Predictive models for accidents on urban links - A focus on vulnerable road users

Jonsson, Thomas

2005

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Bulletin 226

Predictive models for accidents on urban links

A focus on vulnerable road users

Thomas Jonsson

2005

Lund University
Lund Institute of Technology
Department of Technology and Society
Traffic Engineering

Predictive models for accidents on urban links

A focus on vulnerable road users

Thomas Jonsson

Doctoral Thesis

CODEN:LUTVDG/(TVTT-1033)1-142/2005
ISSN 1404-272X

Bulletin 226

Thomas Jonsson

Predictive models for accidents on urban links

A focus on vulnerable road users

Keywords:
traffic safety, accident, crash, prediction, modelling, vulnerable road users, safety, urban, links, speed

Abstract:
Much of earlier work on predictive models for accidents has been focused on rural traffic or urban intersections. This work has aimed at identifying and investigating possible improvements to predictive models for accidents on urban links. A special focus has been on the accidents of vulnerable road users.
The possible improvements investigated have been: a) the use of exposure data for vulnerable road users, b) the use of actual vehicle speeds, c) to separate vehicle accidents into single vehicle and multiple vehicle accidents and model each type separately.
The study involved eight Swedish cities, of which six were included in the development of accident models. In addition to police-reported traffic accidents, data for the models were collected in specific field studies and partly from the cities.
The study shows that the inclusion of exposure data for vulnerable road users in the models for vulnerable road users' accidents greatly improves the predictive ability of the models. Vehicle speeds were found to be very difficult to use in the models as vehicle speeds are highly correlated with most of the other variables and make the model coefficients unstable. The separate modelling of single and multiple vehicle accidents failed partly as the single vehicle accidents were too few for constructing sound models. The models for multiple vehicle accidents developed, however, had a better predictive ability than models for all vehicle accidents. Land use was the next important explanatory variable for most accident types after the exposure variables.
Accident models were developed with one data set and validated with another data set. The models performed very well in prediction by explaining between 71% and 81% of the systematic variation in the validation data. The validation indicated that exponents were 0.5 for both the flows of pedestrians and motor vehicles in models for accidents involving vulnerable road users, and 1.0 for the motor vehicle flow exponent in the models for motor vehicle accidents. For bicyclist accidents the correct exponent for bicyclist flows is likely to be somewhat lower than 0.5, close to 0.35.
The study also recommended practical uses for the models, listed the lessons learned from the modelling as well as proposed a number of topics for further research.

Citation instructions:
Jonsson, Thomas, Predictive models for accidents on urban links – A focus on vulnerable road users., Bulletin 226, Lund University, Lund institute of technology, Department of Technology and Society, Lund, 2005

With support from:
Vägverket

Institutionen för Teknik och samhälle
Lunds Tekniska Högskola
Avdelning Trafikteknik
Box 118, 221 00 LUND, Sverige

Department of Technology and Society
Lund Institute of Technology
Traffic Engineering
Box 118, SE-221 00 Lund, Sweden

Acknowledgements

There are many people without whom this work would not have been what it is today. There is not enough space here to thank them all, but I wish to mention a few.

First of all I wish to thank my supervisors who have been ever patient with me and who have supported me with their extensive knowledge: Christer Hydén, my main supervisor, with his lifelong experience at the front of traffic safety research, Lars Ekman for his enthusiasm and ever brilliant new ideas and Risto Kulmala for his expert support in accident modelling. A special *in memoriam* for Ola Hagring who, unfortunately, is no longer with us, he always added a practical and orderly dimension to the otherwise visionary but chaotic supervision meetings.

In the final phase of the writing I had the privilege of very proficient feedback from the participants at my final seminar. I am ever thankful to Rune Elvik (TÖI, Norway), Stig Danielsson (Linköping University), Ulf Brüde (VTI) and Torsten Berg (SNRA) for this!

In the course of the project I've received much support from several people who deserves much gratitude: Hans Thulin and Alexander Obrenovic for providing me with compiled accident data, Björn Lundberg for assisting me with the field studies, Östen Johansson for the introduction to VITS and Karl Bång for the project management.

A warm thank you to all the present and past colleagues at the department, I don't have words to describe how important it has been for me to work in such an inspiring environment!

Finally an enormous thank you to my family who has not seen as much of me the last few years as they would have liked. My visits have been too few and short due to my dissertation work. Your support means the world to me! I'm sad that my mother is no longer with us so she could also see her second son complete his PhD.

Lund 2005-01-30

With gratitude

Thomas Jonsson

This work has been financed by the Swedish National Road Administration.

Contents

Summary

Background and purpose

Accident models are used to explain and predict the accident occurrence in different traffic environments. The accident models are normally divided into separate models for rural and for urban environments, and separate models for large intersections and for streets including minor intersections (called links). The models are good at predicting the expected number of accidents for a traffic facility but should not be used to calculate change in number of accidents due to reconstruction of the street or other measures causing changes in the variables in the model. This is due to the fact that the models are based on cross sectional data for many links, not statistically controlled before and after studies related to the measures in question. Hence, the relationships found can not be interpreted as causal ones. A better estimate of change in number of accidents is achieved through calculating with the help of the models the expected number of accidents without the measures and after that using knowledge about the effects of the measure to predict the change.

This work has been focused on developing improved accident models for urban links, including minor intersections. The main scopes have been to study how the explanatory and predictive performance of the models can be improved by a) introducing exposure measures for vulnerable road users, b) introducing actual vehicle speeds (instead of speed limits) and c) dividing the vehicle accidents into two groups, vehicle-vehicle and vehicle-single, and developing separate models for the two groups. These have formed the following hypotheses:

- *The degree of explanation of the accident models for vulnerable road users can be greatly improved by including estimates of VRU exposure in the models.*
- *The degree of explanation of accident models can be improved if actual speeds are used instead of speed limit.*
- *The degree of explanation for vehicle accident models can be improved by separating the vehicle accident models into separate models for single vehicle accidents and multiple vehicle accidents.*

Methodology

The study involved eight Swedish cities, of which six were included in the development of accident models. In addition to police-reported traffic accidents, data for the models were collected in specific field studies and partly from the cities.

The field studies were performed according to methods partly developed in this study. VRU exposure was studied through 15 minutes counts and vehicle speeds were measured with handheld radar during the same time period.

The short counts and the resulting uncertainties in the values estimated on the basis of these counts will produce a bias in the estimated parameters of the models, but this bias is likely to be considerably less than 10% of the parameter values.

Accident models were developed using Generalised Linear Modelling with a Quasi-Poisson distribution to take into account that the accident data is overdispersed compared to a true Poisson distribution. Accident models were developed with one data set and validated with another data set.

The following variables were finally included in at least one of the models:

- Vehicle flow per day
- Number of pedestrians per day crossing or walking along the link
- Number of bicyclists per day crossing or walking along the link
- Land use along the link, categorised as one of the following: Residential, Institutional, Commercial, Industrial or None
- Street function, categorised as either Thoroughfare/Entrance, Tangential or Centre
- Visibility, describing the visibility around the link in terms of Good, Medium or Poor
- Existence of a large dedicated crossing for vulnerable road users on the link
- Speed Limit, which is either 30 km/h, Recommended 30 km/h, 50 km/h or 70 km/h
- Existence of any large exits (from shopping centres, large parking places etc.) on the link

Results and discussion

The study shows that the inclusion of exposure data for vulnerable road users in the models for vulnerable road users' accidents greatly improves the predictive ability of the models. Models including exposure data predicted 71-81% of the systematic variation in VRU accidents. Models excluding VRU exposure variables predicted 54-55% of the systematic variation and models including only the variables used in the existing Swedish accident models predicted just 37% of the systematic variation.

When validating the developed models, the models with parameters for exposure preset to values suggested by earlier theoretical and empirical findings predicted the number of pedestrian and vehicle accidents better than the models with freely estimated parameters. This is an unexpected result as in the freely estimated models the parameters are optimised to fit the data, while in the preset models the parameters are forced to take on specific values. The preset parameters were 0.5 for both VRU and vehicle flow exponents in models for VRU accidents, and 1.0 for vehicle flow exponent in the models for motor vehicle accidents. The validation indicates that the 'true' parameters are closer to the preset ones than the freely estimated for the pedestrian and the vehicle accident models. For bicyclist accidents the correct exponent for bicyclist flows is likely to be somewhat lower than 0.5, close to 0.35.

Vehicle speeds were found to be very difficult to use in the models. The inclusion of speed contributed greatly to the degree of explanation of the models, but problems with strong correlations between vehicle speeds and most of the other variables made the model coefficients unstable. The model parameter for speed varied greatly dependent on which other variables were included in the models. Although the large variation in speed parameter, the parameter value never came close to its true value according to the power model (Nilsson 2004), i.e. an exponent of two. Instead the parameter was always estimated to have a negative value indicating decreasing risk with higher speeds.

The separate modelling of single and multiple vehicle accidents failed partly as the single vehicle accidents were too few for constructing sound models. The models for multiple vehicle accidents developed, however, had a better predictive ability than models for all vehicle accidents. When the model parameters were freely estimated for the all vehicle model, the parameter for vehicle flow was estimated as an exponent of 1.38. The exponent for vehicle flow in vehicle accident models is normally estimated to a value close to 1.0. The very high value estimated in the model was probably caused by the mix of single vehicle accidents and multiple vehicle accidents. The freely estimated models for single vehicle accidents also estimated exponents for vehicle flow at values considerably larger than 1.0.

It should also be noted that land use clearly outperformed variables such as link environment and street function, which had in earlier models been used to describe the environments in which the links are located.

The parameter values in the models are estimated for best fit of the models to the data. This does not produce models where the parameter values can be used to predict the causal effect on number of accidents if the street is reconstructed, causing the variables to change in value. Therefore, the models should not be directly used for predicting change in number of accidents due to reconstruction, only to estimate the expected number of accidents. The expected number of accidents can in its turn be used together with known effects of safety measures on number of accidents in order to calculate the change. The expected number of accidents can also be used to identify sites where the number of occurred accidents is unexpectedly high. In this case the predicted number is calculated and compared with how many accidents have occurred.

The models were developed with data from the cities Uppsala, Katrineholm, Danderyd, Linköping, Norrköping and Västerås. The accident rates differed significantly between the different cities, particularly between the smaller ones and the larger ones. It is therefore advisable that the models are tested for a number of links in a city before they are used on a larger scale in that city.

Need for further research

Limiting the field studies to very short time periods at each link produces uncertainties in the variables describing VRU exposure, partly because there is a stochastic variation in the flow, and partly because the systematic variation of VRU flows over the day is not very well known. There is a need for better knowledge about the variation of VRU flows over time.

The division of motor vehicle accidents into separate models for multiple and single vehicle accidents produced ambiguous results. The number of single vehicle personal injury accidents is low enough in urban areas to produce problems when developing reliable accident models. On the other hand, the models for multiple vehicle accidents are in some aspects better than models for all vehicle accidents. More extensive data on motor vehicle accidents enables the division of models according to accident type, and may produce better predictions of accident numbers.

It is known that given a situation where only the speed changes, the number of personal injury accidents changes proportionally to the change in speed to the power of two. Yet, the models developed estimate parameters for speed that go considerably against the known relationship. The confounding circumstances that produce these misestimates deserve further exploration, especially as they would probably increase our understanding of the role of speed in driver's compensatory behaviour and in safety in general.

The accident rate differed significantly between the different cities in the study, particularly between the small cities and the larger ones. This should be studied further to investigate whether the true underlying accident rate really differs or if the differences depend on something else, such as differences in routines between different police districts in accident reporting.

1 Introduction

1.1 Background

Transportation is a necessity for the life we live today, and today's society depends in large on the possibility to travel long distances in a short time. However, our increased mobility also creates problems. In 2003 more than 27 000 people were injured in the Swedish road traffic. Out of these approximately 4500 were severely injured and 500 killed. Two thousand of the severely injured and 134 of the killed received their injuries in urban areas (Table 1). The number of fatalities in road traffic is fairly small compared to many other causes of death, for instance deaths related to cancer and diseases of the circulatory system amount to tens of thousands fatalities per year in Sweden. The fatalities in traffic are however malignant in another way in that they strike fairly evenly across ages from 15 year and up, while most other causes for death occur mostly in the latter years of life (WHO 2004-09-02).

Table 1 Number of injured and killed in urban and rural traffic environments, Sweden 2003 (SCB & SIKA 2004)

	Killed	Severely injured	Slightly injured	Total
Urban environments	134	2003	12107	14244
Rural environments	395	2661	10332	13388
Total	529	4664	22439	27632

In October 1997 the Swedish parliament passed the 'Vision Zero' for the Swedish road traffic. The Vision Zero states in short that it is unacceptable for road users to be killed or severely injured in traffic (prop. 1996/97:137). In the long run the traffic environment should be designed so that road users are allowed to make mistakes in traffic without suffering severe consequences. Considering the traffic safety situation today and the extent of the road network, such a change will require considerable resources and time.

The occurrence of accidents is statistically a concept with a large random variation. Therefore accident history is a tricky source for information on the safety level of individual streets. On a street where during one year several accidents have occurred there may very well be no accidents for several years to follow. Many years of accident data are often needed to make any reliable estimation of the safety level. That is why, in traffic safety work, there is a need for predictive accident models.

Predictive accident models

Instead of relying on accident history alone, predictive accident models are often used. The models calculate the expected number of accidents for a street with the help of information on the characteristic of the street and empirical knowledge about how the number of accidents varies with these characteristics (Vägverket 2001b, Peltola 2000, Greibe 2003).

The calculated expected number of accidents can be used for estimates of number of accidents at streets not yet built or providing a stable estimate of the number of accidents on existing streets for use with calculations of effects of measures etc. The estimated accident numbers are used in the political process of decision making and for evaluation. In the latter case the models can be used to calculate the expected number of accidents without a traffic safety improvement, to be compared with the number of accidents occurring when a measure has been implemented

(Vägverket 2001a). The models, or rather the parameter values for the variables in the models, can not, however, in most cases be used in directly estimating the effect of safety measures.

The development of predictive accident models has intensified during the last decade. Before the beginning of the previous decade most work on accident models was made in the Nordic countries and the UK, and mainly being focused on rural roads. Modeling tools have also evolved from the early modeling using multiple linear regression with an assumption of accidents being Normal-distributed to more advanced generalized linear models with Poisson or Negative Binomial distributions (Maher & Summersgill 1996).

In the beginning of the nineties Swedish predictive accident models were developed for urban and rural intersections (Brüde & Larsson 1992). The intersection models include separate models for pedestrian and bicyclist accidents and for the vehicle-accidents also separate models for separate types of intersections. The models base their calculations of expected number of accidents primarily on exposure measures but for vehicle accidents the mathematical relationships also vary between types of intersection, speed limits and street environment. These models are still used for accident modeling at intersections but with some adjustment for the decreasing trend in accident numbers for the vehicle-accident models (Vägverket 2001b).

For street links including minor intersections the Swedish models consist of accident rates per vehicle km. The accident rates differ between streets of different types, speed limits and number of lanes. Just as for intersections there are separate models for the link-accidents of pedestrians and bicyclists. The number of accidents for vulnerable road users is calculated by taking a parameter times the number of vehicle accidents. The parameters are different for pedestrian respective bicyclists, and also differ between different street types and street environments to take into account the difference in exposure of vulnerable road users (Vägverket 2001b). The actual exposure of vulnerable road users on the link studied is normally not used. This is due to that these data seldom are available.

1.2 Purpose and scope of the study

The purpose of the study was to identify possible improvements in predictive accident models and to construct improved models for urban links.

The work was initiated with a literature survey focusing on predictive accident models and variables affecting traffic safety on a street level.

We know that most of the variation in number of accidents is explained by variation in exposure, i.e. flow. In accident models the exposure of vulnerable road users on the actual link is not used in the existing Swedish models which can be assumed to limit the explanatory power of the models. Therefore one of the scopes of this work was to investigate how estimates of actual exposure can be used to improve the models, and to compare the predicting power of models with and without estimates of vulnerable road users' exposure.

Other topics studied were the need to develop separate models for different types of accidents and whether the use of actual vehicle speeds instead of speed limits could improve the explanatory or predicting power of the models.

Most data required by the models were not available, which has made field studies necessary to gather this data. The field studies have been conducted in eight Swedish cities, out of which six were used for the final modeling. The data collected included exposure of vulnerable road users, vehicle speeds, and street design solutions for the interaction between vehicles and vulnerable road users as well as several other variables describing the traffic environment. In addition some data have been gathered from the municipalities with help from the Swedish National Road and Transport Research Institute (VTI). Accident data have also been compiled by VTI for 5-8 years for the different municipalities. Five years of data have been used for modeling and the additional data have been used for validation studies.

1.3 Structure

Below follows a summary of the structure of this dissertation, but indirectly it is also more or less a chronological description of how the work with the project has been structured.

The chapter ***Theories and concepts in traffic safety*** contains a summary of theoretical concepts and knowledge about how traffic safety is affected by various traffic characteristics. This is part of the result of the literature survey. The other part of the literature survey focused on existing accident models. These are described in the chapter ***Accident models*** with a special focus on the models used in Sweden, Finland and Denmark. The chapter is concluded with a discussion on what can be improved in the Swedish models. This forms a base for the delimitation of the future work.

Methodology describes both methods used for collecting data in the field studies and a description of the methods used for analysing the data obtained. The data used in the modeling are described in the chapter ***Data***.

The results from the modeling are described and analysed in the chapter ***Results and analysis***. The final chapter ***Discussion and conclusions*** deals with the study's contribution to science, implications of the results and identification of needs for future research.

2 Theories and concepts in traffic safety

This chapter is primarily based on a literature survey presented in more detail in Jonsson (2001). The chapter gives a brief summary on different concepts in traffic safety and specifically which variables affect the safety level of urban streets.

2.1 What is traffic safety and how can we measure it?

Every year the Swedish traffic cause the loss of more than 500 lives and more than 27000 are injured (SIKA & SCB 2004). Traffic is a serious threat to public health, but although the risk of dire consequences people choose to travel every day. Transportation is vital to the society and cannot be totally avoided; instead the question has to be: How do we satisfy our need for travel in the safest way? But how do we define the safest way? Two common measures of safety are number of accidents per inhabitant respective number of accidents per travelled kilometre. The first measure is useful for comparing the level of risk associated with different activities. The second estimate is naturally restricted to travel activities but gives on the other hand information on the risk per unit of travel. Predictive accident models traditionally relate the number of accidents to the distance travelled, and this has been the basis for the definition of traffic safety in this work.

The estimate of traffic safety used in this dissertation is the same as defined by Hauer (1999): "…, the safety of a road is measured by the frequency and severity of crashes expected to occur on it."

An essential word in the quote is 'expected'. The number of accidents is never constant, varying traffic conditions and random variation causes the number of accidents per driven kilometre to vary over time even for a street that is not changed in its design. The number of accidents to focus upon is thus an expected number of accidents for given traffic conditions. The expected number of accidents can be seen as the average number of accidents occurring if the traffic conditions were kept constant and the time for studying the number of accidents was significantly long to even out the random variation. Sometimes the expected number of accidents is referred to as underlying true accident rate.

The expected number of accidents is obscured by two factors: underreporting and random variation. The underreporting of accidents is causing problems when measuring number of accidents by looking at police reported accidents, because only about 40% of all personal injury accidents are reported by the police (Englund et al 1998). The underreporting also differs between different types of road users which makes it even more difficult to get a good picture of the accident situation just looking at police reported accidents (Vägverket 2004). The random variation poses problems because of the need of long periods of time for accident data to even out the random variation between years. On the other hand, using accident data from long periods of time poses other problems: the traffic and the road environment seldom remain constant over time as traffic flows and street design change.

Because of the impossibility of getting values of the true underlying number of accidents it is necessary to make assumptions concerning the generalisation of results based on police reported accidents. The main assumption is that whatever effect a variable has on the number of police reported accidents it will also have the same effect on the true underlying accident rate. The correctness of this rests upon the assumption that the police reported accidents are a representative random sample of all accidents occurring. The assumption is violated in two cases as there are two factors covariating with the degree of underreporting; type of road user involved and the severity of the accident (Englund et al. 1998). This calls for caution when trying to extrapolate the number of accidents using the underreporting rate.

In addition to the number of accidents, the number of injured road users is also used to define the safety level. The number of injured road users classified against injury severity defines the severity of accidents.

2.2 Traffic safety concepts

Below follows two conceptual views on traffic safety and the origin of accidents. The concepts provide a better understanding of the process leading to a crash and how to prevent it.

2.2.1 The Haddon Matrix

One view on traffic safety is conceptualised in the Haddon Matrix (Table 2). The Haddon Matrix categorises traffic safety measures in two dimensions. One dimension is the time compared to crash when they are deployed: Before crash, in crash, after crash. The other dimension is categorising measures according to their relation to the road users, the vehicle or the road. (Haddon 1980)

The Haddon matrix can best be described by exemplification:

Imagine an accident at a typical arterial street. The driver is driving under rainy conditions in a fairly old car with worn out tires. When driving through a pool of water the front tires loose contact with the asphalt and the car starts to swerve. The driver fails to regain control over the car and the car crashes into a light pole. The light pole cuts into the car and the driver is struck. Seven minutes later an ambulance and a fire truck arrive and the driver is cut loose from the car wreck and transported to a hospital. The driver goes through surgery but dies.

In the case described several measures can be taken to improve the chances of survival for the driver. In the **pre-crash** phase the **driver** can receive training in driving on slippery road surface and how to regain control over the swerving car. The **road** can be constructed as to prevent pools of water, and the **vehicle** can be equipped with better tires and anti-skidding system. In the crash phase the crash violence to the driver can be reduced by making the light poles along the **road** crashworthy or separating them from the street with fences. The structure of the **vehicle** can be made with a rigid structure protecting the driver. And finally the **driver** can be equipped with a safety belt in order to be kept seated and not thrown out of the car or otherwise exposed to impact. In the **post-crash** phase the common knowledge of **road users** in emergency care can be raised so that the first ones on site can take proper action and more resources can be spent on emergency medical services (**human**) to allow for fast professional treatment of injuries.

Just like the Haddon matrix can be used for categorising traffic safety measures, it can also be used for categorising variables affecting safety.

Variables affecting safety are however more likely to turn up in several of the cells of the matrix, for instance speed has an effect both in the 'Before crash' affecting the possibility of crash avoidance and in the 'In crash' phase on the severity of the crash. Also, on the individual level the exposure and speed are those of the individual road user, while on the street level they are aggregate estimates for all the road users on that street. When working with accident models for streets, aggregate estimates on a street level should be used for modelling safety.

Table 2 The Haddon matrix, Haddon (1980)

Element	Before Crash	In Crash	After Crash
Human	Training Education Behaviour (e.g. drink driving) Attitudes Conspicuous clothing on pedestrians and cyclists	In-vehicle restraints fitted and worn	Emergency medical services
Vehicle	Primary safety (e.g. braking, roadworthiness, visibility) Speed Exposure	Secondary safety (e.g. impact protection) Speed	Salvage
Road	Delineation Road geometry Surface condition Visibility Road safety audit	Roadside safety (e.g. frangible poles) Safety barriers	Restoration of road and traffic devices

2.2.2 Exposure, risk, consequence

Another way to visualise the way in which variables affect traffic safety is through Exposure-Risk-Consequence (Elvik et al. 1997, Nilsson 2004):

1. Exposure – These variables influence, or describe, the level of exposure of road users to risks.
2. Risk – These variables affect the level of risk of an accident per unit of Exposure.
3. Consequence – These variables affect the outcome of an accident.

The safety level is seen as dependent on all these dimensions, where **exposure** describes how much the road users are exposed to events that might lead to accidents. The exposure for road users on links is most often estimated as the traffic volume of or kilometres travelled by the road users. For intersections the exposure can rather be seen as the number of encounters or, more indirect, number of entering vehicles into the intersection.

Risk is defined as the probability per unit of exposure to be involved in an accident. Sometimes the risk is specified as the risk of being involved in a personal injury accident, then the dimensions 'risk' and 'consequence' are somewhat merged. **Consequence** is otherwise defined as the outcome of an accident.

When trying to influence traffic safety, one or several of these dimensions have to be targeted. This view permeates the Swedish Vision Zero (Prop. 1996/97:137) in which the work on traffic safety has shifted towards actions targeting the consequence rather than the risk by 'allowing' accidents resulting in only slight personal injuries, and instead focusing on severe injuries and fatalities.

This view on safety has links to the Haddon matrix. The time-phase before crash relates rather well with exposure and risk while the time-phases in and after crash relates well to factors influencing consequence.

2.3 Variables affecting safety

This section gives an overview of variables affecting the traffic safety on a street level. The variables have been organised according to the classification dimensions: exposure, risk and consequence. The overview is focused on variables affecting the safety on urban street sections including minor intersections.

There has to be made a distinction between variables affecting the traffic safety level for an individual road user and variables affecting the safety level of a street section. The latter case is the focus of this work, but there is usually an overlap between the cases. This influences the survey in a direction where variables affecting the safety on an individual level, or an aggregate level for a larger road network, are only included in case they can be presumed to vary between streets.

Intuitively it might appear like variables affecting individual safety should automatically affect the safety level of many but this is not true. If the road users on a specific street are heterogeneously composed, then differences in the level of risks among different types of road users are evened out when aggregated on the street-level. Thus, if there aren't large differences in the composition of road users between different streets, there is no reason to take such variables as age, experience, use of seat belt etc. into account in the models. On the other hand, if a certain street for some reason has a marked deviation from the average in the composition of road users, the level of safety on that street could be somewhat different than on other streets of the same character but with another composition of road users. However, it is difficult to ascertain these kinds of deviations, and therefore the individual factors have been left out, both because of the difficulties associated with registering different type of road users, but also because of the small effect it may be assumed to have on the total outcome.

It can be noted that this affects the transferability of accident models between different countries or states or even regions. For example, within a country there might be a rather homogenous traffic culture with regard to road users' usage of protective equipment such as seat belt or bicycle helmet, but the likelihood of a road user to use such equipment varies largely from country to country.

2.3.1 Exposure

Several variables affect the exposure of road users to accident prone situations, the first among them being traffic flow. Increasing traffic flow means more road users that are exposed to risk, but also more interactions between road users. Traffic flow tends to be the single variable explaining the most of the variance in traffic accidents (Elvik et al 1997, Brüde & Larsson 1993, Kulmala 1995, Greibe 2003).

Traffic flow here means both vehicle flow and the flow of vulnerable road users. In the case of police reported personal injury accidents involving vulnerable road users there is most often also a motorised vehicle involved which causes both vehicle flow and flow of vulnerable road users to affect the exposure related to this type of accidents.

The flow acts in two ways, both in a primary sense as flow is the base of all exposure, but also in a secondary, passive sense when interactions between road users increase when traffic flow increases. This can be seen as both exposure and risk. When looking at streets the flow along the street is most often used as the measure of exposure, while in intersections sometimes a combination of intersecting flows are used as a measure of exposure (Hauer, Ng & Lovell 1988, Brüde & Larsson 1992). In the case of streets with minor intersections a certain mixture of pure link flows and intersecting flows can be expected to influence the occurrence of accidents. A mixture of accident types will call for different exposure measures.

The relation between vehicle flow and number of personal injury accidents on links is in accident models sometimes considered as being purely linear, while separate studies show a relation where the relation is represented by a power function where the exponent of flow is slightly smaller than 1, Elvik et al's meta studies (1997) show an elasticity of 0,95.

There is no global linear connection between exposure and safety. This is perhaps best shown for bicyclists by Ekman (1996) in a study of their exposure and risks. The study showed that the risk for the individual bicyclist was significantly greater at a small flow of bicyclists than at a large flow (Figure 1) which was explained by the perception of the drivers. When the flow of bicyclists is greater, then the drivers are more alerted as to the presence of bicyclists than when the flow is small. In accordance with the perceived greater probability of encountering a bicyclist, the driver raises his attention. The risk is described as a stair function; at a certain number of bicyclists the drivers start taking more notice of the bicyclists. This knowledge can be used in the modelling to take better account of the exposure.

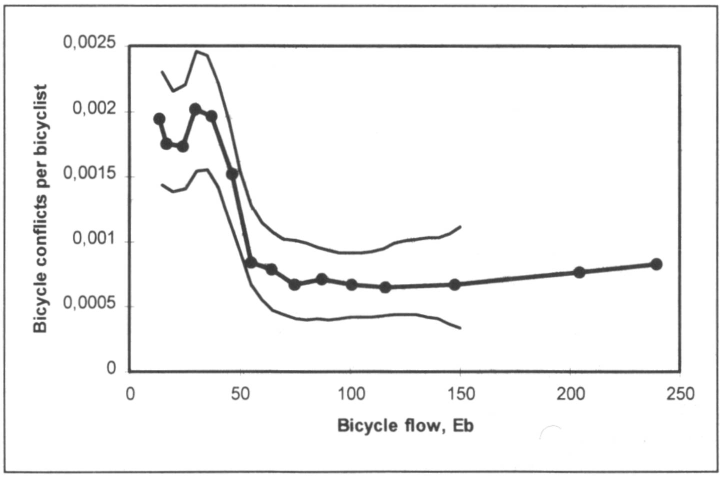


Figure 1 Risk of traffic conflict (risk indicator) versus bicycle flow (Ekman 1996)

Exposure for vulnerable road users (VRUs) is not a highly developed area. The measure of exposure used in traffic safety assessments is mostly the flow of motorised vehicles. The vehicle flow is often registered and stored into databases, and is thus close at hand for use. Flows of pedestrians and bicyclists are not often registered, probably both because they are not as easy to register as motorised vehicles as well as because of the traditional focus on motorised traffic. There are however studies (Gårder 1989) showing a strong connection between the exposure of vulnerable road users and their safety. The number of accidents in intersections was found to be approximately proportional to the square root of both the flows of vulnerable and motorised road users. Thus it is important to use both the flows of vulnerable and motorised road users in accident models for the accidents of VRUs.

Vulnerable road users are usually freer in their movement pattern through the street environment than the motorised road users. For motorised road users exposure is most often measured by total travelled distance or, for intersections, number of incoming vehicles. Vulnerable road users however don't need an intersection to cross a street and they are often more or less separated from the motorised traffic. This calls for a different view on the exposure of vulnerable road users than that of motorised road users.

Vehicle speed is seen as a variable strongly affecting the safety of streets. The effect of speed is mostly related to risk and consequence (see below) but Navon (2001) carries a discussion stating that speed also has an influence on exposure, that the higher the speed, the fewer the interactions will be with other road users.

2.3.2 Risk

Risk is here defined as the probability per unit of exposure to be involved in a personal injury accident. As the focus is on safety on links, the exposure measure is primarily traffic flow. Thus the variables included in this section are those that affect the risk per road user passing through the link.

Speed level

Speed is often considered to be the strongest influencing variable on risk (Nilsson 2004). The connection is however not undebated (Elvik et al 2004). There are also various indicators of speed. In accident models the speed limit is often used as a variable instead of actual speeds. There are however more direct connections established between actual vehicle speeds and traffic safety. Nilsson (2004) has presented a model where the likelihood of a personal injury accident is proportional to the speed to the power of two.

The speed level affects the driver's possibility to stop in a given situation. When a dangerous situation arises, the time available for the driver to react before a crash is given by distance and relative speed between the car and the object of collision. The time is greatly reduced by higher speeds. In a first step the speed influences the distance travelled before the driver even starts to react (Figure 2). In a second phase the driver starts to decelerate. If the speed is low enough the driver will then have time to avoid a collision, otherwise a crash will occur. The violence of the crash is greatly influenced by the initial speed and the reduction in speed during the second phase. A small difference may very well mean the difference between life and death. Two separate levels of speed can be discerned in this case: the initial speed and the speed in the moment of crash. The first is of great importance for the possibility of avoiding an accident altogether. The second one is dependent on the first and determines the violence involved in the crash.

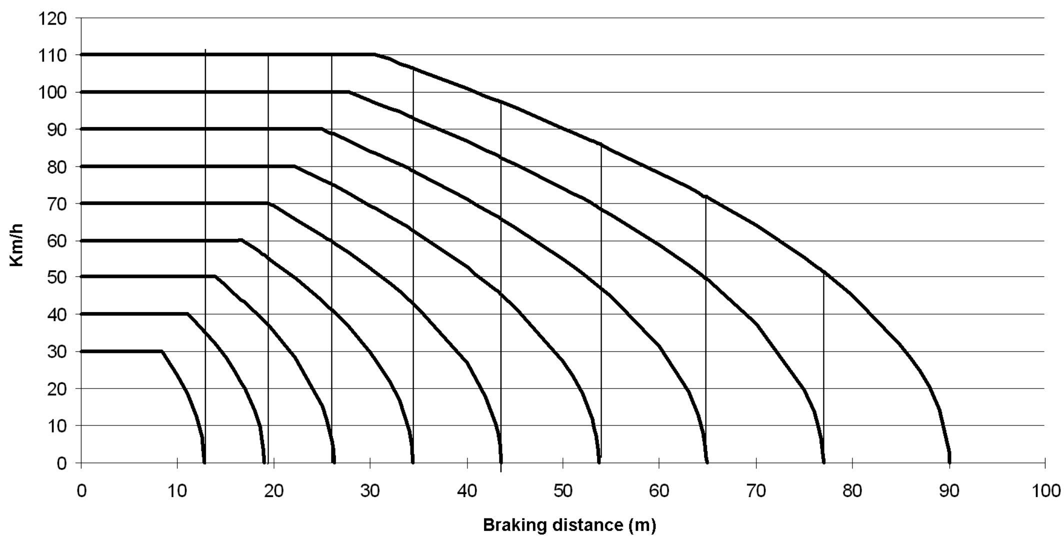


Figure 2 Stopping distance in relation to speed, calculated with deceleration=0.8g and reaction time=1 s (Carlsson 2004)

The speed is the drivers' primary mean for adapting to altered perceived risk, thus compensating for a high or low level of perceived risk by reducing or increasing speed (Wilde 1994). A problem with the speed level is that it is highly correlated with many other variables, such as street design and presence of vulnerable road users. A consequence of this is that the accident risk generally is lower on high speed streets than on low speed streets, not because of the high speeds, but because high speed streets generally have fewer intersections per km, a safer design and less vulnerable road users crossing the street.

Speed dispersion

There are also connections between the dispersion of speeds and the safety (Salusjärvi 1981). These connections are however not as well established as the connection with the average speed. The underlying theory of the connection between the speed dispersion and traffic safety is found in the interaction between vehicles and the expectations of drivers. Increased dispersion of speeds calls for more overtaking in traffic and thus for more situations with a higher risk (Finch et al. 1994). Larger dispersion also increases the probability of unexpected situations. If a driver expects the speeds to be of a certain level, that driver acts on the basis of that view. With a larger dispersion of speeds, the increased likelihood of a large difference between the expected speed and actual speed makes for more unexpected situations which might lead to collisions (Elvik et al. 1997).

Intersections

Intersections can be seen as part of the exposure measure, but given a focus on links with traffic flow as the primary measure of exposure, the intersections instead constitute elements which increase the risk. The risk per driven km is considerably higher close to an intersection compared to other parts of the link (Kulmala 1995).

Accident models for urban main streets often include minor intersections, that is, intersections with smaller, local streets and exits. These are points of conflicts with a high risk of accident. Therefore their number and type have bearing on the number of accidents on the street. A large number of intersections also indicate a more complex environment with crossing traffic interacting with the traffic on the main street.

The intersections are points of increased risk for separated vulnerable road users travelling parallel to the street. Most separated paths for these road users are only separated between the intersections, and when crossing side streets the risk may be even larger at those points than if they wouldn’t have been separated. This is because drivers exiting the main street and crossing the path aren’t as aware of the separated VRUs as they would have been if they had been integrated with the motorised traffic.

2.3.3 Consequence

Consequence is here defined as the outcome of an accident, that is, the number and severity of injured road users per accident. The consequence of an accident can primarily be related to two variables: vulnerability of the road users involved and the collision speed. If the road user is located inside a vehicle, speeds of up to 70 km/h can be sustained in a head-on collision without large probabilities of severe consequences. However, if the road user is totally unprotected, for instance as a pedestrian, even speeds of 40 km/h constitute a considerable risk of severe or fatal injury (Figure 3).

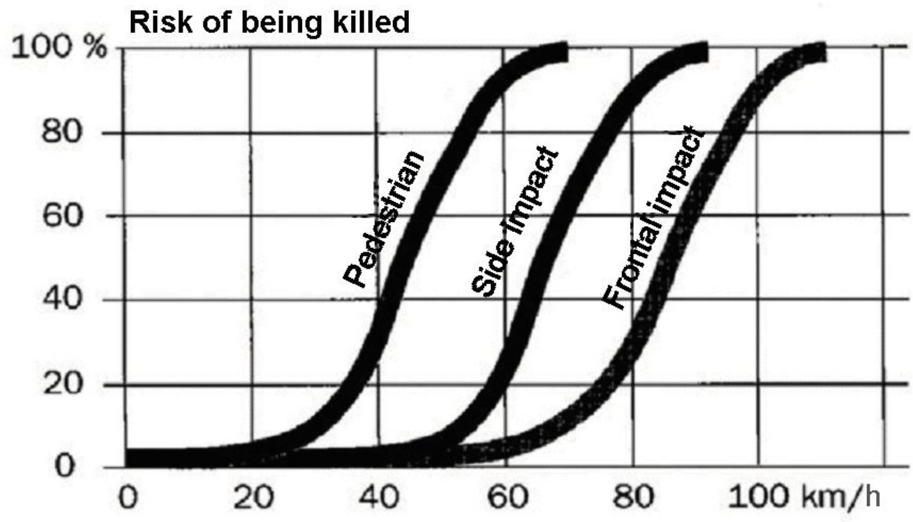


Figure 3 The risk of a road user being killed depending on speeds in various collision types (SveKom 1998)

The connection between the speed level and the risk of a personal injury accident has been shown to follow a quadratic relationship, the number of injury accidents is proportional to the speed level to the power of two (Nilsson 2004). Nilsson also shows that the effect of speed on severity can be described as a power function. Given that a personal injury accident occurs, the probability of a severe personal injury is estimated to be proportional to the speed, and the probability of a fatal accident proportional to the speed to the power of two. The power model also applies to the number of injured (Figure 4).

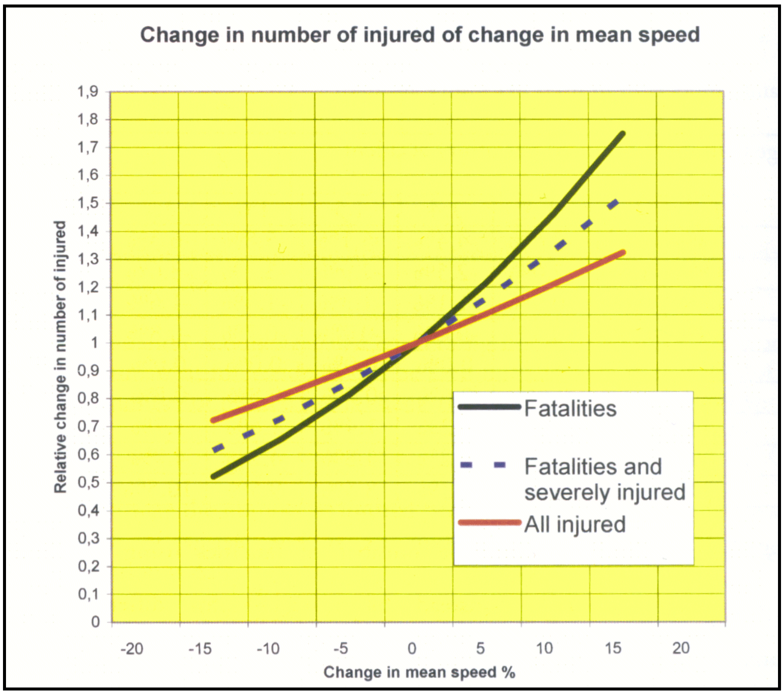


Figure 4 The relation between changes in number of injured and changes in mean speed (Nilsson 2004)

The consequence can also be influenced by the environment surrounding the street. Light poles, signposts, trees and other thin but hard objects close to the street can cut into the car even at fairly low speeds causing severe injuries and fatalities (SveKom 1997).

3 Accident models

This chapter gives an introduction to accident models for urban streets with examples taken from the Swedish, Finnish and Danish models. Also another type of accident models, the DRAG family, is given a brief presentation. The chapter ends with a summary of possible improvements in the Swedish models.

3.1 Introduction

Accident models as well as models for calculating estimates of environmental impact, time delays, vehicle costs and other effects of traffic are used to assess the impact of changes in the transport system in socio-economic calculations. Estimates are based on a set of variables which depend on the investment under analysis.

Through the evaluation of the aspects above (safety, environment and time delays), and the weighing between them, different interests, and conflicts between these interests, are revealed. This visualises the many aspects of traffic. Prioritisation between different interests has to be made and in this the models help in visualising, evaluating and comparing different qualities. To make the basis of decisions as good as possible, all knowledge about the traffic system has to be used. Therefore it is essential that the models are updated and reformed as the knowledge grows. The models also need to be updated with regards to trends in accident development.

Besides supporting socio-economic calculations, the models for traffic safety can also be used as an aid in blackspot identification (Vejdirektoratet 2001). The predicted number of accidents is the used in a comparison with how many accidents occur, and spots where considerably more accidents occur, compared to the predicted number, are considered blackspots. The blackspots are then treated specially with measures improving safety.

Sometimes the models are used to calculate a change in accidents due to reconstruction of a traffic environment. This should be done with caution as the models normally aren't based on before-after studies. The data used in constructing the models are empirical data where various types of traffic environments are compared with each other with regard to traffic safety. For the different types of environments many variables are correlated and the types of streets are categorised by several different variables. When introducing a safety measure into a street a few of these variables will change, but in other aspects it is the same street as before. A safer way to model this is to use the accident models to calculate an estimated number of accidents for a situation without the measure and then use data about accident reductions established by before-after studies to calculate the change in accidents. An example of the problem with using accident models based purely on empirical data is the following:

Example: *A common safety measure is speed reduction through various physical measures. This could for example take the form of a reduction in speed limit from 70 km/h to 50 km/h combined with narrowing of the street. For an arterial street with a tangential function, the speed limit 70 km/h and located in the outskirts of the city, the accident rate according to the Swedish accident models would be 0.216 accidents (including non-injury accidents) per million axle pair km*[1]*. A street with the same location and function but with the speed limit 50 km/h has an accident rate of 0.400 accidents per million axle pair km. No traffic engineer would accept a model saying that this measure would produce almost twice as many accidents. A more reasonable way to calculate the change in accident rate would*

[1] The existing Swedish accident models measure traffic flow in number of axle pairs instead of number of vehicles. Depending on the percentage of heavy vehicles with more than two axles, the number of axle pairs will differ more or less from the number of vehicles.

be to predict the change in speed and use the power model (Nilsson 2004) to calculate the change in number of accidents.

The problem in the example above is that the number of accidents most likely is affected by variables not included in the model. Such variables could be number of intersecting streets and other variables making the traffic environment more complex and more likely to generate accidents. The variables correlate in their turn with the speed limit as when the traffic environment is more complex, the speed limit is generally lower. Since the speed limit is included in the models, that variable also reflects the effects of variables it is correlated with.

Accident models can however be constructed so that they predict changes in accidents fairly well. By using pre-defined effects for variables, their effect on number of accidents can be locked to levels in compliance with theoretical and empirical findings. Through this the effects of other non-preset variables can also, hopefully, receive more logical effects. Tests with this type of modelling have been done in Finland (Peltola 2000) but the models developed still contained some effects for non-preset variables that were not in compliance with effects known from previous studies.

3.2 Applications

The purpose of accident models, both when used in association with socio-economic calculations and black spot identification, is to help making decisions that use resources cost-effectively, illustrated in the two quotes below.

"It (Effektsamband 2000) is intended to be an important support in, with a basis in the political goals, prioritising measures leading to formulated goals and which are socio-economically effective." (Translated from Vägverket 2001a[2])

"Mapping, valuations and analysis of traffic safety is a presupposition for being able to focus measures where the problem is most dominant. This applies both when working with an existing road network and when making prognosis or making consequence assessments for new roads or analysing the importance of changes in the road network." (Translated from Vejdirektoratet 2001[3])

In Figure 5 (Vägverket 2001a) the structure of the basis of decisions about physical measures for the Swedish system is presented. The part where accident models are used is marked grey.

The three different models studied all aim at predicting the amount of accidents for traffic environments. The difference in application is mainly concerning whether the models are optimised for describing the accident rate for different types of roads or optimised for describing the change in accident rate when a street is redesigned. All three models described below aim more or less at both.

The Finnish model, TARVA, is optimised for studying changes in accidents due to changes in road environment. It has fairly simple models for describing how many accidents occur, but uses established effects for various road improvements to calculate the change in number of accidents.

The Danish models uses complicated models with several variables to explain how many accidents occur. The models are however not specifically aimed at prediciting changes in number

[2] Original text: "Den (Effektsamband 2000) avses vara ett viktigt stöd för att utifrån de transportpolitiska målen prioritera åtgärder som leder mot uppsatta mål och som är samhällsekonomiskt effektiva."

[3] Original text: "Kartläggning, värderingar och analyser av trafiksäkerheten är en förutsättning för, att insatsen kan koncentreras där, där problemet är störst. Det gäller både när man arbetar med ett existerande vägnät, och när man gör prognoser eller konsekvensbedömningar av nyanläggningar eller analyserar betydelsen av ändringar i vägnätet."

of accidents although it is stated they can be used to give a hint on the change in accidents due to redesign (Vejdirektoratet 2001).

The Swedish accident models are a bit of both. They use various descriptive variables to predict accident rates for various types of streets. In addition to the prediction of accident rates for the street types, the models use established effects for measures to calculate changes in number of accidents. (Vägverket 2001b & 2001c)

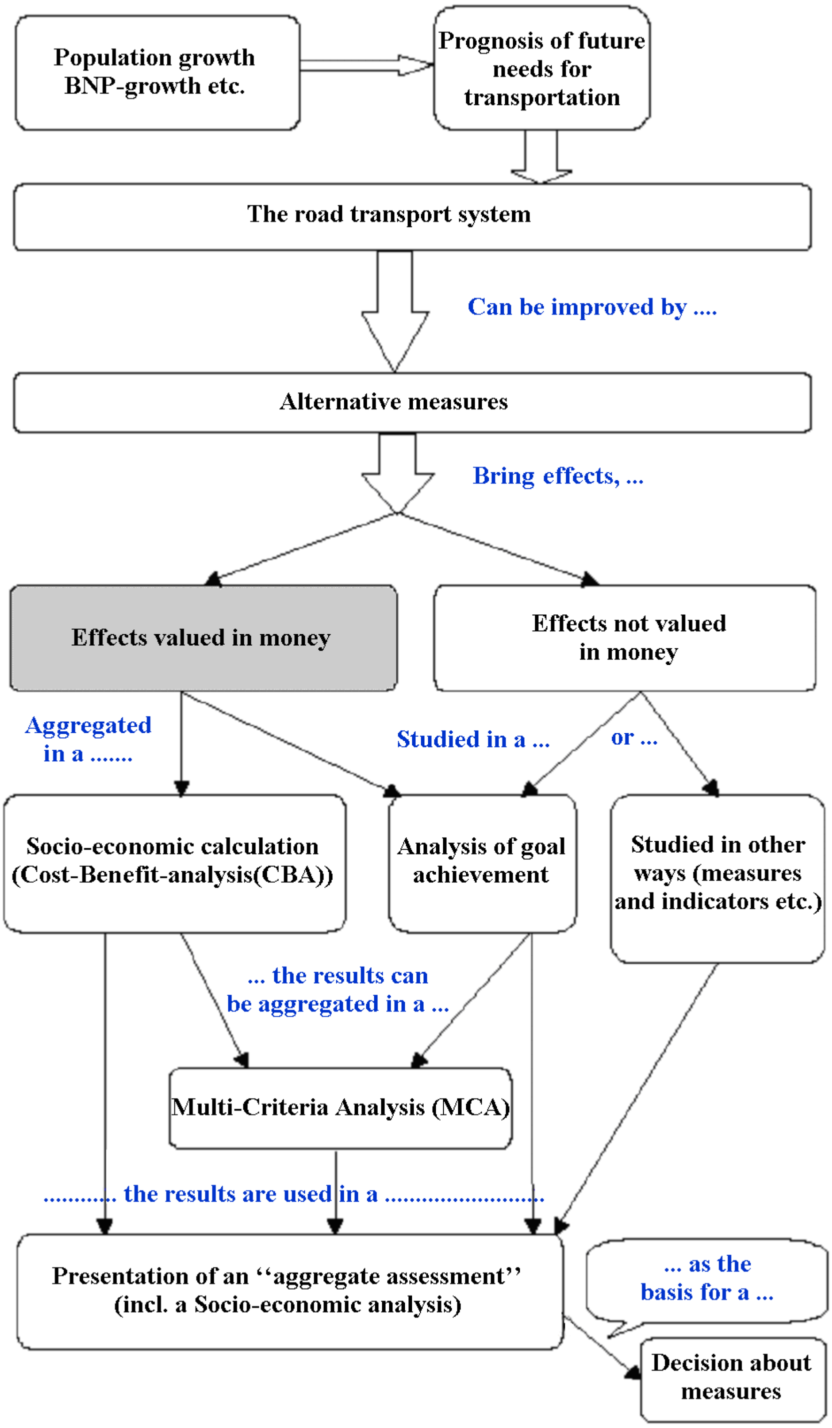


Figure 5 Basis for decision in prioritising of measures (Vägverket 2001a)

3.3 Structure of the models

3.3.1 Accident statistics

A common treat for the three Nordic models described is that they are all empirically based on police reported accidents. All the models also include some form of severity index relating the number of injured, severely injured and/or fatalities among the accidents. While the Danish and Swedish models are based on both injury accidents and property damage only accidents (Vejdirektoratet 2001, Vägverket 2001b), TARVA is based solely on injury accidents (Peltola 2000).

3.3.2 General structure

The difference in application between the three models reflects on the overall structure. TARVA, for example, is built up around a fairly simple procedure for calculating predicted number of accidents without measures and then applying known effects of measures. The Swedish models form part in a larger set of so called effect models for road infrastructure and inherit some characteristics and variables from that concept. The models for calculating the predicted number of accidents use explaining variables that more describe the type of street and location in the city, rather than variables directly influencing the safety, such as vehicle speeds and specific design variables.

Figure 6 shows the process of how TARVA calculates the change in accidents and fatalities due to a measure. In the first step the number of accidents before measure is calculated as a weighed estimate of the predicted number of accidents and accident history for the last five years. The number of accidents is then corrected with the change in traffic and land use to get an estimate of the predicted number of accidents in the after-situation, but without the measure. The effect of the measure is then taken into account by applying an impact coefficient, both in regard to its effect on accident rate and its effect on severity of the accidents.

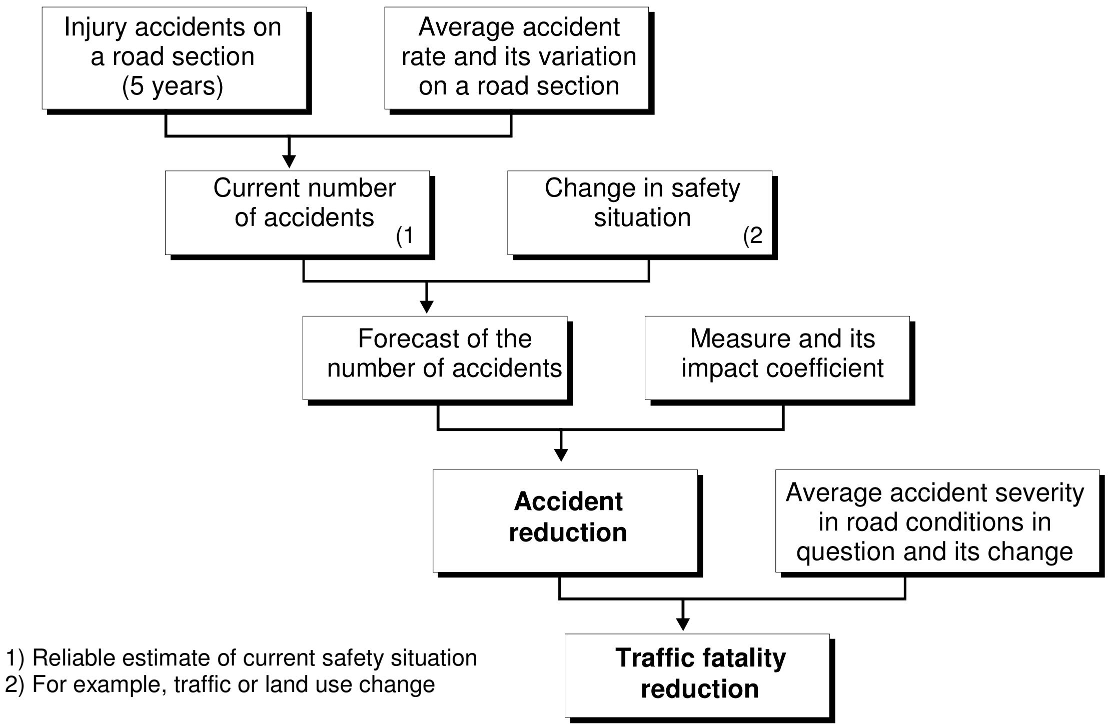


Figure 6 TARVA, organisational diagram for the modelling procedure (Peltola 2000)

3.3.3 Combining predicted values and accident history

All three models have routines for weighing together predicted values from the models with accident history to get more stable estimates of the number of accidents, in the Swedish models this is called 'Adjusted number of accidents'. The balancing weights between observed and predicted number of accidents are determined by the quality of the estimates of the respective number of accidents.

In the Swedish models the weights are determined by the number of predicted accidents. The larger the number of predicted accidents, the stronger the weight for accident history will be. This is based on the assumption that the more accidents that are expected, the more stable the accident history will be. (Vägverket 2001b)

$A(*) = A(N) + K_a *(A(I)- A(N))$ where

$K_a = 0.25*A(N) / (1+0.25*A(N))$

(A(*)=Adjusted number of accidents, A(N)=Predicted number of accidents, A(I)=Observed number of accidents) (Vägverket 2001b)

In the case of the Finnish and Danish models the goodness of fit for the models are also used to calculate weights for how to combine observed and predicted number of accidents.

3.3.4 Division into sub-models

The models are divided into separate models for **rural** traffic environments and for **urban** traffic environments. Rural environments are most often less complicated when it concerns number of intersections per km, number of vulnerable road users and road-side structure, also called road-side friction (Bång, Carlsson & Palgunadi 1995). The models described here are mainly those for urban areas. In the case of the Danish and Swedish models the models for urban areas are singled out and described, while the Finnish model is mainly concerned with rural areas altogether but is still described here as it has a different focus than the other two and thus adds a value to the description.

Urban models are further divided into models for **intersections** respectively **links**. Large intersections are modelled separately and different models are often used for different types of intersections because of the different accident rates dependent on type of intersection, 3- or 4-arm, signalised, non-signalised etc. Minor intersections are normally included in the models for the streets as it would be too much work to make separate calculations for each small exit. The distinction between a minor and a major intersection varies between countries and is dependent on the size of the secondary flow into the intersection. In the Swedish models a major intersection is defined as an intersection with a secondary flow larger than 1000 ADT (Annual Daily Traffic) while in the Danish models the borderline is drawn at 250 ADT.

Intersections are, as mentioned, further divided into three- and four-arm intersections and dependent on type of traffic control type. Separate models are sometimes used depending on **type of road users** and even **type of accident** such as single, rear-end or turning accidents (Kulmala 1995, Vejdirektoratet 2001).

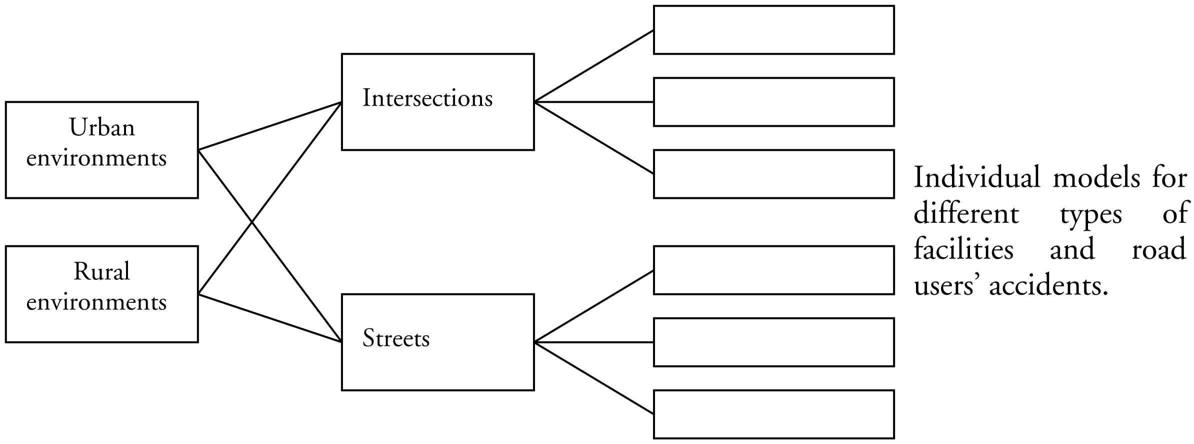


Figure 7 Example of division into sub-models for safety effect models

The division with regard to the type of road user varies between the different models. The Danish models include both a model for all accident types and road users and separate models for separate accident types including one each for pedestrians and bicyclists. The two latter models are however based on a very small amount of data. (Vejdirektoratet 1998)

TARVA separates the accidents for vulnerable road users from those of motorised road users. The effects of the different measures are also divided into effects on accidents for vulnerable road users and for motorised road users.

The Swedish accident models for links have separate models for pedestrian accidents and bicyclist accidents but they all have their base in motorised traffic. The number of accidents for pedestrians and bicyclists is calculated as a percentage of the accidents for motorised traffic. The percentage used is dependent on the traffic environment. In environments with more vulnerable road users the percentage is higher. The variable primarily influencing the number of VRU-accidents in the models is vehicle flow. There are no indicators for the VRU exposure except indirectly through the environment classification.

3.3.5 Variables included in the models

The difference in application between the three models, and the context in which they are used, affect the variables used. While the Finnish model uses traffic flow in the estimation of accidents without measures, and effects of measures to calculate change, the Swedish models use variables describing the type and function of the street. The Danish models use street design variables that more directly relate to factors influencing the risk for, and severity of, accidents. The Swedish and Danish models both use speed limit and some form of indicator for road width. The Danish models also use the variables Number of exits per km, Number of minor crossings per km, Parking facilities and Land use.

3.3.6 Traffic flow in accident models

Flow is the single most important variable for explaining variance in accidents. Thus the inclusion of traffic flow in the models warrants its own section.

The observed relation between number of accidents and traffic flow is close to linear, but for personal injury accidents slightly weaker (exponent less than one) than linear. This reflects in the models as the form of inclusion of traffic flow is found either as a strictly linear relationship, where the number of accidents is considered as being directly proportional to the traffic flow (the Swedish and Finnish model), or power relation where the traffic flow is raised to an exponent less than 1 (the Danish model).

When the relation between accidents and traffic flow is modelled as linear the accident rate is sometimes used instead of number of accidents as the dependent variable in the models. The connection between models for accident rate and models for number of accidents is simple as:

Accident rate = Number of accidents / Vehicle mileage

3.3.7 Model structure

The individual models in Figure 7 can take the form of either plain accident rates or equations where the number of accidents depends on explaining variables such as traffic volumes or environment. The formula below (1) is an example of an equation for calculating number of accidents taken from the Danish model.

$$E(\mu) = aN^{p} \exp\Sigma\beta_j x_{ij} \qquad \textbf{(1)}$$

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 (Greibe 2003)

Below follows the other case, accident rates, taken from the Swedish models (see Table 3).

Table 3 Accident rates and severity for two lane roads

Grundvärden 2kf, dvs. "vanlig väg belagd väg och grus samt ML 50 och 70														
Hast		funktion	MF				C				G			
	miljö	bredd	Ok	SF	AF	EF	Ok	SF	AF	EF	Ok	SF	AF	EF
50	Y	GIF	0,480	0,37	0,063	0,76	0,033	0,8	0,136	0	0,033	0,95	0,276	0
	M		0,560	0,36	0,061	0,76	0,053	0,8	0,136	0	0,053	0,95	0,276	0
	C		0,640	0,35	0,060	0,76	0,076	0,8	0,136	0	0,076	0,95	0,276	0
	Y	Tang	0,400	0,27	0,046	0,76	0,028	0,8	0,136	0	0,028	0,95	0,276	0
	M		0,480	0,26	0,044	0,76	0,045	0,8	0,136	0	0,045	0,95	0,276	0
	C		0,560	0,25	0,043	0,76	0,067	0,8	0,136	0	0,067	0,95	0,276	0
	M	City	0,480	0,25	0,043	0,76	0,057	0,8	0,136	0	0,057	0,95	0,276	0
	C		0,640	0,24	0,041	0,76	0,096	0,8	0,136	0	0,096	0,95	0,276	0
70	Y	GIF	0,320	0,5	0,110	0,65	0,016	0,85	0,187	0	0,016	1	0,590	0
	M		0,400	0,49	0,108	0,65	0,030	0,85	0,187	0	0,030	1	0,590	0
	Y	Tang	0,216	0,35	0,077	0,65	0,011	0,85	0,187	0	0,011	1	0,590	0
	M		0,288	0,34	0,075	0,65	0,022	0,85	0,187	0	0,022	1	0,590	0
	L	< 5.7m	0,456	0,52	0,138	0,64	0,011	0,85	0,315	0	0,009	0,95	0,447	0
		5.7-6.6m	0,416	0,52	0,138	0,64	0,010	0,85	0,315	0	0,008	0,95	0,447	0
		6.7-7.9m	0,376	0,52	0,138	0,64	0,009	0,85	0,315	0	0,007	0,95	0,447	0
		8-10m	0,360	0,52	0,138	0,64	0,009	0,85	0,315	0	0,007	0,95	0,447	0
		10.1-11.5	0,336	0,52	0,138	0,64	0,008	0,85	0,315	0	0,006	0,95	0,447	0
		11.6 -	0,320	0,52	0,138	0,64	0,008	0,85	0,315	0	0,006	0,95	0,447	0
		ML	0,248	0,53	0,133	0,64	0,002	0,85	0,315	0	0,002	0,95	0,447	0

Explanations for Table 3

Road users

MF Motorised vehicles
C Cyclists and mopeds
G Pedestrians

Accident rate and severity

Ok Accident rate per million axle pair km (incl. property damage only)
SF Number of injured and killed per accident
AF Number of seriously injured and killed per accident
EF Number of property only accidents per total number of accidents

Environment

Y Urban area, outer part
M Urban area, between outer and central part
C Urban area, central part
L Rural area

Function

GIF Thoroughfare, entrance route, by-pass
Tang Tangential streets
City City centre streets

3.4 The DRAG family of accident models

The accident models in the DRAG family are based on the division of the occurrence of an accident into the three dimensions exposure, risk and consequence. DRAG stands for Demand for Road use, Accidents and their Gravity. The concept of the DRAG models is to divide the model into separate models for the three phases, explaining each part by its own model.

The DRAG models are more focused on a macro-perspective than the earlier described models. While TARVA and the rest of the national models predict accidents on a level of streets, or street networks, the DRAG models predict accidents for whole regions. The variables used in the models are thus different, focusing more on variables contributing to a variation in exposure, risk or consequence between regions, rather than between streets. The models are either based on cross section data from several cities or regions, or time series data, studying a longer period of time.

The exposure is described as total mileage for cars, and is often calculated from the total fuel consumption in the area studied, or in some cases taken from surveys. Variables used for describing risk can be alcohol and medicine consumption, average speeds in the region, traffic flow and other variables predicting risk on an aggregate level. Speeds are also used for predicting gravity of the accidents, as well is seat belt usage. An interesting result shown is that seat belt usage affect injury accidents more than fatal accidents, due to the fact that the crash violence in fatal accidents often is severe enough to kill the car occupant whether using a seat belt or not. (Gaudry & Lasarre 2000)

3.5 Hypotheses

3.5.1 Exposure for vulnerable road users

Several studies of the fit of accident models show that exposure is the most important variable in explaining the systematic variance in accident numbers (Brüde & Larsson 1993, Greibe 2003, Kulmala 1995). It is very rare, though, that the accident models for VRUs on links include estimates of VRU exposure. In the Swedish accident models for links the exposure used is that for motorised vehicles, even in the models for vulnerable road users. Even though the exposure to motorised traffic flow is one explaining variable for number of accidents with vulnerable road users, it must be secondary to their own flow when it comes to explaining their number of accidents. The most likely reason for the omission of exposure of vulnerable road users is that data on the exposure rarely is available for a given street network.

One of the main aims with this dissertation is to explore how the accident models can be improved by introducing exposure measures for vulnerable road users.

Hypothesis:	*The degree of explanation of the accident models for vulnerable road users can be greatly improved by including estimates of VRU exposure in the models.*
Operationalisation:	Models will be constructed including exposure measures for vulnerable road users. The ability of these models to predict the number of accidents with vulnerable road users will then be tested against that of models including the variables of the existing Swedish models for vulnerable road users.

3.5.2 Vehicle speeds

Most accident models use the speed limit as one of the explaining variables for the accident rate. Actual vehicle speeds offer a finer tuning with a continuous variable. They also have a more direct connection to accidents. On the other hand in empirical data speeds have a tendency to correlate with accidents in a reverse way to most controlled before-after studies. This is due to the fact that speeds correlate with many other variables which in their turn have an effect on accidents. By using variables that describe the complexity of low speed environments (for instance: intersections per km, parking facilities, land use), the hope is for speeds to turn out with increasing effect on accidents, and their severity, when speeds increase.

Hypothesis:	*The degree of explanation of accident models can be improved if actual speeds are used instead of speed limit.*
Operationalisation:	In the fitting of models, vehicle speeds as well as variables describing the complexity of low speed environments will be tried. Alternative models will also be fitted where the effect of speed on accidents is preset to established effects. The fit of the models will be compared between models based on speed limits and the models based on actual vehicle speeds.

3.5.3 Division of models according to accident type

In today's Swedish accident models all accidents with motorised vehicles are combined in the same models. Different accident types however often have different factors affecting their occurrence. Different statistical distributions for each accident type, added together, may also account for some of the over dispersion compared to the Poisson distribution. This should imply that models predicting number of accidents with motorised vehicles should have individual form according to the accident type they are trying to predict in order to achieve the best predictions which is suggested by among others Allain & Brenac (2001) and Qin, Ivan & Ravishanker (2004). A too detailed division however would also produce problems of another kind. The accident data available are most often a limiting factor reducing the practical number of accident types to use or the models for each accident type would be based on too few accidents to produce sound models. Therefore one step in the modelling is to find a good compromise between dividing the models into separate models for distinctly separate accident types and having enough accident data for each type in order to produce sound models with respect to the large random variation in the accident data. This is also dependent on how much data is available. With a large accident material a division into more accident categories can be made.

In this dissertation the models for accidents with motorised vehicles have been divided into two categories; single-vehicle accidents and multiple vehicle accidents. The most important variable, exposure, is fundamentally different in the two cases. In the case of single-vehicle accidents, only one flow of cars influences the exposure of risk, while for multiple vehicle accidents the flow comes into the equation twice. The risk per driven km for being involved in a single-vehicle accident can actually decrease with flow as with increasing flow a potential single-vehicle accident can turn into a head-on crash. The attention towards surrounding traffic can also be expected to be raised as the number of other vehicles in the traffic environment increase.

Hypothesis: *The degree of explanation for vehicle accident models can be improved by separating the vehicle accident models into separate models for single vehicle accidents and multiple vehicle accidents.*

Operationalisation: Models will be constructed for both all vehicle-accidents and for single and multiple vehicle accidents separately. The fit of the models will then be compared to assess whether the division of vehicle accidents into two separate models according to accident type has improved the fit of the models.

4 Methodology

This chapter describes two methods used during the field studies and the modeling process using Generalised Linear Models (GLiM).

The two methods for field studies have been developed for collection of data on the exposure of VRUs and vehicle speeds. The methods can, in short, be described as short observations of the number of pedestrians and bicyclists, and the way they move through the street. The movement of VRUs is described both in terms of crossing behavior and separation of VRUs moving along the street. Measurements of vehicle speeds are carried out with radar equipment parallel with the observations of VRUs.

4.1 Measuring of vehicle speeds

In an earlier project (Ekman 2000) a method for measuring speeds on an area wide scale has been developed. The most characteristic feature of the model is that measurements are made on a large number of spots but to a limited degree at each spot. The reason behind this was the wish for describing speed levels on a whole street network without putting down enormous resources into measurement. If the resources are scarce, this is the way to measure speeds in an area wide way.

When performing the short measurements much time is saved, but the accuracy of the resulting estimates has to be kept under control. The previous testing (Ekman 2000) examined two potential problems:

- Variation of speed levels during the day
- Precision in the estimates of the speeds at the time of the measuring

By performing long time (two weeks) speed measurements with stationary equipment the variation of speeds over the day and week was studied. The speeds were found to be considerably higher during night time, but during day time (8 am to 8 pm) the variations in speed were small. Some precaution should however be made when interpreting this as the street studied did not have any signalised intersections close by the place where the measurements took place, nor was it particularly likely to have congestions.

The main issue in this project is the ability to compare speed estimates on different streets. Since all measuring has been done during daytime, no large problems should arise because of time variation between different times of measuring.

The precision of the estimates was studied, examining the connection between the number of vehicles measured and the precision in the estimates. This was done with a bootstrap procedure checking the confidence intervals for the speed distribution (Ekman 1996&2000). The findings were that for just 30 vehicles measured the precision was often poor. For 50 vehicles measured the precision was better and for 100 vehicles very accurate. The difference in precision is best shown graphically. The graphs below show bootstrap runs with 30, 50 and 100 vehicles respectively. The speeds are from link U8 (Uppsala) which has a standard deviation for speeds of approximately 8 km/h. The confidence interval narrows considerably from 30 to 100 vehicles, but already after 50 vehicles measured the estimates are fairly stable with 90% confidence intervals around 5 km/h.

In this project the whole distribution is not used, but only the mean speeds and the standard deviation (as an indicator of speed dispersion). The precision of the estimate of mean speed can be calculated as the standard deviation divided by the square root of the number of measured vehicles minus one. The standard deviation of speeds on urban streets is normally between 5-10 km/h. From this the size of the 90% confidence interval can be calculated as ±1,96*standard deviation of mean. For the different sample sizes the respective confidence intervals will be: 30 vehicles – ±2,9 km/h ; 50 vehicles - ±2,2 km/h ; 100 vehicles – ±1,6 km/h. Already very few vehicles measured will give us an estimate of the mean speed within a few km/h. It must be noted that the calculations of standard deviation made here naturally produce lower values than if they would have been calculated based on speed measures for 24h. Although the speeds do not vary much during daytime, they are normally higher during night time, and speeds can also be assumed to be less dispersed seen over just fifteen minutes compared to several hours.

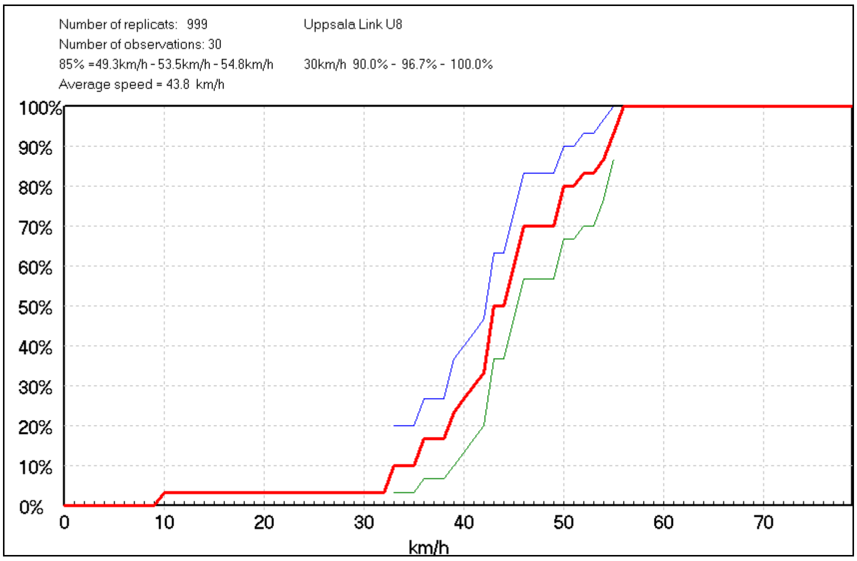


Figure 8 Speed distribution with 90% confidence intervals, sample size=30 vehicles, Link: U8

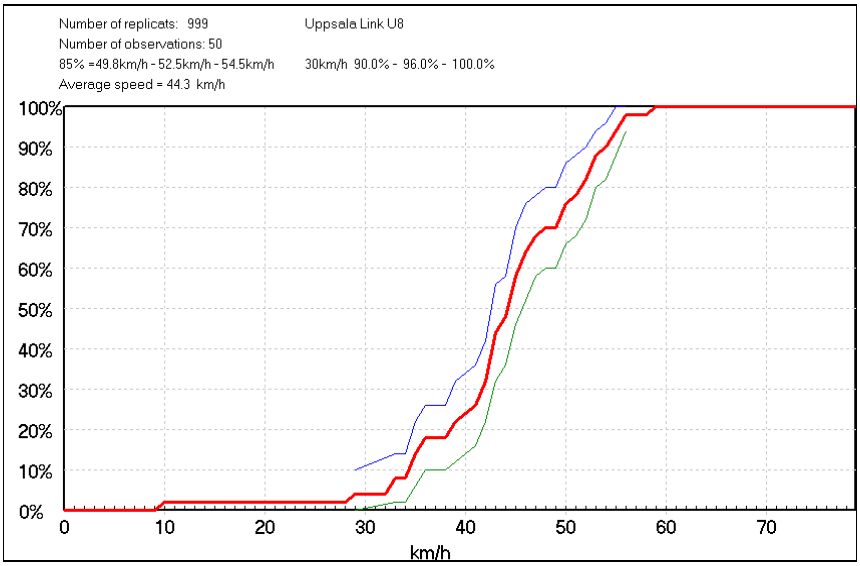


Figure 9 Speed distribution with 90% confidence intervals, sample size=50 vehicles, Link: U8

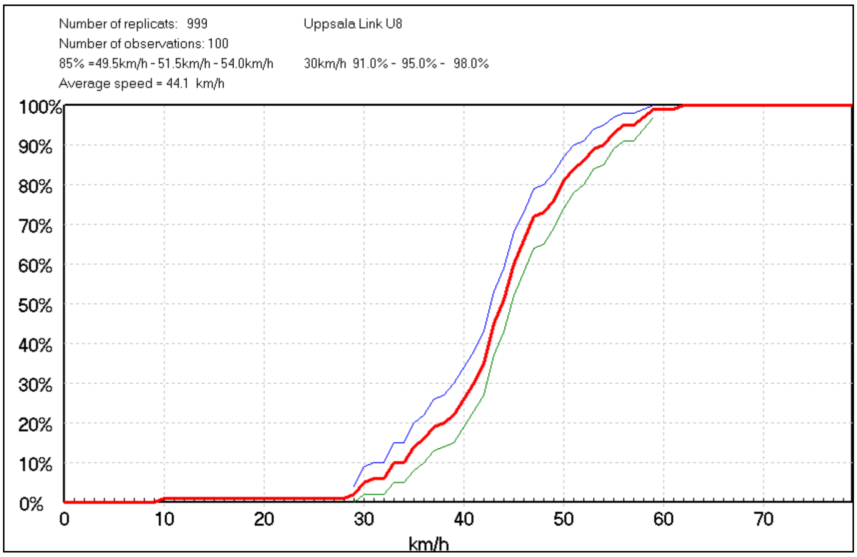


Figure 10 Speed distribution with 90% confidence intervals, sample size=100 vehicles, Link: U8

Spot speed vs. Journey speed

The speeds used are spot speeds. This is the speed normally used in traffic safety work. In other areas of traffic engineering the journey speed is a more appropriate estimate to use, for traffic safety the use of journey speed could in many cases produce very inadequate results. The journey speed is calculated as the distance covered per time over a longer distance. If a driver stops for a red light or for any other reason lowers the speed to a very low level for some time, this has a very large effect at lowering the journey speed. The safety can, however, not be expected to be largely influenced as the speed will still be high for the major part of the travel, and it is the momentary speed when a potential accident occur that influences the risk.

Speeds of free vs. all vehicles

Pasanen (1992) has shown that the speeds influencing the safety situation of VRUs the most are the speeds of so called 'free vehicles'. These are vehicles where the driver isn't limited in the choice of speeds by other drivers in front of the vehicle. For the purpose of linking speeds to number and severity of accidents, it would probably be best to use the speeds of the free vehicles. However, the speeds of all vehicles are more commonly available and easier to obtain. In the earlier tests with the method speeds have been measured for both free and all vehicles. The speeds of free vehicles and those of all the cars differ very little. However, close to signalised intersections, or during congestion, free vehicle speeds can differ largely from all vehicles' speeds. This has not been a large problem in the field studies since very few of all the main streets suffer from congestions, and signalised intersections are not common along the links, as these only contain intersections with local streets, not other main streets. However, in the ends of links, signalised intersections may occur more often. This is not perceived, however, as a large problem since the measurements haven't been carried out close to these.

Field studies

The speeds have been measured with radar at a single spot on the link. The equipment used has been traditional handheld radar guns; although the radar has usually not been held in hand in order not to alert the drivers of the speed measurements. Instead the radar was discreetly placed on a bag or any object in the proximity of the street.

The spots for speed measuring were chosen where speeds were not largely influenced by local speed reducing measures, signalised intersections or other occurrences that might drastically reduce speeds. It might be argued that the speeds most associated with the occurrence of accidents would be the speeds at places where many interactions between road users occur, for instance at pedestrian crossings, minor intersections etc. This would, however, pose other problems. The speeds would have to be measured at many spots along the link, and each link would thus not be associated with one speed, but many. When using the models there would also be a need of estimating the speed at many spots.

At each location 15 minutes of measuring was done, but if 90 vehicle speeds had been measured and the speeds measured seemed stable the measuring was discontinued before the 15 minutes were at an end. On some subsequent links speeds have only been measured for one of them if they have had the same character and conditions. The speeds measured have then been used for both. These occasions were fairly rare and used primarily when the originally selected links had to be divided into several in the field in order to allow observation over the whole link.

The measuring of speeds with handheld radar naturally raises the question whether the speeds may have been influenced by the measuring. Answering this question is not easily done. The handheld radar was positioned in a non-conspicuous and discrete way in order to minimise the risk of influencing the speeds. The radar has rarely been used handheld, only when other alternative positions have been more or less impossible. The speeds have normally been read from the equipment from a distance, enabled by the large display, and noted down. The impressions from the measuring are that vehicle speeds haven't been dropping rapidly when the vehicles approach the measuring spot. The speed has been registered only for a specific spot, but the equipment has allowed the monitoring of the speed of the car for a longer to distance and so allowed for detection of any deceleration (or acceleration).

The use of stationary automatic equipment instead of radar was also considered. With automatic speed measuring equipment, such as pneumatic tubes connected to a logger, the speeds could be measured for a longer period of time without much extra work. The equipment would, however, require substantial time to be installed, and taken down, at each site. The number of such equipments available would also limit the number of streets that could be measured simultaneously. The most important reason for choosing handheld radar equipment was, however, that this could be carried out simultaneously with the manual counts of vulnerable road users. The extra time used because of speed measuring was therefore negligible.

4.2 Measuring of vulnerable road users' exposure

The movement patterns of VRUs differ distinctly from that of motorised road users. The VRUs are able to move much more freely through the traffic environment without keeping to dedicated lanes (Figure 11). This makes the measuring of VRU exposure much more complex than that of the exposure of motorised road users. To be able to describe the exposure of vulnerable road users as accurately as possible a new method has been developed. The method is based on describing both the number of vulnerable road users moving along the street and crossing it, as well as how they do it. How they are moving through the environment is mostly a question of how they interact with, or are separated from motorised vehicles.

Figure 11 Pedestrian movement interlaced with car traffic, Norrköping Link N68

In the initial phase of the field studies the plan was to use snapshot-studies in the measuring of vulnerable road users' exposure. The snapshots discussed were to use the momentary amount of VRUs in the environment of the street as measure of the exposure. During the pilot study it was however discovered that it was possible for observers to do real time counts of VRUs at the same time as measuring speeds. The pilot study was conducted in a smaller city (Vellinge) and in the main study sometimes the counting of VRUs and the measuring of vehicle speeds had to be separated in time due to an overload of tasks for the observers. On most links it was however possible to do both task simultaneously. This was facilitated by the fact that when there are both large flows of vehicles and VRUs, the VRUs are normally very strictly guided to certain paths and thus being easily counted. A second reason for not proceeding with snapshots is that they don't detect environments with many VRUs where they are mainly crossing and not travelling along the street. These VRUs are exposed to the motorised traffic on the street but are only present in the environment during the short time of crossing and thus rarely getting caught in a snapshot.

4.2.1 Number of pedestrians and bicyclists

The number of pedestrians and bicyclists has been counted at all links, but only for fifteen minutes at each link. The road users have been counted if they move in the close proximity of the link. Road users travelling on paths clearly separated from the street, for example by hedges or wide strips of land, have not been counted, as their accidents would likely not be attributed to the link. The numbers have then been adjusted with knowledge about the variation of VRU-volumes over the day. For the adjusting, measures of variation of bicycle flows over the day have been used. The measures are from main streets in Malmö. Figure 12 shows the variation during the day for both streets in general (excluding industrial streets) and for industrial streets. As can be seen in the figure, the flow has peaks during morning and afternoon. The peaks are very high at industrial streets because of the large percentage of daily travel focused to commuting to and from work.

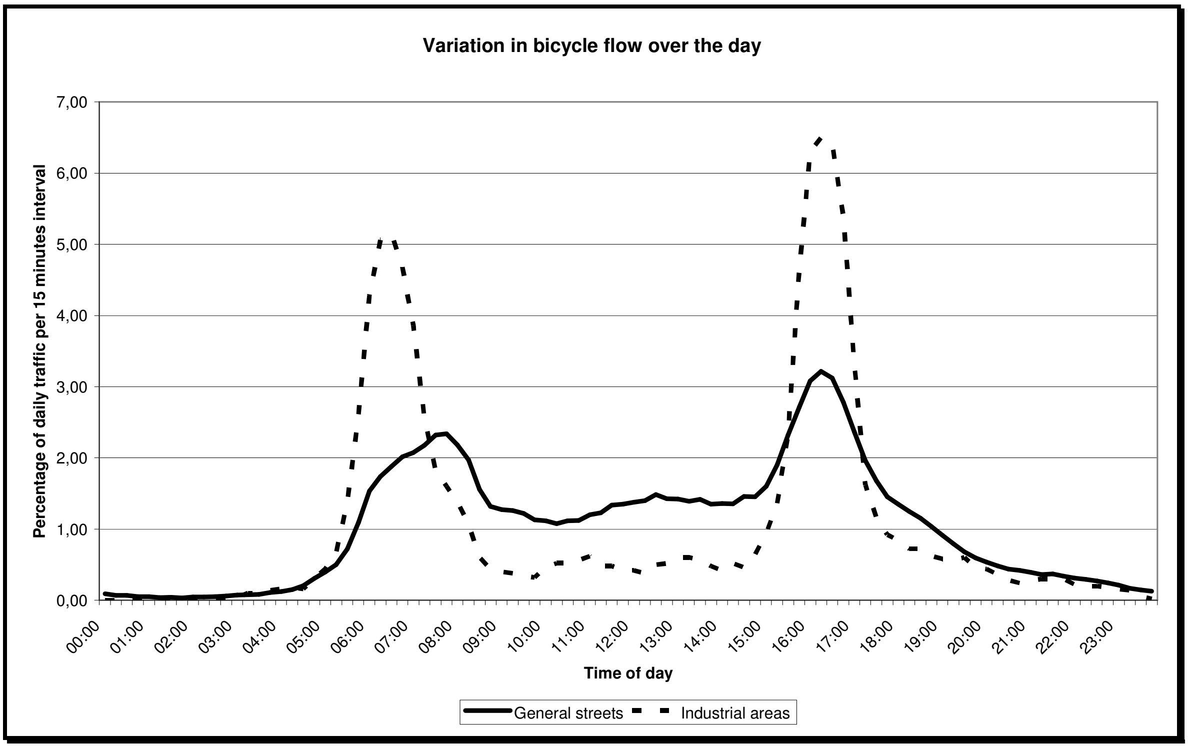


Figure 12 Variation of bicycle flows during the day, main streets in Malmö, divided into Industrial streets and General (all others), floating average over five fifteen minute intervals

The adjusting of fifteen minute counts to daily traffic is an unstable solution. The knowledge of how traffic varies with time of day for different types of streets is not a very well explored area. The details of the variation can be expected to vary between similar types of streets due to activities close to the street, a school or a library may attract flows of vulnerable road users at different times than public service institutions with other opening hours. The adjusting of the counts to daily traffic is therefore imprecise. The rough levels of VRU exposure should however be fairly reliable. During the field studies, the number of vulnerable road users at each link never disagreed totally with what could intuitively be expected for that type of link.

Number of crossing VRUs

The number of vulnerable road users crossing the street is an important variable in explaining the number of accidents on the links. The counting of pedestrians and bicyclists crossing the street has included both those crossing at a minor intersection on the link, at a zebra crossing and those crossing in between. If a road user has crossed the street several times, each time has been counted.

4.2.2 Crossing pattern of vulnerable road users

The pattern by which the vulnerable road users are moving across the street is observed and categorised in the different alternative crossing strategies below. The crossing strategies contain info not only on the way the pedestrians and bicyclists move through the street, but also about the facilities available for crossing. If the road users however don't use a certain facility, then the alternative that the road users actually use has been noted, for example: there is a tunnel under the street but most crossing pedestrians ignore it and cross the street exposed to motorised traffic.

In some cases there are found several frequently used crossing strategies on the same street, mostly in the case of signalised crossings, zebra crossings and freely anywhere. In this case the most 'free' strategy has been chosen, i.e. 'freely anywhere' over 'zebra crossing' and 'zebra crossing' over 'signalised crossing'.

The crossing strategies used were:

Crossing directly inappropriate or impossible

This alternative was used to categorise streets where it's obviously not meant for vulnerable road users to cross and the street prioritises motorised vehicles. An example of this type of street is highly trafficked thoroughfare streets with high speeds and no facilities for the crossing of vulnerable road users. There are most often very limited possibilities for vulnerable road users to even get close to this kind of streets. The streets can also have a design making it difficult or even impossible to cross, for example through steep slopes on the sides or fences barring passage. There may still be some vulnerable road users even in these environments as it is almost impossible to totally shut them out.

Grade separation

The alternative grade separation indicates that crossing vulnerable road users are being led across the street through a tunnel or possibly over a bridge. This alternative is quite close to the previous (crossing directly inappropriate). Grade separation indicates however that there are vulnerable road users in the environment. When this alternative is chosen, only the numbers of vulnerable road users actually crossing the street at grade have been counted. When this crossing strategy has been chosen, the natural way to cross the street is however through the tunnel/bridge, otherwise a different alternative has been chosen.

Canalised to signalised crossing

This alternative is used when the vulnerable road users are either crossing the street at signalised intersections or at a signalised pedestrian crossing.

Canalised to non-signalised zebra crossing

This alternative is used when the dominant form of crossing is at a non-signalised zebra-crossing or at another place clearly identifiable as a place dedicated to VRU crossings. In the latter case the place should be clearly identifiable as a VRU crossing not only to the vulnerable road users but also to the drivers.

Freely anywhere

'Freely anywhere' is used for street environments where vulnerable road users cross freely wherever they want along the street, not at designated places. To move freely across the street is not always a possibility for the bicyclists even in this alternative. However, environments where pedestrians move freely across the street often also have a particular movement pattern for bicyclists. In this type of environment they are relatively rarely separated from the motorised traffic and can move more freely in the street.

4.2.3 Separation of bicyclists along the street

For vulnerable road users travelling along the link the separation has been noted for bicyclists. The reason for having this variable for both bicyclists and pedestrians is that the pedestrians almost always have their own separate space along main streets. In the case where there is no separate space for pedestrians the street is most certainly not intended for pedestrians.

Integrated with motorised traffic

The bicyclists share the same space as the motorised traffic.

Bicycle lane in the street

The bicyclists are located in the street but separated from motorised traffic through having a dedicated bicycle lane.

Bicycle facility separated from motorised traffic

The bicyclists have separated facilities away from motorised traffic. The facilities have to be in the proximity of the street. The reason for this is that bicycle accidents on these facilities otherwise won't be connected to the street in the police records. If the facility is located away from the street, then the street normally has a character of being primarily for motorised traffic and thus instead being categorised as 'Biking along the street directly inappropriate or impossible'.

The bicyclists are normally not totally separated from motorised traffic. At each minor intersection the bicyclists often have to interact with intersecting motorised traffic (Figure 13).

Figure 13 Separated bicycle facility crosses intersecting minor street (left and right)

Biking along the street directly inappropriate or impossible

Just as for the equivalent alternative for crossing strategies this alternative has been used to categorise streets where it's obviously not meant for bicyclists to travel along the street. There may be separate facilities for the bicyclists, but the street can also be located in the outskirts of the city and have a rural character. Still, there might be bicyclists travelling along the street, see Figure 14.

Figure 14 Biker on street categorised as inappropriate for biking

4.3 Accident modelling

4.3.1 Accidents as statistical phenomenon

Systematic and stochastic variation

The variation in accidents can be contributed to both a systematic variation dependent on the variation of road and traffic characteristics influencing the number of accidents, and a purely random part.

The systematic variation in traffic accidents is for instance largely dependent on the amount of traffic. The more intense the traffic, the more accidents the environment tends to generate. The number of accidents at a specific site is not, however, solely dependent on the amount of traffic at the studied site. The occurrence of accidents is firstly due to the existence of traffic flows creating exposure to conflict situations which might lead to accidents. Secondly the number of accidents is dependent on the risk of these conflicts actually evolving into accidents. In an additional third stage, variables affecting the outcome, given an accident, can be studied. These variables affecting the occurrence of accidents and their severity can be said to give rise to systematic variation in the number of accidents.

Not all of the variation is however explainable by way of the variation in traffic characteristics. In a chain of events leading up to an eventual accident, the road users involved have some possibilities to avoid a crash. It is not until several circumstances (e.g. high speed, lack of guard-rail, lack of attention) coincide that an accident actually occurs. It is impossible to make a model which takes into account all possible circumstances. There will always be a certain amount of unexplainable, random, variation in the number of accidents.

Distributional assumption

The view on the statistical phenomenon that traffic accidents constitute, and the modelling of accidents, has evolved over time. Early accident modelling was done with least square estimates with assumptions of the accidents being normally distributed; later accidents have been seen as purely Poisson distributed (Maher & Summersgill1996). Now the most common assumption is that of Poisson distribution at individual sites in combination with a gamma distribution between sites, which result in a negative binomial distribution for the overall (Maher & Summersgill 1996, Kulmala 1995, Allain & Brenac 2001).

In this work the accidents will be modelled as Poisson distributed even aggregated over sites. This assumption of accidents as Poisson distributed over several sites leads to an error as the variance of accidents will usually be larger than the mean. The data will be overdispersed compared to the assumed distribution. The error on parameter estimation in modelling is very small for moderate over dispersion, but test statistics will take on values showing more significant results than they actually are (Maher & Summersgill 1996). To handle this, a Quasi-Likelihood Poisson model is used.

4.3.2 Generalised Linear Models

Generalised Linear Models (GLIMs) have been chosen for the modelling since the dependent variable, the number of accidents, isn't normally distributed, but rather Poisson- or negative binomial-distributed. GLIMs allow for the dependent variable to take on other distributions than the normal distribution, the only restriction being that the distribution belongs to the exponential family.

Generalised linear models will not be described in detail here as there are much better sources available for that (Olsson 2002, Allain and Brenac 2001). Some parts of the GLiM have to be described in order to understand the results generated by them, and to provide some knowledge about the choices made in the modelling.

GLiMs use a link function to generalise the connection between the dependent and the independent variables.

$$g(\mu)=\eta=X\beta$$

In the case of a Poisson distributed dependent variable a logarithmic function is normally used as the link function, which also has been the case in this work.

When a logarithmic link function is used, the model will take on the form:

$$E(Y) = e^{\sum \beta_i x_i}$$

where:

E(Y) is the dependent variable, in this case the expected number of accidents

β_i are parameters to be estimated by the model

x_i are the independent variables

The model design with a sum of βx terms is characteristic for linear models. In the case of a logarithmic link function the exponential relationship will however lead to a multiplicative form since:

$$e^{\beta_1 X_1 + \beta_2 X_2} = e^{\beta_1 X_1} \times e^{\beta_2 X_2}$$

To be able to obtain simple models, the covariates used have been logarithmated. By using logarithmated variables the expression obtained can be simplified using the relation $e^{\beta_1 Ln x_1} = x_1^{\beta_1}$. In the case of categorical variables one of the categories is chosen as base category and obtains the value $e^0=1$ whereas the other categories are assigned dummy values e^β describing their effect on the number of accidents relative to the base category. When a categorical variable is added, then all categories are used in the models, although sometimes one or more have been merged. Table 4 shows an example of a model description. The categorical variable Landuse has the values Residential (B), Insistutional (A), Centre (C), Industrial (I) or None (X). In this particular case, the assigned parameters to A, B and C respective I and X have been very close to each other and so the categories have been merged.

Table 4 Example of model description

Model: Bicycle, no preset				
Deviance	**df**	**Perc.exp.**		
501	381	60%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	2.16E-04		-5.93	0.000
NCXP		0.35	5.50	0.000
Flow		0.76	4.93	0.000
Landuse_ABC	1			
Landuse_IX	0.60		-2.54	0.011
Func_GIF	1			
Func_T	1.36		1.86	0.064
Func_C	1.94		2.27	0.024
Vis_Good	0.71		-2.03	0.043
Vis_Medium	1			
Vis_Poor	1.62		0.96	0.335

On another form, the model can be written:

Bicycle accidents per 5 year and km = $2.16*10^{-4} * NCXP^{0.35} * Flow^{0.76}$ * Landuse * Func * Vis

The link length has been used as an 'offset' variable for all models. This means that in the modelling the number of accidents is rated per link length.

Scaled Deviance (SD)

In general linear models the sum of squares is used in order to optimise the fit of the model. In generalised linear models, the scaled deviance is used instead. Just as the least squares method, the scaled deviance is also a maximum likelihood estimate.

NAG Excel Add-ins

The software used for the modelling has been the NAG Statistical Add-ins for Excel (NAG 2000) from the Numerical Algorithms Group (the developers of the GLIM software). The main advantage of the package is that it is integrated with Excel, making the statistical routines directly available in the database program that was used. A second advantage is the simplicity in using it; the routines are called as ordinary functions in Excel. A disadvantage with the package is that it does not support a Negative Binomial distribution for the dependent variable. A difference compared to large statistical software as SAS is that more manual work is needed with the NAG Add-ins. This provides a better insight into the procedures, but does not provide the quality assurance of the process as large statistical software packages do.

4.3.3 Modelling procedure

The creation of models has been slightly developed throughout the process of modelling. This section describes the final process but in some cases with explanations on the turn of events that led to this.

Separate models are made for four different accident types:

- Multiple vehicle accidents
- Single vehicle accidents
- Bicyclist accidents
- Pedestrian accidents

The development of accident models has followed a step-by-step plan. Some of these steps have occasionally been integrated, but primarily the following procedure has been followed:

1. Construction of modelling variables from gathered data
2. Studies of the individual variables, dependent as well as independent
3. Studies of the correlation between pairs of variables
4. Fitting of GLIM, one variable at a time, the variable presenting the best fit in combination with simplicity and compliance with established empirical effect being chosen. When adding a categorical variable all categories are added.
5. Significance testing of the new model compared to one without the added variable
6. Significance testing of the individual variables in the model
7. Assessing the fit of the model

Step 4-6 are iterated until no more significant variables can be found.

1 Constructing modelling variables

The independent variables used in the models to explain the systematic variation of the dependent variable have been constructed from data gathered in field studies and from the municipalities. The word *construction* refers to the fact that the raw variables gathered sometimes have been standardised or adjusted. One example is the number of intersections which has been standardised into intersections per km as the streets studied don't have the same length. Another example is the flow of pedestrians and bicyclists. These flows have been measured by very short counts at each street and a daily flow has been calculated by taking into account the variation in flow over the day.

The independent variables can be divided into two groups; covariates and group variables. The group variables are classification variables taking on a limited number of discrete values which can be numbers or text. Covariates are all other variables, generally being numerical variables.

Several of the covariates can assume the value zero. In those cases the zero-values have been replaced with relatively small values during the logarithmation, least they would not be logarithmable as the logarithm of 0 does not exist. This is formally not a very correct way of handling zero values, at least not for variables where zero is a fully logical value. To exemplify the latter statement we may look at two different variables that in the data may have zero as value: *Number of pedestrians crossing the street* and *Minor intersections per km*. Even on rural highways where walking and cycling is prohibited there will sooner or later be crossing pedestrians, perhaps due to a car breakdown. The true value of the variable *Crossing pedestrians* should never be zero and so the substitution of zero values with very small values is not totally false. For *Minor intersections per km* on the other hand, the value may very well be zero if there are no minor intersections on the link. An alternative approach would have been to change the variable to a categorical variable, grouping together links with similar number of links per km. This would on the other hand make us lose the continuity of the variable.

In the data used here all covariates which can have the value zero belong to one of two categories: VRU-exposure variables or variables describing number of minor intersections per km. For the VRU-exposure variables all zero-values have been replaced with the value 1 (one vulnerable road user per day). This will give the logarithmated variable the value 0. The variables describing number of minor intersections sometimes assume values lower than 1. The lowest number of minor intersections per km found in the data (besides the zero-values) was 0,425. Zero-values for number of minor intersections were replaced with the value 0,368 [=ln(-1)] so that links with no minor intersections would not receive values higher than a link with minor intersections.

2 Studies of individual variables

The individual variables have been studied in regards of distribution and frequency. For the group variables the frequency is checked to see whether some values are represented by only a few streets. In that case a merging of some values might be preferable.

3 Studies of correlation between variable-pairs

For pairs of variables, the correlation between them has been studied, both numerically through correlation coefficients, and graphic plotting of the variables against one another. This is done before the modelling in order to find out whether the independent variables might actually be dependent on each other. Many of the variables describing the traffic can naturally be expected to covariate with each other. For instance vehicle speeds are covariating with several different street design variables as well as with the presence of vulnerable road users. The results of the correlation studies are used when interpreting the results of the modelling. In case two or more of the variables included in the models are strongly correlated, the interrelation between the variables may cause the relation between the dependent variable and the independents to deteriorate. An indication of this can be seen when adding variables to the model. If the model parameters associated with existing variables change strongly when a new variable is added, then the new variable and the old are strongly correlated which makes the individual effects hard to estimate for the models. When this happens there are several ways to deal with it. One is to simply exclude one of the variables, another is to include both of them but compare the effects with previous knowledge about their effects on the number accidents. In the latter case, a sensitivity check should be performed to see how stable the estimates are.

Example: The speed of vehicles and the exposure of pedestrians have in the theoretical part both been found to influence the number of pedestrian accidents. However, environments with many pedestrians are mostly found in the city centre where speeds are generally low, and so vehicle speeds and the exposure of pedestrians are naturally found to correlate negatively. In the present set of data they have a correlation coefficient of between -0,4 and -0,5. Low speeds might be found to correlate with a high number of accidents, not because of a causal relationship (which should be the opposite), but because of the large number of pedestrians we tend to find in low speed streets.

4 GLIM

When more variables are added to the models, the fit of the model automatically gets better as the fitting procedures gets one more lever to work with in attuning the model. Whether the variable actually adds value to the model is a slightly different issue. The quality of the model is assessed with the deviance explained related to the number of degrees of freedoms used to obtain this explanation (Olsson 2002). This is however not the only criteria used, simple models and models that comply with theoretical effects of variables are preferred.

5 Checking for significance of added variables

To check whether an added variable adds significantly to the degree of explanation the difference in deviance between the model with and without the new variable is compared against a Chi2-distribution with degrees of freedom according to the difference in residual degrees of freedom between the two models. The difference in deviance is divided by the scale factor (see below) to compensate for the effect of over dispersion.

The QL Poisson model uses a scale factor to the variance to compensate for the effect of over dispersion on test statistics. The scale factor is calculated as the sum of squared Pearson residuals divided by the degree of freedoms for the studied model.

6 Significance testing of individual variables in the model

The parameters of individual variables are checked for significance. This is done by checking the test statistic: Parameter Estimate / Standard Error against Tukey's t-distribution. The test statistic is divided by the square root of the scale factor to compensate for the effect of over dispersion on the test statistic (see *Checking for significance of added variables* above).

7 Assessing the total fit of the model

The total fit of the model is assessed by checking how much of the systematic variation is explained. The total amount of systematic variation is calculated as

$$\frac{SD_0 / df_0 - SD_M / df_M}{SD_0 / df_0 - SD_{ME} / df_{ME}} \quad \text{(Kulmala 1995)}$$

where

SD=Scaled Deviance, df=degrees of freedom

0 signifies a null model with only one constant variable

M signifies the studied model

ME signifies a model where all systematic variance is explained, thus is SD_{ME} the Scaled Deviance in the 'perfect' model that explain all systematic. If the SD of a model goes below this level it starts to explain sum of the random variation as well and caution should be taken. The Scaled Deviance for the ME model is estimated by using a well fitted model for the accident type in question and summing up:

$$SD_{ME} = \sum_i^n E(SD_i) = 2\sum_i^n \sum_{y_i} \left[(y_i \log(y_i / \hat{\mu}_i) + \hat{\mu}_i - y_i) \frac{\hat{\mu}_i^{\,y_i} e^{-\hat{\mu}_i}}{y_i!} \right]$$

where:

y_i is the observed number of accidents at link i, $\hat{\mu}_i$ is the expected number of accidents at link i

and n is set to 20 (Formula 21, Kulmala 1995)

4.3.4 The low mean problem

When dealing with a Poisson distributed phenomenon where there are very low means (<0,5), the Scaled Deviance (SD) tends to become small compared to the number of degrees of freedom. A common target for the SD of a model is the number of degrees of freedom, but in the case of low means this may be inadequate as already very crude models can reach an SD below the number of degrees of freedom. (Maher & Summersgill 1996)

In this dissertation the problem with low means, and the resulting low Scaled Deviances, has been solved by not using the degrees of freedom as the target level for the models, but a numerically calculated target level where all systematic variation is explained (Kulmala 1995).

The low Scaled Deviances do not unduly affect the decrease of SD when adding variables to the models (Maher & Summersgill 1996), and so there has not been any need to adjust the Chi2 statistic when adding variables.

Besides the calculation of explained variance, several more tests of the model accuracy are made:

- Residual analysis
- Checking leverages for streets which might have had an unduly large effect on the estimates
- Checking for large variations in parameter estimates between nested models

4.3.5 Mass significance

Testing for significance at the 0.05 level implies a risk of making a faulty judgement 5% of the time. When evaluating many test statistics, each at the specified level, the risk of one of the statements being false increases, unless the problem of mass-significance is taken into account. The way to take mass significance into account is to put more harsh demands on significance when many tests are made

In the course of modelling, mass significance has seldom been a problem from a practical point of view. When adding variables to the models the numbers of tests have been relatively few, only the chosen variable at each step is tested for significance. Here, the levels of significance are also very high, but the situation changes very rapidly when there is nothing more to gain through inclusion of more variables into the models. The results are most often either highly significant or not significant at all. When checking the significance of each variable there is a larger problem; several parameters are tested for significance at the same time and the levels of significance are often poor, at least for the differences between different categories in categorical variables. On the other hand, here the focus has been as much on whether the results are logical, as whether the different categories are significantly different from each other. As long as the addition of the variable improves the model in a statistically significant manner, and the results do not violate previously known, or logical, relations between the categorical variable and the number of accidents, the variable has been accepted. For covariates the significance has been followed more strictly, but they are most often highly significant.

4.3.6 Uncertainties in independent variables

In regression modelling the assumption is that the values of independent variables can be, and have been, estimated accurately without any errors or uncertainty, uncertainties are only allowed in the dependent variable. In reality, however, there are often some uncertainties in the estimates of variable values, especially for variables that are not constant over time, such as traffic flows. These uncertainties result in biased parameter estimates in the models created as shown by Maher (1989). The magnitude of the bias is mainly related to two things; the magnitude of the uncertainty in estimating the variables and the variability between observations, in this case links.

It is possible to estimate the size of the bias if both the magnitude of uncertainty and the variability are known. In our case, the variability between links can be estimated but usually not the magnitude of uncertainty. Hence, the estimates of the size of the bias in our case will remain a matter of speculation without extensive additional studies. Maher, however, estimated the bias caused by uncertainties of car flows due to counts of only 15 minutes to 4% of the parameter value.

In this work, the variables most likely to cause the largest bias are the variables describing VRU exposure. These have been counted for 15 minutes, just as the vehicle flows in Maher's study. From this, the conclusion is drawn that the estimated parameters will probably have a bias smaller than 10%. The uncertainties for the VRU counts are not known, but the variability in counts between links is large, ranging from 0 to a few hundred for the different VRU exposure variables.

5 Data

This chapter describes the data used in the accident modelling. Data has been gathered from several sources. Vehicle flow and street functions have been gathered from the databases of each municipality respectively. Accident data have been compiled by VTI from the national accident database, VITS. Additionally extensive field studies have been carried out in seven Swedish cities. The field studies have been carried out according to the methods described in the chapter "Methods". In the modelling phase of the project some of the data needed was still lacking and some additional field studies were undertaken in order to obtain data on speed limits and parking along the links.

The main purpose for conducting field studies has primarily been to obtain data on the exposure of vulnerable road users and actual vehicle speeds. Data on the exposure of vulnerable road users were rarely available from other sources. Vehicle speeds are often measured, but rarely for the whole street network, only for a limited number of sites.

Information has been gathered about the crossing behaviour of VRUs and about the separation of bicyclists from motorised traffic. Information has also been gathered about the traffic environment, such as number of minor intersections of different types along the street, number of lanes, land use along the street etc. The variables gathered are summarised in Table 5.

Table 5 Variables in database

Basic data	Environment	Intersections	Traffic characteristics	Accident data & injuries (per accident type)
Link-ID	Link length (m)	Number and type of intersections	Vehicle flow	Number of fatal accidents
Street name	Land use	Major crossings for vulnerable road users	Crossing strategy for vulnerable road users	Number of fatalities
Weather and temperature	Visibility	Number of large exits (gas stations, shopping centras etc.)	Separation of bicyclists along the link	Number of accidents with severe injuries
Observer	Number of lanes	Number of small exits (small parkings with only a few vehicles)	Number of crossing pedestrians	Number of severely injured
Date	Separation of lanes		Number of pedestrians travelling along the street	Number of accidents with slight injuries
Time of day	Parking along the link		Number of crossing bicyclists	Number of slightly injured
Miscellaneous additional information about the link	Speed limit		Number of bicyclists travelling along the street	
			Vehicle speeds, mean	
			Vehicle speeds, standard deviation	

5.1 Field studies

Short measurements have been carried out in order to obtain data from as many streets as possible. Each link has been studied for fifteen minutes during which information has been gathered about number of vulnerable road users passing through the environment and how they move through the environment. Vehicle speeds have also been measured, as well as data about the design of the traffic environment.

5.1.1 Municipalities for field studies

Field studies have been carried out first in Vellinge in a pilot study, and later in seven other cities; Uppsala, Helsingborg, Linköping, Danderyd, Katrineholm, Västerås and Norrköping. For the modelling, data has been used from all but Vellinge and Helsingborg. In the case of Vellinge the method for the field studies evolved from the pilot study to the study of the other cities, thus the data was not fully comparable. Helsingborg was excluded as VTI in their parallel studies excluded it and therefore didn't compile accident data for Helsingborg. Originally even more cities were selected for field studies, but had to be excluded because of lack of time and resources.

Table 6 Municipalities where measuring have been carried out and the extent of the main street network

City	Date	Number of links	Total measured link length (km)
Vellinge (pilot study)	021010	13	4,7
Uppsala	030324-030328	111	54,1
Helsingborg	030401-030407	60	28,2
Linköping	030414-030416	95	51,5
Danderyd	030422-030423	52	22,6
Katrineholm	030424-030425	67	27,7
Västerås	030507-030509	49	30,8
Norrköping	030526-030528	71	34,7

5.1.2 Link division

For modelling and measuring purposes the street network of each municipality has been divided into links. Each link is primarily made up by the street between two main street intersections but sometimes the link has been divided into smaller links. This can either be due to the fact that the street is not homogenous between the intersections and therefore shouldn't be modelled as a single homogenous link, or that the street is too long to be surveyed in the field from one spot. On average the links studied have had a link length of 500 m.

In total 505 links were studied in the main study, for the modelling only 393 links have been used. All 60 links in Helsingborg were excluded because of lack of compiled accident data. The remaining excluded links were excluded because of lack of data on vehicle flow.

5.2 Basic data

In addition to the variables collected for the modelling, some more basic data were collected during the measuring for use in the processing of data and construction of the database. These data are briefly described in Table 7.

Table 7 Description of basic data

Code	Description
Link identifier	A code relating to which city and on which link the measuring has been done (ex. U59 is link number 59 in Uppsala)
Street name	Name of the street
Weather and temperature	Weather condition and approximate temperature
Observer	Name of person
Date	Day/Month/Year
Start time, counting of VRUs*	Hour:Minute
End time, counting of VRUs*	Hour:Minute
Start time, measuring of vehicle speeds*	Hour:Minute
End time, measuring of vehicle speeds*	Hour:Minute
Miscellaneous	This field was used to indicate any other circumstances that might be useful when interpreting the data, ex. construction works in progress during the measuring
Photos	Photos were taken at each link to be able to quickly re-check the environment in case of uncertainties in the data or need for clarification.

* The start and stop time for counting VRUs and measuring vehicle speeds were normally the same, but when for some reason the measuring couldn't be done parallel with the counting the VRUs or the measuring of vehicle speeds was finished before the counting of VRUs, the times differed.

5.3 Traffic environment

The variables in this section are used to describe the traffic environment. Each variable's name is followed by an abbreviation used in formulas.

Link length (Length)

For each link in the field studies the link length was measured from maps. The link length is mainly used to standardise other variables, ex. the number of intersections is standardised into number of intersections per km. The average link length is about 500 m and varies mainly between 200 and 800 m. The link length is influenced by the following:

- Distance between major intersections. If the whole distance between two major intersections is fairly homogenous and can be observed from one spot, then that whole distance is used as one link.
- If the whole distance between two major intersections is too long to be observed at the same time, then the street has been divided into several links. However, if the possibility for VRUs to cross has been non-existent then the link has been studied in one go anyhow.
- If the character of the street is shifting somewhere along the distance between two major intersections, then that street is divided into several more homogenous links.

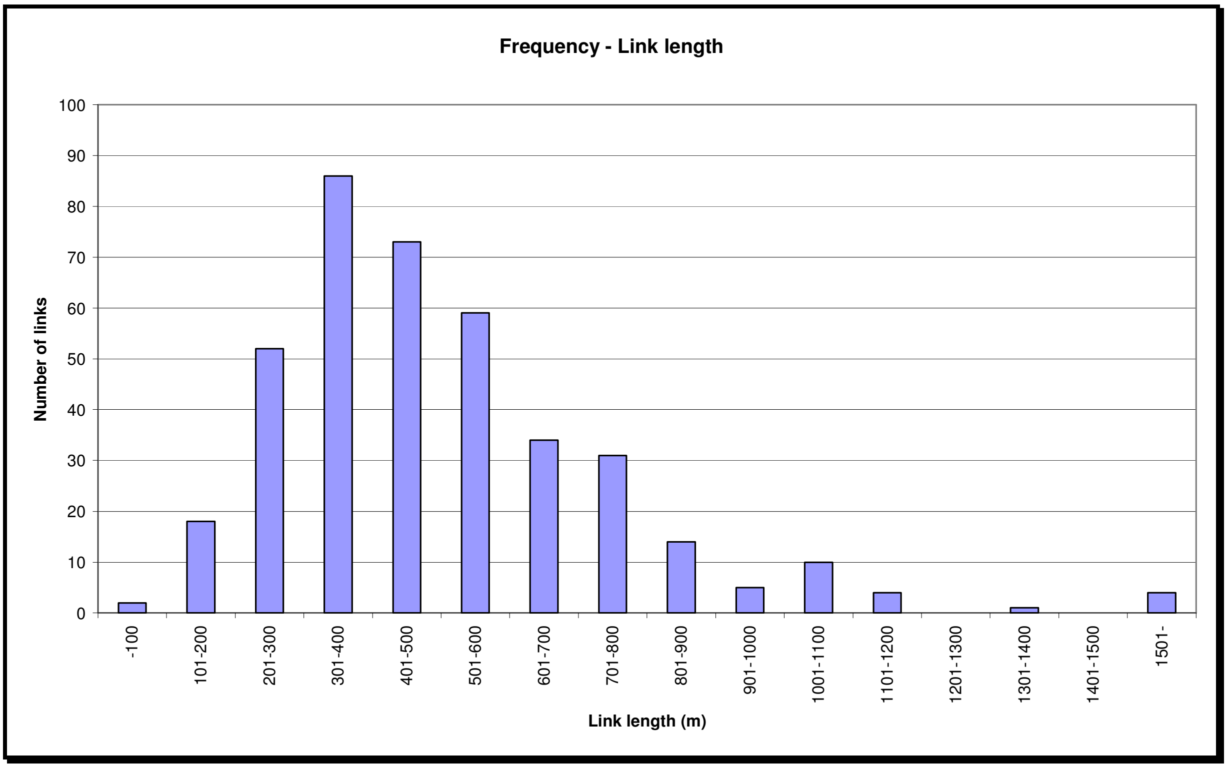


Figure 15 Frequency of link lengths

Land use (Landuse)

The variable describes the type of buildings surrounding the link and has been used with good results in other accident models (Zajac and Ivan 2003, Ossenbruggen, Pendharkar & Ivan 2001). The following categories of land use have been used:

- Institutional
- Residential areas
- Commercial
- Industrial
- None

The alternative Institutional is used when the buildings surrounding the street have some function with people travelling to and from the buildings more continuously than at an industrial street, for example public service, schools, health care etc.

The most common land use along the links studied is residential areas (Figure 16).

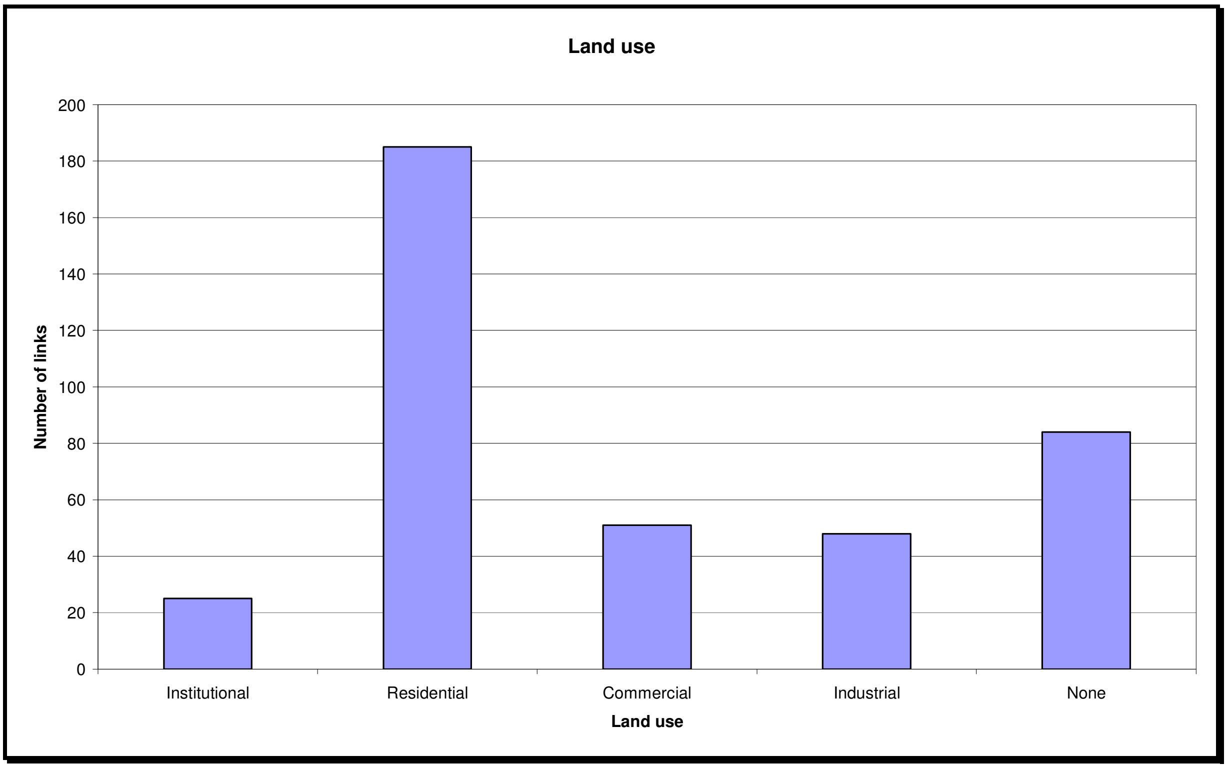


Figure 16 Frequency of different types of land use

Visibility (Vis)

This variable describes the visibility around the link. Visibility was added to the variables to record just before the field studies and so was not thoroughly discussed and tested beforehand. The value 'poor' relates to environments where there are either hedges, houses or similar large objects obscuring the vision and hiding traffic coming from side streets. The value 'good' describes environments where there is hardly any objects close to the street obscuring vision. Environments not falling into either of the categories 'poor' or 'good' were judged as 'normal'.

Only 2% of the links have been judged to have poor visibility, perhaps because main streets with poor visibility tend to get the visibility improved where possible. Good visibility was observed at 47% of the links and normal at 51%.

Figure 17 a&b and Figure 18 a&b illustrate environments assessed as respectively having poor and good visibility.

Figure 17 a&b Examples of environments with poor visibility

Figure 18 a&b Examples of environments with good visibility

Number of lanes (Lanes)

This variable practically always assumes either the value two or four. In some cities however there are streets with three lanes, two in one direction and one in the other. Nine links with three lanes were included among the 393 links used for modelling. Of the other links approximately 80% have two lanes and 20% have four lanes.

Sometimes the width of the street is used in addition to, or instead of, the number of lanes in accident modelling. Unfortunately the street width was not recorded in the field studies and not easily obtainable afterwards so street width was not used in the modelling in this work.

Separation of vehicles in opposite directions (SepV)

This variable describes the separation of vehicles in opposite direction. The variable takes on the values no separation, full separation or mixed. Just a painted line in the middle of the street is not considered as a separation, only physical barriers actually reducing the risk of a skidding car to get over on the opposite lane are considered. The mixed-value is used when within one otherwise homogenous link, the separation is intermittent.

Most of the links (78%) don’t have any separation between the lanes, 17% of the links have separation along the link and 5% have intermittent separation.

Parking along the link (Parking)

This variable describes the occurrence of parking along the link. Three different alternatives have been used:

a) No parking

b) Parking in slots located beside the street

c) Parking in the street reducing the lane width

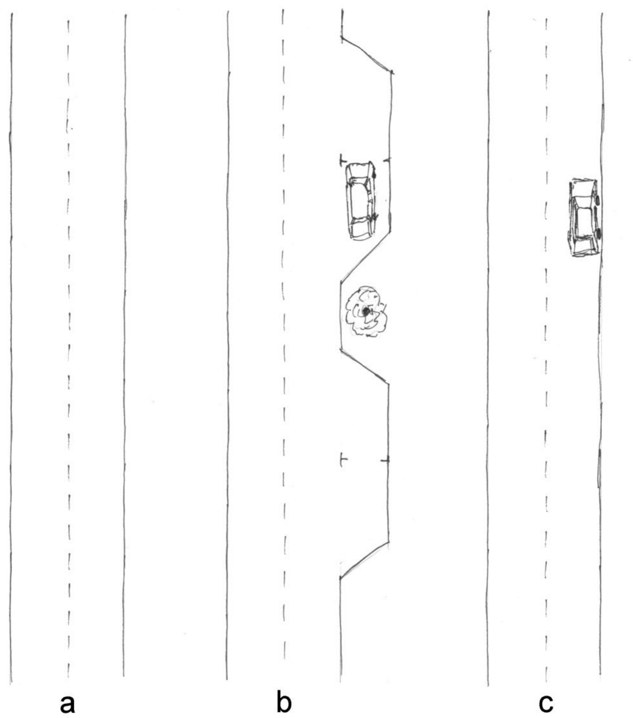


Figure 19 Illustration of parking along the street: a) No parking, b) Parking in slots beside the street, c) Parking in the street reducing the lane width

In those cases where parking is allowed, it has also been recorded whether parking exist on both sides of the street or only one. In the absolute majority of the links, 86%, there is no parking (Figure 20).

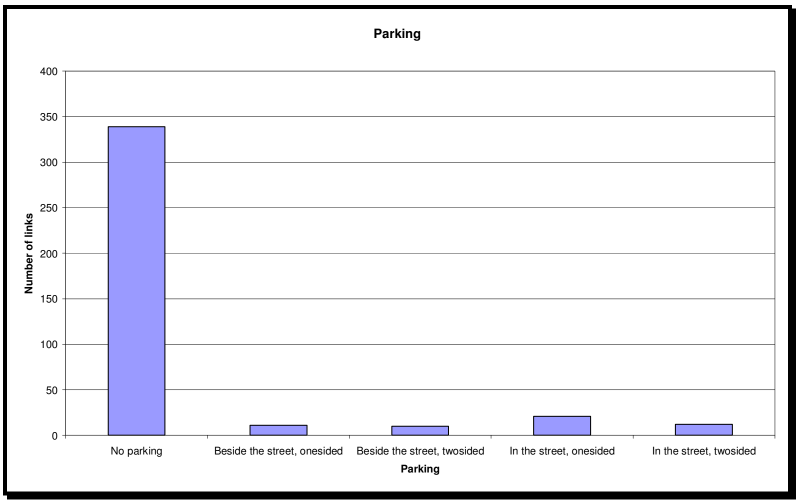


Figure 20 Frequency of parking alternatives on links

Speed limit (SL)

The maximum allowed speed has been registered for each link. The default speed limit in urban areas is 50 km/h which reflects on the links studied, 85% of the links have the speed limit 50 km/h. Remaining links have either the speed limit 70 km/h (9%), 30 km/h (2%), recommended 30 km/h (2%) or a mixture of 30/50 or 50/70 (2%).

Link environment and Function (LEnv, Func)

These two variables are used in the existing Swedish accident models. They describe the environment and function of the street as follows.

Environment is categorised with the names Outer, Middle and Centre, referring to the location within the city. The variable is however not determined by the location, but by the design of the street, the presence of vulnerable road users and whether parking is allowed along the street. For Linköping the actual location of the street was compared to the value for the variable location, 40 out of 95 links had the same location, both as actual location and as the environment-value. The variable environment is thus not to be comprehended as synonymous with location, although the names of the categories indicate it. (Vägverket 2001b)

Most of the links have been categorised as either Outer (37%) or Middle (50%) but there are also a fair amount of links categorised as Centre (13%). (Figure 21)

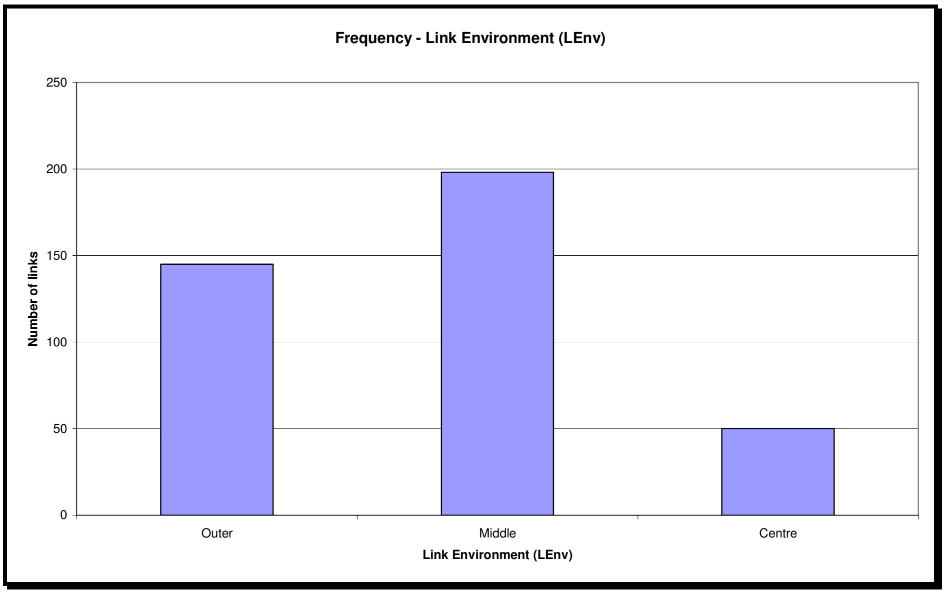


Figure 21 Frequency of links for Link Environment (LEnv)

The variable Function uses the values GIF (Thoroughfare, Entrance, Bypass), Tangential or Centre. The links described with GIF have a large portion of thoroughfare traffic and has large peaks in the variation of traffic flow over the day. Links described with Centre are the opposite with very little thoroughfare traffic and low peak-hour traffic. (Vägverket 2001b)

Most of the links have been categorised as either GIF (55%) or Tangential (40%). Links with a Centre function are rare among the links studied (5%). (Figure 22)

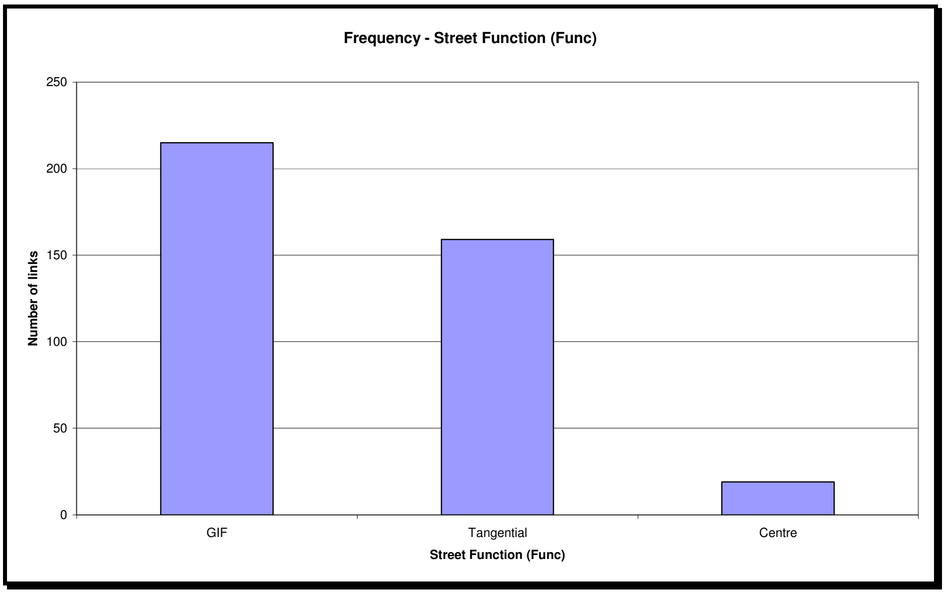


Figure 22 Frequency of links for Street Function (Func)

Minor intersections (X3Km, X4Km, X34Km)

For each link the minor intersections have been registered in regards to number and type. The type has been registered with number of arms and type of regulation. The number of arms has naturally been dominated by three- and four-arm intersections but in one case the number of arms have been five, in this case the intersection has been coded as a four-arm intersection for modelling purposes. The following types of regulation have been used:

- Signalised
- Roundabout
- Stop
- Yield
- Right of way to traffic coming from the right

Most of the links include at least one minor intersection, only 54 out of the 393 links don't have any intersections at all.

Since the minor intersections are intersections between main streets and local streets they often have one dominating street (the main street). Thus it's not surprising to find mostly Yield-regulated intersections (73%). More unexpected is the finding of many intersections with right of way to traffic coming from the right (16%) and signalised intersections (8%). Roundabouts and stop regulated intersections are rare on the selected streets (1%).

Major crossing for vulnerable road users (VRUX)

At some links there are intersecting paths for vulnerable road users. These have been noted down as major crossings for vulnerable road users. These are not just zebra crossings for pedestrians and bicyclists, but places where large number of vulnerable road users can be expected, and the motorised road users are alerted of this by way of signs or specific designs of the crossing.

Major crossings for vulnerable road users occur at 57 of the 393 links.

Large and small exits (LX, SX, LSX)

The number of large and small exits (and entrances) along the link has been registered. The boundary between small and large exits has been defined by the number of parking places served by the exit. If the exit is only serving a few parking places (a villa or a small block) it is defined as a small exit, while exits serving shopping-centres and other larger parking lots are defined as large exits.

Streets with many large exits are generally either located in an industrial area or being a main street where vehicles from parking lots for large blocks are led directly on to the main street. Streets with many large exits exist in all the studied cities except Danderyd (being a smaller urbanisation with many villas).

Streets with many small exits are generally streets through residential areas with villas or small blocks. These streets are mostly found in Danderyd.

Many of the links (44%) don't have any exits, large or small. Most of the other links (41% of total) have between 1-5 exits. Figure 23 shows the distribution when normalised to number of exits (large and small) per km.

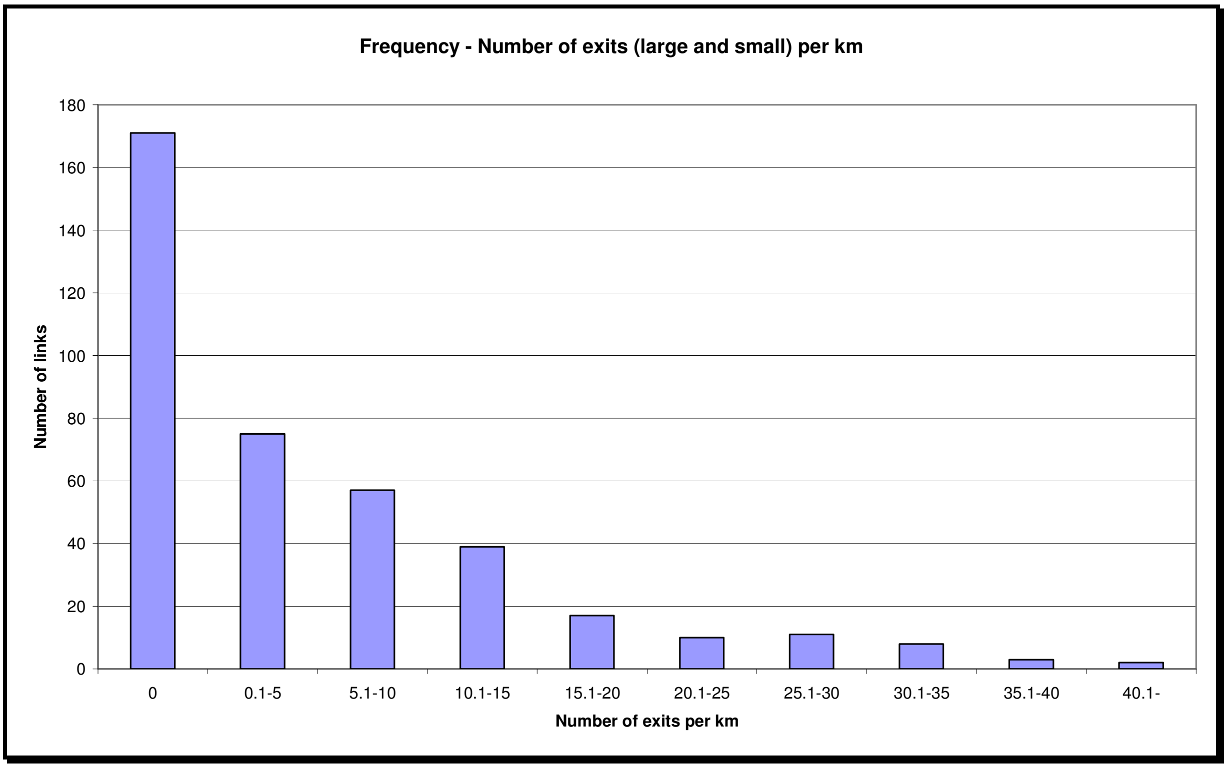


Figure 23 Frequency of links with different number of exits per km

5.4 Traffic characteristics

Flow of motorised vehicles (Flow)

Data on the flow of motorised vehicles on the links have been collected from the municipalities and is specified as AADT (Average Annual Daily Traffic). The vehicle flow ranges mainly between 2000 and 20000 vehicles per day, but as low flows as 800 v/day, and as large flows as 31000 v/day, are found among the links. Figure 24 shows the distribution of vehicle flows among the links.

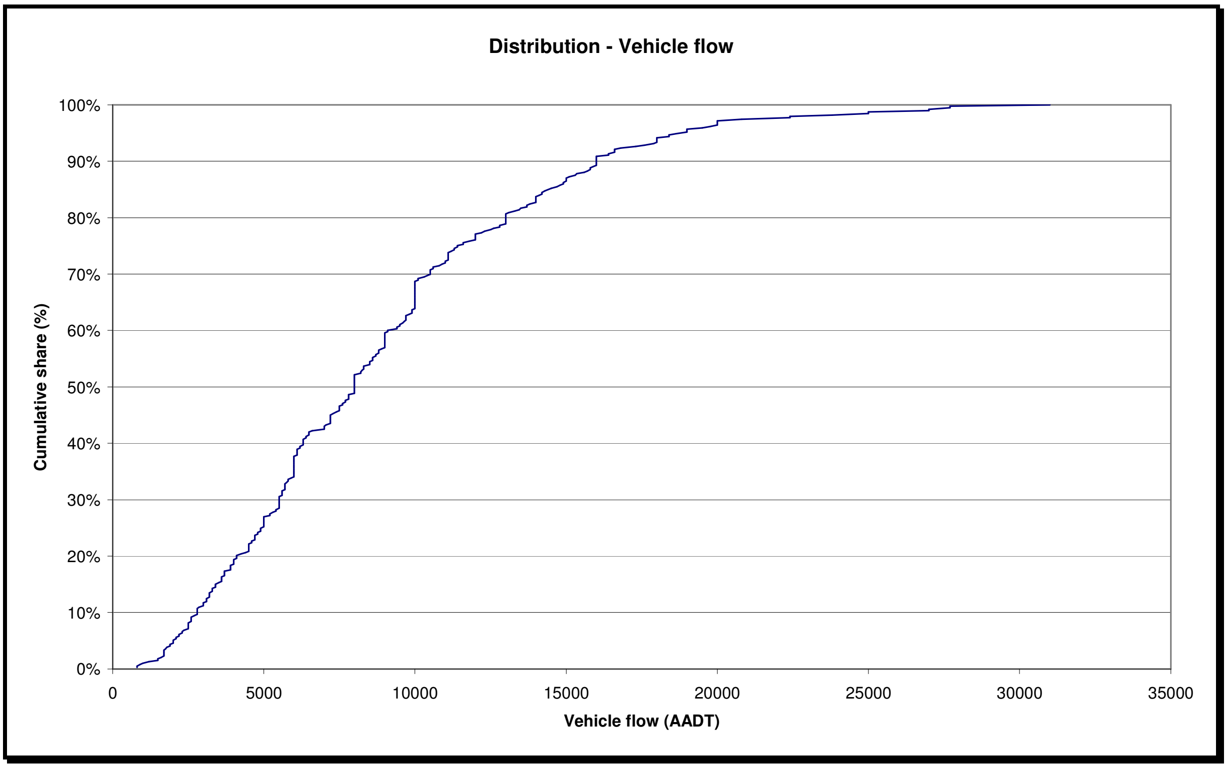


Figure 24 Distribution of Vehicle flows among the links

Vehicle speeds (Speed) and Standard deviation of speeds (SDev)

The vehicle speeds have been measured with handheld radar during the field studies. The arithmetic mean of the measured speeds has been used to describe the speed level. In addition the standard deviation has been calculated and used as an estimate of the dispersion of speeds. The average speeds range mainly between 30 and 60 km/h, but as low average speeds as 18 km/h occur, as well as average speeds up to 72 km/h (Figure 25). The links are normally distributed according to mean speed.

The standard deviation of the speeds at each link range mainly between six and ten km/h (Figure 26) and are normally distributed over the links. Looking at speed distributions separately for each speed limit, it can be seen that the levels differ much between the different speed limits, except for the speed limits 30 km/h and recommended 30 km/h (Figure 27). However, average speeds also differ much between links with the same speed limit.

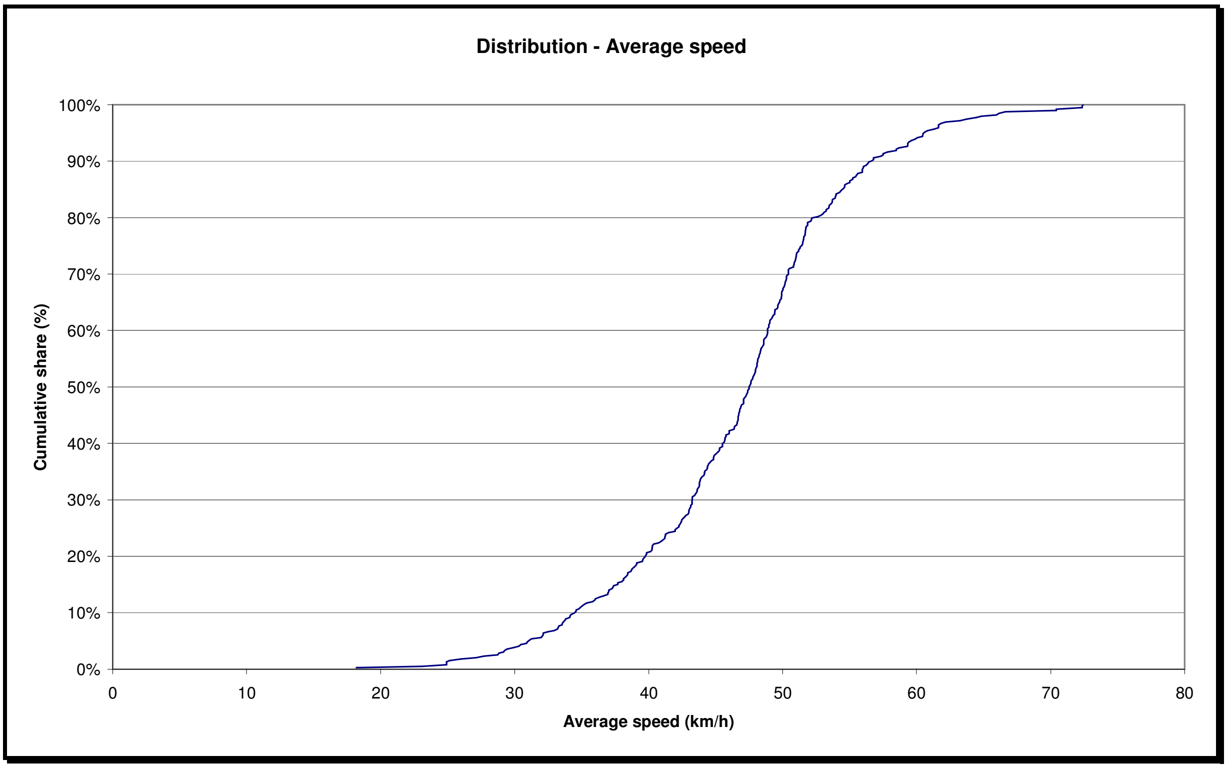


Figure 25 Distribution of Average speeds among the links

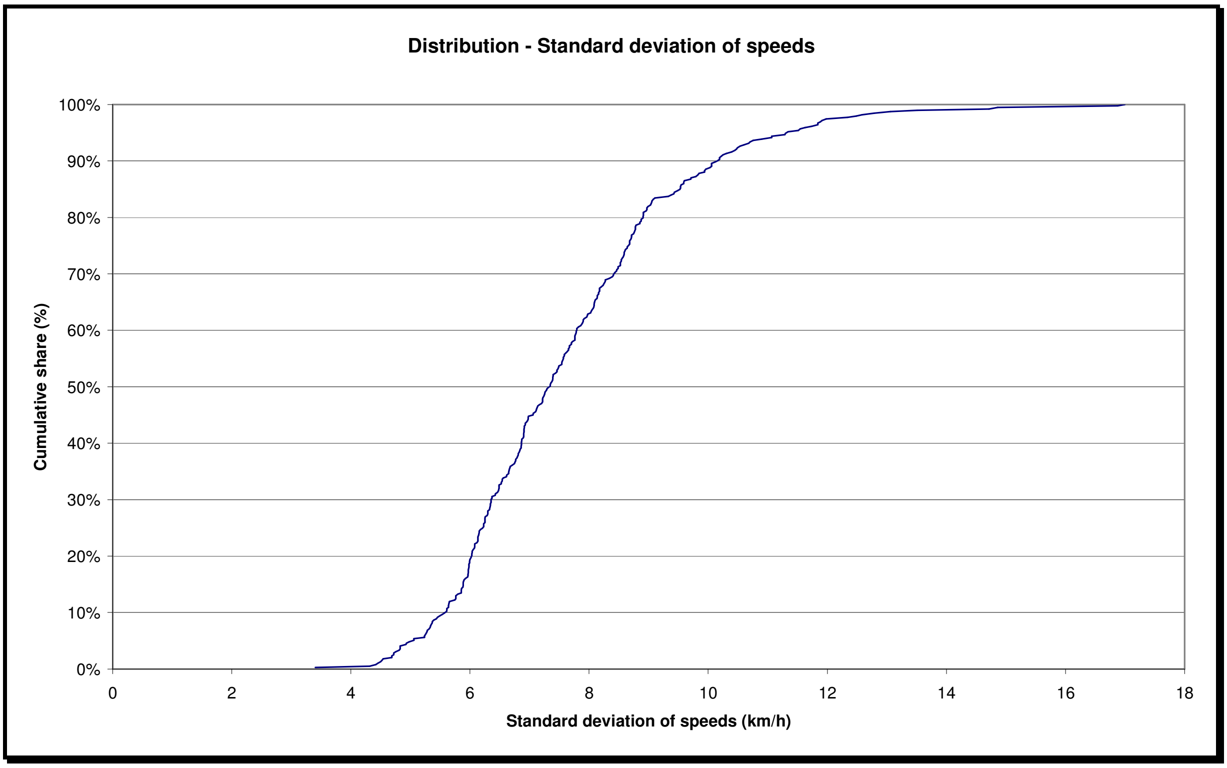


Figure 26 Distribution of Standard deviation of speeds among the links

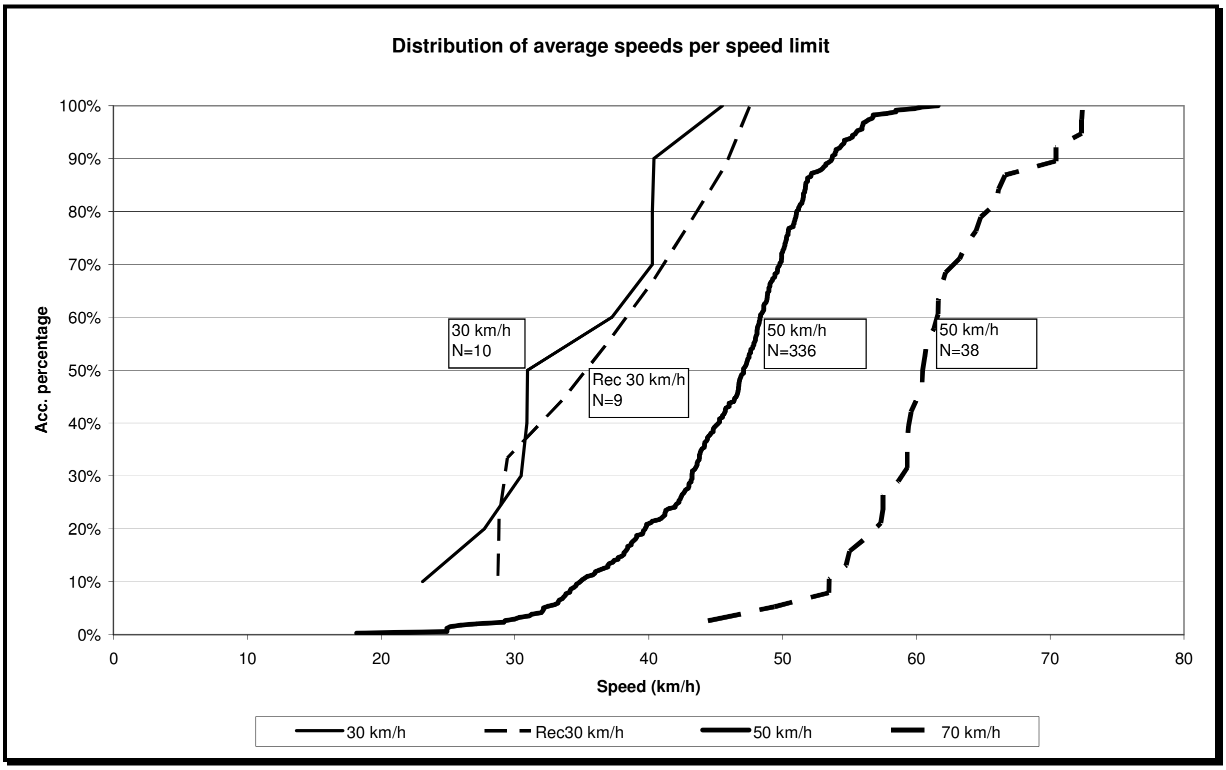
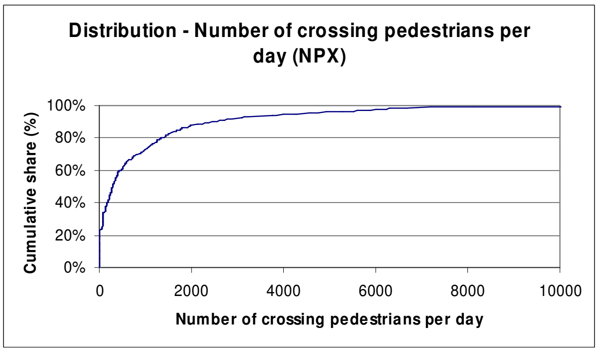
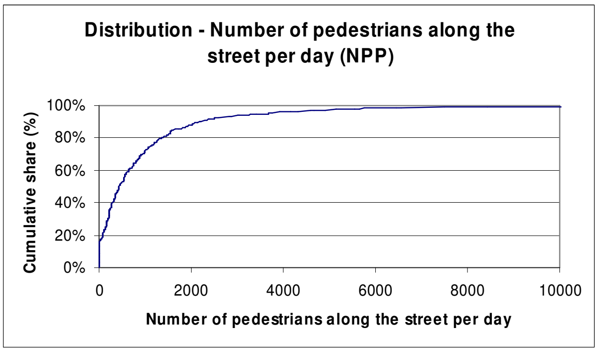


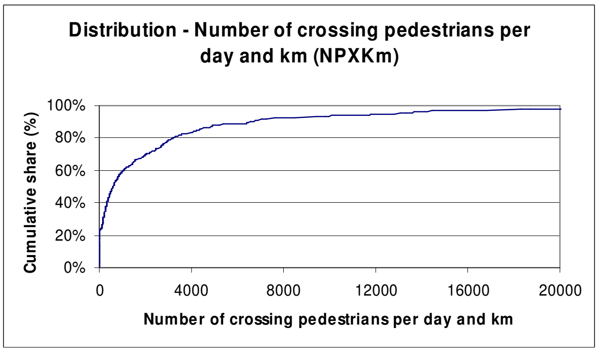
Figure 27 Distribution of average speeds on links for different speed limits

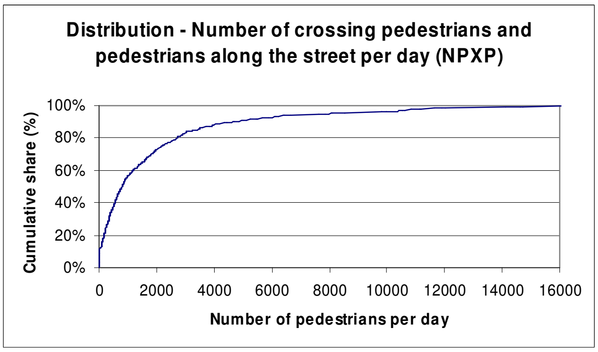
Number of vulnerable road users (NPX, NPXKm, NPP, NPXP, GrNCXP, NCX, NCXKm, NCP, NCXP, GrNPXP)

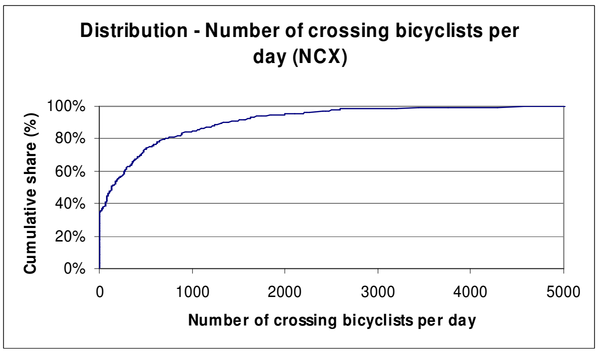
The variables for exposure of vulnerable road users are based on the counts performed in the field studies, normalised to be comparable with each other. In the field studies the numbers of pedestrians and bicyclists crossing (NPX & NCX) and moving along the link (NPP & NCP) have been counted. These counts have then been adjusted with regard to time variation to obtain estimates of daily traffic. To standardise the number of crossing road users, these numbers have also been related to the link length as number of road users crossing per day and km (NPXKm & NCXKm). The numbers of road users crossing and moving along the link have also been combined into aggregate estimates (NPXP & NCXP). This has been done to obtain a single variable for respective road user type, describing the exposure of that type of road user on the link. The aggregate variables have also been classified into categorical variables (GrNPXP & GrNCXP) where links with more than 800 road users per day have been categorised with 'High' exposure and links with fewer than 100 road users per day have been categorised with 'Low' exposure.

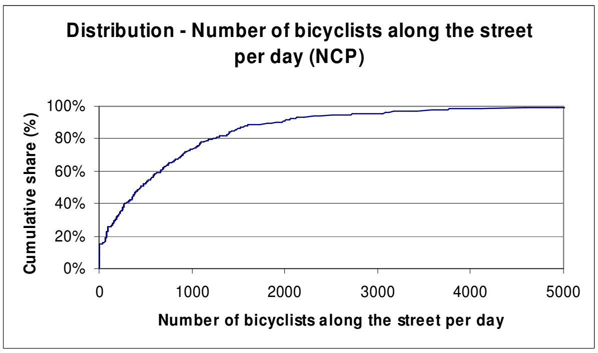


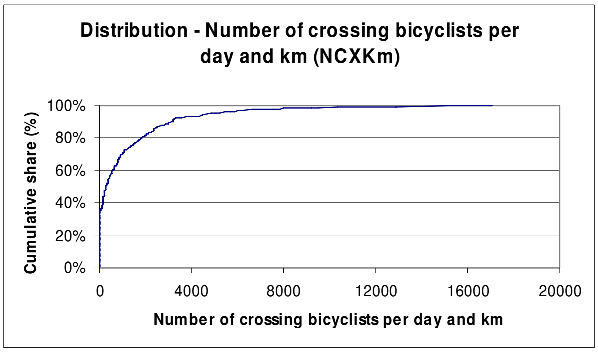












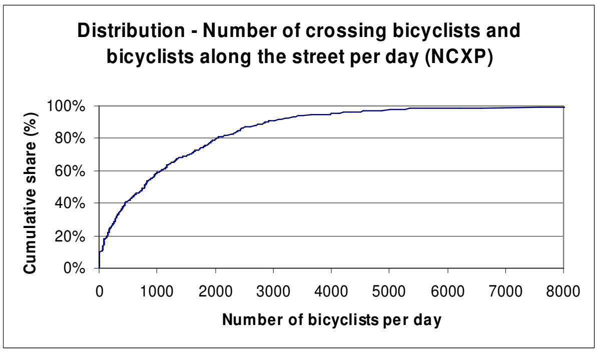


Figure 28 a-h Distribution of variables describing VRU exposure

VRU crossing strategies

The way vulnerable road users cross the street have been registered in form of crossing strategies. The different strategies used are *No crossing*, *Grade separated*, *At signalised crossings*, *At zebra crossings* and *Freely*.

The two most common crossing strategies are Zebra crossing (40%) and Freely (35%) (Figure 29).

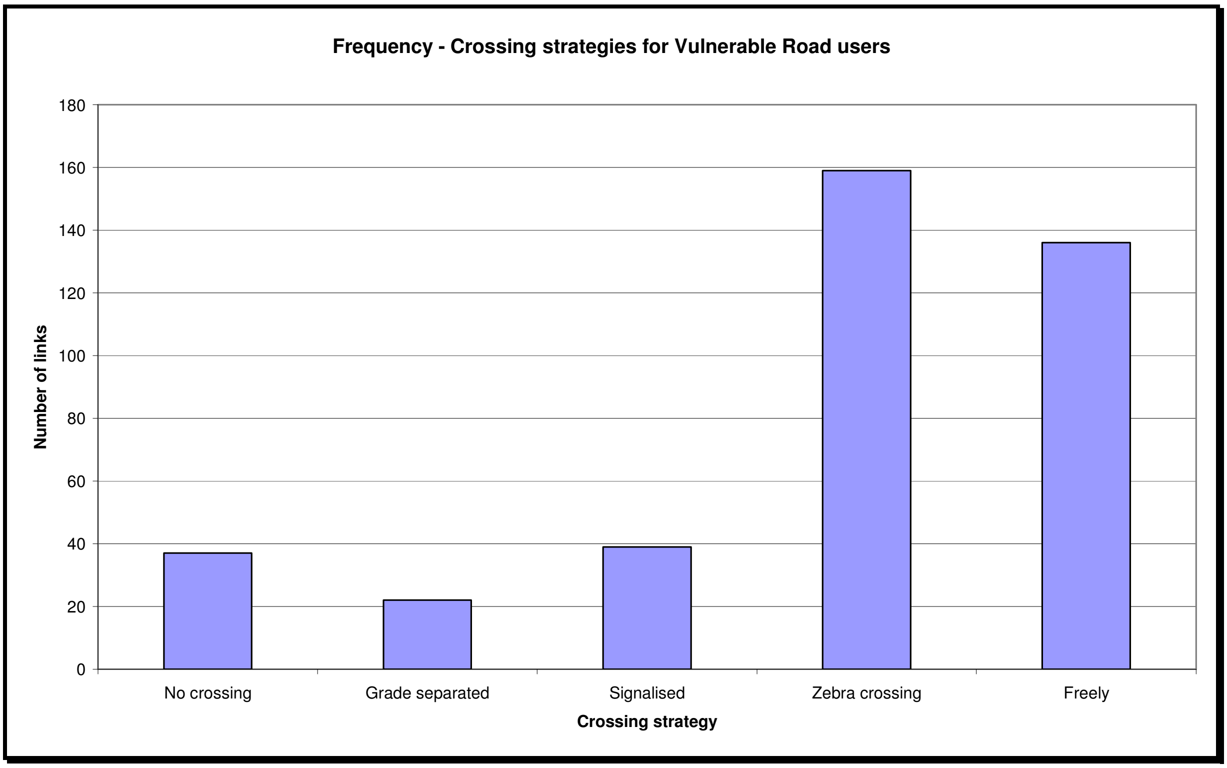


Figure 29 Frequency of crossing strategies

Separation of bicyclists travelling along the link (SepC)

The second variable describing how VRUs are mixed with motorised traffic describes whether bicyclists travelling along the link are separated from, or integrated with, motorised traffic. The variable assumes the following values:

X Biking along the street not appropriate.

Cb Biking is done on a separate path parallel to the street.

Cf Bikers are directed to the street but have a separate lane.

B Bikers are integrated with the motorised traffic

Only one link among the 393 was categorised as Cf, this link has therefore in the modelling been transferred to the B-group. The most common alternative is Cb (67%), separation of bikers to a separate path parallel to the street (Figure 30).

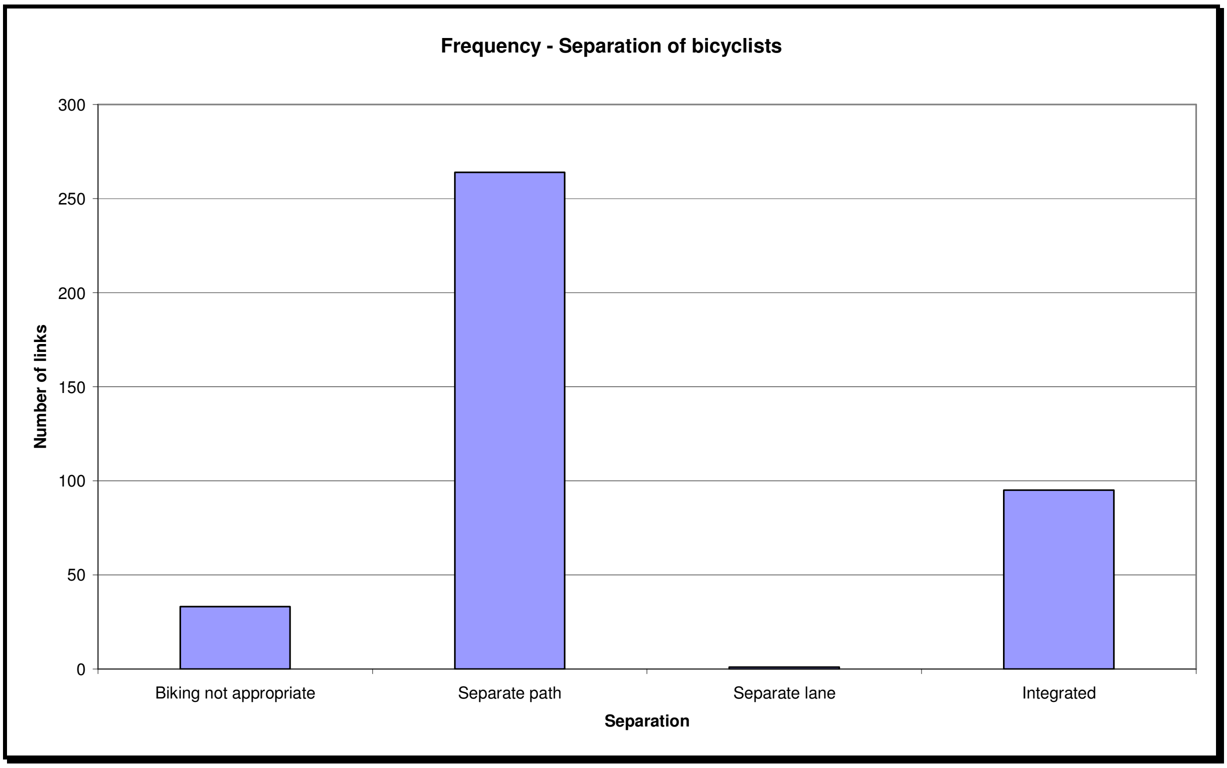


Figure 30 Frequency of bicyclist separation forms

5.5 Accident data

The number of police reported accidents in traffic is considerably lower than the actual number of accidents. This is due to the fact that the official statistics are based on police reported number of accidents and not all accidents are police reported. An estimate is that only 40% of all injury accidents in road traffic are reported to the police (Englund et al 1998). There are also differences between municipalities in regards to the degree of under reporting.

Even though the Swedish political focus in traffic safety nowadays is on the number of severely and fatally injured, data used here also includes accidents with only slight injuries. The reason for including slight injuries is that the severe injuries and fatalities are very rare and therefore too scarce to construct models from.

5.5.1 Injury definitions

Number of injured are separated for people in motorised vehicles, bicyclists and pedestrians. The data also include information on the degree of injury, defining severity in categories: fatality, severe injury and slight injury. The distinction between severe and slight injury is made by the police and mainly reflects whether the injured is likely to be hospitalised or not. The formal distinction is:

*"**Fatality:** deceased within 30 days as a consequence of a traffic accident."*

*"**Severe injury:** the person has as a consequence of a traffic accident received fracture, crush, tear, serious cut, concussion or internal injury. In addition as a severe personal injury is also included any other injury expected to lead to hospitalisation.*

***Slight injury:** other personal injury as a consequence of a traffic accident."*

(Vägverket 2001b)

5.5.2 Accident data for the six municipalities

The compilation of accident data from the national accident data base (VITS) has been carried out by VTI. The data consist of all police reported injury accidents for the years studied. For Uppsala, Katrineholm and Danderyd accident data has been compiled for the years 1994-2001 and for Linköping, Norrköping and Västerås the years 1998-2002. For modelling purposes five years of accident data has been used for each municipality. For the first three municipalities this meant excluding three years of accident data. The years 1994-1996 were chosen to be excluded in order for the remaining five years to be as comparable as possible to the other five years. The remaining three years of accident data for Uppsala, Katrineholm and Danderyd have been used to validate the developed models and compare their predictive ability with that of the existing ones.

Table 8 and Table 9 summarise the number of accident and injured per municipality. The number of accidents and injuries is the largest for Vehicle-Vehicle accidents and smallest for Vehicle-Single accidents. To be able to construct sound models, approximately 200 accidents are needed for each separate accident type (Kulmala 2002). This criterion is satisfied for Vehicle-Vehicle and Bicycle accidents, but not for Pedestrian and Vehicle-Single accidents. This calls for extra attention when constructing models for the P- and VS-accidents.

Table 8 Number of accidents per accident type and municipality, five years of accident data

Accident type	Bicycle	Pedestrian	Vehicle-Vehicle	Vehicle-single
Uppsala	92	35	129	52
Katrineholm	23	7	20	2
Danderyd	9	4	12	3
Linköping	107	38	185	15
Norrköping	70	36	100	11
Västerås	42	10	54	5
Total	**343**	**130**	**500**	**88**

Table 9 Number of injured road users per accident type and municipality, five years of injury data

Accident type	Bicycle			Pedestrian			Vehicle-Vehicle			Vehicle-Single		
Injury severity	F	S	L	F	S	L	F	S	L	F	S	L
Uppsala	0	47	50	1	15	20	0	47	163	0	23	59
Katrineholm	1	3	22	0	3	4	0	3	22	0	1	1
Danderyd	0	3	11	1	1	2	0	2	15	0	0	3
Linköping	0	25	88	1	13	28	0	26	307	0	5	10
Norrköping	0	3	70	1	6	31	0	9	164	0	2	9
Västerås	2	4	37	0	5	6	1	7	75	0	2	3
Total	**3**	**85**	**278**	**4**	**43**	**91**	**1**	**94**	**746**	**0**	**33**	**85**

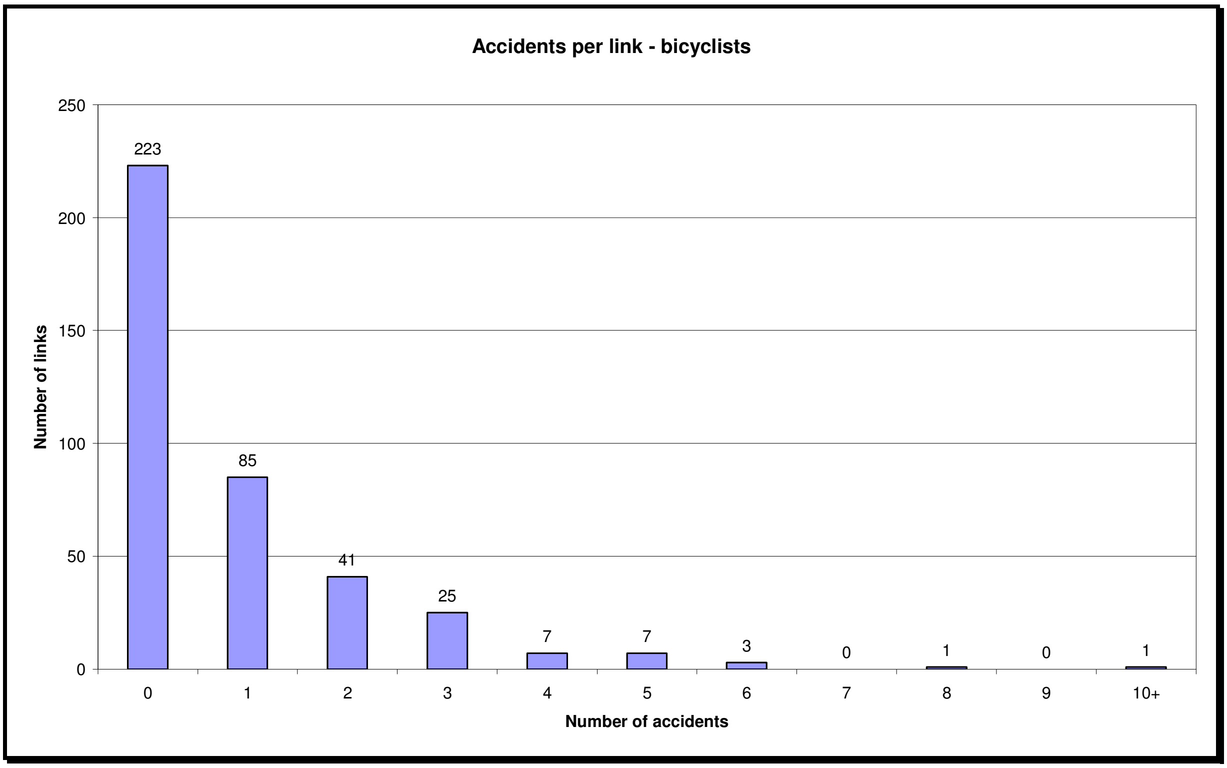


Figure 31 Number of accidents per link - bicyclist accidents

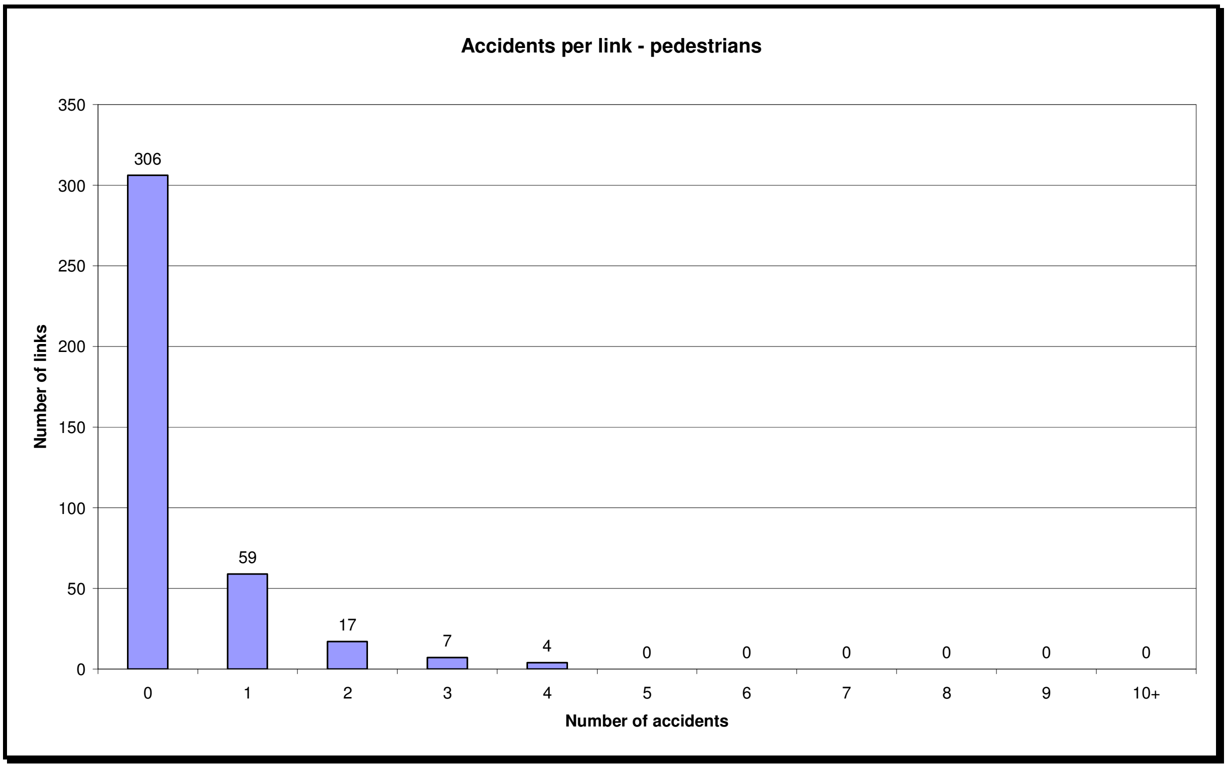


Figure 32 Number of accidents per link - pedestrian accidents

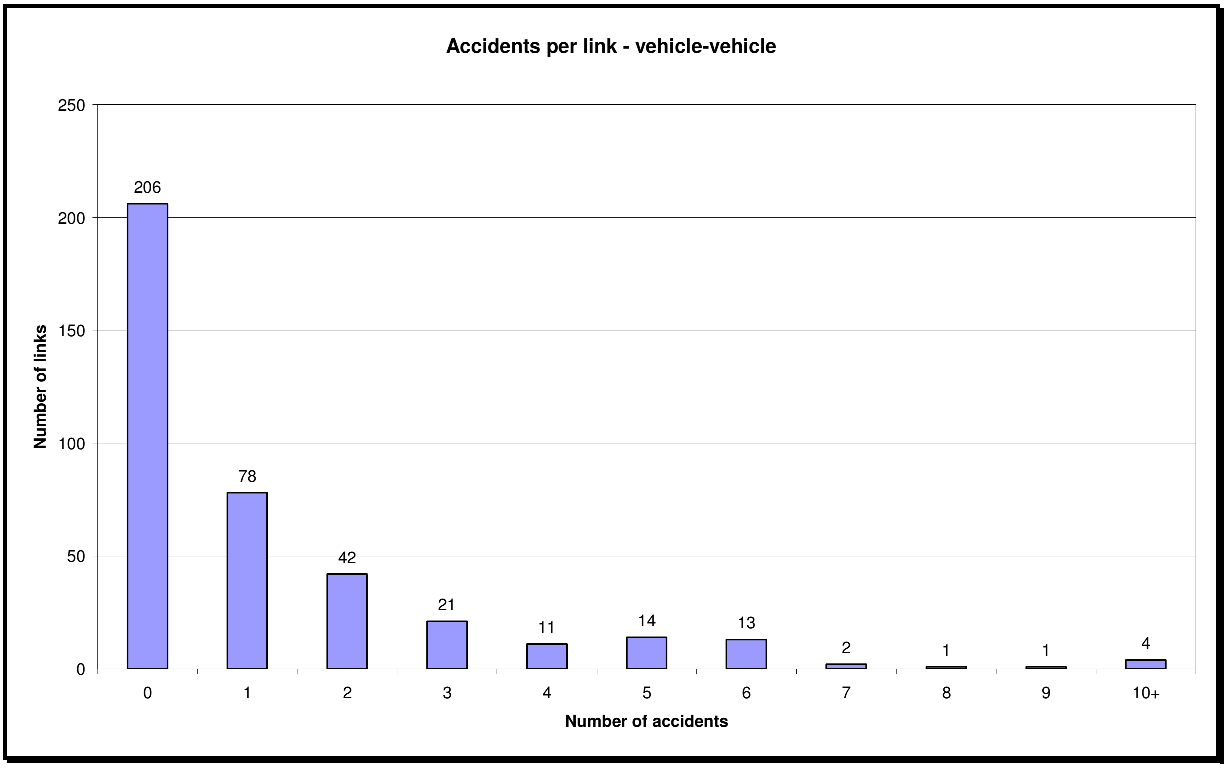


Figure 33 Number of accidents per link - vehicle-vehicle accidents

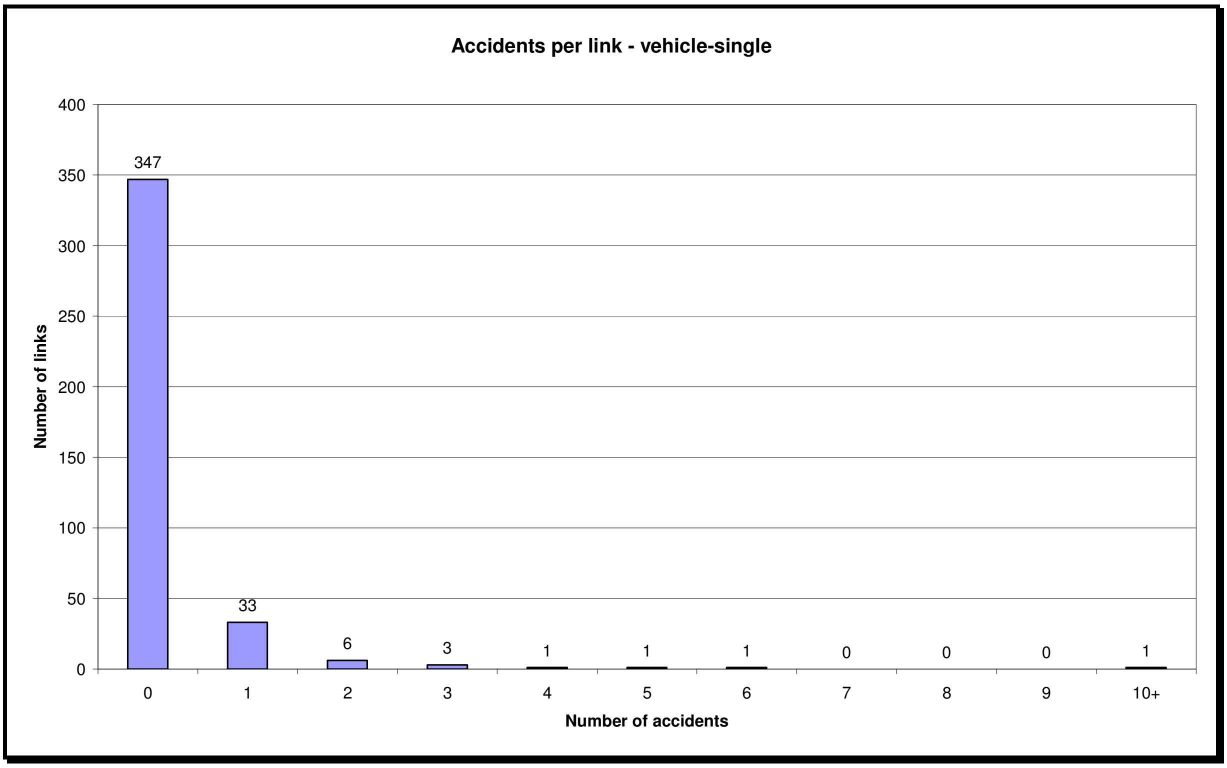


Figure 34 Number of accidents per link - vehicle-single accidents

6 Results and analysis

This chapter is divided into the following primary parts according to the hypothesis to be verified:

- Accident models, the main models developed (6.2)
- Test of actual speed vs. speed limit for explaining accident variation (6.3)
- Test of the contribution of estimates of exposure for vulnerable road users (6.4)
- Test of the contribution of dividing vehicle accidents into single vehicle and multiple vehicle accidents (6.5)

6.1 General modelling results

6.1.1 Correlation between variables

In the modelling 31 variables have been tested for inclusion in the models. Sixteen of them are categorical variables and fifteen are covariates. For easy reference throughout the chapter an overview of the variables is provided in Appendix 1. Many of the variables can be expected to correlate with each other. To be able to take correlations into account when selecting variables for the models the correlations between variables have been studied. Table 10 shows a correlation matrix for the covariates.

The covariates all describe either Vehicle flow, VRU exposure, Vehicle speeds or Intersection density. The covariates correlate to a large extent with the other of the same type, for example number of bicyclists correlate with number of pedestrians and number of 3-arm intersections per km correlate (negatively) with number of 4-arm intersections per km. There are also more or less strong correlations between covariates from different groups:

- Vehicle flow correlates slightly with most other variables, but no strong correlations are found.

- The different estimates of vulnerable road users' exposure correlate strongly between themselves. They are also (negatively) correlated with average speeds. There is a fairly strong correlation between VRU exposure and the number of 4-arm intersections but not between VRU exposure and 3-arm intersections; where there are many VRUs there are often 4-arm intersections, while 3-arm intersections are more correlated with vehicle flow.

- Average speeds are strongly negatively correlated with most estimates of exposure for the vulnerable road users and number of intersections. Average speeds have however a relatively weak correlation with vehicle flows. Dispersion of speeds hardly correlates at all with any other variables.

It is more difficult to give a brief overview over categorical variables, for Landuse there is however some apparent correlation with several other categorical variables. Links located in areas with the land use commercial tend to have low speed limits and a high presence of vulnerable road users. Industrial streets are on the other hand categorised by having high speed limits (50 or 70 km/h) and a different composition and exposure of vulnerable road users (no children or older people and having the flows concentrated to morning and late afternoon). For links without any buildings along them there is never any parking, the visibility is often good, the speed limit is high and the street is often a thoroughfare in the outer parts of the city. For links with residential or institutional land use along the link, other variables don't show any clear pattern, apart from the presence of vulnerable road users which often is high close to institutional buildings.

Table 10 Correlation matrix for covariates used in the modelling, bold figures indicate very strong correlations (>±0.5) and grey figures indicate weak correlations (<±0.1)

Correlation Matrix															
	Flow	**NPX**	**NPXKm**	**NPP**	**NPXP**	**NCX**	**NCXKm**	**NCP**	**NCXP**	**Speed**	**DSpeed**	**X3Km**	**X4Km**	**X34Km**	**DX**
Flow	1														
NPX	0.11	1													
NPXKm	0.07	**0.95**	1												
NPP	0.16	**0.71**	**0.69**	1											
NPXP	0.14	**0.95**	**0.91**	**0.89**	1										
NCX	0.10	0.49	0.39	0.42	0.50	1									
NCXKm	0.07	0.47	0.48	0.41	0.48	**0.88**	1								
NCP	0.17	0.23	0.18	0.46	0.35	0.40	0.32	1							
NCXP	0.16	0.40	0.32	**0.53**	0.49	**0.76**	**0.65**	**0.89**	1						
Speed	0.08	-0.41	-0.45	-0.47	-0.47	-0.32	-0.39	-0.16	-0.27	1					
DSpeed	-0.02	-0.07	-0.07	-0.07	-0.07	-0.02	0.02	-0.09	-0.07	-0.01	1				
X3Km	-0.20	-0.06	-0.05	-0.02	-0.05	-0.11	-0.11	-0.02	-0.07	-0.11	0.01	1			
X4Km	-0.01	0.22	0.23	0.21	0.23	0.25	0.28	0.03	0.14	-0.34	-0.05	-0.21	1		
X34Km	-0.20	0.07	0.09	0.10	0.09	0.04	0.06	0.00	0.02	-0.31	-0.02	**0.82**	0.39	1	
DX	0.21	-0.13	-0.15	-0.17	-0.16	-0.13	-0.17	-0.08	-0.12	0.50	0.02	-0.49	-0.30	**-0.64**	1

Flow	Vehicle flow (AADT)
NPX	Number of pedestrians crossing
NPXKm	Number of pedestrians crossing per km
NPP	Number of pedestrians walking parallell with the street
NPXP	Number of pedestrians crossing and parallell
NCX	Number of bicyclists crossing
NCXKm	Number of bicyclists crossing per km
NCP	Number of bicyclists walking parallell with the street
NCXP	Number of bicyclists crossing and parallell
Speed	Average vehicle speed
DSpeed	Standard deviation of speeds
X3Km	Number of three arm intersections per km
X4Km	Number of four arm intersections per km
X34Km	Number of intersections per km
DX	Average distance between intersections

6.1.2 Scale factors

Because of the over dispersion of traffic accidents compared to a pure Poisson distribution, the test statistics are underestimated and produce too highly significant results. A Quasi-Poisson model is used to account for this (Maher & Summersgill 1996). The Quasi-Poisson model uses a scale factor to adjust the test statistics by inflating the variance, which otherwise should be equal to the mean for Poisson distributed variables.

The first step in calculating the scale factors is to develop models with a good fit for each accident type. These models are then used to calculate Pearson residuals which are summed for the links. The scale factor has been calculated as the sum of the Pearson residuals of the saturated models divided by the remaining degrees of freedom of the well fitted models (Table 11).

The scale factor is used both for the χ^2 tests when checking for significance of added variables and the t-tests when checking significance of individual variables. However, when adjusting the t-statistic, the square root of the scale factor is used since the t-statistic is related to the standard deviation instead of the variance.

Table 11 Scale factors to compensate for over dispersion

	Accidents			Injuries		
	$\Sigma\chi^2$	df	$\Sigma\chi^2$/df	$\Sigma\chi^2$	df	$\Sigma\chi^2$/df
Bicycle	687	379	1.81	754	379	1.99
Pedestrian	441	380	1.16	458	380	1.20
Vehicle-Vehicle	779	381	2.04	1613	381	4.23
Single vehicle	426	380	1.12	556	380	1.46

$\Sigma\chi^2$ = Sum of Pearson residuals, df = remaining degrees of freedom

6.1.3 Outliers

In the process of modelling several links were identified as more strongly influencing the models generated than other links. These links were studied to see whether there could be any unduly influence. Four links were identified as having a character very apart from other links; these four links were part of the European highway E4. The E4 enters Uppsala, and gets a shift in character from a strict highway, through having signalised intersections, to looking almost like an ordinary four lane urban street (Figure 35 a-c). Although only the three outer links (U14, U17 and U20) have a clear rural character, also the fourth (U23) was excluded from the modelling. The reason for this was a combination of the risk for that the behaviour of drivers coming from the outer parts of E4 would be strongly influenced by their behaviour on previous links and the fact that U23 is having an exceptional accident history. Out of the 88 single vehicle accidents on the 393 links, 19 occurred on link U23!

The models presented in the following sections are based on the remaining 389 links.

Figure 35 a-c European Highway E4 in outer part of Uppsala coming from Rural area (upper left, U14), shifting slightly to a more urban form with signalised intersections (upper right, U17) and being transformed to an urban four lane street (lower left, U23)

6.2 Accident models

This subchapter describes the first models developed, from here on referred to as the basic models. There are both models where all variables have been added in the order they have been found to explain the variation in accidents, and models where the variables describing exposure have been included in the models from the beginning with preset parameter values. For vehicle models the variable 'Flow' has been included with the exponent set to 1, for the pedestrian models the variables 'Flow' and 'NPXP' have been included with the exponents set to 0.5 and for bicycle models the variables 'Flow' and 'NCXP' have been included with the exponents set to 0.5.

Variables describing exposure take on a dominant role in the explanation of the systematic variation of accidents.

When deciding on which variable to add to the model in each step, the variable which contribute most to the explanation of the models has been chosen unless it:

1. clearly has only correlative connection without causality, for instance VRU exposure variables when modelling vehicle accidents
2. is assigned parameter values that go strongly against known safety effects
3. has a strong correlation with earlier added variables

The column marked grey in each table over the stepwise addition mark the last included variable.

All accident models are of the form:

Number of accidents per 5 year = A x Link length x dummy variables x covariates$^{\text{Exponent}}$

where A is a constant and dummy variables are the "effects" of categorical variables. The dummy variables are actually estimated as a β value in e^{β}, but to allow easier interpretation of the results the expression has been presented recalculated as Dummy=e^{β}.

In the tables for stepwise addition of variables the deviance for the null model differs between the models without preset exposure parameters and those with. The reason for this is that the null model in the preset models includes the exposure variables in the case of stepwise addition test of significance. When checking for significance the drop in deviance should be calculated as the difference between two models with the only difference being the variable added. In the case of the first variable added in the preset models, the preset variables are already included before the first variable is added and so the 'null model' in this case should include the preset variables. When calculating the variation explained the null model is defined as the model without preset variables as the focus is then on what both preset variables and optimised variables explain together.

A residual analysis has been carried out for all the models, plotting the standardised residuals against each of the independent variables included in the model (Appendix 2). No structure was found in the plots, thus not giving rise to doubt of the adequacy of the models.

6.2.1 Bicyclist models

Most of the variation in bicyclist accidents is described by bicyclist exposure (NCXP). Vehicle flow is also a strong explanatory variable. Other variables included in the model for bicyclist accidents are Landuse, (Street) Function and Visibility (Table 12). The categories of Landuse have been merged, since the parameters for Residential, Institutional and Commercial did not show any significant differences between them, neither did Industrial and None, and so they were merged together into another joint category. (Table 13)

Table 12 Stepwise addition of variables, Model: Bicycle, no preset parameters

Model: Bicycle, no preset						
Var. added	0	1	2	3	4	5
Variable	Null model	NCXP	Flow	Landuse	Func	Vis
Dev	745	585	556	526	510	501
df	388	387	386	385	383	381
dDev		160.0	28.6	30.4	15.9	9.6
ddf		1	1	1	2	2
QP Scale		1.814	1.814	1.814	1.814	1.814
Scaled dDev		88.2	15.8	16.8	8.8	5.3
Chi2-prob.		0.000	0.000	0.000	0.013	0.072

Table 13 Model for bicycle accidents, no preset parameters

Model: Bicycle, no preset				
Deviance	df	Perc.exp.		
501	381	60%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	2.16E-04		-5.93	0.000
NCXP		0.35	5.50	0.000
Flow		0.76	4.93	0.000
Landuse_ABC	1			
Landuse_IX	0.60		-2.54	0.011
Func_GIF	1			
Func_T	1.36		1.86	0.064
Func_C	1.94		2.27	0.024
Vis_Good	0.71		-2.03	0.043
Vis_Medium	1			
Vis_Poor	1.62		0.96	0.335

Road user flows have in other studies been shown to have exponents close to 0.5. Here they differ somewhat with vehicle flow being assigned the exponent 0.76 and NCXP the exponent 0.35. When the exponents for the two exposure variables NCXP and Flow are preset to 0.5, the other variables selected don't differ much (Table 14), probably due to the relatively small forcing needed. The presetting however makes the fit of the model markedly worse. With preset parameters for the exposure only 42% of the systematic variation is explained (Table 15). For the model without presetting, the exponent for NCXP is significantly different on the 0.05 level and the vehicle flow on the 0.10 level. It seems the presetting of exposure exponents to 0.5 is not appropriate for the bicycle model.

Table 14 Stepwise addition of variables, Model: Bicycle, exposure preset to $Flow^{0.5}*NCXP^{0.5}$

Model: Bicycle, exposure preset					
Var. added	0	1	2	3	4
Variable	Null model	Vis	Landuse	SepC	SL
Dev	616	588	576	563	553
df	388	386	383	381	378
dDev		28.0	12.1	13.1	10.0
ddf		2	3	2	3
QP Scale		1.814	1.814	1.814	1.814
adj. dDev		15.5	6.6	7.2	5.5
Chi2-prob.		0.000	0.084	0.027	0.137

Table 15 Model for bicycle accidents, exposure preset to $Flow^{0.5}*NCXP^{0.5}$

Model: Bicycle, exposure preset				
Deviance	**df**	**Perc.exp.**		
563	382	42%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	8.61E-04		-62.00	0.000
NCXP		0.5		
Flow		0.5		
Vis_Good	0.66		-2.46	0.014
Vis_Medium	1			
Vis_Poor	1.86		1.28	0.203
Landuse_AB	1			
Landuse_IX	0.66		-2.13	0.034
Landuse_C	1.36		1.58	0.115
SepC_Mixed	0.75		-1.46	0.145
SepC_Sep	1			
SepC_X	2.10		2.34	0.020

6.2.2 Pedestrian models

As for the bicycle models, the exposure variables turned out being strong explaining variables for the pedestrian models as well (Table 16). Landuse is also a strong explaining variable. In the pedestrian models the categories Residential, Institutional and No land use have been merged as they were assigned almost the same parameter values in the first modelling rounds. Other variables used in the models are Street function (Func), presence of dedicated VRU crossings (VRUX) and, for the non-preset model, the number of intersections per km (X34Km).

Parameter values for VRUX indicate that the presence of a dedicated crossing for vulnerable road users is linked to an increasing number of accidents with vulnerable road users. It should be noted that the crossings here are not grade separated and often don't have any safety measures close by. The effect of the VRUX might either arise from a false feeling of safety in the case of the vulnerable road users, or might be influenced by a correlation with other variables. Meta studies show a best estimate of the effect of a non signalised VRU crossing to increase the amount of pedestrian injury accidents by 19-39% (Elvik et al. 1997). The "effect" in these models is somewhat higher (≈50%). (Table 17)

The pedestrian model without preset parameters explains 93% of the systematic variation in pedestrian accidents.

Table 16 Stepwise addition of variables, Model: Pedestrian, no preset parameters

Model: Pedestrian, no preset							
Var. added	0	1	2	3	4	5	6
Variable	Null model	NPXP	Landuse	Flow	Func	VRUX	X34Km
Dev	439	317	299	286	276	272	268
df	388	387	385	384	382	381	380
dDev		121.5	18.5	12.4	9.8	4.2	4.0
ddf		1	2	1	2	1	1
QP Scale		1.161	1.161	1.161	1.161	1.161	1.161
adj. dDev		104.7	16.0	10.6	8.4	3.7	3.5
Chi2-prob.		0.000	0.000	0.001	0.015	0.056	0.062

Table 17 Model for pedestrian accidents, no preset parameters

Model: Pedestrian, no preset				
Deviance	**df**	**Perc.exp.**		
268	380	93%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	1.60E-05		-5.50	0.000
NPXP		0.38	4.16	0.000
Landuse_ABX	1			
Landuse_C	1.28		1.02	0.310
Landuse_I	0.28		-2.52	0.012
Flow		0.83	3.84	0.000
Func_GIF	1			
Func_T	1.19		0.73	0.466
Func_C	2.47		2.85	0.005
VRUX_No	1			
VRUX_Yes	1.54		1.99	0.047
X34Km		0.23	1.80	0.072

When the exponents for exposure are preset, the number of significant variables in the model decrease and so does the percentage of explained variation, but a level of 83 % is still good. When the exposure is preset the parameter values for street functions GIF and Tangential are very close and so they are merged to one value.

Table 18 Stepwise addition of variables, Model: Pedestrian, exposure preset to $Flow^{0.5}*NPXP^{0.5}$

Model: Pedestrian, exposure preset					
Var. added	0	1	2	3	4
Variable	Null model	Landuse	Func	VRUX	XForm
Dev	321	298	290	286	281
df	388	386	385	384	380
dDev		23.0	8.0	4.0	5.3
ddf		2	1	1	4
QP Scale		1.161	1.161	1.161	1.161
adj. dDev		19.8	6.9	3.5	4.6
Chi2-prob.		0.000	0.009	0.062	0.332

Table 19 Model for pedestrian accidents, exposure preset to $Flow^{0.5}*NPXP^{0.5}$

Model: Pedestrian, exposure preset				
Deviance	**df**	**Perc.exp.**		
286	384	83%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	1.99E-04		-62.92	0.000
NPXP		0.5		
Flow		0.5		
Landuse_ABX	1			
Landuse_C	1.22		0.91	0.361
Landuse_I	0.27		-2.64	0.009
Func_GIF_T	1			
Func_C	1.94		2.39	0.017
VRUX_No	1			
VRUX_Yes	1.49		1.91	0.057

6.2.3 Vehicle-Vehicle models

For the multiple vehicle accidents the vehicle flow is naturally the strongest explaining variable. Other variables used are quite similar to the ones used in EVA, the existing Swedish accident models (Vägverket 2001b). The only new variable is the density of intersections that is identified as the second strongest variable in the models. In the EVA models, there are also the variables Link Environment (LEnv) and Lanes. For the data used here, the number of Lanes is a fairly good explanatory variable but not good enough to be picked for the models. The variable LEnv does not add much explanation at all.

The relation between vehicle-vehicle accidents and vehicle flow in the models is estimated as a flow exponent of 1.4. It is often assumed that the relationship is close to linear for vehicle accidents and vehicle flow (i.e. an exponent of 1.0), but in these models the single vehicle accidents are excluded, and so only accidents involving two or more vehicles remain. The links also include minor intersections and most of the accidents actually occur in these. For intersections the EVA models actually have an exponential relation between incoming vehicle flow and number of accidents, the exponents being 1.45 for yield and stop, and 1.2 for signalised intersections and roundabouts. The vehicle flow on the link might then have the effect of describing both the flow on the link but also indirectly describing the secondary flows incoming in the minor intersections. A high flow on the main street indicates an area with more traffic, and so there will be more incoming traffic from side streets and more interactions in the minor intersections.

Table 20 Stepwise addition of variables, Model: Vehicle-Vehicle, no preset parameters

Model: Vehicle-Vehicle, no preset						
Var. added	0	1	2	3	4	5
Variable	Null model	Flow	X34Km	Func	SL	SepV
Dev	930	718	672	652	628	621
df	388	387	386	384	381	379
dDev		212.5	45.6	20.5	23.7	7.2
ddf		1	1	2	3	2
QP Scale		1.161	1.161	1.161	1.161	1.161
adj. dDev		183.1	39.3	17.6	20.4	6.2
Chi2-prob.		0.000	0.000	0.000	0.000	0.046

Table 21 Model for vehicle-vehicle accidents, no preset parameters

Model: Vehicle-Vehicle, no preset				
Deviance	**df**	**Perc.exp.**		
628	381	61%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	4.00E-06		-9.11	0.000
Flow		1.40	9.70	0.000
X34Km		0.37	4.52	0.000
Func_GIF	1			
Func_T	1.42		2.43	0.016
Func_C	2.51		2.90	0.004
SL_ 30	0.11		-1.52	0.129
SL_Rek30	0.27		-1.81	0.072
SL_50	1			
SL_ 70	1.21		1.08	0.282

When the exponent for vehicle flow is preset to 1, the degree of explanation hardly decreases at all. Even for models only containing the flow (column 1 in Table 20 and column 0 in Table 22) the degree of explanation does not differ much. However, in the case of the freely estimated model, the exponent for flow in a flow only model is 1.18, when intersection density is added the exponent increases to 1.35, indicating a correlation between intersection density and vehicle flow.

Table 22 Stepwise addition of variables, Model: Vehicle-Vehicle, exposure preset to flow[1]

Model: Vehicle-Vehicle, exposure preset					
Var. added	0	1	2	3	4
Variable	Null model	X34Km	SL	Func	SepV
Dev	722	686	660	645	631
df	388	387	384	382	380
dDev		35.5	26.4	15.4	13.6
ddf		1	3	2	2
QP Scale		1.430	1.430	1.430	1.430
adj. dDev		24.9	18.5	10.7	9.5
Chi2-prob.		0.000	0.000	0.005	0.009

Table 23 Model for vehicle-vehicle accidents, exposure preset to flow[1]

Model: Vehicle-Vehicle, exposure preset				
Deviance	**df**	**Perc.exp.**		
631	380	60%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	1.46E-04		-58.00	0.000
Flow		1		
X34Km		0.36	4.43	0.000
SL_ 30	0.11		-1.51	0.132
SL_Rek30	0.29		-1.70	0.091
SL_50	1			
SL_ 70	1.43		2.07	0.039
Func_GIF	1			
Func_T	1.35		2.13	0.034
Func_C	2.45		2.82	0.005
SepV_I	1			
SepV_B	1.72		2.17	0.031
SepV_S	1.33		1.90	0.058

6.2.4 Vehicle -Single models

The vehicle flow is the strongest explaining variable for single vehicle accidents. Among the categorical variables there are some which, when included in the models, make the models never to converge. These variables are: Visibility, Speed Limit, Parking and Function. When investigating the accident occurrence for these variables it can be seen that for each variable there is a category for which there are no accidents occurring, these are: Poor visibility, Speed Limit 30 km/h, double sided parking slots and City Function. All these alternatives indicate a highly urbanised and complex street environment with low speed levels, thus it is not strange that single vehicle accidents do not occur. It is however unfortunate that the models do not handle this as those very variables describe exactly what we are interested in: accident occurrence (or rather non occurrence).

The categorical variable Parking has been merged from having five different levels to only two: Parking Allowed and No Parking, after this it has been technically possible to include the variable Parking in the models.

Table 24Stepwise addition of variables, Model: Vehicle-Single, no preset parameters

Model: Vehicle-Single, no preset			
Var. added	0	1	2
Variable	Null model	Flow	Landuse
Dev	302	269	242
df	388	387	383
dDev		32.2	26.9
ddf		1	4
QP Scale		1.120	1.120
Scaled dDev		28.7	24.0
Chi2-prob.		0.000	0.000

Table 25 Model for vehicle-single accidents, no preset parameters

Model: Vehicle-Single, no preset				
Deviance	df	Perc.exp.		
243	384	57%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	1.30E-07		-5.51	0.000
Flow		1.67	5.37	0.000
Landuse_BC	1			
Landuse_X	0.27		-3.36	0.001
Landuse_I	0.20		-2.77	0.006
Landuse_A	0.32		-1.68	0.093

Early in the addition of the variables the categorical variable Landuse turned out to be a strong explaining variable, but with most peculiar parameter values (Table 25). For residential and CBD land use the single vehicle accident rates are much higher than for industrial, none and institutional land use. In short, the more complex environments Residential and CBD produce more single vehicle accidents than the less complex Industrial and None. It would be expected that the highest single vehicle accident rates would be found where there was no surrounding buildings. This was further investigated in the outlier analysis.

The outlier analysis of links that influenced the model results the most did not reveal considerable peculiarities. In addition to the outlier analysis, a closer look was done at the streets where there were more than one single vehicle accident. Eleven links had more than one single vehicle accident during the years 1997 through 2001; all these were concentrated to Linköping (2), Norrköping (3) and Uppsala (6). The links in Linköping and Norrköping all had two single vehicle accidents while the links in Uppsala had between three and six. The six links in Uppsala were studied even more closely. Three of the links where entrance roads with a character where there could be expected a relatively high number of single vehicle accidents. The other three, however, were consecutive links on a centrally located street, Vaksalagatan, together having 13 single vehicle accidents. This seemed very unlikely and so the original accident data was studied. It turned out that most of the accidents were actually accidents with multiple vehicles, but with only one vehicle recorded in the accident database. These accidents were then removed from the single vehicle accident data.

In the available data there was only detailed enough data to distinguish the faulty single vehicle accidents for Uppsala, Katrineholm and Danderyd. From the national data for the year 2001 (SIKA & SCB 2002) it can be shown that the single vehicle accidents make up 20-25% of the total number of vehicle accidents in the regions where the six cities are located. In the highly urbanised areas studied, the share of single vehicle accidents is probably even lower. Looking at the number of single vehicle accidents in the data used here they make up approximately 10% of the total number of vehicle accidents, except for Uppsala where they before identifying the 'misplaced' accidents made up almost 30%. The problem with lacking vehicle registration in the accident data is probably mostly a problem in the Uppsala data, but without more detailed accident data it is not possible to say for sure.

Because of the problems with identifying which accidents actually are single vehicle accidents, and the overall small number of them, the creation of separate models for single vehicle accidents was more or less made impossible. It seems like single vehicle injury accidents are so rare in urban areas that it becomes difficult to develop any models for them unless the data set is quite extensive.

6.3 Actual speeds vs. speed limit

One of the main hypotheses of this dissertation was to investigate if actual vehicle speeds are better for predicting accidents than speed limits are. This has been tested by fitting accident models to speeds and speed limits respectively and comparing the degree of explanation and the usability of the models. Besides the speed or speed limit, exposure variables have also been used when fitting the models.

6.3.1 Model fit

The first step in the comparison has been to develop models with speed respectively speed limit to the accident data for the four accident types. Besides speed or speed limit, flows of road users have also been used in the models. In the models for pedestrian and bicyclist accidents their respective flows have been used together with vehicle flow. For vehicle accidents only the vehicle flow has been used. In all cases both models with preset parameter values for exposure and freely estimated effects have been developed. In total sixteen models have been developed (4 accident types * 2 variables * 2 types), and these can be found in full in Appendices 3A and 3B. Examples of the models can be found in Table 26 and Table 27.

Speed limit has been used as a categorical variable where one speed limit has been chosen as the base level and the other speed limits are assigned Dummy values estimating their number of accidents in relation to that of the speed limit 50 km/h. Speed has been used as a covariate where the average speeds of the individual links have been logarithmated to achieve a model in the end where the model includes speed as a variable with an exponential relation to the number of accidents (the number of accidents being proportional to the average speed to the power of a parameter). In addition to the freely estimated models, and the models with exposure preset, models with the exponent of speed preset to 2 have also been tried.

Table 26 Model for bicycle accidents - no pre set effect of exposure – Speed limit as explaining variable

Model: Bicyclist accidents, no preset, Speed limit				
Deviance	**df**	**Perc.exp.**		
545	383	48%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	6.56E-04		-5.75	0.000
NCXP		0.39	6.29	0.000
Flow		0.59	4.18	0.000
SL_30	0.19		-1.24	0.215
SL_Rek30	0.88		-0.30	0.767
SL_50	1			
SL_70	0.58		-1.79	0.074

Table 27 Model for bicycle accidents - no pre set effect of exposure - Vehicle speeds as explaining variable

Model: Bicyclist accidents, no preset, Speed				
Deviance	**df**	**Perc.exp.**		
531	385	53%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	2.00E-01		-0.86	0.390
NCXP		0.35	5.67	0.000
Flow		0.56	4.06	0.000
Speed		-1.37	-3.88	0.000

When fitting models to the accident data the speed has turned out negatively correlated with the number of accidents in all models. This can hardly be attributed to any causal relationship, but rather to a correlation between the level of speeds and other variables influencing the number of accidents.

When developing models to study the difference between using speed and speed limit, only exposure measures have been used besides speed or speed limit. In the ordinary development of models however, when all variables have been tried out, the negative exponent for speed shows a strong trend in moving towards zero, or even positive, as more variables are added. This goes well with the theory that other variables correlated with speed are the ones actually making the speed turn out negatively correlated. As more other variables are added to the models, the negative correlation of speed weakens. If all other possible variables affecting safety could be duly included in the models, the effect of speed on personal injury accidents should be quadratic according to the power model of Nilsson (2004).

The results for single vehicle accidents are not very stable because of the low amount of accidents in combination with distribution over speed limits. There are for instance no single vehicle accidents to be found on 30 km/h streets and few on streets with recommended speed limit 30 km/h. Because of this, the comparison is made primarily for the three other accident types.

The degree of explanation is slightly higher for models with actual speeds than for models with speed limits (Table 28 and Table 29). For the bicyclist model the degree of explanation drops markedly when the parameters for exposure are preset, the likely reason being that the 'true' relation does not agree with the preset values.

Table 28 Degree of explanation for models with speed respectively speed limit, models with exponents for exposure decided by optimisation

Accident type	Exponents for exposure decided by optimisation					
	Speed Limit			Speed		
	Deviance	df	Perc.Exp.	Deviance	df	Perc.Exp.
Bicyclists	545	383	48%	531	385	53%
Pedestrians	305	383	71%	295	385	78%
Vehicle-Vehicle	704	384	50%	700	386	51%
Vehicle-Single	258	384	45%	268	386	39%

Table 29 Degree of explanation for models with speed respectively speed limit, models with preset exponents for exposure

Accident type	Exponents for exposure preset					
	Speed Limit			Speed		
	Deviance	df	Perc.Exp.	Deviance	df	Perc.Exp.
Bicyclists	607	385	30%	598	387	34%
Pedestrians	318	385	63%	309	387	71%
Vehicle-Vehicle	709	385	49%	707	387	50%
Vehicle-Single	261	385	43%	269	387	38%

6.3.2 Outliers

Potential outliers among the links have been studied to see if any of them might have unduly influenced the connection between speed, and/or speed limit, and the accident outcome. For each of the 16 models fitted, the 10 links with the strongest leverage on the fitting of the models have been studied. Particular interest has been on whether there are any links with high speeds and low amounts of accidents or low speeds and high amounts of accidents. The conclusion of the study is that there are only a very few potential outliers going against the theory of *higher speeds – more accidents*. Just as many of the links with a high leverage have instead either low speeds and low amount of accidents or the reverse.

After having identified the few links that go against theory, a test has been performed to see whether these might have influenced enough to cause the negative relationship between accidents and speed / speed limit. This has been done by fitting new models and excluding those 3-6 links per accident type that influence in the 'wrong' way. The new models can be found in Appendices 3C and 3D, and can be compared to the original models in Appendices 3A and 3B.

The shift in exponent for the speed is small, only for the single-vehicle accidents does it actually become almost positive. These accidents however are few and the results for speed limit should be considered with care as the models don't converge properly. This is because of the extremely low number of accidents on links with the lower speed limits, 30 km/h and recommended 30 km/h.

6.3.3 Preset effects of speed

It was previously stated that several variables may obscure the effect of speed in the models, making the speed take on a negative exponent in the models. To test whether it at all would be possible to construct sound models with 'true' speed effects, a different approach has been used. In the fitting of models the effect of speed has now been preset to quadratic. Thereafter variables have been added one by one in the same way as before. By presetting the effect of speed, other variables correlated with speed may now be assigned totally new parameter values, and the hope is to find the ones obscuring the effect of speed and to obtain models with fairly good degree of explanation.

Table 31, Table 32 and Table 33 show the models with the exponent of speed preset to 2. The degree of explanation achieved is reduced compared to models with freely estimated parameters, but not drastically. Table 30 shows a comparison with the basic models from 6.2. There is however still the problem with correlation between speed and many of the other variables.

Table 30 Degrees of explanation for basic models and models with exponent for speed preset to 2

	Explained systematic variation		
	Bicyclist accidents	Pedestrian accidents	Vehicle-Vehicle accidents
Basic models, no preset	60%	93%	61%
Models with exponent for speed preset to 2	47%	79%	54%

Table 31 Model for bicyclist accidents, speed preset to Speed^2

Model: Bicyclist accidents, Speed preset				
Deviance	**df**	**Perc.exp.**		
545	380	47%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	6.27E-08		-10.94	0.000
Speed		2		
NCXP		0.41	6.01	0.000
Landuse_B	1			
Landuse_X	0.39		-3.27	0.001
Landuse_C	1.67		2.45	0.015
Landuse_I	0.44		-3.47	0.001
Landuse_A	0.91		-0.35	0.727
Flow		0.74	4.61	0.000
Func_GIF	1			
Func_T	1.50		2.34	0.020
Func_C	3.61		4.22	0.000

Table 32 Model for pedestrian accidents, speed preset to Speed^2

Model: Pedestrian accidents, Speed preset				
Deviance	**df**	**Perc.exp.**		
290	380	79%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	5.72E-09		-9.17	0.000
Speed		2		
NPXP		0.59	5.62	0.000
Landuse_B	1			
Landuse_X	0.97		-0.09	0.932
Landuse_C	1.67		1.92	0.056
Landuse_I	0.20		-3.16	0.002
Landuse_A	1.08		0.24	0.811
Func_GIF	1			
Func_T	1.31		1.10	0.273
Func_C	4.02		4.28	0.000
Flow		0.73	3.30	0.001

Table 33 Model for vehicle-vehicle accidents, speed preset to Speed^2

Model: Veh.-Veh. accidents, Speed preset				
Deviance	**df**	**Perc.exp.**		
672	381	54%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	3.22E-10		-15.97	0.000
Speed		2		
Flow		1.50	10.44	0.000
DX		-0.72	-6.27	0.000
Func_GIF	1			
Func_T	1.58		3.16	0.002
Func_C	4.89		4.96	0.000
SL_30	0.15		-1.34	0.182
SL_Rek30	0.35		-1.46	0.146
SL_50	1			
SL_70	0.88		-0.70	0.486

6.4 Exposure of vulnerable road users

Estimates of vulnerable road users' exposure have been tested to see how the estimates can improve the accident prediction models. The contribution of the estimates has been tested in two ways; first by comparing the contribution to the degree of explanation of the estimates for the vulnerable road users' exposure with that of vehicle flow, the only covariate used for explaining the variation in VRU accidents in existing EVA models. The second check has been done by developing accident models without using the exposure measures, then comparing these with the previously developed models including estimates of exposure. This way it has been examined how much of the systematic variation can be explained without the help of estimates of VRU exposure.

In addition to the two tests, a separate study has been done to investigate further the difference between using the combined exposure variables for both crossing and parallel moving VRUs (NCXP / NPXP) and using variables describing crossing and parallel movement separately. Finally an investigation has been done to see whether the data show any jump in risk at any bicyclist flow, in accordance with Ekman's (1996) findings for traffic conflicts. Ekman showed that the risk decreased as the flow of bicyclists increased, but most of the decrease in risk was focused to an abrupt jump in risk between very low flows and larger flows.

6.4.1 VRU exposure vs. Vehicle Flow

A first crude way to compare the explaining strength of VRU exposure versus vehicle flow is to develop models only including VRU exposure respectively vehicle flow, and also a model containing both. Thereafter the degree of explanation is compared between the models to see the explaining power of each variable and the combination. Descriptions of the models are found in Appendix 4.

For both bicyclist and pedestrian accidents the models with variables describing VRU exposure explain fairly much of the systematic variation by themselves, while those with vehicle flow doesn't explain more than a few percent (Table 34 & Table 35). The combination of VRU exposure and vehicle flow explains slightly more than the sum of the separate models.

Table 34 Degree of explanation for bicyclist models with bicyclist exposure and/or vehicle flow

	NCXP	Flow	NCXP+Flow
Deviance	585	701	556
df	387	387	386
Dev/df	1.51	1.81	1.44
Expl.	38%	3%	46%

Table 35 Degree of explanation for pedestrian models with pedestrian exposure and/or vehicle flow

	NPXP	Flow	NPXP+Flow
Deviance	317	415	308
df	387	387	386
Dev/df	0.82	1.07	0.80
Expl.	65%	4%	71%

In the results above there is one factor making the comparison unjust towards vehicle flow. The null model used for calculating percentage explained does not include the offset variable link length, while all models developed include it (or even more offset variables). For vulnerable road users the use of link length as offset variable, i.e. using accident per length instead of per link, actually increases the scaled deviance. The reason for this is that longer links have fewer accidents with VRUs since the links are normally located in the outer parts of the city and have fewer VRUs passing the link.

If the percentages explained are calculated with a null model including the offset variable link length, the results are still that VRU exposure explains much more than vehicle flow, but the latter variable is seen explaining 12-13% of the systematic variation, instead of 3-4% (Table 36 & Table 37).

Table 36 Degree of explanation for bicyclist models with bicyclist exposure and/or vehicle flow, null model with offset: link length

	NCXP	Flow	NCXP+Flow
Deviance	585	701	556
df	387	387	386
Dev/df	1.51	1.81	1.44
Expl.	43%	12%	50%

Table 37 Degree of explanation for pedestrian models with pedestrian exposure and/or vehicle flow, null model with offset: link length

	NPXP	Flow	NPXP+Flow
Deviance	317	415	308
df	387	387	386
Dev/df	0.82	1.07	0.80
Expl.	68%	13%	73%

Table 38 shows the explained variance for all the different models including only one variable, calculating the variance explained compared to the null model with offset link length. The variables describing the exposure of vulnerable road users all explain fairly much of the variation. Other variables explaining much variation are Landuse and Speed. The latter however appears in the models with an inverse effect to that known, since it is correlated with many other variables affecting safety.

Table 38 Explained variance for bicyclist and pedestrian models with only one variable, null model with offset: link length

	Bicycle models		
Variable	Scaled Deviance	Residual degrees of freedom	Explained variance
GrNCXP	569	386	47%
NCXP	585	387	43%
NPXP	587	387	42%
GrNPXP	601	386	38%
NCXKm	610	387	36%
NPXKm	610	387	36%
NPX	617	387	34%
NCX	619	387	34%
NPP	620	387	33%
NCP	640	387	28%
Landuse	647	384	25%
Speed	670	387	20%
City	679	383	15%
Vis	696	386	12%
X4Km	700	387	12%
Flow	701	387	12%
DX	704	387	11%
XForm	703	384	9%
X34Km	720	387	6%
Func	718	386	6%
LEnv	719	386	6%
SL	717	385	6%
SepC	723	386	5%
Park	731	387	3%
Lanes	730	386	3%
ExitL	733	387	3%
SepV	734	386	2%
VRUX	736	387	2%
DSpeed	742	387	0%
ExitS	743	387	0%
X3Km	745	387	-1%

	Pedestrian models		
Variable	Scaled Deviance	Residual degrees of freedom	Explained variance
NPXP	317	387	68%
NPXKm	322	387	66%
NPX	326	387	63%
GrNPXP	360	386	44%
Speed	361	387	43%
NPP	361	387	43%
Landuse	359	384	43%
NCXKm	369	387	39%
NCX	376	387	35%
GrNCXP	380	386	32%
NCXP	389	387	27%
Func	391	386	26%
City	401	383	18%
XForm	402	384	18%
VRUX	410	387	15%
LEnv	410	386	15%
DX	412	387	14%
Flow	415	387	13%
X4Km	417	387	12%
X34Km	417	387	12%
Park	419	387	10%
NCP	420	387	10%
SL	421	385	8%
SepC	428	386	5%
ExitS	431	387	4%
Lanes	431	386	3%
Vis	432	386	2%
SepV	433	386	2%
DSpeed	434	387	2%
ExitL	438	387	0%
X3Km	438	387	-1%

6.4.2 Models for VRU accidents without VRU exposure

Models have been developed without the use of VRU exposure variables in order to test how much of the systematic variation can be explained without using these variables. The models are here presented both with four and with seven variables. Each addition of a variable has yielded a significantly better degree of explanation for both bicycle and pedestrian accidents (Table 39 & Table 42). However, the separate categorical alternatives frequently yield parameters with non-significant difference between them. The variables have still been used as long as the parameters take on reasonable values but the models are less stable than those with VRU exposure included.

The degree of explanation of the models without VRU exposure variables fall short of the models including VRU exposure. The original bicyclist model that includes a freely estimated effect of bicyclist exposure (Table 13) has a degree of explanation of 60%. The models without bicyclist exposure have a degree of explanation of 42% in the case of the four variable model (Table 40), and 51% for the model with seven variables (Table 41).

The results for pedestrian accidents are similar to the results for bicyclist accidents. With a seven variable model 85% of the variation is explained (Table 44) and 72% with a four variable model (Table 43). This should be compared to the very high degree of explanation of the original pedestrian model (Table 17) which has a 93% degree of explanation.

Table 39 Bicyclist accident models, stepwise addition of parameters, no VRU exposure variables

Model: Bicyclist accidents, no bicyclist exposure								
Var. added	0	1	2	3	4	5	6	7
Variable	Null model	Landuse	Flow	Func	SL	SepC	Vis	X4Km
Dev	745	649	593	575	559	546	532	523
df	388	385	384	382	379	378	376	375
dDev		95.6	56.5	17.7	16.4	12.9	13.9	9.3
ddf		3	1	2	3	1	2	1
QP Scale		1.814	1.814	1.814	1.814	1.814	1.814	1.814
Scaled dDev		52.7	31.1	9.8	9.1	7.1	7.6	5.1
Chi2-prob.		0.000	0.000	0.008	0.028	0.008	0.022	0.024

Table 40 Bicyclist accident model, no VRU exposure variable used, 4 explaining variables

Model: Bicyclist acc., no exposure, 4 var.				
Deviance	**df**	**Perc.exp.**		
559	379	42%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	4.73E-04		-5.29	0.000
Landuse_AB	1			
Landuse_X	0.29		-4.31	0.000
Landuse_C	1.20		0.90	0.366
Landuse_I	0.62		-2.04	0.042
Flow		0.93	5.91	0.000
Func_GIF	1			
Func_T	1.36		1.82	0.069
Func_C	2.87		3.50	0.001
SL_ 30	0.11		-1.62	0.105
SL_Rek30	0.83		-0.43	0.670
SL_ 50	1			
SL_ 70	0.59		-1.64	0.103

Table 41 Bicyclist accident model, no VRU exposure variable used, 7 explaining variables

Model: Bicyclist acc., no exposure, 7 var.				
Deviance	**df**	**Perc.exp.**		
523	375	51%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	8.00E-04		-4.89	0.000
Landuse_AB	1			
Landuse_X	0.36		-3.40	0.001
Landuse_C	1.23		0.93	0.353
Landuse_I	0.73		-1.27	0.205
Flow		0.89	5.68	0.000
Func_GIF	1			
Func_T	1.45		2.22	0.027
Func_C	2.33		2.67	0.008
SL_ 30	0.12		-1.57	0.117
SL_Rek30	0.73		-0.71	0.481
SL_ 50	1			
SL_ 70	0.58		-1.66	0.099
SepC_CbX	1			
SepC_B	0.56		-2.96	0.003
Vis_Good	0.67		-2.32	0.021
Vis_Medium	1			
Vis_Poor	1.53		0.84	0.401
X4Km		0.17	2.27	0.024

Table 42 Pedestrian accident models, Stepwise addition of parameters, no VRU exposure variables

Model: Pedestrian accidents, no pedestrian exposure								
Var. added	0	1	2	3	4	5	6	7
Variable	Null model	Landuse	Flow	Func	XForm	VRUX	DX	Vis
Dev	439	359	336	314	299	291	285	276
df	388	384	383	381	377	376	375	373
dDev		79.7	22.4	22.3	15.3	8.3	5.7	8.9
ddf		4	1	2	4	1	1	2
QP Scale		1.161	1.161	1.161	1.161	1.161	1.161	1.161
Scaled dDev		68.6	19.3	19.2	13.2	7.1	4.9	7.7
Chi2-prob.		0.000	0.000	0.000	0.010	0.008	0.026	0.021

Table 43 Pedestrian accident model, no VRU exposure variable used, 4 explaining variables

Model: Pedestrian acc., no exp., 4 var.				
Deviance	**df**	**Perc.exp.**		
299	377	72%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	1.77E-04		-3.86	0.000
Landuse_B	1			
Landuse_X	0.66	-0.42	-1.22	0.222
Landuse_C	2.19	0.79	3.11	0.002
Landuse_I	0.25	-1.39	-2.71	0.007
Landuse_A	1.39	0.33	1.02	0.307
Flow	2.51	0.92	3.79	0.000
Func_GIF	1			
Func_T	1.21	0.19	0.80	0.423
Func_C	4.32	1.46	4.72	0.000
XForm_Ö	1			
XForm_F	0.61	-0.50	-2.01	0.046
XForm_X	0.15	-1.87	-2.32	0.021
XForm_S	0.84	-0.17	-0.61	0.541
XForm_P	0.34	-1.08	-1.63	0.103

Table 44 Pedestrian accident model, no VRU exposure variable used, 7 explaining variables

Model: Pedestrian acc., no exp., 7 var.				
Deviance	**df**	**Perc.exp.**		
276	373	85%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	1.64E-05		-4.72	0.000
Landuse_B	1			
Landuse_X	0.81	-0.21	-0.56	0.577
Landuse_C	2.03	0.71	2.87	0.004
Landuse_I	0.27	-1.32	-2.52	0.012
Landuse_A	1.16	0.15	0.44	0.663
Flow	2.76	1.01	4.22	0.000
Func_GIF	1			
Func_T	1.25	0.22	0.94	0.347
Func_C	4.65	1.54	4.79	0.000
XForm_Ö	1			
XForm_F	0.63	-0.46	-1.86	0.063
XForm_X	0.20	-1.62	-1.98	0.048
XForm_S	0.86	-0.15	-0.53	0.594
XForm_P	0.45	-0.80	-1.20	0.231
VRUX_Yes	1.87	0.63	2.88	0.004
VRUX_No	1			
DX	0.54	-0.61	-2.76	0.006
Vis_Good	1.75	0.56	2.46	0.014
Vis_Medium	1			
Vis_Poor	0.32	-1.15	-1.04	0.301

6.4.3 Combined exposure variables vs. divided variables

The combined variables of pedestrians respectively bicyclists crossing and moving along the street have in the modelling been shown to be better predictors for VRU accidents than their components. A natural question is how a model with the two components stands up to a model with the combined variable. To test this, models have been developed with the combined variable in one model and the two separate variables in the other. The models developed show that the combined variables (Table 45 & Table 47) describe at least as much variation as the separate variables in combination (Table 46 & Table 48)

The most likely explanation is that the model form for inclusion of VRU exposure should be additive. When adding the two variables crossing pedestrians (NPX) and pedestrians along the street (NPP) the model automatically yields a multiplicative form, while intuition says that the number of pedestrians crossing and walking along the street each should have their own additive effect.

Table 45 Model for bicyclist accidents, explaining variable: NCXP

Model: NCXP				
Deviance	**df**	**Perc.exp.**		
585	387	38%		
Parameter Estimates				
Parameter	Estimate	Standard Error	t value	Scaled t-value
Constant	-2.57	0.34	-7.61	-5.65
NCXP	0.47	0.05	10.12	7.52

Table 46 Model for bicyclist accidents, explaining variables: NCX and NCP

Model: NCX+NCP				
Deviance	**df**	**Perc.exp.**		
584	386	37%		
Parameter Estimates				
Parameter	Estimate	Standard Error	t value	Scaled t-value
Constant	-1.33	0.22	-6.13	-4.55
NCX	0.18	0.03	6.88	5.11
NCP	0.18	0.03	5.29	3.93

Table 47 Model for pedestrian accidents, explaining variable: NPXP

Model: NPXP				
Deviance	**df**	**Perc.exp.**		
317	387	65%		
Parameter Estimates				
Parameter	Estimate	Standard Error	t value	Scaled t-value
Constant	-5.09	0.57	-8.90	-8.26
NPXP	0.66	0.07	9.12	8.47

Table 48 Model for pedestrian accidents, explaining variables: NPX and NPP

Model: NPX+NPP				
Deviance	**df**	**Perc.exp.**		
323	386	61%		
Parameter Estimates				
Parameter	Estimate	Standard Error	t value	Scaled t-value
Constant	-3.63	0.43	-8.36	-7.76
NPX	0.41	0.07	5.55	5.16
NPP	0.12	0.07	1.81	1.68

6.4.4 Test of discrete levels of risk for bicyclists

Ekman (1996) has shown that the number of traffic conflicts per crossing bicyclist decreases as the flow of bicyclists increase. According to Ekman, the decrease in risk is due to an increased awareness of bicyclists for drivers. As the flow of bicyclists increase and reach a certain level, the drivers anticipate the presence of bicyclists at the location. The decrease in risk is mainly focused to a big decrease between low and high flows (Figure 36). To investigate whether the accident data for bicyclists show the same tendencies, and should be modelled with distinctly different risk levels for low and high bicyclist flows, the number of accidents per bicyclist passing through the environment (AccC/NCXP) has been plotted against the bicyclist flow. To produce more stable results the numbers of accidents and cyclists have been aggregated for several links with roughly the same flow of cyclists. Figure 37, Figure 38 & Figure 39 show the aggregated estimates with an aggregate interval of 100, 200 and 500 cyclists respectively. As the interval of aggregation is increased, the levels of accident rate are stabilised.

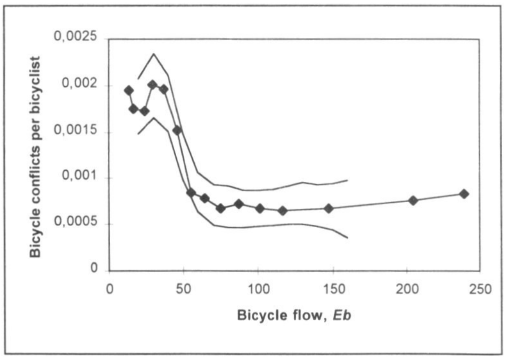


Figure 36 Bicycle conflicts per bicyclist (Ekman 1996)

The risk per bicyclist is decreasing as the flow increase, but the decrease shows a more continuous trend, not a distinct drop. For very low levels there might however be a much higher risk. This result is, however, not totally reliable as the method for measuring number of VRUs only produces rough estimates of the flows of vulnerable road users. A few VRUs more or less passing through the environment during the short time of measuring can easily change which 100-interval the link will belong to. As the interval is increased, the risk for very low bicyclist flows is more in harmony with the overall trend of slowly decreasing risk. As a result the relation between bicyclist flow and accidents is rather modelled with a continuous variable for flow, than with a categorical variable describing level of bicycle flow.

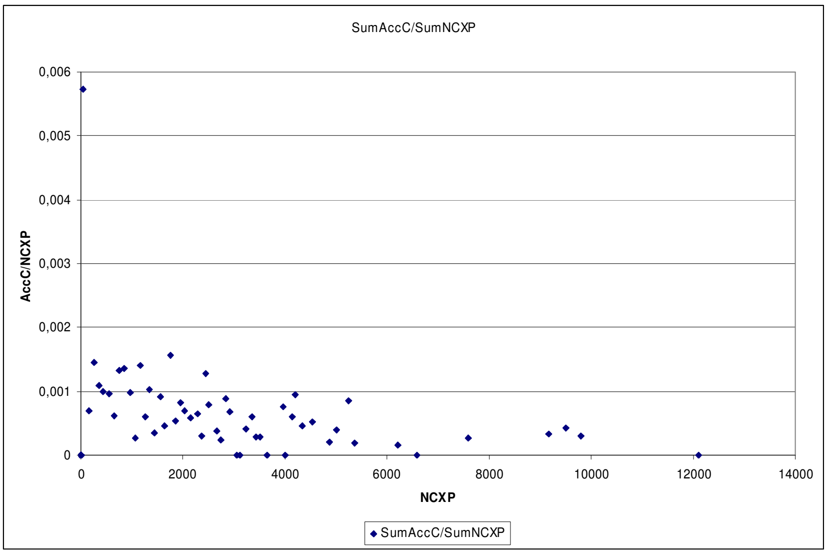


Figure 37 Accident rate plotted against bicyclist flow, links aggregated over intervals of 100 bicyclists/day

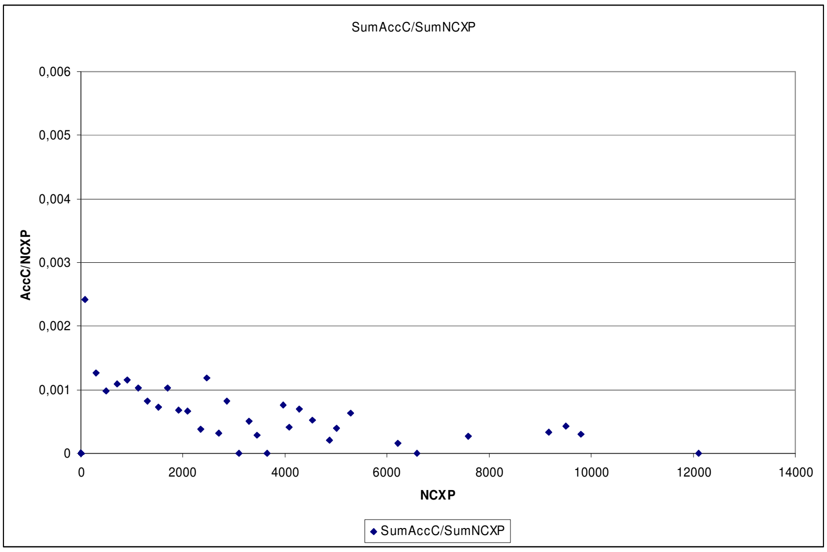


Figure 38 Accident rate plotted against bicyclist flow, links aggregated over intervals of 200 bicyclists/day

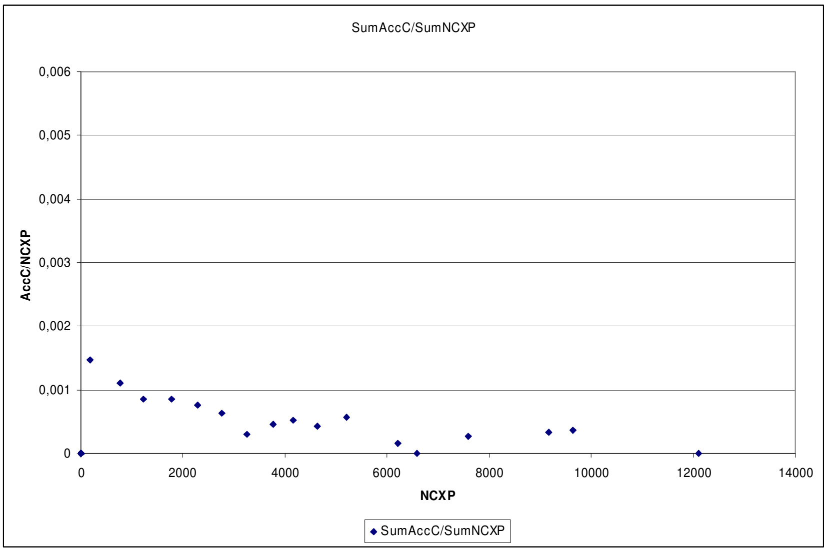


Figure 39 Accident rate plotted against bicyclist flow, links aggregated over intervals of 500 bicyclists/day

6.5 Dividing vehicle accidents into single vehicle and multiple vehicle accidents

The original testing strategy for the division of vehicle accident models into separate models for single and multiple vehicle accidents was to compare the degree of explanation for an all vehicle accident model with that of a combined model. As the number of single vehicle accidents in the data has been very small, and the models developed not totally reliable, the comparison has been limited to a study of the fit of an all vehicle accident model with that of the multiple vehicle accident model to see if the model fit is improved.

The model developed for all vehicle accidents is displayed in Table 50 and should be compared to the multiple vehicles accident model in Table 21. The all vehicle model include the same variables as the multiple vehicles model with the addition of ExitL, presence of busy exits. The parameters differ to some degree, but on the whole the models are fairly the same. The degree of explanation is only slightly higher for the multiple vehicles model, most likely due to the fact that the single vehicle accidents excluded only made up 15% of all the vehicle accidents. The multiple vehicles model explains 61% of the systematic variation (for multiple vehicle accidents), while the all vehicle model explain 59% (for all vehicle accidents).

Table 49 Model for all vehicle accidents, stepwise addition of variables, no preset

Model: All vehicle accidents, no preset							
Var. added	0	1	2	3	4	5	6
Variable	Null model	Flow	X34Km	SL	ExitL	Func	SepV
Dev	965	721	663	644	632	618	610
df	388	387	386	383	382	380	378
dDev		244.5	57.9	18.7	12.4	13.6	7.8
ddf		1	1	3	1	2	2
QP Scale		1.814	1.814	1.814	1.814	1.814	1.814
Scaled dDev		134.8	31.9	10.3	6.8	7.5	4.3
Chi2-prob.		0.000	0.000	0.016	0.009	0.024	0.116

Table 50 Model for all vehicle accidents, no preset

Model: All vehicle accidents, no preset				
Deviance	**df**	**Perc.exp.**		
618	380	59%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	4.52E-06		-10.21	0.000
Flow		1.38	10.85	0.000
X34Km		0.41	5.55	0.000
SL_ 30	0.09		-1.76	0.080
SL_Rek30	0.46		-1.52	0.130
SL_50	1			
SL_ 70	1.05		0.31	0.758
ExitL_Yes	1			
ExitL_No	1.40		2.70	0.007
Func_GIF	1			
Func_T	1.25		1.71	0.088
Func_C	2.18		2.62	0.009

In addition to the freely estimated model, a model with the parameter value for flow has been pre-set to an exponent of 1. Although the models freely estimated found an optimised model with the exponent of 1.38 (Table 50), the model with the exponent forced to 1 offer almost as good a degree of explanation, 55% compared to 59%. In comparison with the preset model for multiple vehicle accidents (Table 23), which had a degree of explanation of 60%, the all vehicle accident models have fairly the same. Whether the parameter value for flow is pre-set or not, the models for all vehicle accidents are almost as good as those for multiple vehicle accidents only.

Table 51 Model for all vehicle accidents, stepwise addition of variables, exposure preset to flow[1]

Model: All vehicle accidents, exposure preset				
Var. added	0	1	2	3
Variable	Null model	X34Km	SL	ExitL
Dev	726	681	658	644
df	388	387	384	383
dDev		45.1	22.6	14.1
ddf		1	3	1
QP Scale		2.004	2.004	2.004
Scaled dDev		22.5	11.3	7.1
Chi2-prob.		0.000	0.010	0.008

Table 52 Model for all vehicle accidents, exposure preset to flow[1]

Model: All vehicle accidents, exposure preset				
Deviance	**df**	**Perc.exp.**		
644	383	55%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	1.71E-04		-57.49	0.000
Flow		1		
X34Km		0.40	5.27	0.000
SL_30	0.10		-1.61	0.109
SL_Rek30	0.48		-1.37	0.172
SL_50	1			
SL_70	1.17		0.96	0.339
ExitL_Yes	1			
ExitL_No	1.41		2.62	0.009

6.6 Comparison between existing and developed models

6.6.1 Comparability between models

One of the main aims of this dissertation has been to develop models that can better predict the amount of accidents and injured on urban streets than the existing models. To assess how well the new models describe and predict the number of accidents a comparison has been made between the models previously described and models only including the variables used in the existing Swedish accident models (EVA). The actual Swedish models haven't been used for the comparison because of problems with comparability between the models. Some issues that complicate the comparison are listed below. Because of the many problems with a straight comparison of the predictions of the EVA models and the models developed here, the comparison has been made between the models developed in this project and models developed in the same way but with only the variables used in the EVA models.

Vehicle km vs. Axle pair km

The existing models use accident rates with accidents per million axle pair km, while the newly developed models use accidents per vehicle km. Depending on the percentage of heavy vehicles, such as lorries and buses, these variables can differ by up to 5-10%.

Property damage accidents vs. Injury accidents

The existing models include both injury accidents and property damage only accidents, while the models developed in this work only include personal injury accidents. This makes a comparison close to impossible for vehicle accidents. For the vulnerable road users' accidents the problem should be smaller as most accidents including VRUs lead to a personal injury. The EVA models however base the number of VRU accidents on the number of vehicle accidents which makes even this comparison problematic at best.

Accident data

The accident data available for comparison is the accident data for 1994-1996 from Uppsala, Katrineholm and Danderyd, since eight years of accident data were compiled for these. Although the data are for the years 1994-1996, the years not included in the data for developing the new models, the municipalities still constitute half of the municipalities used for the models. This gives the new models a clear advantage in the comparison. On the other hand, the years for the accident data (1994-1996) corresponds better with the accident data years the existing models are based on (1995-1998) than those the new models are based on (1997-2001 / 1998-2002).

Most of the problems with differing link characteristics between the links used for developing the existing and the new models are listed below.

Speed limit 30 km/h

The existing accident models do not have any accident rates for streets with the speed limit 30 km/h, therefore no predicted number of accidents can be calculated with the EVA models for links with speed limits 30 km/h or recommended 30 km/h.

6.6.2 Comparing developed models with models only containing EVA variables

In order to check specifically for the additional value of the new variables used in this dissertation the models in chapter 6.2 are compared to models developed in a similar way but restricted to variables included in today's EVA models. The models are presented in

Table 54, Table 55 and Table 56. A comparison of the models with regard to degree of explanation is found in Table 53. Restricting the models to only using EVA variables naturally decrease the degree of explanation achieved. For the vehicle-vehicle models the degree of explanation decreases only from 61% to 54% while for the bicyclist and pedestrian models the degrees of explanations are reduced drastically, from 60% to 31% and from 93% to 54% respectively. Since the EVA models are focused mostly on motorised traffic it is understandable that the variables included in the models are focused mostly on such variables as describe the variation in vehicle accidents. It is interesting to see that obviously quite different variables than those of EVA describe the variation in VRU accidents.

Table 53 Degree of explanation for basic models and models with only EVA variables

	Explained systematic variation		
	Bicyclist accidents	Pedestrian accidents	Vehicle-Vehicle accidents
Basic models, no preset	60%	93%	61%
Models with only EVA variables	31%	54%	54%

Table 54 Model for vehicle-vehicle accidents, no preset parameters, only EVA variables

Model: Vehicle-Vehicle, only EVA variables				
Deviance	**df**	**Perc.exp.**		
673	382	54%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	1.10E-05		-8.52	0.000
Flow		1.33	9.35	0.000
Func_GIF	1			
Func_T	1.54		2.97	0.003
Func_C	2.92		3.39	0.001
SL_30	0.13		-1.43	0.154
SL_Rek30	0.33		-1.54	0.125
SL_50	1			
SL_70	0.93		-0.44	0.663

Table 55 Model for bicyclist accidents, no preset parameters, only EVA variables

Model: Bicyclist model, only EVA variables				
Deviance	**df**	**Perc.exp.**		
599	382	31%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	3.78E-04		-5.70	0.000
Flow		0.94	6.34	0.000
SL_ 30	0.13		-1.50	0.134
SL_50Rek30	1			
SL_ 70	0.38		-3.08	0.002
Func_GIF	1			
Func_T	1.44		2.22	0.027
Func_C	3.50		4.42	0.000
LEnv_CM	1			
LEnv_Y	0.66		-2.39	0.017

Table 56 Model for pedestrian accidents, no preset parameters, only EVA variables

Model: Pedestrian model, only EVA variables				
Deviance	**df**	**Perc.exp.**		
331	382	54%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	4.83E-05		-4.43	0.000
Func_GIF_T	1			
Func_C	6.14	1.81	5.50	0.000
Flow	2.91	1.07	4.38	0.000
LEnv_CM	1			
LEnv_Y	0.50	-0.70	-2.21	0.028
SL_ 30	0.51	-0.68	-0.70	0.485
SL_Rek30	1.97	0.68	1.44	0.152
SL_ 70	0.42	-0.88	-1.61	0.109

6.7 Validity of the models

To test the validity of the models developed they have been tested against accident data not used in the modelling. For Uppsala, Katrineholm and Danderyd eight years of accident data were compiled, while only five years of accident data were used for the modelling. The remaining data, for the years 1994-1996, have been used for the validation of the models.

The models developed have been used for predicting the number of accidents on the links in Uppsala, Katrineholm and Danderyd. The scaled deviance has then been calculated for the fit of the predicted number of accidents to the occurred number of accidents in the validation data.

The calculated scaled deviance is then used for calculating how much of the systematic variation is explained. This is done in the same way as when assessing the fit of the models in the modelling phase, although now new Null models and estimations of how much of the variation is systematic have been calculated for the validation data (1994-1996).

For the models to have a good predicting power, they should be able to explain most of the systematic variation in the validation data. Table 57 and Table 58 show the percentage of systematic variation explained for the basic models. For the freely estimated models, the degrees of explanation are ranging between 71 and 77 %, except for the model predicting the total number of vehicle accidents which has 60% explained variation. It seems as if the model combining both single and multiple vehicle accidents has weaker ability to predict the number of accidents than the other models.

Basic models

Looking at the models with exposure preset the results are more widely spread. The bicyclist model only explains 53% of the systematic variation in the validation accident data. The preset pedestrian and vehicle-vehicle models explain approximately 80% of the systematic variation, and the all vehicle model explains 75% of the systematic variation which is slightly better than for the freely estimated models. The preset model for all vehicle accidents is considerably better than the same model with freely estimated parameters. The preset models seem to predict accidents better than the freely estimated ones. The explanation for this is probably that the 'true' parameters are actually closer to the preset values than the freely estimated. This, however, does not seem to be true for the bicyclist accidents.

Table 57 Percentage of systematic variation explained, models without preset parameter values

Models without preset				
	Bicyclist	Pedestrian	Vehicle-Vehicle	Vehicle -All
Null model	263	164	304	322
Scaled deviance for estimated model	196	104	192	238
Full model	168	83	158	183
Percentage explained systematic variation	71%	74%	77%	60%

Table 58 Percentage of systematic variation explained, models with preset parameter values for exposure

Models with preset parameters for exposure				
	Bicyclist	Pedestrian	Vehicle-Vehicle	Vehicle -All
Null model	263	164	304	322
Scaled deviance for estimated model	212	99	188	219
Full model	168	83	158	183
Percentage explained systematic variation	53%	81%	79%	75%

VRU models without VRU exposure variables

When modelling VRU accidents using only variables from the existing EVA models, the predicting capability of the models is rather poor, only 37% of the systematic variation is explained for each type of accident (Table 59). For models where also new variables (except VRU exposure) have been used the predicting capability is somewhat better, slightly more than 50% for the best models (Table 60). This is still much poorer than the models with VRU exposure included. The main difference between the models with only EVA variables and the models with just VRU exposure variables excluded is due to the variable Landuse. The four variable models without VRU exposure include the most important EVA variables and the variable Landuse. The variable Landuse seems to add a lot of predictive capability to the models compared to using only EVA variables.

Table 59 Percentage of systematic variation explained, VRU accident models with only EVA variables

VRU models with only EVA variables		
	Bicyclist	Pedestrian
Null model	263	164
Estimated model	228	134
Full model	168	83
Percentage explained systematic variation	37%	37%

Table 60 Percentage of systematic variation explained, VRU accident models without VRU exposure

VRU models without VRU exposure				
	Bicyclist 4 variables	Bicyclist 7 variables	Pedestrian 4 variables	Pedestrian 7 variables
Null model	263	263	164	164
Estimated model	211	226	120	122
Full model	168	168	83	83
Percentage explained systematic variation	55%	39%	54%	52%

Looking at the degrees of explanation in Table 60 there is one odd value, the degree of explanation for the bicyclist model with seven variables. All the other models predict between 52-55 % of the systematic variation while this model only predicts 39 %. One reasonable explanation is that one, or more, of the three variables added compared to the model with four variables has corrupted the model. This explanation is even more likely considering that the difference between the bicyclist models with freely estimate exposure parameters (Table 57) and preset exposure parameters (Table 58) above is that in the model with preset exposure parameters, that predicts accidents poorly, the variable Separation of bicyclists (SepC) is included instead of Street function (Func). The variable SepC is one of the three variables added in the bicycle model with seven variables compared to the model with four variables (Table 40 & Table 41).

One reason for the variable SepC to work poorly in predicting bicyclist accidents is if separation of bicyclists is implemented in a different way in different cities, and the cities used in the validation study differ from the whole. Table 61 shows how bicyclists are distributed on different forms of separation or integration in the different cities. There are clear differences between the different cities. The three cities used in the validation study, Danderyd, Katrineholm and Uppsala, mark themselves out by hardly having any links where biking is not recommendable or directly prohibited. Those three cities instead have bicyclists integrated with motorised traffic for a large portion of the street network.

Table 61 Distribution of links by bicyclist separation in the different cities

	Danderyd	Katrineholm	Linköping	Norrköping	Uppsala	Västerås
Mixed	37%	15%	9%	24%	45%	3%
Separate	62%	82%	78%	63%	54%	69%
No Biking	2%	3%	13%	13%	1%	28%

To test whether the variable SepC is the variable that makes the bicyclist model have a very poor prediction, a new preset model was developed where SepC was replaced by Speed limit (the next best variable after SepC). This new model was able to explain 60 % of the systematic variation in the bicyclist accidents in the validation data. This is better than the model with SepC which explained 53 %, but not as good as the models for the other accident types where 75-81 % of the systematic variation is explained.

The reason behind the poor predicting capability of the bicyclist model with preset parameters for exposure variables is probably dependant on a mix of having SepC in the model and the exponent for bicyclist exposure being significantly different from the preset value 0.5.

Models with parameters for speed preset

When the parameter value for speed is preset to an exponent of 2 in the models, the models have a lower degree of explanation for the validation data compared to the freely estimated models, but they still explain more than half of the systematic variation (Table 62). The degree of explanation is however considerably lower than the degree of explanation for the basic models.

Table 62 Percentage of systematic variation explained, models with exponent for speed preset to 2

Models with exponent for speed preset to 2			
	Bicyclist	Pedestrian	Vehicle-Vehicle
Null model	263	164	304
Estimated model	210	109	211
Full model	168	83	158
Percentage explained systematic variation	56%	68%	64%

7 Discussion and conclusions

7.1 Verification of hypotheses

7.1.1 Exposure of vulnerable road users

The hypothesis to be verified: *The degree of explanation of the accident models for vulnerable road users can be greatly improved by including estimates of VRU exposure in the models.*

Variables describing the exposure of vulnerable road users have been shown to be very good explanatory variables for the variation in their accidents. The variables NCXP and NPXP, describing the total number of bicyclists respectively pedestrians are the variables with highest explanatory power. These variables are however more difficult to interpret than the separate variables for VRUs crossing and moving along the street, as the combined variables NCXP and NPXP present aggregates of VRUs both crossing and walking along the link.

The most likely reason for the combined variables to work better than the separate ones can be found in the model form. VRUs crossing respectively moving along the street are exposed in separate ways and should intuitively have an additive form of inclusion in the models. The model form however includes the variables on a multiplicative form. When instead combining the separate variables into one summed variable, the model form becomes additive with regard to VRU flow.

Another reason for the combined variables to function better than the separate ones may be that the field studies have been very short and a combination of the estimates of road users forms a more stable estimate.

When comparing the explanatory power of vehicle flow with that of the combined VRU exposure variables, the vehicle flow explains very little of the variation by itself, while the VRU exposure variables by themselves can explain approximately 40% of the variation of the bicyclist accidents and two thirds of the variation of pedestrian accidents. When developing models without using the VRU exposure variables the degrees of explanation have been considerably lower compared to when using the exposure variables. The number of variables needed in these models in order to get satisfactory levels of explanation is high and the parameter values for categorical variables often do not differ significantly from each other. The variables used in the models without exposure have typically been such variables as indirectly describe the number of vulnerable road users in the street environment, for example land use and street function. The results, however, show that these indirect descriptors of VRU exposure do not work nearly as well as the more direct exposure variables (NCXP and NPXP) in explaining the variation of VRU accidents.

7.1.2 Actual speeds vs. Speed limit

The hypothesis to be verified: *The degree of explanation of accident models can be improved if actual speeds are used instead of speed limit.*

Actual speeds are naturally related to the speed limit on the link in question. The data also shows a clear general difference in actual speed levels between links with different speed limits. There are however also large variations among links with the same speed limit, and an individual link with the speed limit 30 km/h, or recommended 30 km/h, can very well have a higher average speed than an individual link with the speed limit 50 km/h.

Models with the average of actual speeds as one of the explaining variables tend to generate slightly better degree of explanation than models with speed limit. The differences are however small and if the effect of actual speeds is set to be decided by optimisation in the modelling, the effect generally takes on a negative effect on the number of accidents, i.e. with higher speeds the number of accidents decrease. This can hardly be due to a causal relationship, but rather depends on the fact that speeds are correlated with many other variables affecting the safety outcome. This is to be expected as speed choice is the main manner with which the driver tries to adapt to the surrounding road and traffic environment and any changes in it.

The parameter values for the speed limit variable, however, correspond in most cases to theoretical and empirical evidence of the relationship between speed and accidents better than those for the actual speeds.

If the effect of speed is preset to its true value according to the power model (Nilsson 2004) and the other variables are introduced one by one, it is still possible to achieve models with a relatively high degree of explanation. The effects of other variables can however be contradictory to previous theoretical findings, probably because of the correlation with the preset speed variable and other variables.

It must be noted that the poor results for vehicle speed in the accident modelling is for the speed variable used in this study: average speed of all vehicles, measured at a spot along the link where speeds are relatively undisturbed by traffic signals or other phenomena that may considerably lower speeds. If the study would have included other estimates of vehicle speed, the results might have been different.

7.1.3 Separating Vehicle-Vehicle and Vehicle-Single accidents

The hypothesis to be verified: *The degree of explanation for vehicle accident models can be improved by separating the vehicle accident models into separate models for single vehicle accidents and multiple vehicle accidents.*

The separation of vehicle accident models into separate models for single and multiple vehicle accidents has been difficult to study due to the unexpectedly small number of single vehicle accidents and some deficiencies in the accident registration. The data on single vehicle accidents has not been very reliable, and models based on it have produced odd parameter estimates. It has been possible to achieve slightly higher degrees of explanation for multiple vehicle accident models compared to accident models for all vehicle accidents. Only 15% of all the vehicle accidents in the data are single vehicle accidents, and this explains why the increase in degree of explanation is very small when these accidents are excluded.

When applying the different models on the validation accident data the freely estimated vehicle-vehicle accident model predicted the number of vehicle-vehicle accidents better than the freely estimated all vehicle accident model was able to predict the total number of vehicle accidents. This is an indication that a split into separate models for separate accident types can produce more stable models. The preset models did however not differ much in explanatory power compared to the freely estimated basic models.

7.2 Contribution to scientific knowledge

Earlier accident modelling has more focused on rural roads than urban streets. The results in this dissertation give a deeper insight into statistical modelling of accidents on urban links, and specifically modelling of vulnerable road users' accidents.

7.2.1 Assessing the safety situation for vulnerable road users

The results show that in order to be able to properly assess the safety situation for vulnerable road users we need knowledge about their exposure. In the validation of the accident models for vulnerable road users the models with exposure variables for vulnerable road users explained more than 70% of the systematic variation compared to 37% with models of earlier origin. The estimates of exposure for VRUs are the variables that explain the main part of the systematic variation in number of accidents between different streets.

The relation between VRU exposure and their accidents is often considered to take the form of accidents being proportional to the exposure to the power of 0.5. For pedestrian exposure and accidents the results in this dissertation complies, but for bicyclist exposure and accidents the exponent takes on a significantly lower value than 0.5.

Besides the flows of vulnerable road users, the vehicle flow and the land use along the street are the most important variables for explaining the variation of VRU accidents. As most of the injured VRUs in the data used have been hit by a motorised vehicle the vehicle flow is a natural variable for describing the risk for the VRUs.

The land use plays a less obvious part than the flows of VRUs and vehicles in describing the amount of accidents. One of the phenomena described by land use is the variation of exposure over the day. The land use 'industrial' characterises a link with extra high peaks in flows of road users concentrated to start and end of working hours. The land use commercial on the other hand characterises a link where the traffic is more evenly spread over the hours of the day. This is of course not the only difference between the different land uses, but it is one that can explain differences in risks due to concentration of the traffic to certain hours. It should also be noted that land use clearly outperformed variables such as link environment and street function, which had in earlier models been used to describe the environments in which the links are located.

The short measurements of the exposure of vulnerable road users have been shown to contribute significantly to the possibility to explain and predict the safety situation of vulnerable road users. Still, the short measurements have also been shown to have weaknesses. When there are very few vulnerable road users in the environment only fifteen minutes is too short a period to count the flows. The variation over the day of VRU flows is not well established and varies between individual streets and so the extrapolation of short measures to daily flows is very 'shaky'. Better estimates of VRU exposure would most likely have made it possible to create models explaining even more of the systematic variation of the VRU accidents than the models developed in this work.

In order to describe the exposure of vulnerable road users both VRUs crossing and VRUs moving along the street should be included in the counting. VRUs crossing the street are exposed to the traffic on the main street while VRUs moving along the street are exposed to turning and crossing traffic in minor intersections on the link. It was a new finding that it is better to use the sum of VRUs crossing and moving along the street in the models than VRUs crossing and VRUs moving along the street separately.

7.2.2 The use of preset parameters for exposure variables

When comparing the fit of developed models between models with preset parameters for exposure variables and models with freely estimated parameters, the models with freely estimated parameters have a slightly better fit. This is an expected result as the preset models are forcing parameters to take on specific values instead of letting the estimation process optimise the parameters for best fit. However, when validating the models against new data the preset models for pedestrian and vehicle accidents are better at predicting the numbers of accidents than the freely estimated models. The conclusion drawn is that the 'true' values for the exposure parameters are closer to the preset parameter values than to the freely estimated ones. The preset parameters were 0.5 for both VRU and vehicle flow exponents in models for VRU accidents, and 1.0 for vehicle flow exponent in the models for motor vehicle accidents. For bicyclist accidents the 'true' exponent for bicyclist exposure is likely to be somewhat lower than 0.5, close to 0.35.

7.2.3 The use of speeds in predictive accident models

Average speeds have been shown to be difficult to include in predictive accident models. The drivers' choice of speed is influenced by the characteristics of the street environment and the traffic. Because of this the speed level on a street correlates with most of the other variables used in the models and the 'true' underlying relation between speed and number of accidents is never found when freely estimating the parameter values of the model. Instead the variable Speed is included with parameter values totally contradicting all empirical and theoretical evidence about the relationship between speed and accidents. When variables describing the street environment and the traffic are added, the parameter values for speed moves towards the established relationship but never come close.

7.3 Implications for practice

7.3.1 Model choice

Which of the models developed are the best for predicting accidents? Several different models have been developed for each type of accident, and with various results. This section gives a discussion of the differences and a recommendation as to which of the model to use.

Pedestrian and bicyclist models

For bicyclist accidents the model with bicyclist exposure included and no parameters preset is recommended (Table 63), while for pedestrian accidents the model with pedestrian exposure included and the exposure parameters preset to exponents of 0.5 for vehicle flow and pedestrian exposure is recommended (Table 64).

Models without variables describing VRU exposure have a poor ability to predict the number of VRU accidents. In order to be able to make sound statements about the number of VRU accidents, these variables have to be included. Models with only the existing EVA variables are very poor at predicting VRU accidents and should not be used.

Models with the use of a preset relation (the so-called power model) between the number of accidents and average speeds show fairly good prediction strength. The results in the modelling process with freely estimated parameters, however, indicate that speed is a very unstable explanatory variable. When other variables are added the exponent of speed varies widely. Speed is the drivers' main tool to respond to perceived changes in risk; when the street environment changes, so does the drivers' speed. As a result the other variables included in the models correlate strongly with speed and their parameter values change considerably when speed is included. On the whole, speed is not recommended as a variable to use in predictive accident models although its contribution to the degree of explanation can be large.

When the parameters of the models are estimated freely the exponents for VRU exposure receive values lower than 0.5 and those of vehicle flow values larger than 0.5. When presetting both exponents to 0.5 the bicyclist model lose in predictive ability. The predictive ability of the pedestrian model on the other hand is improved. For describing bicyclist accidents the model with freely estimated parameters is clearly the best. For describing variation in pedestrian accidents, on the other hand, the model with preset parameters for exposure variables is the best. The pedestrian model with preset parameter values for exposure both predicts the number of accidents better and comply with earlier findings about the relation between pedestrian exposure and number of pedestrian accidents.

Table 63 Model for bicycle accidents, no preset parameters

Model: Bicycle, no preset				
Deviance	**df**	**Perc.exp.**		
501	381	60%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Mean	2.16E-04		-5.93	0.000
NCXP		0.35	5.50	0.000
Flow		0.76	4.93	0.000
Landuse_ABC	1			
Landuse_IX	0.60		-2.54	0.011
Func_GIF	1			
Func_T	1.36		1.86	0.064
Func_C	1.94		2.27	0.024
Vis_Good	0.71		-2.03	0.043
Vis_Medium	1			
Vis_Poor	1.62		0.96	0.335

Table 64 Model for pedestrian accidents, exposure preset to $Flow^{0.5}$*$NPXP^{0.5}$

Model: Pedestrian, exposure preset				
Deviance	**df**	**Perc.exp.**		
286	384	83%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Mean	1.99E-04		-62.92	0.000
NPXP		0.5		
Flow		0.5		
Landuse_ABX	1			
Landuse_C	1.22		0.91	0.361
Landuse_I	0.27		-2.64	0.009
Func_GIF_T	1			
Func_C	1.94		2.39	0.017
VRUX_No	1			
VRUX_Yes	1.49		1.91	0.057

Vehicle accidents

The model recommended for predicting number of vehicle accidents is the model for all vehicle accidents with the exponent for vehicle flow preset to 1.0 (Table 65).

The division of vehicle accidents into separate models for single vehicle accidents and multiple vehicle accidents might work well in many cases, but the single vehicle accident data has in this data set been too scarce for fitting appropriate models for this accident type. Therefore the model including all vehicle accidents is recommended.

Models with preset parameter values for speed are not recommended for much the same reason as for the accident models for vulnerable road users; the correlation of speed with other variables is complex and produces unstable estimates of parameter values.

Table 65 Model for all vehicle accidents, exposure preset to flow[1]

Model: All vehicle accidents, exposure preset				
Deviance	**df**	**Perc.exp.**		
644	383	55%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Mean	1.71E-04		-57.49	0.000
Flow		1		
X34Km		0.40	5.27	0.000
SL_30	0.10		-1.61	0.109
SL_Rek30	0.48		-1.37	0.172
SL_50	1			
SL_70	1.17		0.96	0.339
ExitL_Yes	1			
ExitL_No	1.41		2.62	0.009

7.3.2 Application of the models

When using the models it is important to know their strengths and weaknesses. The models presented in this dissertation are based on cross-sectional empirical data from 389 links. This type of data is good for developing models explaining how the number of accidents varies between existing links. The models will not, however, reflect the causal relationships between the dependent variables and the number of accidents. The independent variables often correlate with each other which makes it impossible to identify any causal relationship with the number of accidents. Despite this, decision makers and traffic planners do not appreciate models with coefficients contradicting knowledge about causal relationships, and for this reason models with "false" coefficients are not trusted and used even for purposes, where they would perform well. In the development of models consideration has therefore been taken not to include variables when their corresponding parameter values assigned contradict established knowledge about their relationship with accidents. They should still not be used in predicting changes in accidents due to reconstruction of the street or other measures causing changes in the variables in the model. In that case a better estimate is to calculate the expected number of accidents without the measures and then use knowledge about the effects of the measure to predict the change.

Calculation of expected number of accidents

The models presented throughout the dissertation calculate the number of accidents per five year and km. This is important to note as the predicted number of accidents has to be divided by 5 and multiplied by the link length in order to get a predicted annual number of accidents. Below are the models given on equation form:

AccC = $2.16\times10^{-4}\times Length\times NCXP^{0.35}\times Flow^{0.76}\times Landuse\times Func\times Vis/5$

AccP = $1.99\times10^{-4}\times Length\times NPXP^{0.5}\times Flow^{0.5}\times Landuse\times Func\times VRUX/5$

AccV = $1.71\times10^{-4}\times Length\times Flow\times X34Km^{0.40}\times SpeedLimit\times ExitL/5$

Where:

AccC is the expected number of bicyclist accidents per year, AccP is the expected number of pedestrian accidents per year and AccV is the expected number of vehicle accidents per year.

The variables NCXP and NPXP are the total number of bicyclists and pedestrians, respectively, passing through the link environment in an average annual day. Flow is the Annual Average Daily motorised Traffic (AADT) for the link.

Landuse, Func, Vis, VRUX and ExitL are all categorical variables that are assigned dummy values depending on the value of the variable. The Dummy values for respective model are found in Table 66, Table 67 and Table 68.

Table 66 Bicyclist model - Dummy values for categorical variables

Bicyclist model - Dummy values for categorical variables		
Categorical variable	**Value**	**Dummy value**
Landuse	Residential, Institutional, Commercial	1.00
	Industrial, None	0.60
Street function	Thoroughfare/ entrance	1.00
	Tangential	1.36
	Centre	1.94
Visibility	Good	0.71
	Medium	1.00
	Poor	1.62

Table 67 Pedestrian model - Dummy values for categorical variables

Pedestrian model - Dummy values for categorical variables		
Categorical variable	Value	Dummy value
Landuse	Residential, Institutional, None	1.00
	Commercial	1.22
	Industrial	0.27
Street function	Thoroughfare/ entrance, Tangential	1.00
	Centre	1.94
VRUX (larger VRU crossing)	Yes	1.49
	No	1.00

Table 68 Vehicle model - Dummy values for categorical variables

Vehicle model - Dummy values for categorical variables		
Categorical variable	Value	Dummy value
Speed Limit	30 km/h	0.10
	Rec. 30 km/h	0.48
	50 km/h	1.00
	70 km/h	1.17
Presence of large exit	Yes	1.00
	No	1.41

7.3.3 Need for data on vulnerable road users' exposure

The main finding of this dissertation is that estimates of vulnerable road users' exposure are needed to make a good assessment of the safety situation for the vulnerable road users. This calls for putting more resources into measuring the flows of pedestrians and bicyclists. Many municipalities count bicyclists more or less on a regular basis but rarely as extensively as for the motorised vehicles.

The short measurements used in this project have added greatly to the assessment of the safety situation for vulnerable road users. The short measures are however unstable and longer measurements are recommended. In addition to prolonging the measurements at each link they can also preferably be split over several times during the day to counteract the hazard of extrapolating a daily flow from measurement from only one time of the day.

Studies of the variation of VRU flows over the day indicate that there are significant peaks during morning and late afternoon. As the peaks often are short it can be hazardous to do short measures close in time to the peak as the flow varies abruptly at the beginning and end of the peak. When doing measurements during these times it can be difficult to take proper account to time variation.

7.3.4 Vulnerable road users' safety

The results show that with more vulnerable road users on the link, the risk per VRU decreases. The number of accidents is proportional to the VRU flow to the power of an exponent of 0.5 or even less. This is in agreement with earlier findings for pedestrians and bicyclists (Jacobsen 2003, Ekman 1996). It indicates that it would be beneficial with a planning strategy where vulnerable road users are concentrated to fewer and busier paths to make the motorised road users more aware of them.

7.4 Implications for research

7.4.1 Lessons learned

The strength of correlation

The main lesson learned is that statistical modelling of accidents is a highly complicated task. The correlation between the independent variables violates the idea of independency between variables and makes it difficult to fit models with both a high degree of explanation and model parameters that agree with existing knowledge about the effect of the variables on the number of accidents. This was naturally also expected, but some relations are unexpectedly strong.

To minimise the risk for corrupted models due to correlation between the explanatory variables, a thorough control of correlations between variables must be done. In addition, the parameter values for the different variables should be kept under watch as more variables are included in the models. If the parameter values vary much when a new variable is added then there is reason to believe that there may be an unduly correlation between the added variable and the variable for which the parameter value changed.

Accident data consistency

From the occurrence of traffic accidents to the inclusion of the accidents in the data base used for this project there are many steps in the recording and compilation phases where data corruption may occur. One example of this is the lack of registration of several road users in accident reports in Uppsala, making multiple vehicle accidents appear as single vehicle accidents in the data. All accident data has to be thoroughly checked to minimise the risk for faulty data corrupting the models.

Parking

In other studies the variable parking has turned out useful in predicting number of accidents. The links studied however, indicate that parking is rarely allowed on main streets in Swedish urban areas thus not making it a very useful variable in the accident prediction models.

7.4.2 Future research needs

Measuring VRU exposure

The method used here for gathering data on vulnerable road users' exposure has been successful. The estimates of exposure produced have been very useful in predicting the variation of the accidents of vulnerable road users. The method can, however, be improved.

Limiting the field studies to very short time periods at each link produces uncertainties in the variables describing VRU exposure, partly because there is a stochastic variation in the flow, and partly because the systematic variation of VRU flows over the day is not very well known. There is a need for better knowledge about the variation of VRU flows over time. How much can the variation differ between different streets of the same type, and of different types?

The uncertainties in the estimation of the exposure variables produce a bias when estimating the model parameters. This bias is both dependent on the magnitude of the uncertainties in the estimates and how much the exposure varies between links. How large bias can be expected?

If the variation of VRU exposure over the day varies in different ways than motorised traffic does, how will this then affect the safety situation? The number of VRU accidents is dependent of both VRU exposure and vehicle flow. If these flows peak in different times of the day, how will this affect the number of accidents? Will this separation in time produce fewer or more accidents?

Dividing accidents into separate models by accident type

The division of motor vehicle accidents into separate models for multiple and single vehicle accidents has produced ambiguous results. The number of single vehicle personal injury accidents is low enough in urban areas to produce problems when estimating stable models. On the other hand, the models for multiple vehicle accidents are in some aspects better than models for all vehicle accidents. If more data can be obtained on motor vehicle accidents, a division of models according to accident type is promising and may produce better predictions of accident numbers.

The use of speed in accident models

It is known that given a situation where only the speed changes, the number of personal injury accidents change proportionally to the change in speed to the power of two. Yet, the models developed estimate parameters for speed that go considerably against the known relationship. The confounding circumstances that produce these misestimates deserve further exploration, especially as they would probably increase our understanding of the role of speed in driver's compensatory behaviour and in safety in general. The use of other indicators of speed than average speeds of all vehicles must also be explored.

Differences in accident rates between cities

There is a large difference in accident rates between different cities. Throughout the modelling a categorical variable has been used to identify the different cities in order to see if there are significant differences in accident rates between the cities. The smaller cities, Katrineholm and Danderyd, generally had lower accident rates than the larger cities. Even when the full models were fitted and the main differences in characteristics of the streets were accounted for, the difference in accident rates was still considerable. The city with the lowest accident rates, Danderyd, had roughly only half the accident rates compared to the large cities.

Is the accident rate generally higher on links in larger cities than in smaller cities? The results seem to indicate it. The results are however based only on five cities, as one of the cities, Västerås, has only been included with half the street network (all in the outer part).

Another possible reason behind the differences in accident rates is differences in how many of the accidents are reported by the police. Depending on differences in routines, experience and resources between different police districts the degree to which accidents are properly reported may vary.

Crossing strategies for vulnerable road users

Within the field studies, data was gathered regarding the crossing strategies of vulnerable road users, i.e. how they cross the street. Many earlier studies have been done concerning differences in risk levels for different solutions, for example differences in risk when crossing on pedestrian crossings and where there are no crossings. Studies show different risk levels for different solutions, but in the modelling the variable crossing strategies (XForm) has consistently had very weak explanatory strength for VRU accidents. No clear explanation for this has been identified.

Separation of bicyclists

The variable Separation of bicyclists has been included in two of the developed models for bicyclist accidents. In both models the variable has added significantly to the fit of the model to the modelling data. When the models, however, are used to predict the number of accidents in the validation data, the prediction strength is very poor compared to other models. It would seem that the variable Separation of bicyclists is not good for predicting the number of bicyclist accidents. This might however also be due to different strategies in the different cities regarding bicyclist separation from, or integration with, motorised traffic. The cities used in the validation studies have markedly different strategies for bicyclist separation compared to the three cities not used in the validation study.

Land use

The variable Landuse often comes out as a strong explaining variable for the number of accidents. The largest difference can be found between residential areas, industrial areas and Commercial areas (CBD, Central Business District). The number of accidents is generally higher in CBDs, and lower in industrial areas, compared to in residential areas. It might be argued that the exposure is higher in CBDs and lower in industrial areas, but the exposure is already accounted for in the models. Therefore there are some other reasons for the strong connection between land use and the number of accidents. The role of land use in predictive accident models needs further investigation.

References

Allain E., Brenac, T. (2001) Modèles linéaires generalises appliqués à l'étude des nombres d'accidents sur des sites routiers – Le modèle de Poisson et ses extensions, Recherche transports sécurité No 72, France

Brüde U., Larsson J. (1992) *Trafiksäkerhet i tätortskorsningar / VTI-meddelande 685,* VTI, Linköping, Sweden

Brüde U., Larsson J. (1993) *Models for predicting accidents at junctions where pedestrians and cyclists are involved. How well do they fit?*, Accident analysis and prevention 25 p499-509 Vol. 25, No.5 1993, Pergamon Press, Oxford, United Kingdom

Bång K.L., Carlsson A., Palgunadi (1995) *Development of speed-flow relationships for Indonesian rural roads using empirical data and simulation,* Transportation Research Record. 1995. (1484) pp24-32, Washington DC, USA

Carlsson G. (2004) *Kunskapssammanställning – Dödsolyckor och Hastighet*, NTF – National-föreningen för Trafiksäkerhetens Främjande, Stockholm, Sweden

Danielsson S. (1999) *Statistiska metoder vid analys av trafiksäkerhet,* Matematiska institutionen, Linköpings universitet, Linköping, Sverige

Ekman L. (1996) *On the Treatment of Flow in Traffic Safety Analysis,* Bulletin 136 Institutionen för Trafikteknik, LTH, Lund, Sweden

Ekman L. (2000) *Sänkt hastighet i bostadsområden: önskan eller verklighet?,* Bulletin 180 Institutionen för Teknik och samhälle, LTH, Lund, Sweden

Elvik R., Borger Mysen, A., Vaa, T. (1997) *Trafikksikkerhetshåndbok,* TÖI, Oslo, Norway

Elvik R., Christensen P., Amundsen A. (2004) *Speed and road accidents – An evaluation of the Power model,* TÖI, Oslo, Norway

Englund A., Gregersen N. P., Hydén C., Lövsund P., Åberg L. (1998) *Trafiksäkerhet – En kunskapsöversikt,* KFB & Studentlitteratur, Lund, Sweden

Finch D. J., Kompfner P., Lockwood C. R., Maycock G. (1994) *Speed, speed limits and accidents,* TRL, London, United Kingdom

Gaudry M., Lasarre S. et al. (2000) *Structural road accident models – The international DRAG family*, Elsevier Science, Oxford, United Kingdom

Greibe P. (2003) *Accident prediction models for urban roads*, Accident analysis and prevention Vol. 35 No.2 p273-285, 2003, Pergamon Press, Oxford, United Kingdom

Gårder P. (1989) *Pedestrian safety at traffic signals: A study carried out with the help of a traffic conflicts technique,* Accident analysis and prevention Vol. 21, No.5 1989, Pergamon Press, Oxford, United Kingdom

Haddon W. (1980) *Address options for the prevention of motor vehicle crash injury*, Israel Journal of Medical Sciences. 1980. 16(1) p 45-68, Jerusalem, Israel

Hauer E., Ng J. C. N., Lovell J. (1988) *Estimation of safety at signalised intersections* Transportation Research Record 1185, Washington D.C., USA

Hauer E. (1999) *Safety in Geometric Design Standards*, Department of Civil Engineering, University of Toronto, Canada

Jonsson T. (2001) *Effektmodeller för trafiksäkerhet i tättbebyggt område*, Bulletin 203, Institutionen för Teknik och samhälle, Lunds Universitet, Lund, Sweden

Jacobsen P.L. (2003) Safety in numbers: more walkers and bicyclists, safer walking and bicycling, Injury Prevention 2003;9:205–209, Montreal, Canada

Kulmala R. (1995) *Safety at rural three- and four-arm junctions – Development and application of accident prediction models*, VTT Technical Research Centre of Finland, Espoo, Finland

Maher M.J. (1989) Fitting predictive equations to accident data when there is uncertainty in the traffic flow estimates, Traffic management and road safety. Proceedings of seminar H held at the 17th PTRC Transport and planning summer annual meeting, University of Sussex, September 11-15 1989. Volume p323 pp171-81, PTRC Education research services Ltd, United Kingdom

Maher M.J., Summersgill I. (1996) *A comprehensive methodology for the fitting of predictive accident models*, Accident analysis and prevention 28 p281-296 Vol. 28, No.3 1996, Pergamon Press, Oxford, United Kingdom

NAG (2000) NAG Statistical Add-ins for Excel, The Numerical Algorithms Group Limited, Oxford, United Kingdom

Navon D. (2001) *The paradox of driving speed*, Accident analysis and prevention Vol. 35 No.3 p361-367, 2001, Pergamon Press, Oxford, United Kingdom

Nilsson G. (2004) Traffic *safety dimensions and the power model to describe the effect of speed on safety*, Bulletin 211, Institutionen för Teknik och samhälle, Lunds Universitet, Lund, Sweden

Olsson U. (2002) *Generalised Linear Models – An Applied Approach,* Studentlitteratur, Lund, Sverige

Ossenbruggen P.J., Pendharkar J.,Ivan J. (2001) *Roadway safety in rural and small urbanized areas,* Accident analysis and prevention 33 p485-498 Vol. 33, No.4 2001, Pergamon Press, Oxford, United Kingdom

Pasanen E. (1992) *Driving speeds and pedestrian safety; A mathematical model* Technical University of Helsinki, Helsinki, Finland

Peltola H. (2000) *Background and Principles of the Finnish Safety Evaluation Tool, TARVA* 13th ICTCT workshop, Corfu, Greece

Prop. 1996/97:137 (1996) *Nollvisionen och det trafiksäkra samhället*, Sveriges Regering, Stockholm, Sweden

Qin X., Ivan J.N., Ravishanker N. (2004) Selecting exposure measures in crash rate prediction for two-lane highway segments, Accident analysis and prevention Vol. 36 p183-191, 2004, Pergamon Press, Oxford, United Kingdom

Salusjärvi M. (1981) The speed limit experiments on public roads in Finland, VTT Publications 7/1981, Technical Research Centre of Finland, Espoo, Finland

SCB & SIKA (2002) *Vägtrafikskador 2001,* Statistics Sweden, Stockholm, Sweden

SCB & SIKA (2004) *Vägtrafikskador 2003,* Statistics Sweden, Stockholm, Sweden

SveKom (1997) *Farligt nära,* Svenska kommunförbundet, Stockholm, Sverige

SveKom (1998) *Lugna gatan,* Svenska kommunförbundet, Stockholm, Sverige

Vejdirektoratet (1998) *Uheldsmodel for bygader – Del 2: Model for strækninger*, Notat 59, Vejdirektoratet, Copenhagen, Denmark

Vejdirektoratet (2001) *Håndbog i trafiksikkerhedsberegninger – Brug av uheldsmodeller og andre vurderingar*, Rapport 220, Vejdirektoratet, Copenhagen, Denmark

Vägverket (2001a) *EFFEKTSAMBAND 2000 – Gemensamma förutsättningar*, Publikation 2001:75, Vägverket, Borlänge, Sverige

Vägverket (2001b) *EFFEKTSAMBAND 2000 – Nybyggnad och förbättring*, Publikation 2001:78, Vägverket, Borlänge, Sverige

Vägverket (2001c) *EFFEKTSAMBAND 2000 – Nybyggnad och förbättring - Handledning*, Publikation 2001:80, Vägverket, Borlänge, Sverige

Vägverket (2004) *Olycksrapport Skåne - 2003*, Publikation 2004:120, Vägverket, Borlänge, Sverige

WHO Mortality data base, Sweden 1999, http://www3.who.int/whosis/menu.cfm?path=mort, information gathered 2004-09-02, Released: 25 February 2003

Wilde G.J.S. (1994) Target risk, PDE Publications, Ontario, Canada

Zajac S.S., Ivan J.N. (2003) *Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut*, Accident Analysis and Prevention 35 p369–379, Vol. 35, No.3 2003, Pergamon Press, Oxford, United Kingdom

Verbal sources

Kulmala R., 2002, Discussion on number of accidents needed to construct sound models

Appendices

Appendix 1 Overview of dependent variables

Appendix 2 Residual plots for accident models

Appendix 3 Models – Actual speeds vs. Speed limit

Appendix 4 Models describing VRU accidents with only VRU exposure and/or vehicle flow

Appendix 1 Overview of dependent variables

Dependent variables used in the modelling with short descriptions and, for categorical variables, the categories.

Categorical variables

Variable	Description	Categories					
City	Dummy variable for the city	Uppsala	Katrineholm	Danderyd	Linköping	Norrköping	Västerås
Landuse	Landuse along the street	Residential (B)	Institutional (A)	Central Business District (C)	Industrial (I)	No buildings (X)	
Vis	Visibility	Good	Medium	Poor			
Lanes	Number of lanes	2	3*	4			
SL	Speed limit	30 km/h	Rec. 30 km/h	50 km/h	70 km/h		
Park	Parking along the street	No parking (NP)	Parking in the street, one side (PG1)	Parking in the street, both sides (PG2)	Parking slots, one side (PF1)	Parking slots, both sides (PF2)	
LEnv	Link environment	Outer (Y)	Middle (M)	Inner (C)			
Func	Street function	Thouroughfare /entrance (GIF)	Tangential (T)	City (C)			
SepV	Separation of vehicles	None (I)	Mixed (B)	Separated (S)			
XForm	Crossing form for VRUs	Freely (F)	Canalised to non-signalised crossings (Ö)	Canalised to signalised crossings (S)	Grade separated (P)	No crossing (X)	
SepC	Separation of bicyclists	Separate bicycle path (Cb)	Mixed traffic (B)	No biking (X)			
GrNPXP	Group variable for NPXP	Low	Medium	High			
GrNCXP	Group variable for NCXP	Low	Medium	High			
ExitL	Existance of large exits	Yes	No				
ExitS	Existance of small exits	Yes	No				
VRUX	Existance of dedicated VRU crossing	Yes	No				

* Merged with 2 in the models

Covariates

Variable	Description
Flow	Vehicle flow per day (AADT)
NPX	Number of pedestrians crossing the street per day
NPXKm	Number of pedestrians crossing the street per day and km
NPP	Number of pedestrians walking along the street per day
NPXP	Number of pedestrians crossing and walking along the street per day
NCX	Number of bicyclists crossing the street per day
NCXKm	Number of bicyclists crossing the street per day and km
NCP	Number of bicyclists walking along the street per day
NCXP	Number of bicyclists crossing and walking along the street per day
Speed	Average vehicle speeds (km/h)
DSpeed	Standard deviation of vehicle speeds (km/h)
X3Km	Number of three arm intersections per km
X4Km	Number of four arm intersections per km
X34Km	Number of intersections per km
DX	Average distance between intersections (km)

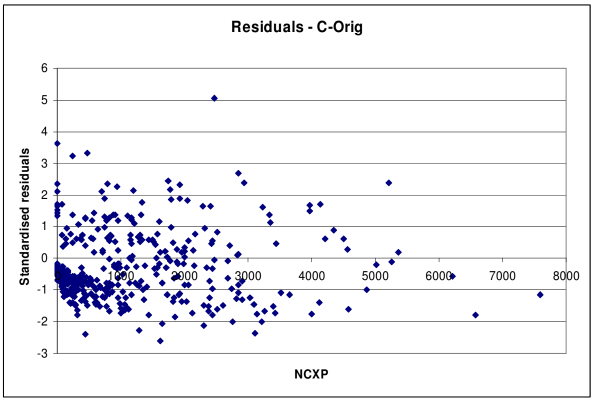
Appendix 2 Residual plots for accident models

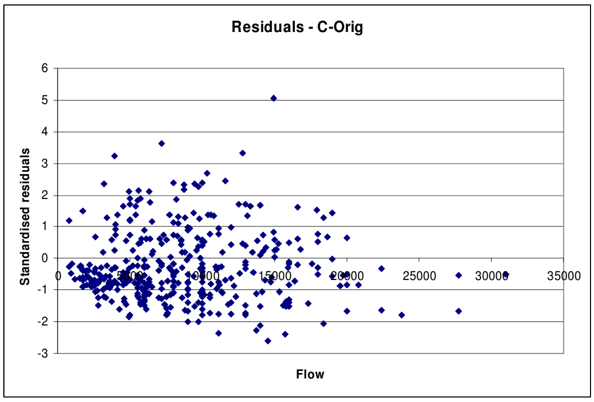
Key to category numbers for categorical variables

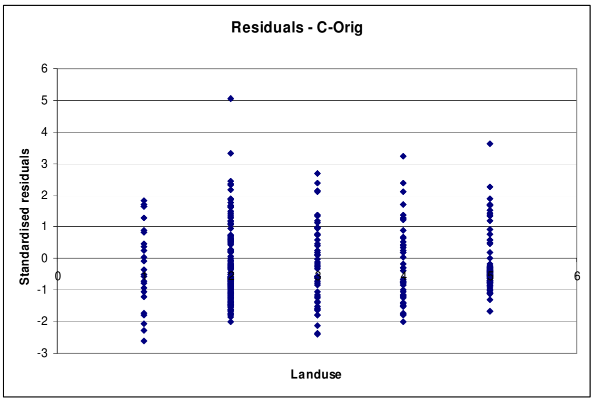
	1	2	3	4	5
Landuse	A	B	C	I	X
Func	GIF	T	C		
Vis	G	M	P		
SepC	Mix	Sep	X		
VRUX	Yes	No			
SL	30	Rek30	50	70	
SepV	None	Mix	Sep		

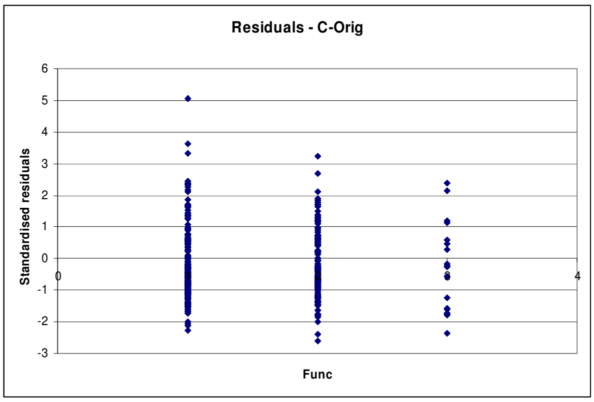
Variables used in the models

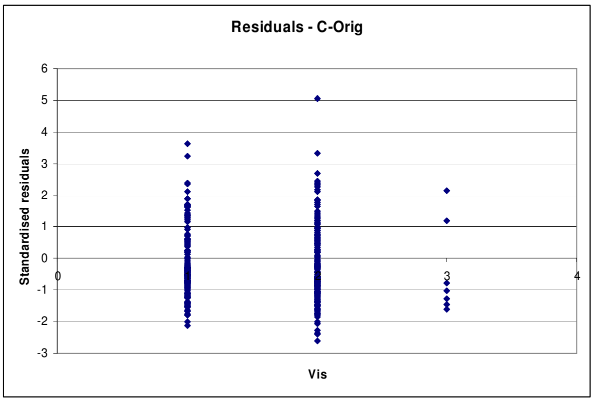
C	Orig	NCXP+Flow+Landuse+Func+Vis
	Preset	(NCXP+Flow+)Vis+Landuse+SepC
P	Orig	NPXP+Landuse+Flow+Func+VRUX+X34Km
	Preset	(NPXP+Flow+)Landuse+Func+VRUX
VV	Orig	Flow+X34Km+Func+SL
	Preset	(Flow+)X34Km+SL+Func+SepV

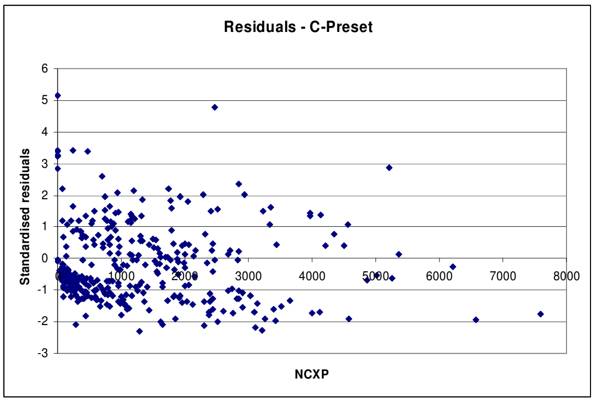

Residuals - C-Orig
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
1000
2000
3000
4000
5000
6000
7000
8000
NCXP

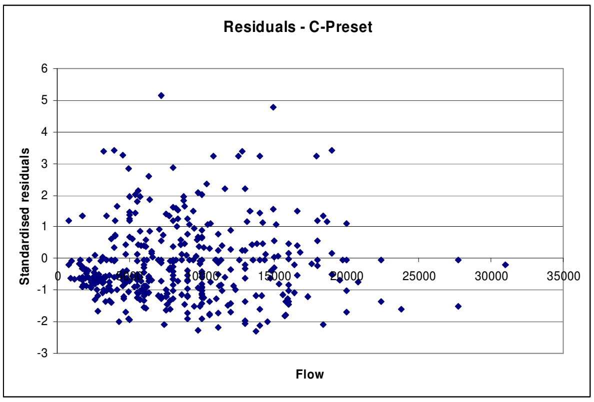

Residuals - C-Orig
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
5000
10000
15000
20000
25000
30000
35000
Flow

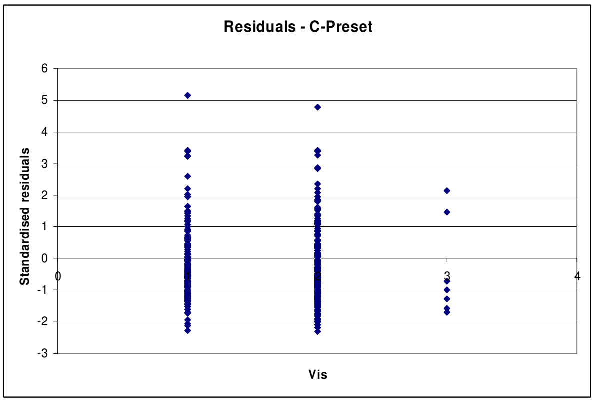

Residuals - C-Orig
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
6
Landuse

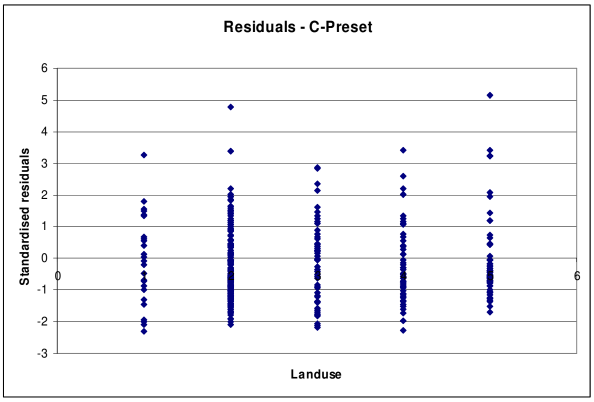

Residuals - C-Orig
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
4
Func

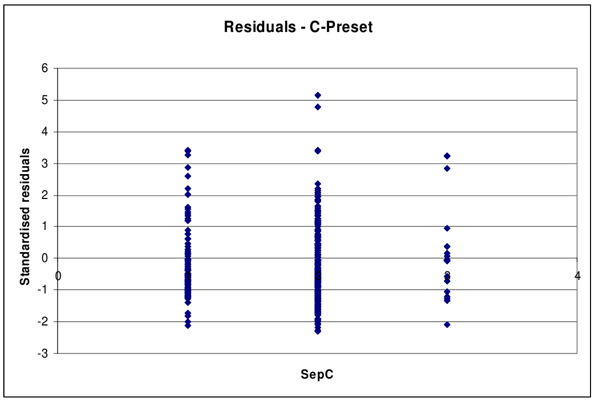

Residuals - C-Orig
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
3
4
Vis

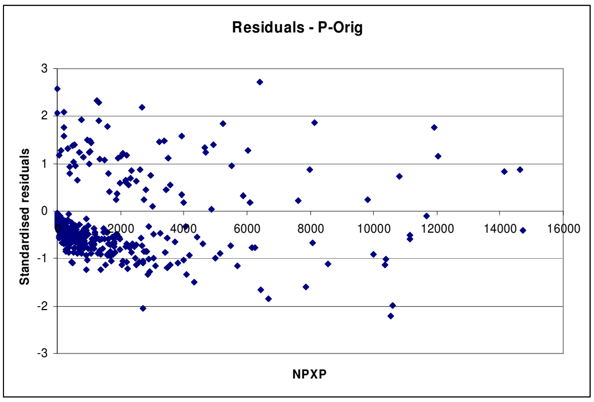

Residuals - C-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
1000
2000
3000
4000
5000
6000
7000
8000
NCXP

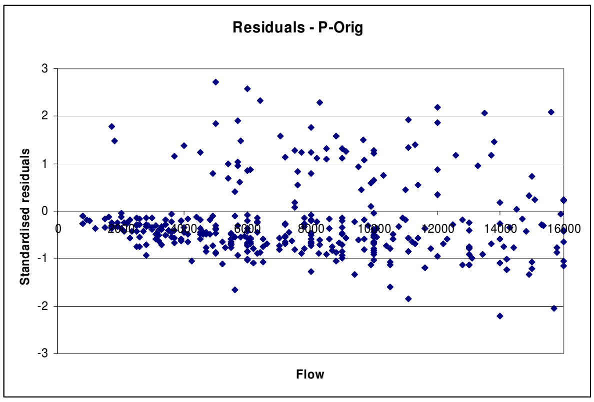

Residuals - C-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
5000
10000
15000
20000
25000
30000
35000
Flow

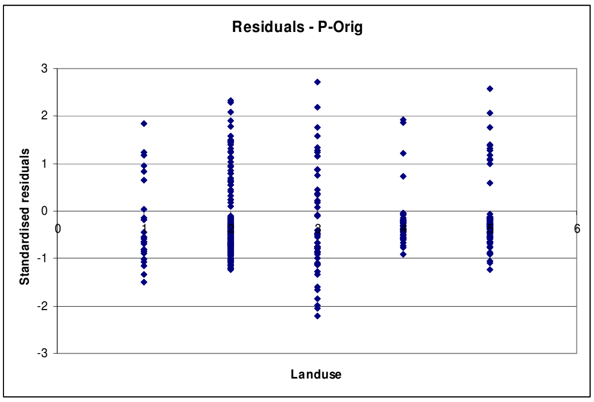

Residuals - C-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
3
4
Vis

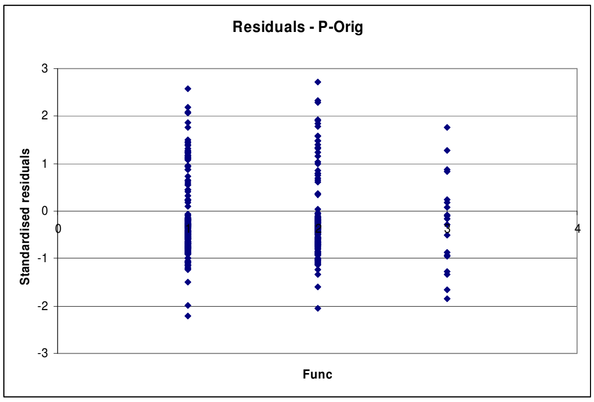

Residuals - C-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
6
Landuse

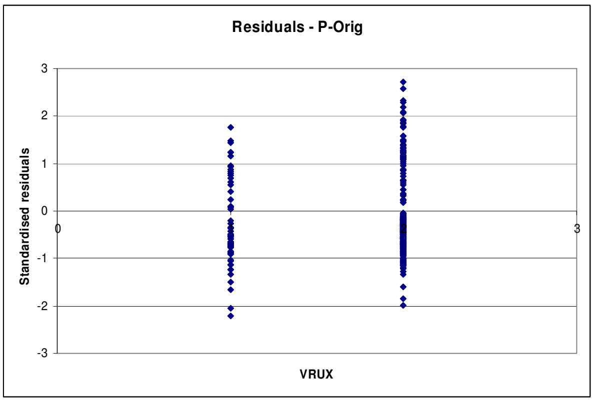

Residuals - C-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
0
4
SepC

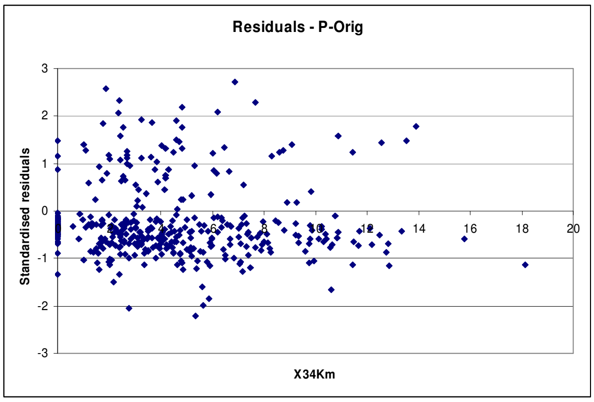

Residuals - P-Orig
Standardised residuals
3
2
1
0
-1
-2
-3
2000
4000
6000
8000
10000
12000
14000
16000
NPXP

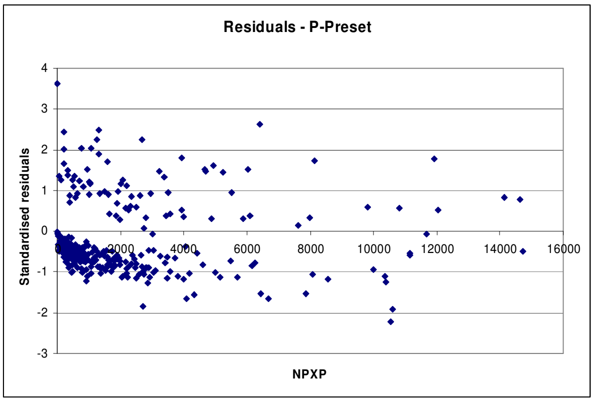

Residuals - P-Orig
Standardised residuals
3
2
1
0
-1
-2
-3
0
16000
Flow

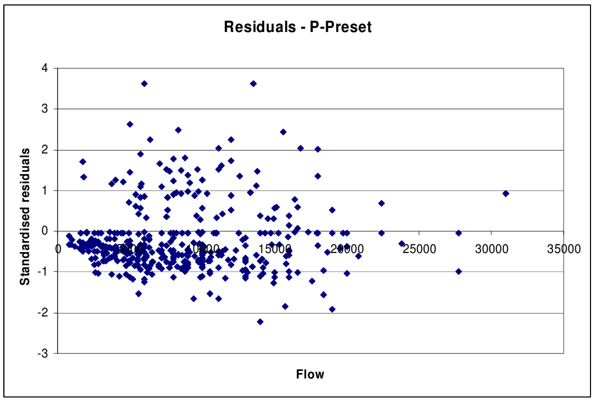

Residuals - P-Orig
Standardised residuals
3
2
1
0
-1
-2
-3
0
6
Landuse

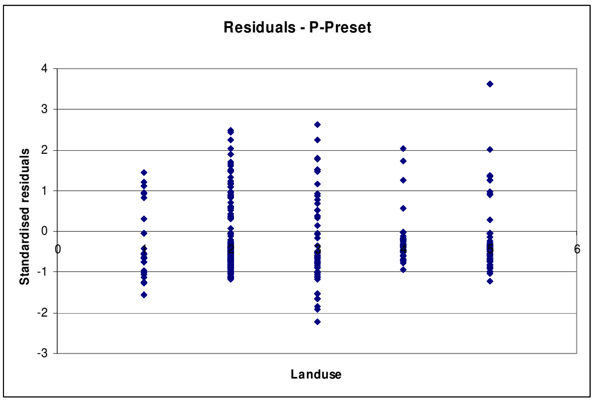

Residuals - P-Orig
Standardised residuals
3
2
1
0
-1
-2
-3
0
4
Func

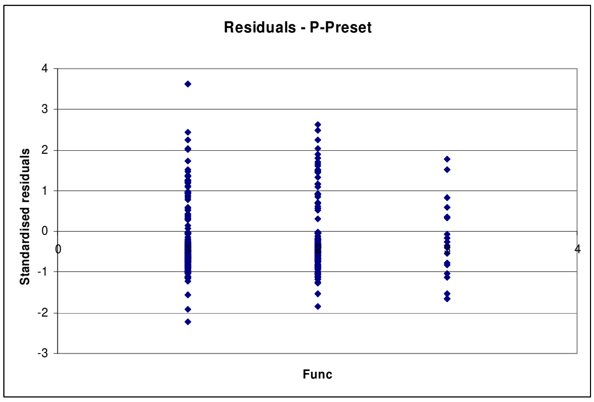

Residuals - P-Orig
Standardised residuals
3
2
1
0
-1
-2
-3
0
3
VRUX

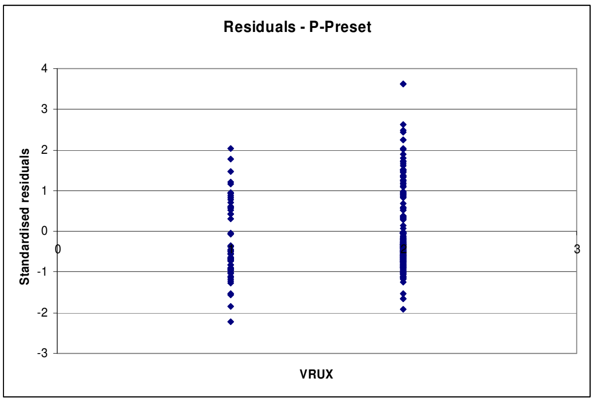

Residuals - P-Orig
Standardised residuals
3
2
1
0
-1
-2
-3
14
16
18
20
X34Km

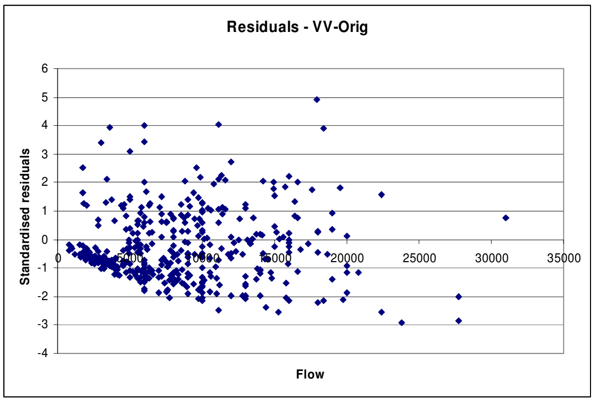

Residuals - P-Preset
Standardised residuals
NPXP
0 2000 4000 6000 8000 10000 12000 14000 16000
4 3 2 1 0 -1 -2 -3

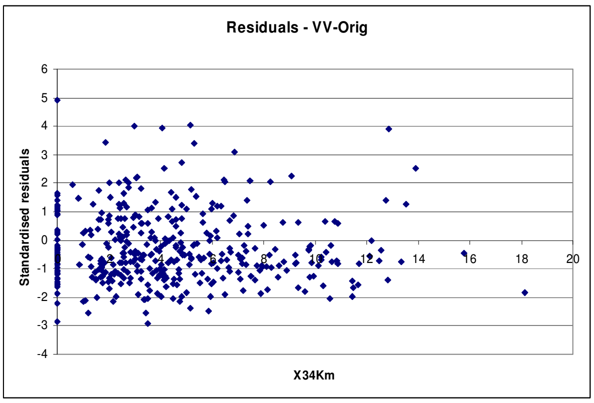

Residuals - P-Preset
Standardised residuals
Flow
0 5000 10000 15000 20000 25000 30000 35000
4 3 2 1 0 -1 -2 -3

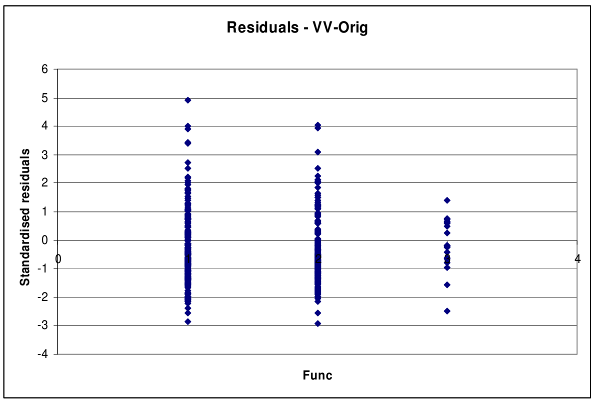

Residuals - P-Preset
Standardised residuals
Landuse
0 6
4 3 2 1 0 -1 -2 -3

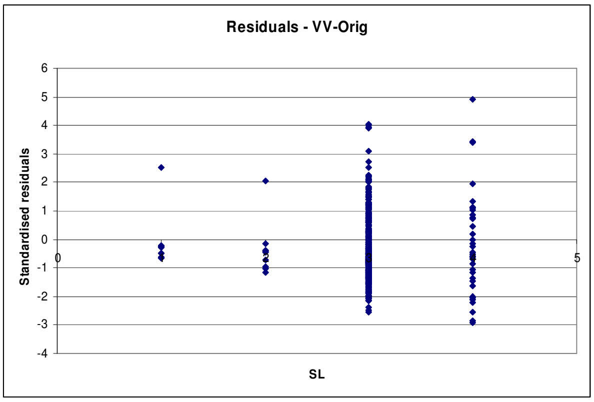

Residuals - P-Preset
Standardised residuals
Func
0 4
4 3 2 1 0 -1 -2 -3

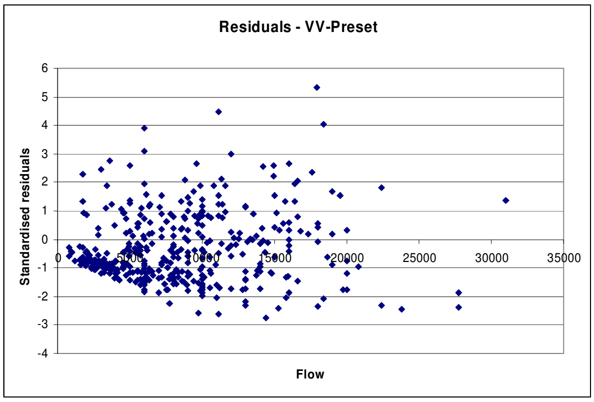

Residuals - P-Preset
Standardised residuals
VRUX
0 3
4 3 2 1 0 -1 -2 -3

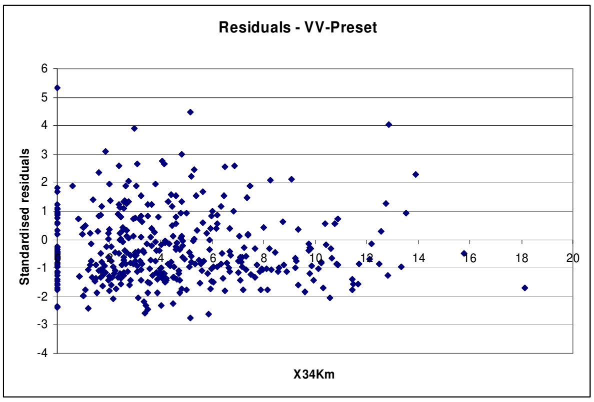

Residuals - VV-Orig
Standardised residuals
Flow
0 5000 10000 15000 20000 25000 30000 35000
6 5 4 3 2 1 0 -1 -2 -3 -4

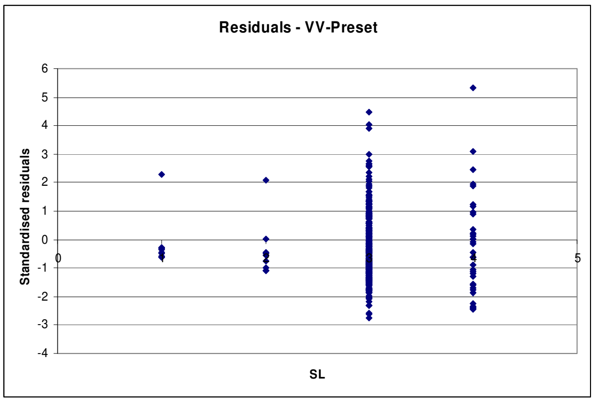

Residuals - VV-Orig
Standardised residuals
X34Km
0 2 4 6 8 10 12 14 16 18 20
6 5 4 3 2 1 0 -1 -2 -3 -4

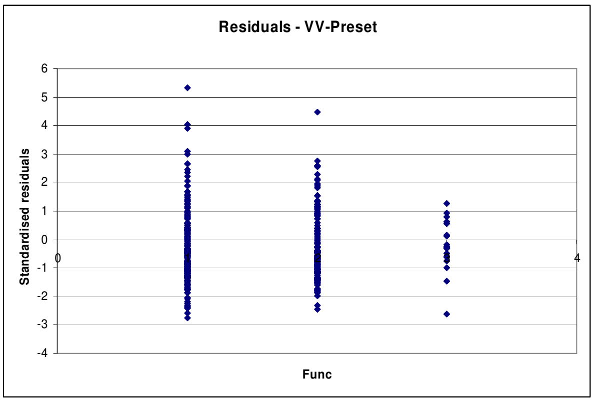

Residuals - VV-Orig
Standardised residuals
Func
0 4
6 5 4 3 2 1 0 -1 -2 -3 -4

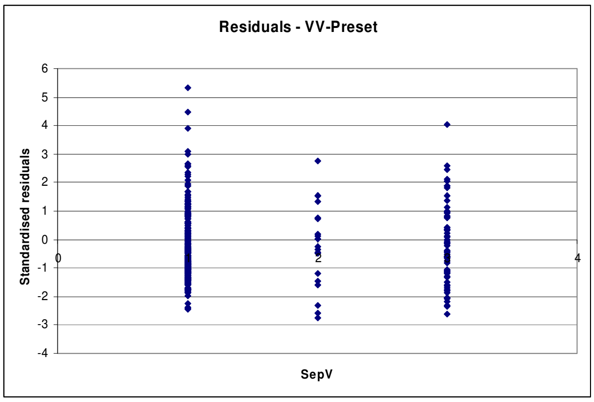

Residuals - VV-Orig
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
-4
0
5
SL


Residuals - VV-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
-4
0
5000
10000
15000
20000
25000
30000
35000
Flow


Residuals - VV-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
-4
0
4
6
8
10
12
14
16
18
20
X34Km


Residuals - VV-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
-4
0
5
SL


Residuals - VV-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
-4
0
4
Func


Residuals - VV-Preset
Standardised residuals
6
5
4
3
2
1
0
-1
-2
-3
-4
0
2
4
SepV

Appendix 3 Models – Actual speeds vs. Speed limit

A. Models for comparing speed vs. speed limit - potential outlier links not removed – models without preset

Original models

Bicyclists

No preset / Speed limit

Deviance	df	Perc. Exp.
545	383	48%

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-7.33	0.95	-7.74	-5.75
NCXP	0.39	0.05	8.47	6.29
Flow	0.59	0.11	5.63	4.18
SL_ 30	-1.67	1.00	-1.67	-1.24
SL_Rek30	-0.13	0.32	-0.40	-0.30
SL_ 70	-0.54	0.22	-2.41	-1.79

Pedestrians

No preset / Speed limit

Deviance	df	Perc. Exp.
305	383	71%

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-9.21	1.63	-5.66	-5.25
NPXP	0.59	0.08	7.52	6.98
Flow	0.50	0.19	2.72	2.53
SL_ 30	-0.23	0.72	-0.32	-0.30
SL_Rek30	0.56	0.35	1.60	1.49
SL_ 70	0.15	0.40	0.37	0.35

Vehicle-Vehicle

No preset / Speed limit

Deviance	df	Perc. Exp.
704	384	50%

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-9.90	0.86	-11.49	-8.04
Flow	1.19	0.09	12.78	8.94
SL_ 30	-1.86	1.00	-1.85	-1.30
SL_Rek30	-1.04	0.50	-2.06	-1.44
SL_ 70	-0.11	0.11	-0.96	-0.67

Vehicle-Single

No preset / Speed limit

Deviance	df	Perc. Exp.
258	384	45%

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-14.59	2.44	-5.97	-5.64
Flow	1.49	0.26	5.69	5.38
SL_ 30	-7.44	43.87	-0.17	-0.16
SL_Rek30	0.59	0.59	0.99	0.94
SL_ 70	-1.08	0.41	-2.63	-2.48

Bicyclists

No preset / Speed

Deviance	df	Perc. Exp.
531	385	53%

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-1.61	1.39	-1.16	-0.86
NCXP	0.35	0.05	7.63	5.67
Flow	0.56	0.10	5.47	4.06
Speed	-1.37	0.26	-5.23	-3.88

Pedestrians

No preset / Speed

Deviance	df	Perc. Exp.
295	385	78%

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-2.87	2.40	-1.20	-1.11
NPXP	0.41	0.08	4.89	4.54
Flow	0.67	0.18	3.62	3.36
Speed	-1.70	0.46	-3.66	-3.40

Vehicle-Vehicle

No preset / Speed

Deviance	df	Perc. Exp.
700	386	51%

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-6.79	1.09	-6.25	-4.37
Flow	1.24	0.09	13.67	9.56
Speed	-0.94	0.21	-4.36	-3.05

Vehicle-Single

No preset / Speed

Deviance	df	Perc. Exp.
268	386	39%

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-10.09	3.00	-3.36	-3.17
Flow	1.32	0.25	5.25	4.96
Speed	-0.80	0.59	-1.35	-1.28

B. Models for comparing speed vs. speed limit - potential outlier links not removed – preset models

Preset models

Bicyclists

Preset / Speed limit

Deviance	df	Perc. Exp.
607	385	30%

Parameter	Estimate	Standard Error	t value	Scaled t-value
		Parameter Estimates		
Mean	-7.26	0.06	-127.23	-94.47
SL_ 30	-1.67	1.00	-1.67	-1.24
SL_Rek30	-0.17	0.32	-0.53	-0.39
SL_ 70	-0.38	0.21	-1.80	-1.34

Bicyclists

Preset / Speed

Deviance	df	Perc. Exp.
598	387	34%

Parameter	Estimate	Standard Error	t value	Scaled t-value
		Parameter Estimates		
Mean	-3.04	0.97	-3.15	-2.34
Speed	-1.12	0.25	-4.40	-3.27

Pedestrians

Preset / Speed limit

Deviance	df	Perc. Exp.
318	385	63%

Parameter	Estimate	Standard Error	t value	Scaled t-value
		Parameter Estimates		
Mean	-8.43	0.10	-88.40	-82.06
SL_ 30	-0.21	0.71	-0.29	-0.27
SL_Rek30	0.62	0.35	1.80	1.67
SL_ 70	0.05	0.37	0.13	0.12

Pedestrians

Preset / Speed

Deviance	df	Perc. Exp.
309	387	71%

Parameter	Estimate	Standard Error	t value	Scaled t-value
		Parameter Estimates		
Mean	-3.23	1.38	-2.35	-2.18
Speed	-1.38	0.37	-3.72	-3.46

Vehicle-Vehicle

Preset / Speed limit

Deviance	df	Perc. Exp.
709	385	49%

Parameter	Estimate	Standard Error	t value	Scaled t-value
		Parameter Estimates		
Mean	-8.14	0.05	-157.83	-110.37
SL_ 30	-1.90	1.00	-1.89	-1.32
SL_Rek30	-1.03	0.50	-2.06	-1.44
SL_ 70	-0.03	0.11	-0.29	-0.20

Vehicle-Vehicle

Preset / Speed

Deviance	df	Perc. Exp.
707	387	50%

Parameter	Estimate	Standard Error	t value	Scaled t-value
		Parameter Estimates		
Mean	-4.89	0.83	-5.92	-4.14
Speed	-0.85	0.21	-3.95	-2.76

Vehicle-Single

Preset / Speed limit

Deviance	df	Perc. Exp.
261	385	43%

Parameter	Estimate	Standard Error	t value	Scaled t-value
		Parameter Estimates		
Mean	-10.04	0.13	-75.16	-71.01
SL_ 30	-7.44	41.33	-0.18	-0.17
SL_Rek30	0.58	0.59	0.98	0.93
SL_ 70	-0.89	0.40	-2.22	-2.10

Vehicle-Single

Preset / Speed

Deviance	df	Perc. Exp.
269	387	38%

Parameter	Estimate	Standard Error	t value	Scaled t-value
		Parameter Estimates		
Mean	-7.56	2.29	-3.30	-3.12
Speed	-0.68	0.59	-1.14	-1.08

C. Models for comparing speed vs. speed limit - potential outlier links removed – models without preset

Original models

Bicycle

No preset / Speed limit

Deviance	df
536	379

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-6.23	0.94	-6.66	-4.95
NPXP	0.33	0.04	8.24	6.11
Flow	0.52	0.11	4.83	3.59
SL_ 30	-9.38	45.34	-0.21	-0.15
SL_Rek30	-0.67	0.50	-1.32	-0.98
SL_ 70	-0.23	0.23	-1.00	-0.74

No preset / Speed

Deviance	df
544	381

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-3.10	1.51	-2.05	-1.52
NPXP	0.27	0.04	6.66	4.94
Flow	0.55	0.10	5.28	3.92
Speed	-0.81	0.31	-2.62	-1.94

Pedestrian

No preset / Speed limit

Deviance	df
294	379

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-8.77	1.66	-5.28	-4.90
NPXP	0.57	0.08	7.00	6.50
Flow	0.47	0.19	2.49	2.31
SL_ 30	-0.19	0.72	-0.26	-0.24
SL_Rek30	0.39	0.42	0.92	0.85
SL_ 70	0.31	0.41	0.76	0.70

No preset / Speed

Deviance	df
288	381

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-3.74	2.52	-1.48	-1.38
NPXP	0.41	0.09	4.77	4.43
Flow	0.63	0.19	3.35	3.11
Speed	-1.38	0.50	-2.76	-2.56

Vehicle-Vehicle

No preset / Speed limit

Deviance	df
665	378

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-10.56	0.90	-11.78	-8.24
Flow	1.26	0.10	13.03	9.11
SL_ 30	-1.82	1.00	-1.82	-1.27
SL_Rek30	-2.26	1.00	-2.26	-1.58
SL_ 70	0.04	0.12	0.34	0.24

No preset / Speed

Deviance	df
677	380

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-8.61	1.21	-7.11	-4.97
Flow	1.31	0.10	13.78	9.64
Speed	-0.63	0.24	-2.66	-1.86

Vehicle-Single

No preset / Speed limit

Deviance	df
243	381

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-14.27	2.50	-5.70	-5.38
Flow	1.46	0.27	5.40	5.10
SL_ 30	-7.44	43.94	-0.17	-0.16
SL_Rek30	-0.31	1.01	-0.30	-0.29
SL_ 70	-1.06	0.42	-2.55	-2.41

No preset / Speed

Deviance	df
252	383

Parameter Estimates

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-12.30	3.17	-3.88	-3.66
Flow	1.23	0.26	4.79	4.53
Speed	-0.02	0.66	-0.03	-0.03

D. Models for comparing speed vs. speed limit - potential outlier links removed – preset models

Preset models

Bicycle

Preset / Speed limit

Deviance	df
588	381

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-7.25	0.06	-126.07	-93.61
SL_ 30	-8.52	32.29	-0.26	-0.20
SL_Rek30	-0.72	0.50	-1.42	-1.06
SL_ 70	-0.39	0.21	-1.84	-1.36

Preset / Speed

Deviance	df
589	383

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-3.38	1.00	-3.38	-2.51
Speed	-1.03	0.26	-3.92	-2.91

Pedestrian

Preset / Speed limit

Deviance	df
306	381

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-8.48	0.10	-86.02	-79.84
SL_ 30	-0.16	0.71	-0.23	-0.21
SL_Rek30	0.42	0.42	0.99	0.92
SL_ 70	0.23	0.37	0.62	0.57

Preset / Speed

Deviance	df
301	383

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-4.48	1.51	-2.96	-2.75
Speed	-1.05	0.40	-2.61	-2.42

Vehicle-Vehicle

Preset / Speed limit

Deviance	df
673	379

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-8.16	0.05	-156.46	-109.41
SL_ 30	-1.88	1.00	-1.88	-1.31
SL_Rek30	-2.25	1.00	-2.25	-1.57
SL_ 70	0.12	0.11	1.10	0.77

Preset / Speed

Deviance	df
689	381

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-6.03	0.92	-6.54	-4.57
Speed	-0.55	0.24	-2.31	-1.62

Vehicle-Single

Preset / Speed limit

Deviance	df
246	382

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-10.07	0.14	-73.28	-69.23
SL_ 30	-7.41	41.33	-0.18	-0.17
SL_Rek30	-0.33	1.01	-0.33	-0.31
SL_ 70	-0.87	0.40	-2.16	-2.04

Preset / Speed

Deviance	df
253	384

Parameter	Estimate	Standard Error	t value	Scaled t-value
Mean	-10.59	2.54	-4.16	-3.93
Speed	0.09	0.65	0.14	0.13

Appendix 4 Models describing VRU accidents with only VRU exposure and/or vehicle flow

Bicyclist accidents

Model: Bicyclist accidents, Flow				
Deviance	**df**	**Perc.exp.**		
701	387	3%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	8,22E-03		-4,16	0,000
Flow		0,59	4,72	0,000

Model: Bicyclist accidents, NCXP				
Deviance	**df**	**Perc.exp.**		
585	387	38%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	7,65E-02		-5,65	0,000
NCXP		0,47	7,52	0,000

Model: Bicyclist accidents, NCXP & Flow				
Deviance	**df**	**Perc.exp.**		
556	386	46%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	9,61E-04		-5,56	0,000
NCXP		0,42	6,95	0,000
Flow		0,52	3,85	0,000

Pedestrian accidents

Model: Bicyclist accidents, Flow				
Deviance	**df**	**Perc.exp.**		
415	387	4%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	9,50E-04		-4,48	0,000
Flow		0,72	4,29	0,000

Model: Bicyclist accidents, NCXP				
Deviance	**df**	**Perc.exp.**		
317	387	65%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	6,14E-03		-8,26	0,000
NPXP		0,66	8,47	0,000

Model: Bicyclist accidents, NCXP & Flow				
Deviance	**df**	**Perc.exp.**		
308	386	71%		
Parameter Estimates				
Parameter	Dummy	Exponent	Scaled t-value	t-prob.
Constant	8,18E-05		-5,45	0,000
NPXP		0,60	7,50	0,000
Flow		0,53	2,75	0,006