ACCIDENT MODELS

Harri Peltola & Risto Kulmala
VTT, Technical Research Centre of Finland

1. BACKGROUND

By accident models we mean equations where the expected number of injury accidents on a road section or at a junction is expressed as a function of the traffic and road characteristics on that section or junction. We have developed accident models in Finland for two main reasons:

i) to gain more insight into the relationships between traffic and road conditions and the number of accidents

ii) to estimate the expected number of accidents on a particular road section more reliably than by just counting the observed number of police reported accidents

In this paper we are going to concentrate on models done for the latter reasons, accident prediction models.

In Finland, the traffic safety objectives are expressed in the number of fatalities. The number of traffic deaths should be halved in almost ten years. Usually the number of fatal accidents is too low for reliable model building, but the number of injury accidents is sufficient for modelling. Hence we use the number of injury accidents together with the average severity of accidents (deaths/injury accidents) to estimate how much some action contributes to road safety improvements.

2. DATA TO BE USED WHEN MAKING MODELS

Accident models are based on the number of police-reported injury accidents that have occurred usually during the last five years. Sometimes we model the number of a certain type of accidents (e.g. those involving motor vehicles only, unprotected road users, animals). Quite often we model the total number of injury accidents. We usually make different models for junctions and road sections. For road sections, we are interested in the number of accidents per vehicle mileage and for junctions the number of accidents per incoming vehicles.

In Finland accident models can quite easily be done for rural highways. All accident and traffic data, and almost all road data from rural roads have been recorded in our road register. In addition to accident data, we have in our registers the following traffic and road data (some examples):

- time and location - AADT, cars - pavement width
- consequences - AADT, heavy vehicles - curviness
- accident type - vehicles entering - hilliness
  junction
Due to the lack of above mentioned data on urban networks, we did not develop accident models for urban streets.
The average annual daily traffic (AADT) on Finnish main roads is about 4 400 automobiles. The total length of the main roads is 13 300 kilometres. Annually about 1 700 injury accidents occur on these roads. The accident rate is 8 injury accidents / 100 million automobile kilometres and the death rate is 0.9 deaths/ 100 million automobile kilometres. The accident densities are 13 injury accidents and 1.4 deaths /100 road kilometres/year.

The biggest problems in modelling data are caused by missing flow data for unprotected road users and the lack of adequate information of the land use along the road. Using the number of inhabitants near the road in the models has been shown to deal with the latter problem.

3. MAKING ACCIDENT MODELS

We have developed accident models by using generalised linear modelling techniques, and the GLIM-software. Usually the models have been done for homogenous road sections. When making the first models, junctions were not separated from road sections. In later models, this separation has been done.

In the beginning we developed quite complicated accident models. The models fitted very well the data. Still there were problems with the "effects" i.e. parameter values of some variables. Because of complicated internal correlations, the "effects" of some variables differed remarkably from those known from many before and after studies. For example, many factors correlate with the speed limit. Two such factors are the number of unprotected road users and the land use along the road. These factors are not routinely coded in our road register, so their effects are reflected in the “effect” of speed limit.

After these difficulties, we attempted to develop some models with preset effects of some important factors (speed limit and the existence of road lighting and pedestrian/bicycle path).

Among road planners and other users of the accident models, there was uncertainty whether same kinds of problems arising from internal correlations would still exist. Hence we tried to make very simple accident models.

The purpose was to use the models together with the accident history to estimate the expected number of accidents in the past. This kind of information we can use for example when predicting the safety effects of some road improvements. When we attempt to understand the relationships between traffic and road conditions and the number of accidents, we can use more complicated models. Even then, relationships identified of the models must be separately tested by e.g. before and after studies.

4. MODEL COMPARISONS

When having several different models, we tried to do some comparisons to find out their strengths and weaknesses. We made the following tests:
A how well different models fit with the accident data they were fitted in
B how well different models fit accident data from a year different from those they were fitted in. Different kinds of time periods were used in the comparisons.

With different models we mean here:
- complicated accident models
- complicated accident models with some preset effects (speed limit ...)
- simple accident models.

When comparing accident models with accident data for a different year from those used in modelling, we also used the accident history alone as a predictor. Examples of different kinds of models are presented in Table 1.

Table 1. Examples of different kinds of models (Motor vehicle accidents on single carriageway main roads outside urban areas, road sections).

"Complicated model":
E= 0.156 * A1 * A2 * A3 * A4 * A5 * A6 * MILEAGE, where
E= expected number of injury accidents per year
MILEAGE= motor vehicle mileage as millions of kilometres/year
A1= 1.000 if speed limit = 50 km/h
A1= 0.619 if speed limit = 60 km/h or 70 km/h
A1= 0.662 if speed limit = 80 km/h
A1= 0.604 if speed limit = 100 km/h
A2= exp (0.00091 * (percentage of lighted road length))
A3= exp (-0.005882 * percentage of road length where 300 meter sight distances) A4= exp (0.0279 * percentage of heavy vehicles)
A5= exp (0.0748 * (busy private road junctions/road km))
A6= 1.127 if paved road, width of pavement under 6.9 meters
A6= 1.046 if paved road, width of pavement at least 6.9 meters
A6= 1 if the road is not paved (gravel road)

"Preset model":
E=0.1315*B1 *B2*B3*B4*B5*B6*MILEAGE, where
B1= 0.780 if speed limit = 50 km/h
B1= 0.850 if speed limit = 60 km/h
B1= 0.993 if speed limit = 70 km/h
B1= 1.000 if speed limit = 80 km/h
B1= 1.250 if speed limit = 100 km/h
B2= 1 - (0.1 * (percentage of lighted road length/100))
B3= exp (-0.009952 * percentage of road length where 300 meter sight distances) B4= exp (0.01485 * percentage of heavy vehicles)
B5= exp (0.1368 * (busy private road junctions/road km))
B6= 1.201 if paved road, width of pavement under 6.9 meters
B6= 1.110 if paved road, width of pavement at least 6.9 meters
B6= 1 if the road is not paved (gravel road)

"Simple model": E= 0.0173 * MILEAGE
NOTE: The “simple model” would not be so simple, if the comparison should include more variable surroundings. In the comparison only motor vehicle accidents on single carriageway main roads outside urban areas were included (road sections).
5. RESULTS

5.1 Fit for the data used in modelling

When developing the complicated accident models, it is very obvious that the model prediction fits quite well the accident data they were fitted from. It was not so sure that the models with some preset effects would fit very well the actual accident history.

Two preset effects were the effect of speed limit and the effect of a pedestrian/bicycle lane. We tested how well the preset models predict the number of accidents on roads with different speed limits.

The preset accident models were tested with the data from 10 highways where during five years had occurred 4 600 injury accidents. The number of accidents was predicted by the models separately for every road section, and the prediction was compared to accident history. The results are summarised in Table 2.

Table 2. The error of a preset accident model as a percentage of the accident history. The observed number of accidents (accident history, N) is also shown.

<table>
<thead>
<tr>
<th>Speed limit</th>
<th>Motor vehicles only</th>
<th>Unprotected road users</th>
<th>Animals</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 80 km/h</td>
<td>-24% N=388</td>
<td>+38% N=124</td>
<td>(+75%) N=2</td>
<td>-8%</td>
</tr>
<tr>
<td>80 km/h</td>
<td>-2% N=1238</td>
<td>+46% N=294</td>
<td>+111% N=27</td>
<td>+9%</td>
</tr>
<tr>
<td>100 km/h</td>
<td>+19% N=1993</td>
<td>+120% N=264</td>
<td>-14% N=259</td>
<td>+25%</td>
</tr>
<tr>
<td>Total</td>
<td>+7%</td>
<td>+70%</td>
<td>-1%</td>
<td>+16%</td>
</tr>
</tbody>
</table>

+ = the model predicted too many accidents
- = the model predicted too few accidents

We can see from Table 2 that the preset accident models have serious problems. When presetting some values, you can lose the good correlation between the predicted and observed number of accidents. The total number of motor vehicle and animal accidents was quite well predicted by the model. Nevertheless, there are great differences when looking at different speed limits. The predicted number of accidents involving unprotected road user (pedestrian, bicycle) was systematically higher than the observed. Because of the lack of traffic volumes for unprotected road users, the modelling of these accidents is quite difficult.

It seems that presetting the effect of speed limit and a pedestrian/bicycle lane causes bias when comparing the model to the history. So preset accident models are not good when attempting to estimate the expected number of accidents with the model.
This time we could not do the same calculations for the accident models that were done without presetting. In the past calculations the errors had not been this large.

5.2 Fit with the data from different years than used in modelling

When having all the important variables available, it is easy to make an accident model that fits the data well. And if the accident model is good, it can even predict the number of accidents in the future quite well.

To test this, we compared the accident model done for years 1987-1991 to the number of accident years 1992 and even 1985 - 1986. We did the comparisons only for one accident and road type: motor vehicle accidents on paved rural highways outside junctions.

The models described here were developed for over 5 000 homogenous road sections, the average length of which was 2.7 km. A total of 4 700 injury accidents occurred on the studied network in 1987 - 1991.

The accident predicting models compared were:

I accident history, the number of accidents unchanged *)
II accident history, the accident risk unchanged (accident rate)
III preset accident model (the effects of speed limit, pedestrian/bicycle lane and road lighting preset)
IV quite a complicated accident model

*) Homogenous road sections were so short that there were many sections with no accidents during the five years. Prediction by history fitted much better when the average risk was used as a predictor instead of zero accidents on those sections. So when having 0 accidents, the average risk was used instead.

The goodness of the predictions was estimated by counting how big proportion of the systematic, non-random variation the model can explain (the degree of explanation). The main results of the comparison are presented in Table 3.

Table 3. The degree of explanation when the predicted numbers of accidents are compared to the observed number of accidents in one year and three years.

<table>
<thead>
<tr>
<th>Prediction model</th>
<th>Degree of explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 year data</td>
</tr>
<tr>
<td>I accident history</td>
<td>27,0%</td>
</tr>
<tr>
<td>II risk history</td>
<td>42,0%</td>
</tr>
<tr>
<td>III preset model</td>
<td>42,1%</td>
</tr>
<tr>
<td>IV complicated model</td>
<td>42,8%</td>
</tr>
</tbody>
</table>
All the models predict much better the accident data of three years than just one year. In the number of accidents of one year there is so much random variation that it can not be predicted very well.

The accident history is not a good way of predicting accidents even if you should replace the prediction by the average risk when having no accidents. The number of accidents per road section is so low that there seldom happens accidents on the same sections they happened the year before.

The average risk is quite a good prediction at least for one homogenous road group and one accident group (automobile accidents on paved rural highways outside junctions).

The prediction does not improve very much when making the models more complicated. Preset and complicated accident models are almost as good in predicting accidents. The model without presetting is slightly better.

To study the importance of different explanatory factors, we can study how much of the accident variation different variables can explain. For instance at junctions, traffic flow variables (functions of incoming flows from major and minor road) explained 52 - 78% of the systematic variation of numbers of accidents. All the other statistically significant variables explained an additional 6 - 9% (Kulmala 1995).

We have had similar results concerning vehicle kilometres in explaining the number of accidents on road sections outside junctions. Hence, in terms of percentage of variation explained, traffic flows and mileage are quite in their own class. Naturally, this must not be understood as meaning that the other factors have only minor impact on accidents but to mainly reflect the large variation in vehicle flows and mileage - this variation really dominates the overall variation in the group of explanatory variables.

6. MAIN CONCLUSIONS

The number of accidents during one year can not be predicted well. Hence there is no meaning to make conclusions from accident data for one year when trying to identify hazardous road sections.

Motor vehicle mileage explains most of the variation of motor vehicle accidents. Adding more explaining variables does not improve the model very much. One important reason for this is that the accident models were done for homogenous group of road sections.

One can conclude that you can use quite simple accident models when estimating the expected number of accidents on a particular road section. To understand and illustrate the relationships between traffic and road conditions and the expected number of accidents, you probably need more complicated models.

In practice this means that reliable estimates of exposure are necessary for developing good accident models. This is true for motor vehicles, for which we have adequate flow data, but also for unprotected road users and animals, for which we have almost no exposure data at all.
This brings up a question: why to use a complicated accident prediction model when a simple one is just as good.

7. FROM CONCLUSIONS TO ACTIONS

At the beginning of 1990’s The Finnish Road Administration and VTT concluded that the prediction of accident saving due to road improvements should be done in two phases: 1) evaluation of the current safety situation on an existing road combining simple accident models and accident history 2) the safety effect of road improvements can be evaluated using the current safety situation and safety impact factors based on most reliable research results available around the world (Figure 1).

The principles of evaluating the safety effects of road improvements presented in figure 1 have been used in Finland since 1994. The calculations are done based on extensive and consistent database including all the public roads in Finland. The calculations are done with the help of evaluation programme called TARVA.

Using the TARVA – programme the safety effects of road improvements can be evaluated easily and using same data and definitions for the whole Finland. The minimum input is which measure and where it is implemented. There are almost 100 predetermined measures in the programme and own measures can be defined by the user if needed. Also the implementation costs can be entered but the average costs for measures (per km or per measure) are used, if these values are not entered.

The results of the calculations are the current safety situation on the modified road network, safety effects of improvements in all and by measure and economic figures to describe the effectiveness of safety improvements. The TARVA – programme has been used for evaluating all the safety effects of road improvements on public roads in Finland for more that ten years.

Figure 1. Evaluation of Traffic Safety Effects of Road Improvements
8. ACKNOWLEDGEMENTS

Accident model studies for rural highways have been commissioned by the Finnish Road Administration (FinRA).

9. LITERATURE


TARVA – home pages at web: http://www.tarva.net/tarvaintro.asp