PERGAMON

Accident Analysis and Prevention 35 (2003) 273–285

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Accident prediction models for urban roads

Poul Greibe

Danish Transport Research Institute, Knuth Winterfeldts Allé, DK-2800 Kgs. Lyngby, Denmark

Received 23 May 2001; received in revised form 16 November 2001; accepted 4 December 2001

Abstract

This paper describes some of the main findings from two separate studies on accident prediction models for urban junctions and urban road links described in [Uheldsmodel for bygader-Del1: Modeller for 3-og 4-benede kryds. Notat 22, The Danish Road Directorate, 1995; Uheldsmodel for bygader- Del2: Modeller for strækninger. Notat 59, The Danish Road Directorate, 1998] (Greibe and Hemdorff, 1995, 1998).

The main objective for the studies was to establish simple, practicable accident models that can predict the expected number of accidents at urban junctions and road links as accurately as possible. The models can be used to identify factors affecting road safety and in relation to 'black spot' identification and network safety analysis undertaken by local road authorities.

The accident prediction models are based on data from 1036 junctions and 142 km road links in urban areas. Generalised linear modelling techniques were used to relate accident frequencies to explanatory variables.

The estimated accident prediction models for road links were capable of describing more than 60% of the systematic variation ('percentage-explained' value) while the models for junctions had lower values. This indicates that modelling accidents for road links is less complicated than for junctions, probably due to a more uniform accident pattern and a simpler traffic flow exposure or due to lack of adequate explanatory variables for junctions.

Explanatory variables describing road design and road geometry proved to be significant for road link models but less important in junction models. The most powerful variable for all models was motor vehicle traffic flow.

Keywords: Accident prediction models; Generalised linear modelling; Road accidents; Junction; Road links

1. Background

Traffic planners and engineers in Denmark have used accident prediction models since the early 1980s. The models used were based on data from state and county roads, which are generally located in rural areas. An increasing need for models that could also be used for urban roads stimulated research on accident prediction models for urban junctions and road links beginning in the early 1990s.

The main objective of this study was to formulate practicable accident prediction models which would describe the expected number of accidents at junctions and road links in urban areas as accurately as possible. Furthermore, the models were to be used to identify factors affecting safety, e.g. geometry, land use, etc. Unlike the used models for rural roads, which were based only on traffic flows, this study was to include additional explanatory variables in the modelling.

E-mail address: Poul@greibe.dk (P. Greibe).
URL: http://www.dtf.dk

2. Description of data

In order to develop the models, detailed information on accident data, traffic flow and road design was required. Since the official Danish national road database only holds data on state and county roads (mainly rural), a close co-operation was established with a number Danish municipalities in order to collect data from urban roads.

Accident data were collected from the official accident statistics database covering all police recorded accidents. The data included all possible data on the accident, including a small text description of the nature and location of each accident. All accidents were related to the specific road link or road junction by use of the road ID number. A 5 years accident period was used (1987–1991 for junctions and 1990–1994 for road links).

Traffic flow counts were collected from the municipalities and converted to AADT counts. Unfortunately, the traffic counts did not always include heavy truck volumes and volumes of vulnerable road users, so these data fields were left blank.

0001-4575/02/$ – see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0001-4575(02)00005-2

Data on junction and road link geometry were collected partly by visual inspections and video recordings, and partly by use of local road databases at the municipalities.

The road network had to be divided into junctions and road links, since the objective was to make separate models for junctions and road links. Junctions with an incoming traffic flow on secondary arms of less than 250 and 500 AADT for 3- and 4-arm junctions, respectively were considered as part of the road link. These limits were chosen since most traffic count databases used had these as a minimum level for traffic flows. As a result of this definition, road links in this study may have a number of minor junctions or accesses along the road section. Variables describing the number of minor junctions and accesses are therefore included in the road link models.

2.1. Road link data

Data from 142 km urban road were collected and separated into 314 individual homogeneous road sections, giving an average road link length of approximately 450 m (S.D. = 285 m).

Data on the following variables were collected for road links:

(a) traffic flow (motor vehicles, heavy vehicles and vulnerable road users);
(b) length of road section;
(c) speed limit;
(d) one/two-way traffic;
(e) number of lanes;
(f) road width;
(g) speed reducing measures;
(h) number of minor crossings/exits/side roads;
(i) cyclist facilities;
(j) footway;
(k) central island;
(l) parking facilities;
(m) bus stop;
(n) land use.

The traffic flow data were considered as a continuous variable while all other data were converted into class variables with two–six levels. See Appendix A for a complete list of variables.

A total of 1058 police recorded accidents can be related to the 314 road links, 523 of which were personal injury accidents and 535 damage only accidents. The personal injury accidents resulted in 30 deaths and 554 injured persons.

Even though the road links only have small access or exit roads, more than 40% of the accidents can be related to crossing manoeuvres.

The accident distribution for road links can be found in Appendix A.

2.2. Junction data

The study included data from 1036 urban junctions. The number of each junction type is as follows: 85 (3-arm, signalised), 547 (3-arm, non-signalised), 250 (4-arm, signalised), 154 (4-arm, non-signalised).

Other junction types, e.g. roundabouts, have not been treated since there was an insufficient number of them.

Data on the following variables had been collected for junctions:

(a) traffic flows (motor vehicles, heavy vehicles and vulnerable road users);

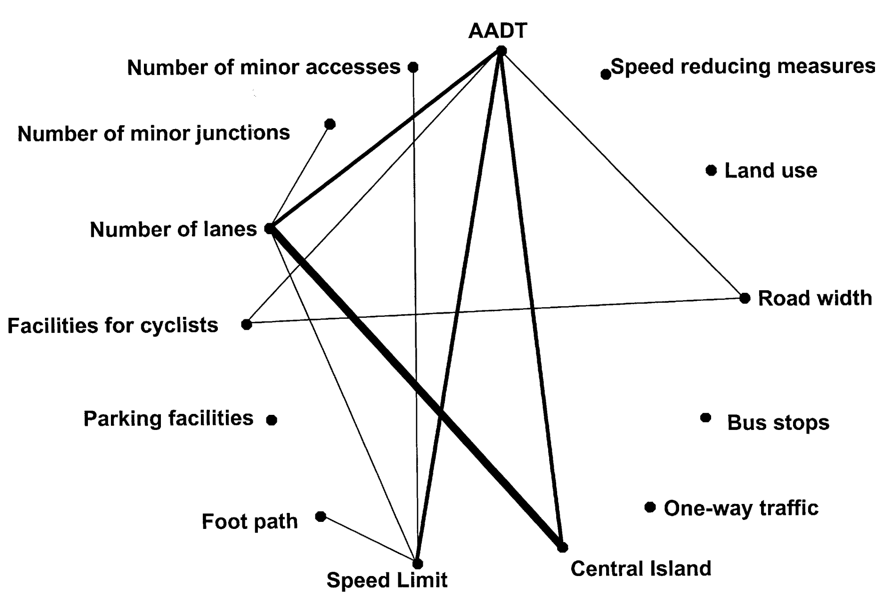

Fig. 1. Illustration of correlation matrix for road link data. Thick lines indicate 'strong' correlation ($\rho > 0.6$) and thin or no lines indicate 'weaker' correlation among variables. For example, the number of lanes and the presence of a central island correlate strongly.

(b) number of lanes;
(c) traffic island;
(d) turning lane;
(e) bicycle facilities;
(f) signalised/non-signalised;
(g) number of arms.

The traffic flow data were considered as a continuous variable while all other data were converted to class variables with two–seven levels. See Appendix A for a complete list of variables.

A total of 2534 police recorded accidents can be related to the 1036 junctions. The majority of accidents occurred in 4-arm signalised junctions and more than 50% of the 3-arm non-signalised junctions had zero observed accidents.

The accident distribution for junctions can be found in Appendix A.

In general, data on traffic flows for vulnerable road users and heavy vehicles were difficult to obtain, since these are rarely counted. This unfortunately also applies to junctions and for road links.

2.3. Correlation among data

Analysis of the data collected from junctions and road links showed some correlation among the explanatory variables.

Fig. 1 illustrates some general correlations within the data for road links based on the correlation matrix for the data-set. The motor vehicle traffic flow variable (AADT) correlated strongly ($\rho > 0.6$) with other explanatory variables, e.g. the presence of a central island, the number of lanes, road width, speed limit etc. The same conditions were found for data on junctions.

Due to the strong correlation among variables in the data-sets, it was not expected to be possible to estimate the true safety effect from a single road geometry variable.

3. Model structure

The models relate the number of observed accidents to traffic flow and road design. Generalised linear modelling techniques were used to fit the model, and the distributions of accident counts were assumed to follow a Poisson distribution. The regression analyses were performed by use of the GENMOD procedure in SAS.

Whether it is reasonable to assume that accident counts are Poisson distributed is a recurrent issue. The main advantage of the Poisson distribution is its simplicity, e.g. the variance is equal to its mean. However, difficulty arises concerning the phenomenon of "overdispersion" when the observed variance is actually greater than the mean. Overdispersion does not affect the coefficient estimates but does cause their standard errors to be underestimated (Miaou et al., 1992). Recent studies have proved that the negative binomial distribution might be more appropriate because it allows greater variance in the data and thereby deals with the overdispersion.

In this study, the Poisson distribution was chosen over the negative binomial simply because it is easier to use and, as the study was planned and initiated in the early 1990s, the Poisson was the only practicable solution with the statistical software available at that time. However, the final estimations were performed by use of the GENMOD procedure, which does support a number of different distributions, but subsequent estimations proved, as in other studies (Kulmala, 1995), that the values of the model parameters were almost the same regardless of the distributional assumption. It was therefore decided to stick to the pure Poisson model. It should be mentioned that no corrections were made for any possible overdispersion.

3.1. Model structure

Different ways to relate accident frequencies to traffic flows have been investigated in a number of previous studies, e.g. Satterthwaite (1981), Hauer et al. (1988), Bélanger (1994), Brüde and Larson (1993), Golias (1992), Hall (1986), Taylor et al. (1996), Maher and Summersgill (1996).

For road links the general opinion is that accident frequencies can be described by a flow function raised to a power. Often the flow function consists simply of the motor vehicle traffic flow along the link (AADT), but some studies (e.g. Summersgill et al., 1996) also include flows for pedestrians along or across the link.

The flow function for junctions are more complex since there is a number of individual traffic flows, especially when turning flows and flows for pedestrians and cyclists are included. In most cases, as in this study, the only flow data available will be the incoming motor vehicle traffic from each arm of the junctions, leaving limited possibilities to use complicated flow functions.

A number of TRL studies (e.g. Layfield et al., 1996) on junctions proved that models with more complicated flow functions based on data from individual traffic flow (including pedestrian flows) predict the accidents frequencies quite well. However, such models require data on turning flows, etc. which put a strain on practicality.

Recent studies (Mountain et al., 1998) include variables that allow for changes in risk over time in order to take any possible trend into consideration. This would ensure that the models do not become rapidly outdated. They found a 6% annual decrease in risk per year for junctions. However, to estimate annual changes in risk, large time series are required.

Previous Danish studies on accident prediction models (Wass et al., 1983; Kronbak and Greibe, 1994) have investigated models based on different flow functions of some with flows for cyclists and moped drivers as well. The general conclusion was that the models shown below are suitable for Danish conditions, and since these new models for urban roads were to be used along with existing models for rural roads, it was decided to use the same model structure here as well.

The model structure for road links

$$E(\mu) = aN^p \exp \sum \beta_j x_{ij}$$

where $E(\mu)$ is the expected number of accidents (accidents per year per km), N the motor vehicle traffic flow (AADT), x variables describing road geometry or environment of the road a, p, β_j are estimated parameters and for junctions

$$E(\mu) = aN_{\mathrm{pri}}^{p1} N_{\mathrm{sec}}^{p2} \exp \sum \beta_j x_{ij}$$

where $E(\mu)$ is the expected number of accidents (per year), N_{pri} the incoming motor vehicle traffic flow (AADT) from the primary direction, N_{sec} the incoming motor vehicle traffic flow (AADT) from the secondary direction, x variables describing road geometry, a, $p1$, $p2$, β_j are estimated parameters.

For non-signalised junctions, the primary direction represents the two arms where traffic has the right of way. For signalised junctions, the primary direction at the junction represents the two arms with the highest traffic volumes. It can not be rejected that the primary direction in signalised junctions represents two arms which are not opposite in the junction. This will be very seldom though.

3.2. Modelling procedure

At first, all variables were included in the model for the regression analysis. Insignificant variables were excluded one by one, starting with the least significant variables. Insignificant variables were identified based on likelihood ratio statistics and standard errors of the estimated parameter values. At the same time, the correlation matrix was studied and if two variables were correlated strongly with each other, one variable was excluded from the model on the condition that the model fit did not suffer significantly. If a variable turned out to be significant but did not contribute significantly to the decrease in deviance, the variable was excluded from the model.

Some explanatory variables were simplified, e.g. if a model showed that the parameter values for speed limits of 60 and 70 km/h did not differ much, the two categories were combined. All variables were also tested solely along with the traffic volume in order to identify any ‘hidden’ effects that did not show when other variables were included.

Separate models were estimated for: all accidents, all injury accidents, certain types of accidents (single accidents, rear-end accidents, crossing accidents, turning accidents, etc.). The purpose of modelling certain types of accidents was to examine the effect of road design factors on different accident types.

Individual models for accidents involving two-wheelers were also examined. Some of these are presented in this paper.

3.3. Goodness-of-fit

The total variation in the accident count consists of a random part (presumed Poisson distributed) and a systematic part. The model’s ‘goodness-of-fit’ was measured by how much of the systematic variation the model could explain and will be referred to as “percentage explained”. The method is proposed by (Kulmala, 1995) as being suitable for Poisson models. The percentage explained value are estimated on the basis of the scaled deviance of the studied model, the zero model (a model with only one constant parameter) and the expected scaled deviance of a model describing the total systematic variation.

Some problems arise in using this method when the average number of accidents is low (<0.5). Even though this study had a number of sites having zero accidents, we did not have mean values below 0.5. However, when modelling certain types of accidents, e.g. single accidents, mean values for some junction groups were below 0.5.

In general, the results (as described below) showed ‘percentage explained’ values for the models in the area of 30–80%. The best models were produced for road links.

4. Results

4.1. Junctions

Since there are considerable differences in the accident pattern and risk at different junction types, individual models were produced for four individual types of junctions, 3- and 4-arm, signalised and non-signalised, respectively.

Individual models for each junction type were estimated with all explanatory variables included. There was difficulty in achieving stable and significant explanatory variables in the junction models. This might be due to the fact that the safety effect of the individual explanatory variables is relatively small, and the models allow only for small variations in the estimated coefficients without becoming insignificant.

For all junction models, the variable for motor vehicle traffic proved to be the strongest variable, representing more than 90% of the systematic variation described by the entire model. Only few other explanatory variables were significant.

As previously mentioned, strong correlation within the data-sets caused interpretation difficulties of the explanatory variables. The “true” safety effect of the explanatory variables could therefore not be estimated properly.

To test whether the number of junctions in the data set was sufficient, models based on 40, 80, 120, 160 and 200 junctions were estimated. Serious estimation problems arose when the models were based on less than 80 junctions. The estimated parameter values did not change considerably when more than 100 junction were used. The standard error decreased as more junctions were added but the geometry variables were still not significant and it was estimated that

Table 1
Estimated parameter values for 4-arm signalised junctions[a]

Parameter		Estimated value
a		1.13×10^{-4} [0.41;4.04]
$p1$		0.56 [0.44;0.68]
$p2$		0.47 [0.37;0.56]
β_1: Number of lanes in	2	1.29 [1.08;1.56]
primary direction	>2	1
β_2: Number of lanes in	2–3	0.83 [0.72;0.96]
secondary direction	>3	1

[a] Values of 95% confidence limits are shown in brackets.

the number of junctions would have to be increased by three times in order to see any effect. Since it was only practicable to increase the number of junctions by a few hundred, the idea of adding more junctions to the data was rejected, as this would not help to make geometry variables significant.

An example of an accident prediction model for 4-arm signalised junctions with variables describing motor vehicle flow and other explanatory variables are presented in Table 1. The only significant variables, besides traffic flow were β_1 and β_2 describing the number of lanes.

The model results shown in Table 1 has a percentage explained value of 60%, which is only 1% better than a similar model that excludes explanatory variables describing road geometry. Equivalent results were found for other junction types. By far the most powerful variable was that of motor vehicle volumes. In general, variables describing junction design improved the models by only 1–4%.

The results for junctions indicate that simple models (only motor vehicle flow-functions) are almost as good as more complicated models. The reason for this must be the complicated internal correlations in junction design data or lack of good descriptive variables on road design.

Models with explanatory variables describing traffic volumes for vulnerable road users were only 4–5% better than models without these variables. The reason for this might also be strong correlation between motor vehicle traffic flows and traffic flows for vulnerable road users, or the fact that less than 5% of accidents involving vulnerable road users are reported to the police.

Results from the simple flow function models (only holding motor vehicle traffic flows) are presented in Table 2.

The best models were estimated for 4-arm signalised junctions, probably because this junction type had the highest number of observed accidents.

4.1.1. *Safety effect of signal control*

To evaluate the influence of signal control (traffic light) on the total number of observed accidents, a model was created with a two-level variable with or without signal control. The effect is given by the estimated coefficients.

The signal control variable was not significant in the model, which indicates that the expected total number of accidents is very similar for signalised and non-signalised junctions with the same flow function. In other words, non-signalised junctions are in general just as safe as signalised junctions with the same amount of incoming motor vehicles. This applies for an incoming motor vehicle traffic flow between 5,000–18,000 and 5,000–22,000 for 3- and 4-arm junctions, respectively.

With respect to different accident types, the figures for single and rear-end accidents are significantly higher in signalised junctions than in non-signalised junctions. The number of crossing accidents is, however, lower in signalised junctions. The estimated safety effects are shown in Fig. 2.

It was expected that signalised junctions would have more rear-end accidents and fewer accidents with crossing road users but also the number of single accidents are higher at signalised junction. The reason for this is unknown but one reason could be the fact that signalised junctions are more complex with more obstacles and other equipment that drivers can run into.

Other effect studies on signal control (Elvik et al., 1997) finds by use of meta-analysis on a number of before–after studies a 15% reduction in accidents for 3-arm junction and 30% reduction for 4-arm junctions.

These figures can not be confirmed in this study. However, considering only personal injury accidents (not shown in Fig. 2), a non-significant reduction of 9 and 12% were found for 3- and 4-arm junctions, respectively.

4.1.2. *Cyclist safety at junctions*

As 32% of the accidents in junctions involve cyclists or moped riders, special models to predict these accidents were formulated. Unfortunately, only 4-arm signalised junctions had sufficient traffic flow data to permit regression analysis.

The model below is estimated on a subgroup of junctions comprising only 41 4-arm signalised junctions. The results were

$$E(\mu_{\text{cyc/mop}}) = 8.01 \times 10^{-8} N_{\text{tot}}^{0.91} C_{\text{tot}}^{0.86}$$

Table 2
Estimated parameter values for junction models[a]

Junction type	a	$p1$	$p2$	'Percentage explained'
3-Arm, non-signalised	1.04×10^{-5} [0.41;3.43]	0.69 [0.57;0.82]	0.60 [0.48;0.72]	47
3-Arm, signalised	1.34×10^{-5} [0.07;23.4]	0.88 [0.58;1.19]	0.33 [0.14;0.50]	32
4-Arm, non-signalised	7.12×10^{-4} [1.16;43.7]	0.30 [0.15;0.44]	0.55 [0.34;0.76]	29
4-Arm, signalised	1.08×10^{-4} [0.51;3.08]	0.53 [0.42;0.64]	0.52 [0.44;0.60]	59

[a] Values of 95% confidence limits are shown in brackets.

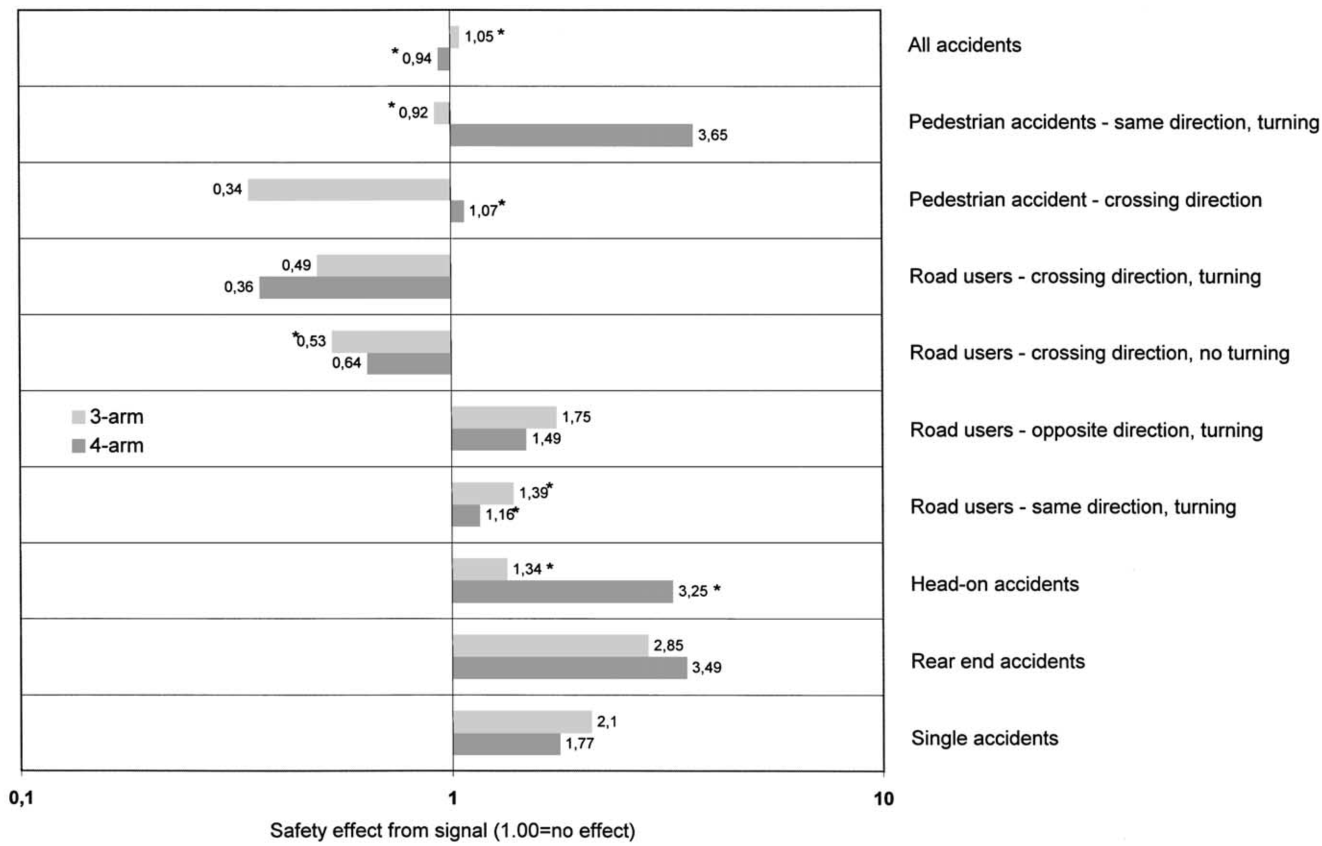

Fig. 2. Safety effect for the number of accidents at signalised junctions compared to non-signalised junctions. 1.00: no effect. (e.g. signalised 3-arm junctions have 2.85 times more rear-end accidents than non-signalised 3-arm junctions with the same traffic flows). Figures marked with an * are not significant.

where $E(\mu_{\text{cyc/mop}})$ is the expected number of accident involving cyclists or moped riders per year, N_{tot} the incoming motor vehicle traffic flow (total sum, AADT), C_{tot} the incoming cyclist/moped traffic flow (total sum, AADT).

Explanatory variables describing facilities for cyclist (cycle track/path) were not significant and stable in the model. The variable for traffic flows for cyclists and moped riders improved the 'percentage explained' value by 5% from 70 to 75%. It seems that although the explanatory variable for cyclist and moped traffic flows are statistically significant, they do not improve the models substantially. In view of the difficulty of obtaining flow data for cyclists and moped riders, it can be questioned whether it is worth using explanatory flow data variables for cyclists and moped riders in the models.

4.2. Road links

Road links with a speed limit of 30, 40 or 80 km/h were excluded from the data sets since these road sections were represented by a total of less than 20 km, and furthermore differed considerably from other road sections. Only sections with a speed limit of 50, 60 or 70 km/h are represented in the following models.

As mentioned earlier, generalised linear modelling techniques were used to fit the models and the variation in accidents was assumed to follow a Poisson distribution.

Separate models were estimated for:

(a) all accidents (injury and damage only accidents);
(b) injury accidents;
(c) certain types of accidents (single accidents, rear-end accidents, head-on accidents, accidents involving turning traffic, accidents involving pedestrians, accidents involving parked vehicles, etc.).

The purpose of modelling certain types of accident, was to examine the effect of road design factors on different accident types. For reasons of space, only results for models based on all accidents (injury and damage only accidents) will be presented in the following sections.

4.2.1. Simple flow function model

First a simple model containing only variables for motor vehicle traffic flow was tested. The result was

$$E(\mu) = 2.44 \times 10^{-3} N^{0.75}$$

where $E(\mu)$ is the expected number of accidents (per km per year), and N the motor vehicle traffic flow measured as AADT.

This model had a 'percentage explained' value of 30%. As found in other studies (Hemdorff, 1996; Summersgill and Layfield, 1996) the traffic volume is raised to a power of less than 1, in this case 0.75.

4.2.2. Models with explanatory variables

A model with other explanatory variables included was tried and the results are given below. Note that variables for vulnerable road user traffic flows were not included since data on this were insufficient.

The model below describes the total number of accidents (injury and damage only accidents). Only variables that proved to be significant are included. The model (simplified) has the following structure:

$$E(\mu) = aN^p \beta_{1,i}\beta_{2,i}\beta_{3,i}\beta_{4,i}\beta_{5,i}\beta_{6,i}$$

where $E(\mu)$ is the expected number of accidents (per km per year), N the motor traffic flow (AADT), β_{1-6} class variables-represents certain road design or road layout parameters, a, p, $\beta_{n,i}$ are estimated parameters.

The estimated parameters are shown in Table 3.

As an example, the estimated accident frequency for a road link with an AADT of 4000, speed limit of 50 km/h, road width of 5.0–7.5 m, 5–40 exits per km, 0–5 minor crossings per km, limited parking, situated in a residential neighbourhood would be

$$\begin{aligned} E(\mu) &= 6.09 \times 10^{-4} \times 4000^{0.80} \times 2.25 \times 0.83 \times 1.00 \\ &\quad \times 0.75 \times 1.00 \times 1.58 \\ &= 1.03 \text{ accidents per km per year} \end{aligned}$$

An interpretation of the variable values, e.g. β_4 (number of minor side roads), indicates that the accident risk rises as the number of minor side roads per km increases. A road link with >10 side roads per km (value 1.25) has a 1.7 times higher accident risk than a road link with 0–5 side roads per km (value 0.75). Furthermore, road sections in shopping districts have approximately 2.4 times higher risk than roads in areas with scattered housing.

The 'percentage explained' value for the model above is 69%, which is considerable higher than a model without the other explanatory variables (30%).

Similar models have been estimated for different accident types. In general, variables describing traffic flow, land use, number of minor crossings, parking facilities and speed limits proved to be the most important variables in the models.

Since the models did not include exposure data on vulnerable road users, it must be expected that other variables, e.g. land use and speed limit, to a certain extend explain the level of exposure for vulnerable road users. This relates to the fact that land use and speed limit turn out to be among the most descriptive variables in the data-set.

4.2.3. Cyclist safety on road links

A simple flow model describing the number of accidents involving cyclists and moped riders as a function of the traffic flows for motor vehicles, cyclists and mopeds were tested. The results are given below

$$E(\mu_{\text{cyc/mop}}) = 4.44 \times 10^{-5} N_{\text{car}}^{0.68} N_{\text{cyc/mop}}^{0.49}$$

where $E(\mu_{\text{cyc/mop}})$ is the expected number of accident involving cyclists and moped riders (per km per year), N_{car}

Table 3
Estimated parameter values[a]

Parameter		Estimated value
a		6.09×10^{-4} [1.90;19.5]
p		0.80 [0.68;0.92]
β_1: Speed limit	50 km/h	2.25 [1.66;3.05]
	60 km/h	2.85 [2.07;3.90]
	70 km/h	1
β_2: Road width	5.0–7.5 m	0.83 [0.70;0.99]
	8.0–8.5 m	0.68 [0.57;0.81]
	9.0–15.0 m	0.80 [0.67;0.95]
β_3: Number of exits per km	0 and >40	1
	5–40	1
β_4: Number of minor side roads per km	0	0.72 [0.58;0.89]
	0–5	0.75 [0.64;0.88]
	5–10	1
	>10	1.25 [1.05;1.48]
β_5: Parking	Prohibited	1.19 [1.48;2.11]
	Rarely	1
	Bays at kerb	1.77 [1.48;2.11]
β_6: Land use	Shops	2.44 [1.81;3.28]
	Blocks of flats	1.56 [1.21;2.03]
	Industrial/residential/neighbourhood	1.58 [1.32;1.89]
	Scattered housing	1

[a] Values of 95% confidence limits are shown in brackets.

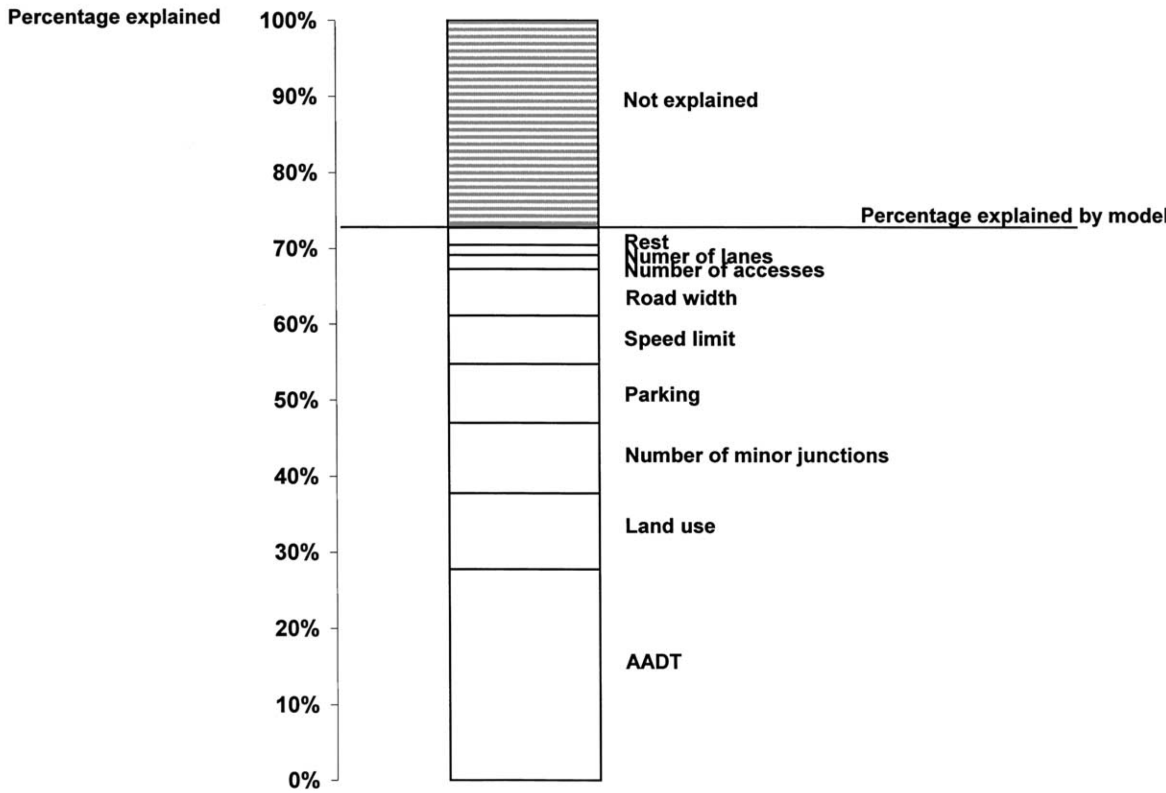

Fig. 3. Percentage of systematic variation explained by individual variables in the road link model.

the motor vehicle traffic flow (AADT), $N_{\mathrm{cyc/mop}}$ the traffic flow (AADT) for cyclists and moped riders.

The model was estimated on a data sub-set comprising only 26 road sections.

4.2.4. Percentage explained by variables in road link models

A model with all variables included gives a 'percentage explained' value of 72%. Fig. 3 illustrates how much the individual variables contribute to the overall 'goodness-of-fit'.

As mentioned earlier, the 'percentage explained' value indicates the level of systematic variation explained by the model. The AADT variable is the most powerful variable (describing approximately 30%), followed by variables for land use, number of side roads, etc.

Even though all variables are represented in the models, 30% of the systematic variation remains unexplained. A likely explanatory variable that could increase the 'percentage explained' value might be the actual speed of the motor vehicles. We know from other studies that speed is one of the most significant factors in traffic safety. Another likely explanatory variable that could increase the 'percentage explained' value could be better exposure data for vulnerable road users and better exposure data for vehicles entering the road link. As mentioned earlier a high proportion of accidents at road links involved vehicles entering the link from side roads or accesses and/or accidents with vulnerable road users.

The figure also illustrates that it is possible to explain more than 50% of the systematic variation by use of the four–five most important variables, and that additional explanatory variables do not improve the models very much.

4.2.5. Explanatory variables and their safety effect

One of the objectives of this study was to evaluate the safety effect of different explanatory variables, e.g. road layout, etc. Despite the strong correlation among the variables in the data-set, which may lead to incorrect conclusions, the effects derived by the models are given in the following sections. The list below is based on Table 3 and results on different types of accidents, which are not presented elsewhere in the article. Only variables that proved to be significant are shown.

4.2.5.1. AADT. The motor vehicle traffic flow is the most important variable in the models. The results showed that accident frequencies are related to the AADT raised to power of 0.8–1.0. However, for some accident types, e.g. rear-end and single accidents, the parameter value were 1.23 and 0.52, respectively. These deviations correspond with results found by (Summersgill and Layfield, 1996).

4.2.5.2. Speed limits. The modelling showed that road links with high speed limits tend to have lower accident risk. This does not mean that high speeds in general are safer, rather the results for this variable illustrate the correlation

Table 4
Description of road types

Road type	Land use	Minor side road per km	Speed limit (km/h)
A–D	Shopping/city centre	0–10	50–60
E	Blocks of flats/residential housing/industry	0–10	50–60
F	Blocks of flats/residential housing/industry	0–10	70
G	Blocks/residential housings/industry	>10	50–60
H	Scattered	–	50–60
I	Scattered	–	70

problems within the data set. High speed roads tend to have few vulnerable road users and to be situated in sparsely built-up areas.

4.2.5.3. Number of lanes. Road links with only one lane (no marked centre line) have more accidents involving motor vehicles going in the same direction than road links with two or more lanes.

4.2.5.4. Road width. Apparently, road links with a road width (from kerb to kerb) of 8–8.5 m have the lowest accident risks for most accident types.

4.2.5.5. Speed reducing measures. Road links with speed reducing measures have a higher risk of single accidents. Even though it is well-known that speed reducing measures usually improve safety, the explanatory variable describing

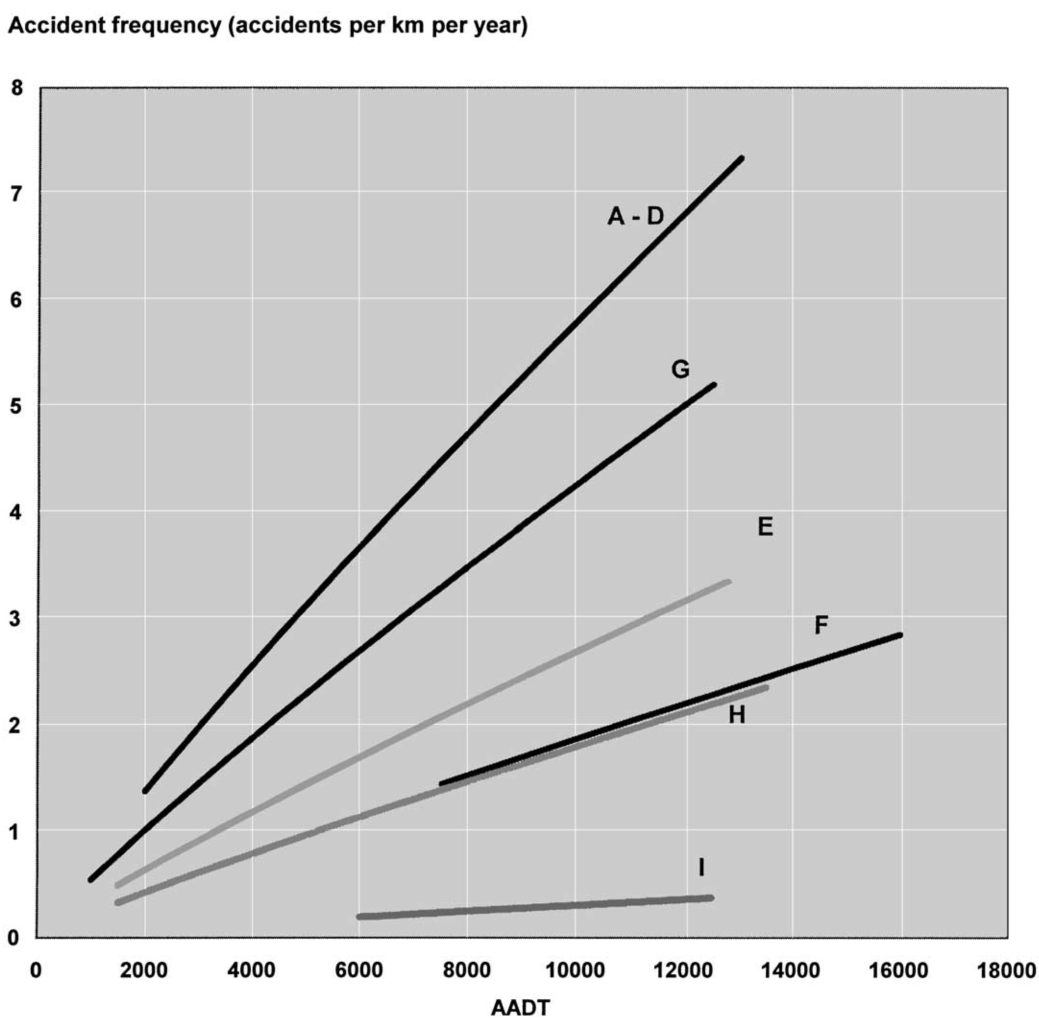

Fig. 4. Accidents per km per year as a function of the AADT for six urban road types.

the presence of speed reducing measures in the models was not significant in most cases. It should be noted that only a few km of road with speed reducing measures were represented in the data material.

4.2.5.6. Number of accesses. The relation between accident risk and the number of accesses (exits from private properties, parking places, etc.) seems to be an inverted 'U-shape'. Roads with no accesses and roads with a large number of accesses have the lowest accident risk, while roads with a medium number of accesses have the highest accident risk.

4.2.5.7. Number of minor side roads. The explanatory variable for the number of minor side roads proved to be very important in the models. In general, the more side roads, the higher the accident risk.

4.2.5.8. Parking conditions. Road links with parked motor vehicles along the roadside (at kerb) or in marked parking bays have the highest accident risk, particularly for accidents involving pedestrians, accidents involving motor vehicles from access roads or minor side roads, and for accidents involving parked vehicles. Other studies, (Elvik et al., 1997) also find that marked parking bays increases risk.

4.2.5.9. Land use. The road environment (type and function of buildings along the road) has a considerable influence on the accident risk. Shopping streets and city centre roads have significantly higher accident risk than, for example, residential roads in less densely built-up areas. In general, the lower the building density, the lower the accident risk. Since exposure data for vulnerable road users were not included in the models, it must be presumed that the variable 'land use' to some extent also represents the level of pedestrian and cyclist activity.

4.2.5.10. Urban road types. Based on the collected data, six urban road link types have been defined, and accident prediction models have been estimated for each road type. For practical reasons only six road types were defined, based on the most important variables found from the models. The road type definitions are therefore based on information on land use, number of minor side roads and on the speed limit.

The six road types are shown in Table 4.

Fig. 4 illustrates the expected average accident frequency for the six urban road types based on accident prediction models.

The six road types were produced by starting from 23 road types, and then grouping the roads step by step until all individual road types had significant parameter estimates, while at the same time attempting to keep an overall logical and functional classification of the road type. From a statistical viewpoint road type F and H could possibly be grouped as one road type, but since these roads are very different in geometry, etc. it would not be consistent with the overall logic to put these roads in the same group.

5. Conclusion and discussion

The purpose of the study was to formulate accident prediction models which would describe the expected number of accidents at junctions and on road links in urban areas as accurately and practicably as possible. Furthermore, the objective was to identify factors affecting safety at junctions and on road links by use of the explanatory variables in the models.

Based on data from 1024 junctions and 142 km of road links, a number of models were estimated in two separate studies.

In general, the most powerful variables in the models were those describing the motor vehicle traffic flow, particularly for junctions. Additional explanatory variables in the junction models did not improve the 'percentage explained' value considerably. For road links, however, additional explanatory variables describing the road environment, number of minor side roads, parking facilities and speed limit proved to be significant and important variables for predicting the number of accidents.

The models produced had a 'percentage explained' value in the area of 40–80%, with junction models seems to be at the lower end. This indicates that modelling accidents is less complex for road links than for junctions.

The safety effect from signal control in 3- and 4-arm junctions was examined and the results showed that non-signalised junctions in general are just as safe as signalised junction with the same traffic flows. The accident distribution differs though: for example, signalised junctions have fewer crossing accidents but more rear-end accidents.

A high proportion of accidents in urban areas involve cyclists or moped riders. Models estimating this accident group were tested including explanatory variables describing exposure data for cyclists and moped riders. Models with exposure data for cyclists and moped riders were only slightly better than models without.

A major problem was strong internal correlation within the data. Variables describing traffic flow tend to correlate strongly with other variables like road width, number of lanes, etc. Hence, the safety effects from a single explanatory variable were difficult to estimate since it may be affected by other variables in the model. Variables describing traffic flows for vulnerable road users (cyclist and mopeds) were also included in some of the models. They were significant but did not contribute considerably to the 'percentage explained' values, mainly due to strong correlation with motor vehicle traffic flow variables.

Another well-known problem in safety analysis is the relatively small number of observed accidents in the data, which may cause problems in the statistical studies. In the study on junctions for example, less than 50% of the 3-armed non-signalised junctions had any police reported accidents, which limited modelling possibilities. Efforts were made to estimate accidents involving cyclists or moped riders, but since less than 10% of these accidents are reported by the police, the reliability of the data is limited, which complicates the modelling.

In general, the safety effect from factors like road geometry, road environment, etc. can be estimated in various ways. The most reliable way is by use of controlled ‘before–after’ studies. However, ‘before–after’ studies require a large number of sites and long study periods. An alternative is to make multidimensional cross-tabulation of accident rates by different safety factors, a so-called ‘with-or-without’ study, e.g. accident rates for junctions with or without signal control. However, it is only possible to cross-tabulate for a limited number of variables/factors and these variables cannot be continuous. Furthermore, comparing accident rates at different sites can be complicated since differences in geometry, etc. can rarely be explained by a few variables.

The use of models has some advantages over the above-mentioned study methods. Models relate the number of accidents to selected factors that can be explained by either continuous or class variables. In addition, in the models, the accident number is assumed to follow a certain statistical distribution, e.g. the Poisson or negative binomial distribution.

However, the safety effects from various factors found in this study were not always absolutely comparable to the safety effects found in other studies, e.g. traditional before–after studies. The reason for this could be the internal correlation problems within the data sets as mentioned earlier. It is therefore recommended to interpret the safety effect of a single variable with caution. Nevertheless, the accuracy of safety effects found by modelling must be considered better than that of ‘with-or-without’ studies, but worse than that of ‘before–after’ studies.

Currently, the practical use of the developed accident prediction models (black spot identification, safety effects from changes in traffic flow, etc.) described in this paper are being tested and demonstrated in a number of Danish municipalities. In relation to this, a handbook on the use of accident prediction models has been published (Greibe and Hemdorff, 2001).

Appendix A

Number of junctions and road links according to the number of accidents

Number of accidents	Road links	Junctions			
		3-Arm		4-Arm	
		Signalised	Non-signalised	Signalised	Non-signalised
0	70	19	278	25	40
1	62	13	132	22	34
2	41	8	61	28	25
3	28	13	29	30	17
4	21	11	19	21	11
5	18	6	12	27	17
6	18	4	9	18	3
7	15	4	4	15	4
8	11	0	2	13	2
9	10	2	1	14	1
10	3	1	0	7	0
11	3	2	0	7	0
12	4	1	0	3	0
13	3	0	0	3	0
14	1	1	0	2	0
15	3	0	0	4	0
16	1	0	0	3	0
17	0	0	0	2	0
18	1	0	0	1	0
19	0	0	0	1	0
20	0	0	0	2	0
21	0	0	0	0	0
22	1	0	0	1	0
23	0	0	0	0	0
24	0	0	0	1	0

Class variable list for road link models

Explanatory variable	Value	Road data		Accident data		
		Road length (km)	Travelled vehicle (million km)	Total (5 years)	Frequency (per km per year)	Rate (per million vehicle km)
Speed limit	30–40 km/t	4.5	4.0	15	0.67	0.74
	50 km/t	97.5	185.5	767	1.57	0.83
	60 km/t	20.8	53.1	204	1.96	0.77
	70 km/t	13.2	47.8	54	0.82	0.23
	80 km/t	6.3	37.5	18	0.57	0.10
One-way street	Yes	3.3	5.8	15	0.91	0.52
	No	139.0	322.2	1043	1.50	0.65
Number of lanes	1	35.3	42.5	188	1.06	0.89
	2	97.2	233.9	805	1.66	0.69
	4	9.8	51.6	65	1.33	0.25
Road width	5.0–6.0 m	33.6	49.8	203	1.21	0.82
	6.5–7.5 m	37.2	68.7	275	1.48	0.80
	8. 0–8. 5m	38.1	82.0	249	1.31	0.61
	9.0–15.0 m	33.3	127.5	331	1.99	0.52
Speed reducing measures	Yes	16.3	16.8	81	1.00	0.96
	No	126.0	311.1	977	1.55	0.63
Accesses per km	0	27.4	83.3	130	0.95	0.31
	0–10	29.2	84.0	245	1.68	0.58
	10–20	35.0	82.0	320	1.83	0.78
	20–30	18.0	29.5	174	1.93	1.18
	30–40	12.7	24.6	102	1.60	0.83
	>40	20.0	24.6	87	0.87	0.71
Side roads per km	0	26.2	96.6	140	1.07	0.29
	0–5	47.2	102.0	285	1.21	0.56
	5–10	48.1	93.6	425	1.77	0.91
	>10	20.8	35.7	208	2.00	1.16
Cyclist facilities	None	58.5	81.2	314	1.07	0.77
	Lane	17.2	31.0	106	1.23	0.68
	Track	66.6	215.8	638	1.92	0.59
Footway	Yes	124.6	282.6	1010	1.62	0.71
	No	17.6	45.4	48	0.54	0.21
Median divider	Yes	12.4	63.8	149	2.41	0.47
	No	129.9	264.1	909	1.40	0.69
Parking	Bays at kerb	14.8	38.2	240	3.25	1.26
	Prohibited	54.0	144.9	384	1.42	0.53
	Rarely	73.5	144.8	434	1.18	0.60
Bus stop	Yes	98.2	234.4	833	1.70	0.71
	No	44.0	93.5	225	1.02	0.48
Land use	Shopping	4.7	10.2	85	3.65	1.66
	Block of flats	9.3	26.6	139	2.99	1.04
	Residential	72.5	132.2	548	1.51	0.83
	Scattered housings	47.7	139.0	218	0.91	0.31
	Industry	8.0	19.9	68	1.69	0.68
Total		142.3	328.0	1058	1.49	0.65

Class variables for junctions

Explanatory variable	Value	3-Arm				4-Arm			
		Non-signalised		Signalised		Non-signalised		Signalised	
		Number	Accident rate	Number	Accident rate	Number	Accident rate	Number	Accident rate
Lanes in primary direction	1	68	0.08	0	–	35	0.27	0	–
	2	376	0.08	14	0.09	91	0.18	40	0.25
	3	75	0.08	35	0.12	19	0.15	107	0.19
	4	17	0.08	8	0.13	5	0.08	33	0.18
	5	9	0.09	3225	0.12	4	0.14	39	0.21
	6	1	0.22	5	0.11	0	–	27	0.18
	7	1	0.09	1	0.25	0	–	4	0.24
Lanes in secondary direction	1	100	0.06	2	0.06	15	0.20	3	0.22
	2	436	0.09	38	0.10	134	0.18	112	0.18
	3	10	0.13	38	0.14	4	0.24	93	0.18
	4	1	0.00	5	0.11	0	–	24	0.24
	5	0	–	2	0.13	1	0.29	13	0.25
	6	0	–	0	–	0	–	4	0.22
	7	0	–	0	–	0	–	1	0.34
Turning lanes in primary direction	Yes	97	0.09	71	0.12	26	0.15	214	0.20
	No	450	0.08	14	0.10	128	0.20	36	0.19
Turning lanes in secondary direction	Yes	14	0.10	48	0.13	4	0.25	139	0.20
	No	533	0.08	37	0.09	150	0.18	111	0.18
Island at primary direction	Yes	113	0.10	72	0.12	43	0.16	191	0.20
	No	434	0.08	13	0.09	111	0.20	59	0.18
Island at secondary direction	Yes	71	0.10	45	0.12	44	0.21	139	0.21
	No	476	0.08	40	0.12	110	0.17	111	0.18
Total number		547	0.08	85	0.12	154	0.18	250	0.20

References

Bélanger, C., 1994. An estimation of the safety of four legged unsignalized intersections. In: Proceedings of the 73rd Annual Meeting of Transportation Research Board, Washington, DC.

Brüde, U., Larson, J., 1993. Models for predicting accidents at junctions where pedestrians and cyclists were involved. Anal. Prevention 25 (5), 499–509.

Elvik, R., Mysen, A.B., Vaa, T., 1997. Trafikksikkerhetshånbok-3. udgave. TØI.

Golias, J.C., 1992. Establishing relationships between accidents. Anal. Prevention 24 (6), 689–694.

Greibe, P., Hemdorff, S., 1995. Uheldsmodel for bygader-Del1: Modeller for 3-og 4-benede kryds. Notat 22, The Danish Road Directorate.

Greibe, P., Hemdorff, S., 1998. Uheldsmodel for bygader- Del2: Modeller for strækninger. Notat 59, The Danish Road Directorate.

Greibe, P., Hemdorff, S., 2001. Håndbog i trafiksikkerhedsberegninger-Brug af uheldsmodeller og andre vurderingsmetoder. Rapport 220, The Danish Road Directorate.

Hall, R.D., 1986. Accidents at 4-arm single carriageway urban traffic signals. TRRL, Contractor Report 65.

Hauer, E., Jerry, C.N., Lovell, J., 1988. Estimation of Safety at Signalized Intersections, Transportation Research Record 1185.

Hemdorff, S., 1996. Introduktion til VISplet. Notat 29. The Danish Road Directorate.

Kronbak, J., Greibe, P., 1994. Uheldsmodeller for signalregulerede vejkryds. Technical University of Denmark.

Kulmala, R., 1995. Safety at rural 3- and 4-arm junctions. Development and application of accident prediction models. VTT Publications 233.

Layfield, R.E., Summersgill, I., Hall, R.D., Chatterjee, K., 1996. Accidents at urban priority crossroads and staggered junctions. TRL report 185.

Maher, M.J., Summersgill, I., 1996. A comprehensive methodology for the fitting of predictive accidents models. Acci. Anal. Prevention, Vol. 28 (3).

Miaou, S.-P., Hu, P.S., Wright, T., Rathi, A.K., Davis, S.C., 1992. Relationship between truck accidents and highway geometric design: a poisson regression approach. Transport Research Record 1376.

Mountain, L., Maher, M., Fawaz, B. 1998. The influence of trend on estimates of accident at junctions. Acci. Anal. Prevention 30 (5).

Satterthwaite, S.P. 1981. A survey of research into relationships between traffic accidents and traffic volumes. TRRL Supplementary Report 692.

Summersgill, I., Kennedy, Janet V., Baynes. D., 1996. Accidents at three-arm priority junctions on urban single-carriageway roads. TRL report 184.

Summersgill, I., Layfield, R.E., 1996. Non-junction accidents on urban single-carriageway roads. TRL report 183.

Taylor, M.C., Hall, R.D., Chatterjee, K., 1996. Accidents at 3-arm traffic signals on urban single-carriageway roads. TRL report 135.

Wass, C., Jørgensen, N.O., Jørgensen, E., 1983. Model for uheld i vejkryds. Vejdatalaboratoriet, Danmarks Tekniske Højskole.