of number of accidents is that the number of accidents in a given period of time at a site is Poisson distributed with a certain mean. This is widely accepted, I think. The trouble is we do not know what the mean is for any site. We might expect that on motorways the design is pretty uniform. So once one has taken the length of a section into account and knows the flow on the section, is there much else that one can find out about it? Is there anything else that is likely to affect the number of accidents?

If there were not then the scatter about the fitted line should be just pure Poisson—there should be nothing else other than just pure random variation about it. So the question is: how good would a model assuming a pure Poisson error structure be in allowing one to estimate the underlying mean rate of accidents at that site? But it turns out that that scatter of points, those residuals, those errors, are *not* just pure Poisson. There must be other variables at work; this phenomenon is what is known as *over-dispersion*. So even in a case as simple and clean a problem as motorway sections with a very uniform design, there is over-dispersion at work.

What has now become the standard approach to fitting models to data of this type, where other variables affect the rate of accident occurrence, is to employ a negative binomial error structure. In this case the number of accidents at a given site is pure Poisson but about a mean which varies from site to site because of the other variables.

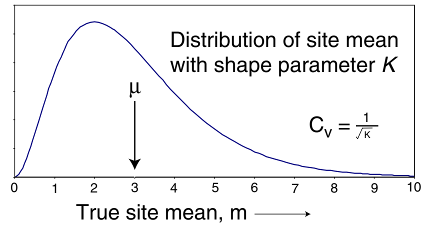
What we seek to do is to predict that mean value. If one looks at a graph like this then one can think that the predictive model in terms of the flow and any other characteristics will give a value of what we would expect the mean number of accidents to be at each site. But the actual number of accidents at the site, the true mean, is then distributed about that. The quality of the modelling process (that is, the quality of the predictive process) can be measured by the parameter C_v, the coefficient of variation, the ratio of the standard error to the mean. For the motorway data, we get a figure for C_v of about 25%. You will need to bear that in mind when we come to other issues later.

Gamma model for mean accidents

Prior site mean $\mu = f(\text{flow})$

True site mean $m = \mu \exp(z)$

z represents the other unobserved explanatory variables



There are therefore other things which we do not generally know about that affect the numbers of accidents, that affect the mean, at any site. We cannot observe these other things, we cannot measure them, but nevertheless they are at work.

Junction accident models

Everything I have said so far has been about road links, about stretches of road and motorways or whatever. Relatively speaking they are quite simple to deal with. Much more difficult, relatively, are investigations of junctions; this is because of the conflicts of different movements that take place in junctions.

Predictive models for junctions

- TRL junction accident studies
- Studies on accidents at:
  - 4-arm roundabouts (1984)
  - urban 4-arm signalised junctions (1986)
  - urban T junctions (1996)
  - urban crossroads (1996)
  - urban 3-arm signals (1996)
  - rural dual carriageways (1998)
  - mini roundabouts (1998) etc
- All used same methodology

It seems to me that the benchmark for modelling of numbers of accidents at junctions was really set by TRL in its work for the Department for Transport over many years in a long series of TRL junction accident studies. It started way back in the mid 1980s with studies of accidents at 4-arm roundabouts, then rural T-junctions and 4-arm signals. They all used the same sort of methodology, and that, as I say, has become pretty much the norm for anybody else trying to do that in other parts of the world.

Methodology

- Sample of sites (number from 78 to 900)
- Accidents from STATS19 (PIAs)
  - total : 1500 - 2500 over 4 or 5 year period
  - disaggregated by type and location (eg arm)
- Turning flow data (vehicles and peds)
  - manual 12/16 hour count, to give AADT
- Geometric (design) variables
  - egentry width, ICD, CID, presence/absence
- Regression to fit *Y* to *Q*s, *G*s
  - negative binomial error structure
  - log-linear

Let me just say a little bit about how that modelling is done. First of all, one decides on the type of junction, such as 4-arm roundabout, let us say. One selects a sample of such roundabouts and from STATS19 finds out how many accidents and of what type there have been over the last 4 or 5 years. Then by site visits one measures the vehicle and pedestrians flows and turning movements, and the design variables, the junction geometry and so on.

Three levels of model

- Level 1
  - total accidents versus total flow
  - example: $\mu = k Q^{0.68}$
- Level 2
  - disaggregated by accident type and arm of junction
  - separate model, related to relevant flows
  - example: $\mu_{c\text{-}e} = k Q_e^{0.68} Q_c^{0.36}$
- Level 3
  - as for level 2, but with design variables added
  - example: $\mu_a = k Q_e^{1.7} \exp(20 C_e - 0.1e)$

A specialised type of regression modelling is then used to fit a model to the data. In the TRL junction accident studies, modelling was conducted at three different levels. The coarsest level *(Level 1)* involved just taking all of the accidents at a junction and relating that to some measure of the total flow passing through the junction. So for example one might just have that the number of accidents, ì, is proportional to some power of the total flow going into the junction.

For *Level 2*, the accidents were disaggregated by type—these might be entering-circulating accidents or approaching accidents or right angle accidents—and according to which arm of the junction they occurred on. (For example there may be a collision between a vehicle coming in from one arm and a vehicle circulating across it). The number of accidents of that type was then related to the relevant flows and a model fitted like the example on the slide, with number of accidents related to entering flow rate to some power times circulating flow rate to some power.

The third level of modelling, what the modellers were really after, was the same as *Level 2* but where analysts added the geometric or design variables in as well. The objective was to find out which features of the design of a junction made some junctions safer, or other junctions less safe, so that design practice could be improved, so as to reduce accident numbers in the future.

How good are the models?

- Level 3 more complex, requires more data
  - but should give better fit, and better insight
  - sum up predictions over arm, accident types
- Typical measures of quality of fit:
  - level 1: $C_v = 0.40$
  - level 2: $C_v = 0.35$
  - level 3: $C_v = 0.30$
- Better – but not as much as expected?
- Total flow dominant explanatory variable

Now *Level 3* models are clearly more complex than the others, a lot more work has to go into getting the data in the first place, fitting the models and then applying them afterwards. One would expect a better fit because one employs a better insight into how those particular accidents were caused.

If one returns and predicts the total number of accidents at the junction one can sum up over the different accident types and over the arm of the junction, if that is appropriate. It is interesting to ask the question; 'how good are the predictive models?' Once one has them in place and can use them to predict the total number of accidents that are likely to occur at a junction, how good is the result?

Let us return to the coefficient of variation, the C_v measure relating standard error related to the mean (remember that for the simple motorway data, I said that it was about 25%). I have looked at a few of the studies that I did during my time at TRL, many years ago now, and found that typical sorts of values for C_v were, roughly, about 0.4 (40%) for *Level 1*, 35% for *Level 2* and about 30% for *Level 3*. So one certainly gets a better fit, a better predictive quality, with the more detailed type of modelling; but perhaps not quite as much better as one might have hoped for in going through that process.

Certainly what seems clear is that incorporating some measure of the total flow provides a lot of information. What can be added in terms of explanatory power from including the pattern of that flow and the design variables is perhaps less marked.

The use of predictive models

Once we have a predictive accident model, in principle we can say, that given what we know about this particular site (that is, what type of site it is, the flow that is passing through it, and possibly some of the details of its design) then we can make a prediction about how many accidents there are likely to be at such a site.

What use are the models?

- Junction design
  - incorporated into ARCADY etc, SafeNET
  - influences of design on delay and safety
- Blackspot identification
  - rank sites by Y_{obs} - μ_{pred} instead of by Y_{obs}
  - potential for accident reduction
  - Euro RAP, and McGuigan (1981)
- Evaluation of treatment effectiveness
  - before and after comparisons

What do we do with such models? There are three broad ways in which they can be used. The first is in junction design, helping the traffic engineer to design for safety as well as for operational things like delays and capacity and queues and so on. They have been incorporated into ARCADY, PICADY and OSCADY, the junction design programs that traffic engineers use all the time, and also into a network model, SafeNET, that amalgamates all of those. So that is the first thing: to help the traffic engineer to design for safety as well as for operational objectives.

The second purpose is for identifying accident cluster sites. Typically, what the local authority does is to identify problem sites (cluster sites) by looking at the total number of accidents in the last few years—usually the last three years or so—and ranking the sites with respect to that. But it may be far more sensible, far more efficient, to rank them instead by the observed number of accidents minus the predicted number of accidents from one's predictive accident model. That difference, that excess of *observed* over *predicted*, is then referred to as PAR, the 'potential for accident reduction' (the term that I think was originally coined by David McGuigan in 1981, but it is also used now in the European Roads Assessment Program).

The third purpose for predictive accident models is in the evaluation of treatment effectiveness, i.e. in *before* and *after* comparisons that examine how much an applied treatment has reduced the number of accidents. This is what I am going to concentrate on for the rest of my talk this evening.

Before and *after* studies and the estimation of treatment effectiveness

North Lanarkshire data

Number of sites N_k with k accidents 1992-94

k	0	1	2	3	4	5	6	7	8	9	11	13
N_k	7411	1645	341	117	38	26	13	7	2	1	1	1

	92 - 94	95 - 97	change
Whole network	3136	2799	-11%
Cluster sites	458	233	- 49%

To start with I am going to show you some real data from the North Lanarkshire region of Scotland from some years ago; this was data that a part time Napier MSc student who was working for North Lanarkshire was using as the basis of his dissertation. There are about 9600 sites—stretches of road or junctions—in North Lanarkshire. For the three years of data analysed it was found that at 7411 sites there were no accidents, at one site there were 13 accidents, at another 11 accidents, and so on.

North Lanarkshire data

Number of sites N_k with k accidents 1992-94

k	0	1	2	3
N_k	7411	1645	341	117

	92 - 94	95 - 97	change
Whole network	3136	2799	-11%
Cluster sites	458	233	- 49%

North Lanarkshire are interested then in deciding where they should apply remedial treatment to try and alleviate some of these problem sites. They define 'cluster sites' as any site with 4 or more accidents in the last three years. In the 3 year period 1992-1994 there were 458 accidents at those cluster sites; in the following 3 years (1995-1997) this went down to 233, a reduction of almost 50%. In the network as a whole there was some improvement generally; about 3100 accidents reducing to 2800; one might conclude that there was a downward trend with a reduction of 11%. From that one would think that the general 11% reduction accounts for some of the 49% in the clusters, and so the other 38% must then be attributable to the remedial treatment that has been applied at those cluster sites.

Regression to the mean

- Bias by selection
- Sites chosen on basis of high $Y = m + \varepsilon$
- Top sites tend to have both:
  - high systematic component (mean m)
  - high positive random component (error ε)
  - systematic component persists
  - … but random component does not
- Exaggerated estimate of treatment effectiveness, unless corrected for

Now that is a fine argument except that at that time North Lanarkshire did not have any safety budget. They therefore were not able to carry out any remedial treatment, and so we have some understanding and explaining to do. What is going on here? That 38% reduction over and above the trend was obviously due to something else, and this is an example of what is known as the ‘regression to mean’ problem. It arises because of the bias in the selection process. We have selected those sites that had a high number of accidents—the ‘cluster sites’. We did not just choose a random selection of sites and apply treatment to them, and the result is what gives rise to *regression to the mean*. Sites are chosen on the basis of a large y-value, a high number of accidents in that three year period; this is made up (or we can think of it as being made up) of the true mean number of accidents, m, plus a Poisson component ε, that is the random element. The sites with a high number of accidents (the ones with the thirteen and the eleven and the nine accidents and so on) tend to have high m and a high positive ε so they qualify typically on both counts.

Now the m (the systematic component, the true mean underlying the accident frequency at that site), persists, of course. One might reduce it if one applied some treatment, but otherwise it persists. The random component does not persist, this is random, just good luck or bad luck, and therefore the expected value of that component in the following period will be zero.

So that is why one gets this reduction, because one has the m part, the true mean, persisting, but there is no reason why the original high positive random component should persist, and you get this drop automatically. That is why one gets that sort of effect that we see with the North Lanarkshire data. It is just bias by selection. One has picked out those sites that have had that high number of accidents in the before period. No treatment was applied to the junction, and one can see clearly that all of the accident reduction over and above the trend was due to *regression to the mean.* Of course, if one does apply treatment then one will have two things mixed up together and one will need then to extract one from the other. If one does not allow for *regression to the mean* there will be an exaggerated estimate of the effectiveness of the remedial treatment. In this case there was no remedial treatment yet the data *look* improved.

What is regression to the mean?

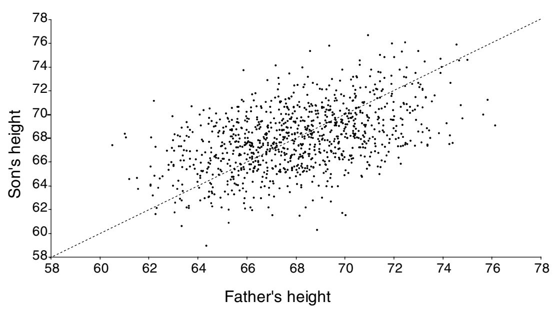
What is regression to the mean?

- Sir Francis Galton, (1822 – 1911), eugenicist, biometrician, statistician, observed:

Tall fathers tend to have sons who are also tall – but who are not as tall as themselves

Let us look a little at regression and *regression to the mean* because it is not always obvious just where the term comes from. ‘Regression’ is going back. The term *regression to the mean* comes from Sir Francis Galton founder of Eugenics. Of the many studies that he and his colleagues conducted the one that is best known is where he measured fathers’ and sons’ heights, and observed that tall fathers tend to have sons who are also tall, but not as tall as themselves.

Galton height data



The sort of data he might have gathered can be plotted like this; the father’s height is plotted on the x-axis and son’s height on the y-axis, with the dotted line showing where y = x. One can see that for tall fathers the sons' heights tend not to be as great as the fathers' heights but to be above average for the whole population. So there is *regression to the mean* as one moves from one generation to the next because in producing the son's height there is a systematic component coming through from the father in the genes but there is then a random component as well. Again we have the effect of a true systematic component carrying through and a random component being added on.

RTM appears in other places, too

- Golf tournaments:
  - the players who score best in the first round tend, on average, to score well in the second round too - but not as well as they did in the first

Regression to the mean can arise in other places as well. I play a little golf (very badly) and it struck me that if one looks at golf tournaments then one can very regularly observe that the players who scored best in the first round tend, on average, also to score well in the second round, but not as well as they did in the first.

Just to demonstrate this, and without any fixing or anything, I looked at the 2006 US Open, just a few weeks ago. They play the first two rounds and then a cut is made. Those who do not 'make' the cut go home, and the others who do make the cut stay on and play rounds 3 and 4.

2006 US Open Golf Tournament

Pos	Name	Rounds 1-2	Rounds 3-4
1	Steve Stricker	139	149
2	Colin Montgomerie	140	146
3	Geoff Ogilvy	141	144
4	Ken Ferrie	141	147
5	Jim Furyk	142	144
6	Padraig Harrington	142	145
7	Phil Mickelson	143	143
8	Arron Oberhauser	143	148
9	Jason Dufner	143	153
10	Graeme McDowell	143	154

I looked at the leader board, the top ten players after the first two rounds; Steve Stricker was the leader, Colin Montgomery was in second place and so on. The table shows their combined scores in rounds 1 and 2. Then I looked to see how they did in rounds 3 and 4. You can observe that only Phil Mickleson did as well in the later two rounds as in the first. Everybody else did worse, about 5.5 strokes worse, on average, in the second two rounds than in the first two rounds; and this was not because the weather suddenly deteriorated.

Again, it is a case of *regression to the mean*: the fact is that, yes, these are good golfers but there are still differences between them. Tiger Woods is certainly a better player than I am and he is certainly a better player than Jason Dufner: but on any day there is good and bad luck, little bits of things that go the player's way or work against them; so there is clearly a random component at work. One has a systematic component—the true skill of the player—and added to that is the random component. We are just following through those that were top of the list at the end of round two, so again we have the same phenomenon at work; *regression to the mean*, bias by selection.

In the same way once a site has been selected on the basis of the *before* accident total, the number of accidents will reduce automatically, on average, even if nothing is done there. That is the problem: a simple *before* and *after* comparison will tend to be flawed because of the bias by selection. I am going to illustrate the nature of the problem and then show what can be done about it using predictive accident models by reference to a couple of recent studies to do with the effectiveness of safety cameras. This is the third element of the title of this talk; 'Accidents, exposure and cameras'.

Effectiveness of safety cameras

Effectiveness of safety cameras

- Before and after comparisons of accidents at safety cameras
  - EPSRC study (Liverpool and Napier): 62 sites
  - DfT study: 216 sites
- Disaggregate observed change into:
  - trend
  - flow changes
  - RTM (using Empirical Bayes Method)
  - camera effect (due to speed reduction)

There were two studies: one was an EPSRC study carried out by Liverpool University and Napier University where we had 62 sites where cameras had been installed; the other one was carried out last year using a subset of data from a DfT study on camera partnerships.

The aim of the analysis was to see how the number of accidents at these camera sites changed *before* to *after* and then to try to disaggregate that and break the change down into components.

- One component is due to time trend—just what would be going on generally, nationally and regionally over that period from the before period to the after period.

- Another arises from possible flow changes. If one applies some treatment which then diverts traffic away from the site, perhaps not so much with speed cameras but with speed humps, one might very much get a re-routing effect.

- *Regression to the mean* of course; and we wish to try and allow for that.

- What is left over is then hopefully the genuine effect of cameras in reducing accidents, because of the speed reduction they cause.

The problem

- Cameras installed at sites with high number of accidents
  - at least eight PICs/km in three years
  - with at least four involving fatality or serious injury
- Accidents would tend to reduce even if nothing done
  - bias produced by selection criterion (RTM)
  - therefore necessary to allow for that in the analysis methodology
  - and also allow for other possible effects

The problem is that of course cameras are installed at sites that have a high number of accidents in the specified *before* period. It is likely—or at least possible—that one will have the regression to mean problem arising. To install a camera the criteria or guidelines were that there had to have been at least eight collisions causing personal injury (PICs) in the last three years, with at least four involving a fatality or a serious injury. The second part in particular is a stringent requirement, so there is a strong possibility that accidents would reduce in the *after* period following the installation of the camera even if nothing had in fact been done or even if the camera had absolutely no effect whatsoever because of the bias produced by the selection criterion. So we need to try and allow for that in the analysis. That is the problem.

One of the ways to allow for regression to mean is a technique known as the '*Empirical Bayes*' method which involves the use of a predictive accident model. What we are trying to get at is the true mean number

The solution: empirical Bayes method

- What is expected value of mean *m*
  - given observed value of no. accidents *x*?
- Bayes' Theorem
  - prior distribution for *m* from predictive accident model
  - combine with observed *x*
  - to give posterior estimate of *m*
- Depends on model and the shape *K*

$$\hat{m} = \alpha\,\mu + (1-\alpha)x \qquad \text{where: } \alpha = \left(1 + \frac{\mu}{K}\right)^{-1}$$

of accidents at the site. All we have got is the observed number of accidents, the x, but we are trying to infer something about the *true mean* value for that site.

What we do is to use Bayes' theorem from statistical inference. We have a predictive accident model of the type that I was discussing earlier, which says, for such and such a type of site carrying such and such a level of flow (or whatever else goes into the model), we would expect there to be so many accidents per year. On the other hand we have an observation on the number of accidents at that particular site where the camera was installed. We are trying to merge these two pieces of information together. We take a prior distribution, in Bayesian terminology, from the predictive accident model, we bring in the observation of the actual number of accidents and then we move to a posterior distribution for this m (for the true mean). It turns out that the result depends not just on the predictive accident model; you also need to know something about the *quality* of the model as measured by the C_v, which I was talking about earlier, or an equivalent parameter, K, a shape parameter. Just having the predictive model on its own is not sufficient.

Verification of performance of EBM

- Simulation studies
  - generate set of sites, each with observed *q*
  - model gives μ = *f*(q) as predicted value
  - generate true *m* from distribution with mean μ and shape *K*
  - generate accident frequency *y* as Poisson (*m*)
  - select sites with highest *y* for "treatment"
  - apply EB method to estimate mean *m*
  - compare with known true *m*
  - results show that method is unbiassed and insensitive to precise distribution of *m*

When one does have those two bits of information—the prediction from the model and the observation from that actual number of accidents at that site—then it is possible to bring the two together in a simple weighted average, of just a weight α times the prediction, plus 1-α times the observed number of accidents. But α (the weight) depends on the quality of the model that you have; it depends on the shape parameter or equivalently on the C_v value that I have talked about before. That is, one has to have those two things: the predictive model and a measure of its quality.

Results from EPSRC study

- Predictive model: $\mu = 0.9q^{0.6}\, L \exp(n/L)$ with $K = 1.9$
  - for PIAs on 30 mph, single carriageways
  - allow for national trend from before to after
  - allow for change in flow at site
  - allow for RTM
- Produces estimate (and CI) on camera effect – separately for PIAs and KSIs

For the two studies mentioned above, we had quite a simple model for predicting the personal injury accidents at the sites, which were all 30 mile per hour single carriageways. We allowed for national time trend from the *before* to the *after* period and we allowed for any changes in the flow at that site and, of course, using the *Empirical Bayes* method, we allowed for *regression to the mean*.

PIAs: up to 500m from camera

Total observed accidents	1305
Observed change in accidents	-26% (-36,-13)
Change due to trend	+4% (-1, 9)
Change due to RTM	-5% (-7,-3)
Change due to change in flow	-5% (-10,-1)
Change due to change in speed	-19% (-30,-8)

(95% confidence intervals in brackets)

We got results like these in the slide. I am just picking out the results for personal injury accidents (PIAs) within 500 m of the camera, a 1 km length in all, for each site.

At the 62 sites in the study, we found from the number of accidents that in the first place the observed change was -26% from *before* to *after*. We were aiming, in the analysis, to break that down into the different components; so +4% was due to time trend, -5% due to *regression to the mean*, -5% due to change in flow, and then the real camera effect, the bit left over, was -19%. So in that case we see the regression to mean part, the 5%, was really quite small, not of any enormous concern.

KSIs: up to 500m from camera

Total observed accidents	211
Observed change in accidents	-34% (-51, 11)
Change due to trend	+5% (-10, 1)
Change due to RTM	-18% (-25,-9)
Change due to change in flow	-5% (-8,-2)
Change due to change in speed	-6% (-21, 12)

(95% confidence intervals in brackets)

But if we now look not at all accidents but at killed and seriously injured accidents (albeit obviously with smaller numbers of accidents), then we find that the change due to regression to mean is 18%: much more of a substantial effect. If *regression to the mean* had not been allowed for there would be an exaggeration of the camera effect—the change in accidents due to speed would not just be the 6%, but there would also have been the 18% as well. So one would get a rather different measure of the effectiveness of the camera in that case.

Camera partnership: 4-year evaluation

- UCL/PA study for DfT
  - data from camera partnerships
  - 4-year report published December 2005
- Appendix H to investigate RTM effect
  - EB method applied to subset of data
  - 216 sites for which data & model available
  - urban sites (30 and 40 mph limits)

We also applied the same sort of methodology to a subset of the data from the evaluation reports of camera partnerships by UCL and PA Consulting over several years, the four year report of which was produced just before Christmas 2005. We were invited to investigate this subset of the data and apply the *Empirical Bayes* method to it (the subset being those sites for which there was an appropriate model).

From the results for personal injury collisions, (PICs) we get a camera effect of 16% and a regression to mean effect of 7%. If the analysis is done in terms of what would have been expected in the *after* period, that is if a trend had continued, then one is relating what one actually observes to what would have been the case in the *after* period if the camera had not been

Results – for PICs

PICs/site/year: before 4.65, after 3.22 (-31%)

Overall reduction =	1.43	=	0.37	+	0.31	+	0.75
			trend		RTM		camera
	31%	=	8%	+	7%	+	16%

relative to what would have been: 25% allowing for trend; 19% allowing for trend + RTM

installed. For this case the results are 25% reduction in accidents due to effectiveness of the camera if one allows for trend and 19% if you allow for trend and for *regression to the mean.* Not an enormous difference, so for those circumstances, *regression to the mean* is not a huge problem, but when one turns to the fatal and serious collisions (FSCs), then the picture is substantially different.

Results – for FSCs

PICs /site/year: before 1.05, after 0.48 (-54%)

Overall reduction =	0.57	=	0.10	+	0.36	+	0.11
			trend		RTM		camera
	54%	=	10%	+	34%	+	10%

relative to what would have been: 50% allowing for trend; 19% allowing for trend + RTM

Related to what would have been the case in the *after* period, if one does not allow for *regression to the mean*, we now get an estimate of 50% reduction due to cameras whereas if one does allow for trend and *regression to the mean* then you get a more modest 19% or so.

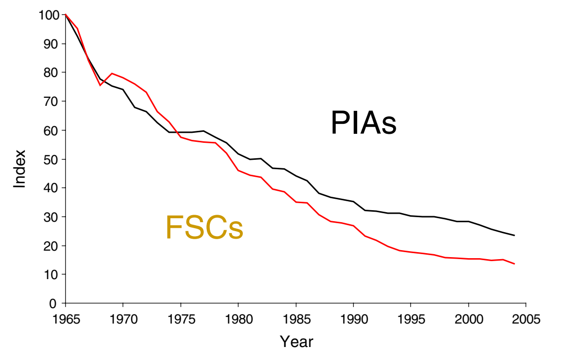
Points to note

- Requirements for predictive model
  - shape parameter K or C_v needs to be known
  - explanatory variable values need to known
  - any other variable used in selection process?
- Ageing of models
  - sometimes, model quite old
  - eg TRL models – some data 20 years old
  - needed to adjust predictions in EPSRC/DFT studies by use of γ^t factor

The points that I want to highlight here are that to apply the *Empirical Bayes* method, one needs a predictive model (obviously), but one also needs the value of the shape parameter or the C_v value: the measure of how good, how precise the predictions are. Of course, to be able to apply the predictive model the values of the relevant explanatory variables for any site are needed. Indeed there is then a question which arises as to whether there are any other variables which are used in the selection process. Are camera sites chosen just because of the number of accidents in the before *period* or are they based on any other criterion as well, such as there being a speeding problem?

A particular issue that I want to raise here is about the ageing of models, particularly regarding some of the TRL models that I mentioned earlier that were developed in the mid 1980s based on data from the early 1980s. Clearly these are getting quite old now. So we felt that it was essential that we modify the models, bring them up to date as it were, by allowing for the things that had happened in the meantime—the improvements in the risk factors. So we adjusted the model by means of what I have referred to as a ãt factor, a geometric type decay or exponential decay in the risk over time. So if the model was say, 10 years old, and one was then applying it here then one would have some discounting, some geometric decay factor, raised to the appropriate number of years.

Change in accident rate over time



If one looks at the change in accident rate over 40 years, starting with an index of 100 for 1965, one can see how the accident rate per vehicle km for personal injury accidents and for fatal and serious collisions, have both fallen dramatically over the period. Even taking 5 or 10 years one will get an appreciable ageing of that process. It seemed essential to us to try and allow for that in using a model which had been fitted on data which was now some way out of date.

I am not sure whether we did it in the right way or not. Let us take a typical model of the TRL junction

Ageing of models

$$\mu_a = k\,Q_e^{1.7}\exp(20\,C_e - 0.1e)$$

- Does ageing affect only the risk *k* ?
- Can we assume the effect of flow remains constant: $Q_e^{1.7}$?
- Can we assume the design effects remain constant ?
- How should we check and update models ?
- Can we adjust *k* from national trends ?

accident study (the figure shows a model for approaching accidents at 4-armed roundabouts). Here there is some coefficient k which had a numerical value in the original model-fitting process; there is the entering flow Q_e to some power (1.7 in this case) and then there are some geometrics (entry curvature and entry width) included. Those are the relevant explanatory variables for that particular type of accident. If one thinks about this ageing process for these models, can we assume that it affects only the coefficient k? That is really what we were doing by just applying that geometric decay factor. Is it safe to assume that the effects of flow on accidents remain the same—that is, entry flow to the same power, 1.7? Similarly, can we assume that the design effects remain the same—i.e. the same variables enter the model, and that they enter in the same way?

Future modelling needs

- If we are to carry out new modelling studies
  - what type should they be?
  - aimed at what application?
  - do we need detailed "level 3" models?
  - is improvement in accuracy sufficent?
  - or will coarser models (eg level 1) suffice?
  - should they simply "update" existing models?
  - would simple rate models suffice?
  - make use of new types of flow data?

So the question arises—if one is going to use these predictive accident models how can we have confidence that we are using them correctly given that the models have been fitted quite some time ago? How should we update the models or check them? What should we do to try and ensure that they are suitable for the purpose that we want to put them to? Can we just, for example, adjust k in those models as we did from national trend data (that would be rather simple to do, at least) or is something more complex required?

What I have been arguing here is that predictive accident models are useful and necessary for some purposes: accident cluster site identification and for evaluation of remedial treatments. But of course we need them to be of good quality or at least to know exactly what their qualities are. If we are looking to have good quality predictive accident models for future use how should we go about things? What type of models should they be?

Should they be the very detailed Level 3 type junction accident models of the TRL type or would a lesser quality do, still hopefully accurate and unbiased, but with less precision? As long as we know what the precision is then we can cope with the implications. So should we just make do with coarser models, should they even just be the simple rate models that I mentioned at the beginning, at least for lengths of road?

When the TRL junction accident studies were done possibly the most difficult part of the data collection process was the flow information. Accident information can be acquired easily from STATS19 for the last four or five years, but actually to get the flow information required a site visit, manual counts with people standing by the side of the road for 12 or 16 hours and then scaling the counts up to a whole day. Even then it was just a simple snapshot, using those counts that were done on that particular day to represent the flow that there had been over the four or five year period that the accidents had occurred in. This was thus quite a difficult part of the data collection process.

These days there are increasingly greater sources of flow data. As I mentioned before, the TRADS2 database gives information for the motorway and trunk road network. It has day-by-day, hourby-hour counts automatically so there are very comprehensive data measurements available. There are all sorts of new sources of data on flows becoming increasingly available whether from automatic counters by the side of the road, in the road's surface, beacons or in-vehicle devices. There are things like the Norwich Union pay-as-you-drive insurance scheme with a GPS device put in the car so that the position of the car is tracked at all times. If that is to be, in time, in a substantial proportion of the vehicle fleet then there is a vast amount of flow data coming in, which could, in principle, be used to keep the flow measurements up to date and with the same sort of level of detail and quality as you have for the accident data in STATS19. Therefore there is the potential for keeping the flow information very much more up to date; and this could, it seems to me, be used in the future to update or to develop new predictive accident models.

Summary

Summary

- Tried to give an overview of the need of exposure measures to interpret accident statistics …
- ... and to show the nature and application of predictive accident models …
- … without getting into too many technical aspects of statistical methodology…
- …and to show why I've found it such a fascinating subject over the years!

What I have tried to do in this lecture is to give an overview of this very broad, and in places quite technical, subject. I did not get heavily involved in the statistical modelling because I did not really think that you would all necessarily welcome a description of the fine details of it. I have therefore tried to cover at least some of the concepts, like *regression to the mean*, in what I hope was a reasonably accessible way, and to show you the uses of predictive accident models. Finally I have tried to show the sort of things that I have found have made it a fascinating subject over many years.

Acknowledgements

- Ruben Smeed
  – first work on traffic safety modelling
- Rod Kimber
  – and colleagues involved in TRL studies
- Ezra Hauer
  – regression to the mean
- Linda Mountain
  – colleague on several studies

Acknowledgements

A vast amount of work has been done in this area over many years by all sorts of people. I have certainly not attempted to give rigorous referencing but I would just like to mention a few people.

One is Ruben Smeed whose early work on the subject I have mentioned. He seemed to me to take traffic safety onto the level of being a science and something that one tries to analyse and think about and develop models for. In my very early days as a lecturer at Leeds University he came and gave that seminar which made such a big impact on me at the time. The work that he started up was then continued at UCL through the capable hands of people like Richard Allsop, Heather Ward and several others there so that whole strand, I think, deserves a considerable amount of acknowledgement.

At TRL I should mention in particular Rod Kimber for his work and that of many of his colleagues who are involved in the junction accident studies; Geoff Maycock, Janet Kennedy, Ian Summersgill, Richard Hall and many others. It seems to me that they really set the standard for predictive accident modelling.

Then Ezra Hauer at Toronto University, retired now, tirelessly sought to explain *regression to the mean* at a time when he was faced with a rather sceptical and only halfunderstanding audience. But he banged away at it year after year until he had achieved his goal of getting it accepted as a genuine phenomenon. I met him several times at international conferences and he always seems to be so fired up and yet so patient in trying to explain the concept to anybody. I have a lot of regard for him and I would like to acknowledge his work.

Finally I would like to mention Linda Mountain at Liverpool University with whom I have worked on several predictive accident model related studies over the years. I thank her for her enthusiasm, insight, hard work, support and tolerance!

Thank you for listening; I hope you have got something out of it. And thank you all for coming.

Discussion

Rod Kimber

Well, thank you very much Mike. That is an excellent overview of a subject which I have found fascinating—for years—and which has been at the centre of a lot of safety policy and decision making. I am sure there are some interesting questions to follow.

Question

Thank you very much for a very lucid exposition. I am not quite sure whether this is a question but I hope you will want to respond to it. In the context you referred to, there is a great deal of public concern and discussion about these issues, and that means that a lot of people (including the media of course) are bandying around statistical findings with, regrettably, a limited understanding of the statistics behind them. It is a difficult area we work in.

I think there is a particular risk of believing that because we have this *Empirical Bayes* method for adjustment for *regression to the mean*, which you have set out and which I absolutely endorse, that therefore *regression to the mean* can always be adjusted for in practice. You mentioned a number of further research challenges in improving our capability to do this and I would like to mention another one.

What we need is a model or prior distribution to tell us what would have happened at the population of the sites where we have done something, had we not done it, and this model or distribution needs to represent what I like to call 'candidate sites'. These are sites at which we might have done what we have done elsewhere. There has been a tendency in the literature on *regression to the mean* to think that all sites of the relevant kind in an area constitute a suitable population for this purpose. Thus if one has a measure for reducing accidents at roundabouts that have some design fault, then an appropriate population or model is a population or model for all roundabouts in that area.

But typically the sites at which anyone would ever carry out the measure is a subset of the roundabouts, and a subset for which the measure is appropriate, and therefore, as Mike himself mentioned, the sites are not simply selected as the ones with the most accidents, they also have to be suitable. They have to be able to lend themselves to the measure. Ezra Hauer, the world guru on this topic who, as Mike has mentioned, has himself recently written that if a group of professionals select some sites they will have a mixture perhaps of technical and political criteria for their selections, and the sites that are selected as candidates are very unlikely to be similar to the sites that are not selected as candidates—*even before picking the ones that have a lot of accidents among the candidates*. I have not had the opportunity to establish this by personal research, but I believe there will be a tendency for the candidate sites to have higher true mean number of accidents than the noncandidate sites.

If we take the whole population including the non-candidate sites in an *Empirical Bayes* estimate of *regression to the mean* we would be using a distribution on the low side and we will over-estimate the *regression to the mean*. That is an additional difficulty to the ones mentioned and I think this is important in the public debate. We do have to remember that because this method exists in principle this does not mean it can easily be done. Identifying the model or distribution that is going to be used in the calculation is genuinely difficult in practice and I believe sometimes (at the present state of knowledge) impracticable. That ought to be borne in mind in the public debate about situations where *regression to the mean* ought to be allowed for. An analogy from another field is that because any good firm of consultant engineers can design a suspension bridge with a mile long span, that does not mean I can take them to any mile wide estuary and say 'build me a bridge here'.

Mike Maher

There have been discussions about this over quite an extended period of 2005, and certainly I would not disagree with anything that has been said about the need for the predictive model to be appropriate for the sites under consideration. In fact one of the reasons, why we could only analyse 216 sites was that there were some other sites (not in 30 or 40 mph limits) where we agreed that it would not have been safe to have applied the *Empirical Bayes* method because there was the secondary selection criterion which related to the speeds—it had to be demonstrated that there was a speeding problem at the sites and we felt that speeding in 30 and 40 mph speed limits was endemic and that therefore did not really change the nature of the selection significantly. We certainly agreed that it was not safe to apply the same principle for the 50 and 60 mph sites. So I would not really disagree with you at all, and I certainly did not intend to give the impression that because the *Empirical Bayes* method was there, that solved all problems. If 'candidate sites' are different, then it should be possible to characterise how they are different and make sure that these characteristics are included as explanatory variables in any future modelling work.

Question

On the face of it, this regression to mean error only occurs because we do not sample the data correctly in the first place, so why do we not just count for 5 years, rather than 3 years, or 6 years or 10 years and then the problem will go away, will it not?

Mike Maher

Well, it is true that the longer the *before* period, the lesser is the problem of the regression to mean, relatively speaking. So, yes, if you went back 6 years, 9 years, 10 years, the regression to mean effect would be smaller. But I think most traffic safety engineers are reluctant to do that, because usually there have been other changes at the site over that extended period. They would be, I think, reluctant to go back more than 4 years typically.

Rod Kimber

The interesting thing is that these mean values that you look at are themselves time-dependent, for other reasons, and so there are other components of variation. Therefore selecting a time window that fits is very much a compromise between how much one understands about these other reasons and the available data. It is a fascinating issue and it can not be walked away from. It is not that these things are small refinements, they really are important.

Question

I want to follow up the discussion of time-varying factors and return to Mike's γ_t discounting factor. Of course, it is not time itself that plays a role here; time in this context is being used as a proxy for time-varying co-factors (covariates) that are not at the moment measured. Not being an expert in the safety area, the thought that occurs to me is this: has there been any work undertaken that has attempted to measure what those time-varying covariates might be?

The things that strike me as candidates would be vehicle technology, occupant restraint or medical treatment technology, things which will certainly at the least affect the balance between minor and serious and fatal injuries. I think there might be merit in exploring the extent to which the time discounting factor could be more fully explicated in terms of measurable covariates. It seems to me that this is another dimension of the future challenge—trying to get a handle on what it is that has changed over time, that has dragged the profiles down as you have described. It is also important at a policy level because it helps us to understand how effective the transport engineering and traffic engineering interventions have been relative to other factors bearing on accident causation and impact.

Mike Maher

It seems particularly difficult to be able to isolate and measure the effect of any particular single intervention. You will see studies in the past which have attempted to estimate the effect of the introduction of drink-drive legislation or seatbelt wearing or crash helmet wearing or whatever, and it always seems in those studies that it is particularly difficult to see the step change. In most cases you will see a steady stream: an improvement in braking, in the use of airbags in cars and all sorts of vehicle design improvements as well as all the work that has been going on in identifying the poor parts of the road network. I would not be particularly confident about being able to isolate individual parts—to separate that downward drift out into bits that you could attribute to each of those many different things that have been going on over those last 30 or 40 years. Perhaps others who are braver and better could do it, but I certainly would not feel confident.

Rod Kimber

Perhaps I could just add to that. I suspect that there are many influences, and some of them (and the ones that have been pointed out in the question are the ones that one might go for) can be seen almost in a deterministic fashion. For example one can improve the inside of a vehicle to the extent that one removes a certain type of head injury, and as soon as the vehicle fleet has that modification incorporated then one can actually make some kind of estimate of the effect on that type of accident, on that type of injury. But as Mike says, when one actually looks at the whole spectrum of influences there are lots of others which are very difficult to get at, and they have also got time trends built into them. To infer them by disaggregation, top down, is difficult.

Response

Can I come back with just one thought, again looking at this almost as an outsider? It seems to me that one of the types of effects, say of improvements in vehicle technology, would be a migration between categories of accident, so that accidents which hitherto, with unsafe vehicles, might lead to fatalities will now lead to more minor injuries. That suggests to me that there might be virtue (and this may have been done so forgive me if it has) in looking at models in which one is simultaneously trying to estimate the frequency of different categories of accidents, allowing for correlation between the unobservables over time. Then one is not estimating separate models for different categories of accidents, but one is estimating a simultaneous system which accounts for the fact that these data may be subject to similar varying and unobserved temporal influences. It might be that that is do-able, technically, and it might be interesting to take a look at that.

Mike Maher

Yes, very interesting, absolutely.

Question

I am left wondering what are the risks of not accounting for the *regression to the mean*, in practice. In the public sector where there are more potential sites for intervention measures than there is money to actually make interventions whether or not one should account for regression to mean in the prediction of accident saving. One could adopt a policy of accounting for the *regression to the mean* or one could adopt a policy of not bothering. In each case one would be pulling off the accident sites with the highest clusters of accidents. When one cannot afford to address all those sites, will one not actually simply address the priority sites that the budgets will allow? Is it likely in practice, in, say, a local authority's or central authority's road safety program, to make a significant difference to what sites it actually tackles?

Mike Maher

There are two different problems that I was trying to keep distinct. There is the question of accident cluster site identification, the potential for accident reduction and the difference between the observed and the predicted. On the other hand, there is the question of *regression to the mean* appearing and corrupting the estimate of the effectiveness, *post hoc*, of the treatment that has been applied. If I understand you right, you are saying that if a local authority does not take any account of *regression to the mean* in its evaluation of effectiveness of the treatment, is there any impact on its program of application of remedial treatment? The same sites would tend to be selected anyway. So in that sense, no, it's not going to have any detrimental effect on the way in which the authority would select those sites for treatment and use its road safety budget. The authority may get an estimate of how well it has used that budget which is, in fact, somewhat exaggerated, but apart from that I do not see that it's obviously going to change the way in which it would do the selection or the whole program of remedial work.

Response

So it's more likely to influence the strategic approach policy towards accident potential?

Mike Maher

Yes.

Question

My question is about the presence of serial correlation in modelling the accidents. You have mentioned STATS19 data; the generalised linear models, and the Poisson and negative binomial techniques that are widely used for modelling accidents. When we extract the accident information from STATS19 data the effect of time series and serial correlation comes into account and that affects the t values of the explanatory variables at a significant level. Can you throw light—on some other suitable techniques for example—on handling this serial correlation in modelling the accidents?

Mike Maher

Obviously you see a serial correlation in accident frequencies if you look at numbers of accidents over the day of the week; clearly there is a seasonal pattern. You will see a trend over the years. The serial correlation that you will observe in the accident frequencies is almost entirely due to the serial correlation in the explanatory variables. There is variation in the flows over the week—the weekend flows are far less than the weekday flows. If you go down to the level of the flows through the day the trend there might be rising over the years. So it seems to me that serial correlation in the observed frequencies is attributable directly to the systematic effects—these explanatory variables. I am not convinced that there is then any residual serial correlation in the error structure, in the Poisson or whatever else one might call it. So from what I have seen, I am not convinced that one will go really wrong at all in ignoring the serial correlation in the error structure.

Rod Kimber

Thank you very much Mike. It has been an extremely interesting and stimulating lecture. You have expounded it very clearly and put your finger on the key issues. Thank you very much indeed.