

The final publication is available in: Accident Analysis and Prevention, 91, 2016, 175-182.

10.1016/j.aap.2016.03.005

Safety-in-numbers: Estimates based on a sample of pedestrian crossings in Norway

Rune Elvik

Institute of Transport Economics

Gaustadalleen 21, NO-0349 Oslo, Norway

E-mail: re@toi.no

ABSTRACT

Safety-in-numbers denotes the tendency for the risk of accident for each road user to decline as the number of road users increases. Safety-in-numbers implies that a doubling of the number of road users will be associated with less than a doubling of the number of accidents. This paper investigates safety-in-numbers in 239 pedestrian crossings in Oslo and its suburbs. Accident prediction models were fitted by means of negative binomial regression. The models indicate a very strong safety-in-numbers effect. In the final model, the coefficients for traffic volume were 0.05 for motor vehicles, 0.07 for pedestrians and 0.12 for cyclists. The coefficient for motor vehicles implies that the number of accidents is almost independent of the number of motor vehicles. The safety-in-numbers effect found in this paper is stronger than reported in any other study dealing with safety-in-numbers. It should be noted that the model

explained only 21 percent of the systematic variation in the number of accidents. It therefore cannot be ruled out that the results are influenced by omitted variable bias. Any such bias would, however, have to be very large to eliminate the safety-in-numbers effect.

Key words: safety-in-numbers; cyclists; pedestrians; pedestrian crossings; accident prediction model

1 INTRODUCTION

A transport policy objective in many countries is to curtail growth in the use of cars and promote walking or cycling. Important reasons for adopting this policy objective are to reduce global warming and improve public health. Pedestrians and cyclists have a higher injury rate per kilometre of travel than car occupants (Bjørnskau 2015). If more people walk or cycle, one may expect the number of traffic injuries to increase. A counterargument is that the injury rate for pedestrians and cyclists is not constant, but subject to a “safety-in-numbers” effect, which means that the larger the number of pedestrians or cyclists, the lower the injury rate for each pedestrian or cyclist. If sufficiently strong, this protective effect may to a large extent counteract and perhaps eliminate the increase in the number of injuries that would otherwise be expected when there is more walking or cycling.

Unfortunately, there are many problems in estimating the safety-in-numbers effect and particularly in determining the causality of the effect. In the first place, data on the number of pedestrians or cyclists tend to come from short-term counts that may be associated with considerable uncertainty (Kröyer 2015). In the second place, the reporting of accidents involving pedestrians or cyclists in official accident statistics is very low, in particular for cyclists (Lahrmann 2015). In the third place, nearly all studies of safety-in-numbers rely on cross-sectional data, which make it difficult to establish causal relationships. In a recent review, Elvik and Bjørnskau (2016) concluded that no studies of safety-in-numbers have controlled adequately for all relevant confounding variables and that one cannot conclude that these studies have uncovered a causal relationship.

A fourth problem is that there are many ways of developing and fitting the accident prediction models by means of which the safety-in-numbers effect is estimated (Hauer 2015) (note: the word “effect” is used as shorthand only and does not necessarily imply a claim of causality). Results may vary depending on, for example, which variables a model includes and how the statistical relationship between these variables and the number of accidents is modelled.

While it is difficult to establish causal relationships in a single study, replication of studies may, as the number of studies grows, reveal consistent patterns that may at least suggest causality. Consistency in the relationship between a cause and its effect (same cause, same effect) is one of the oldest criteria of causality. If a safety-in-numbers effect has been reproduced consistently in a range of countries and during a long period of time, that at least shows that it reflects a general tendency, which is robust with respect to the many differences between the individual studies.

The main objective of this paper is to estimate the safety-in-numbers effect in a sample of pedestrian crossings in Norway. Part of the sample was used in a previous study (Elvik, Sørensen and Nævestad 2013), but it has now been enlarged as explained in the next section. The paper tries to implement the stepwise approach to regression modelling in road safety proposed by Hauer (2015).

2 DATA AND METHOD

2.1 Sample of pedestrian crossings

A data set consisting of 389 marked pedestrian crossings in and close to Oslo has been created by merging four data sets. The four data sets are described in reports by Amundsen and Sætre (2009), Sætre et al. (2010), Sørensen et al. (2010) and Sørensen and Nævestad (2012). These pedestrian crossings were selected for detailed safety inspections for one or more of the following reasons:

1. Accident history: crossings with a history of accidents were selected.
2. Accident severity: crossings where accidents were severe, in particular where there had been fatal accidents, were selected.
3. Speed limit: crossings located on roads with a speed limit of 50 or 60 km/h were selected.
4. Complaints: crossings for which the public had made complaints were selected.

The pedestrian crossings are not representative of all pedestrian crossings in Oslo and its suburbs. In particular, the mean number of accidents per crossing is likely to be considerably higher than for a typical pedestrian crossing in Oslo and its suburbs. The variables recorded for each pedestrian crossing are listed in Table 1.

Table 1 about here

There are three groups of variables. The first group consists of dependent variables, i.e. variables whose values are influenced by the other groups of variables listed in

Table 1. These variables include the count of injury accidents and the counts of injured road users according to injury severity.

The second group of variables describe traffic volume. Traffic volume is indicated both by means of estimates of the annual average daily number of cars, pedestrians or cyclists and by means of counts made when the pedestrian crossings were inspected. AADT is almost always estimated on the basis of short-term counts. As shown by Mensah and Hauer (1998), using an average value for traffic volume, rather than the volume prevailing at the time of each accident, may lead to bias when estimating the relationship between traffic volume and the number of accidents. AADT is, however, very often the only available data on traffic volume.

In most of the pedestrian crossings, traffic counts were made when the crossings were inspected. These counts were typically made during daytime on weekdays and for a period of six hours. Based on the counts, the number of pedestrians and cyclists crossing the road during the maximum hour was estimated. One could argue that these estimates might be more strongly related to the number of accidents than AADT, since most accidents involving motor vehicles and either pedestrians or cyclists happen in daytime when hourly traffic volume is higher than at night.

The third group of variables listed in Table 1 are various characteristics of the road layout and traffic control at the pedestrian crossings. This includes the number of directions from which vehicles may approach a pedestrian crossing (arms: an indicator of the number of traffic movements a pedestrian or cyclist must attend to when crossing the road), the number of lanes, the presence of a refuge, the presence

of traffic signal control, speed limit and the 85th percentile speed of approaching motor vehicles.

2.2 Analytic choices

Hauer (2015) emphasises the importance of making all analytic choices when developing an accident prediction model explicitly and stating the reasons for the choices that were made. Unless models are developed this way, one cannot know whether the final model was the best possible model, given the available data and the intended use of the model, or whether it was inferior. In this paper, the main analytic choices are:

1. Are the independent variables so highly correlated that there could be collinearity problems in developing a model?
2. Which set of variables describing traffic volume is most closely related to the count of accidents?
3. Which of the other independent variables should be included in a model?

2.3 Correlations among variables

To help answer the first question, a correlation matrix (Pearson correlation coefficients) was estimated. It is shown in Table 2. Most of the correlations are minor or moderate. Only three correlations are quite strong. These are the correlations between the two measures of traffic volume, which are highlighted in bold italics in Table 2.

Table 2 about here

Based on these correlations, a choice must be made between the two sets of traffic volume variables. To guide the choice, an analysis employing cumulative residuals plots (CURE-plots) (Hauer and Bamfo 1997, Hauer 2015) has been made.

2.4 Choice between variables indicating traffic volume

Negative binomial regression models were developed in order to determine which of the two sets of variables describing traffic volume were best suited for analysis. To fit the models, traffic volume variables were transformed to natural logarithms. A value of 1 was added to pedestrian and cyclist counts to avoid taking the logarithm of zero. The first model was based on the counts made when the pedestrian crossings were inspected. Figure 1 shows the CURE-plot based on the fitted values of the model.

Figure 1 about here

It is seen that model predictions are quite poor. For long stretches of the predicted values, the residuals are consistently negative or positive and have large values. The CURE-plot shows that the model fits the data badly. A similar model based on the AADT-values for traffic volume is shown in Figure 2.

Figure 2 about here

The variables were again transformed to natural logarithms and the value of 1 added for pedestrians and cyclists. This CURE-plot shows that the model fits the data well. The cumulative residuals oscillate around the value of zero and do not stray as far away from this value as the residuals in Figure 1. It is therefore concluded that the model based on AADT-estimates for the traffic volume variables is satisfactory.

2.5 Variables to be included in the model

Which variables to include in a model in addition to the traffic volume variables is decided according to whether the variables are safety-related and improve the precision of estimates of safety based on the model (Hauer 2015). The precision of estimates is assessed in terms of how well the model fits the data. The variables considered for inclusion in the model are: number of arms (number of directions from which vehicles can enter a pedestrian crossing), number of lanes, presence of refuge, presence of traffic signals, and either speed limit or 85th percentile speed.

Figure 3 shows how the mean number of accidents per pedestrian crossing depends on the number of directions from which vehicles can enter the crossing. It is seen that the mean number of accidents increases as the number of directions vehicles can enter from increases.

Figure 3 about here

There are only three values for the number of directions variable: 2, 3 and 4.. A linear function fits the data very well and with only three values for the variable, there is no point in searching for more complex functions.

Figure 4 shows the relationship between the number of lanes and the mean number of accidents per pedestrian crossing. When testing various functions in Excel, a second degree polynomial was found to best fit the data. As shown figure 4, this function does not fit the data very well.

Figure 4 about here

It is clear that the relationship between the variables is not linear. In Figure 4, the data point to the right has a large influence on the shape of the curve. This data point is, however, based only on two pedestrian crossings. When fitting the model, the two rightmost data points in Figure 4 will have a small influence. It is still useful to test whether including the number of lanes squared improves the fit of a model, as even the data points for 1, 2, 3, and 4 lanes suggest a non-linear relationship.

The presence of a refuge is a binary variable. The mean number of accidents in crossings without a refuge is 1.200. The mean number of accidents in crossings with a refuge is 1.374. This difference is not statistically significant. The presence of a refuge will nevertheless be included in order to test whether it improves the fit of the model. The presence of traffic signal control is also a binary variable. The mean number of accidents in crossings without traffic signals is 1.101. The mean number of accidents in crossings with traffic signals is 2.860. This difference is statistically significant. It is clear that the simple bivariate relationships between a refuge and accidents, and between traffic signals and accidents, are seriously confounded by, for example, traffic volume.

The final variable to be considered is speed. This can be entered either as the posted speed limit, which has the values of 30, 40, 50 or 60 km/h, or as the 85th percentile speed of approaching vehicles. The vast majority of crossings have speed limits of either 50 or 60 km/h. There was no clear relationship between speed limit and the number of accidents, which suggests that each speed limit should be entered as a dummy variable. When 85th percentile speeds were compared for pedestrian crossings having 0, 1, 2, ..., 11 accidents, no clear relationship emerged. Speed will

therefore be included in the model as a set of dummy variables for speed limit. The mean value for speed limit was 52.2 km/h and the mean value for the 85th percentile speed was 44.8 km/h.

2.6 Fitting the model in stages

The model was fitted in stages, in which the first stage included the traffic volume variables only. In the next stages, one new variables was added at each stage. The reason for developing the model in these stages, adding one new variable at each stage, is to assess the stability of the regression coefficients across the different model specifications. More specifically, any model will have omitted variables. Omission of a variable may cause bias in the estimates of regression coefficients. However, not including a variable believed to be relevant will not always cause bias. One sign that an omitted variable could cause bias is that the regression coefficients for the variables already included in a model change value when the omitted variable is added to the model. Thus, if regression coefficients remain unchanged as more variables are added to a model, this indicates that the omitted variables did not cause bias, since their addition to the model does not influence the values of the regression coefficients for the variables that were already included. This criterion is obviously quite weak, and it by no means ensures that a model is not affected by omitted variable bias. It does, however, weaken the argument for rejecting a model because it does not including everything that is known or believed to influence safety.

3 RESULTS

The models were fitted by means of negative binomial regression with a log link function. Regression coefficients were estimated by means of the maximum likelihood technique. Table 3 presents the models that were developed. The models are based on data for 239 of the 389 pedestrian crossings, as data were missing on pedestrian and cyclist volume for 150 crossings. Goodness-of-fit is indicated by the overdispersion parameter listed at the bottom of the Table. If a model explains all systematic variation in the number of accidents, the value of the overdispersion parameter is zero. A positive value indicates that there remains systematic variation not explained by the model.

Table 3 about here

All coefficients have the expected sign, but are mostly not statistically significant. With a few exceptions, the coefficients change little as new variables are added to the models. The addition of number of lanes squared changed the value of the coefficient for number of lanes considerably, indicating that the relationship was non-linear as indicated by the exploratory analysis. The coefficient for motor vehicle volume dropped in value when speed limits were added to the model and again when traffic signals were added to the model. At stage 7, the coefficient for motor vehicle volume had a much smaller value than at stage 1. The coefficients for pedestrian volume and cyclist volume remained more stable across model specifications. The final value of the coefficient for motor vehicle volume suggests that the number of accidents is almost independent of volume, which is quite surprising. The

coefficients for pedestrian and cyclist volume are also quite small and suggest a dramatic safety-in-numbers effect.

The models were, however, not successful in explaining systematic variation in the number of accidents. Figure 5 shows that the variables included in the model with the smallest value of the overdispersion parameter explained only about 21 percent of the systematic variation in the number of accidents.

Figure 5 about here

Traffic volume explained only a little more than 10 percent of the systematic variation in the number of accidents. The other independent variables, put together, explained slightly less than 11 percent of the systematic variation in the number of accidents.

4 DISCUSSION

The objective of this paper was to determine whether there is a safety-in-numbers effect at pedestrian crossings in Oslo and its suburbs. The results, if taken at face value, are striking and suggest a very strong safety-in-numbers effect. All the regression coefficients have smaller values than found in almost all other studies in the safety-in-numbers literature.

Nevertheless, the estimated coefficients are not entirely unprecedented. Geyer et al. (2006) estimated a coefficient for motor vehicles of 0.15, which is within the range of values found in this study. Harwood et al. (2008), in one model, estimated a coefficient of 0.05 for motor vehicle volume. With respect to pedestrian volume, the

lowest coefficient reported in the literature is 0.18 (Hall 1986). For cyclist volume, Turner et al. (2006) reported a coefficient of 0.09 in one of their models.

Although studies can be found that have estimated coefficients for traffic volume close to those found in this study, it is clear that the estimates are very low. It is therefore appropriate to examine if it is reasonable that the values should be so low, in particular the coefficient for motor vehicle volume. Figure 6 shows the relationship between AADT for motor vehicles and the number of accidents in the 239 pedestrian crossings that were included in the final model.

Figure 6 about here

There is a very weak correlation between motor vehicle volume and the number of accidents. The empirical integration routine proposed by Hauer and Bamfo (1997) was used to identify a suitable function to describe the relationship between motor vehicle volume and accidents. The following function described the empirical integral well:

$$\text{Empirical integral function} = 0.1746 \cdot \text{AADT}^{1.2451} \quad (R^2 = 0.9912)$$

This implies that the function relating motor vehicle volume to accidents should be (the function is the first derivative of the empirical integral):

$$\text{Function for motor vehicle volume} = 0.2174 \cdot \text{AADT}^{0.2451}$$

The exponent (0.2451) is close to the one found in the model including traffic volume variables only (0.190). This suggests that the value of the coefficient estimated in negative binomial regression is not wrong, but simply reflects the fact that there is a weak relationship between motor vehicle volume and the number of

accidents. Moreover, it is not uncommon that the estimated value of a regression coefficient is slightly attenuated as more variables are added to a model.

One difference between the models fitted in this paper and most previous studies, is that motor vehicle volume, cyclist volume and pedestrian volume were included in the same model. This is not common. Thus, Daniels et al. (2010) had data on traffic volume for light cars, heavy cars, motorcycles, mopeds, bicycles and pedestrians. In the models that were developed, however, only motor vehicle volume and cycle volume were included when cycle accidents was dependent variable and only motor vehicle volume and pedestrian volume when pedestrian accidents was the dependent variable (the models included additional variables not referring to traffic volume). A sensitivity analysis was therefore made, including only two terms for traffic volume (in addition to the other variables included in the most comprehensive model). In the model with motor vehicles and pedestrians, the coefficients were 0.04 for motor vehicles and 0.13 for pedestrians. The coefficient for pedestrians had a marginally larger value than when cyclist volume is also included (0.066; see Table 3). In the model including motor vehicles and cycles, the coefficients were 0.07 (motor vehicles) and 0.15 (cycles), again marginally larger values than when pedestrian volume is also included (0.048 and 0.12; see Table 3). The standard errors of the coefficients are too large to conclude that they really are different.

Given the fact the final model explained only a little more than 21 percent of the systematic variation in the number of accidents, there must be important variables that have not been included in the model. Therefore, bias due to omitted variables cannot be ruled out. For example, none of the models fitted in this paper included

data on drinking and driving, which has been found to be more common on low-volume roads than on high-volume roads (Vanlaar 2008). If drinking and driving is more common at pedestrian crossings with low traffic volume, not accounting for this will bias the regression coefficients for traffic volume, since they will partly include the effect of a factor that both varies according to traffic volume and influences the number of accidents.

Another relevant variable that ought to have been included, is the volume of heavy vehicles. Prato et al. (2014) found a statistically significant positive coefficient for heavy vehicle volume using bicycle accidents as dependent variable. When a heavy vehicle is involved, injuries to cyclists and pedestrians are likely to be more severe than when it is not involved.

5 CONCLUSIONS

The main conclusions of the research reported in this paper can be summarised as follows:

1. A very strong safety-in-numbers effect was found in a sample of 239 pedestrian crossings in Oslo and its suburbs by means of negative binomial regression analysis.
2. The regression coefficients were 0.05 for motor vehicle volume, 0.07 for pedestrian volume and 0.12 for cyclist volume. These values are lower than in nearly all other studies of the safety-in-numbers effect.
3. The model explained only about 21 percent of the systematic variation in the number of accidents. It therefore cannot be ruled out that if more variables

could have been included in the analysis, estimates of the safety-in-numbers effect would have changed.

SOURCE OF FUNDING

The research presented in this paper was funded by the Research Council of Norway, grant number 210486.

REFERENCES

- Amundsen, K. S., Sætre, K. 2009. Kvalitetssikring av gangfelt i 60-soner i Stor-Oslo distrikt, Asplan Viak for Statens vegvesen Region øst, Sandvika.
- Bjørnskau, T. 2015. Risiko i veitrafikken 2013-2014. TØI rapport 1448. Oslo, Transportøkonomisk institutt.
- Daniels, S., Brijs, T., Nuyts, E., Wets, G. 2010. Explaining variation in safety performance of roundabouts. *Accident Analysis and Prevention*, 42, 393-402.
- Elvik, R., Bjørnskau, T. 2016. Safety-in-numbers: A systematic review and meta-analysis of evidence. Forthcoming in *Safety Science*, Special Issue: Cycling Safety.
- Elvik, R., Sørensen, M. W. J., Nævestad, T.-O. 2013. Factors influencing safety in a sample of marked pedestrian crossings selected for safety inspections in the city of Oslo. *Accident Analysis and Prevention*, 59, 64-70.
- Geyer, J., Raford, N., Pham, T., Ragland, D. R. 2006. Safety in numbers. Data from Oakland, California. *Transportation Research Record*, 1982, 150-154.

- Hall, R. D. 1986. Accidents at four-arm single carriageway urban traffic signals. Contractor Report 65. Crowthorne, Berkshire, Transport and Road Research Laboratory.
- Harwood, D. W., Torbic, D. J., Gilmore, D. K., Bokenkroger, C. D., Dunn, J. M., Zegeer, C. V., Srinivasan, R., Carter, D., Raborn, C., Lyon, C., Persaud, B. 2008. Pedestrian safety prediction methodology. NCHRP Web-only document 129: Phase III. Transportation Research Board, Washington D. C.
- Hauer, E. 2015. The art of regression modelling in road safety. New York, Springer.
- Hauer, E., Bamfo, J. 1997. Two tools for finding what function links the dependent variable to the explanatory variable. Proceedings of ICTCT workshop 1997 in Lund, Sweden.
- Kröyer, H. R. G. 2015. Accidents between pedestrians, bicyclists and motorised vehicles: Accident risk and injury severity. Doctoral dissertation. Lund University, Department of Technology and Society.
- Lahrman, H. 2015. Randomized trials and self-reported accidents as a method to study safety-enhancing measures for cyclists – two case studies. Paper submitted to Road Safety on 5 Continents (RS5C), Rio de Janeiro, 2016.
- Mensah, A., Hauer, E. 1998. Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. Transportation Research Record, 1635, 37-43.

Prato, C. G., Kaplan, S., Rasmussen, T. K., Hels, T. 2014. Infrastructure and spatial effects on the frequency of cyclist-motorist collisions in the Copenhagen region.

Artikler fra Trafikdage på Aalborg Universitet (ISSN 1603-9696).

Sætre, K., Knag, Ø. R., Amundsen, K. S. 2010. Kvalitetssikring av gangfelt på fylkesveger i Bærum kommune i 50 og 60-soner, Asplan Viak for Bærum kommune, Sandvika.

Sørensen, M. W. J., Mosslemi, M., Akhtar, J. 2010. Kvalitetssikring av gangfelt i 50 soner i Oslo, TØI rapport 1058. Oslo, Transportøkonomisk institutt.

Sørensen, M. W. J., Nævestad, T-O. 2012. Kvalitetssikring av 75 gangfelt i Oslo. TØI rapport 1231. Oslo, Transportøkonomisk institutt.

Turner, S. A., Roozenburg, A. P., Francis, T. 2006. Predicting Accident Rates for Cyclists and Pedestrians. Land Transport New Zealand Research Report 289. Wellington, Land Transport New Zealand.

Vanlaar, W. 2008. Less is more: The influence of traffic count on drinking and driving behaviour. Accident Analysis and Prevention, 40, 1018-1022.

LIST OF FIGURES AND TABLES

Figure 1:

CURE-plot for model including traffic volume indicated by counts only

Figure 2:

CURE-plot for model including traffic volume indicated by AADT only

Figure 3:

Number of directions from which vehicles may enter a pedestrian crossing

Figure 4:

Relationship between number of lanes and mean number of accidents per pedestrian crossing

Figure 5:

Percent of variance explained by the final model

Figure 6:

Relationship between motor vehicle volume and number of accidents

Table 1:

Variables recorded for each pedestrian crossing

Table 2:

Correlation matrix

Table 3:

Models estimated

Figure 1:

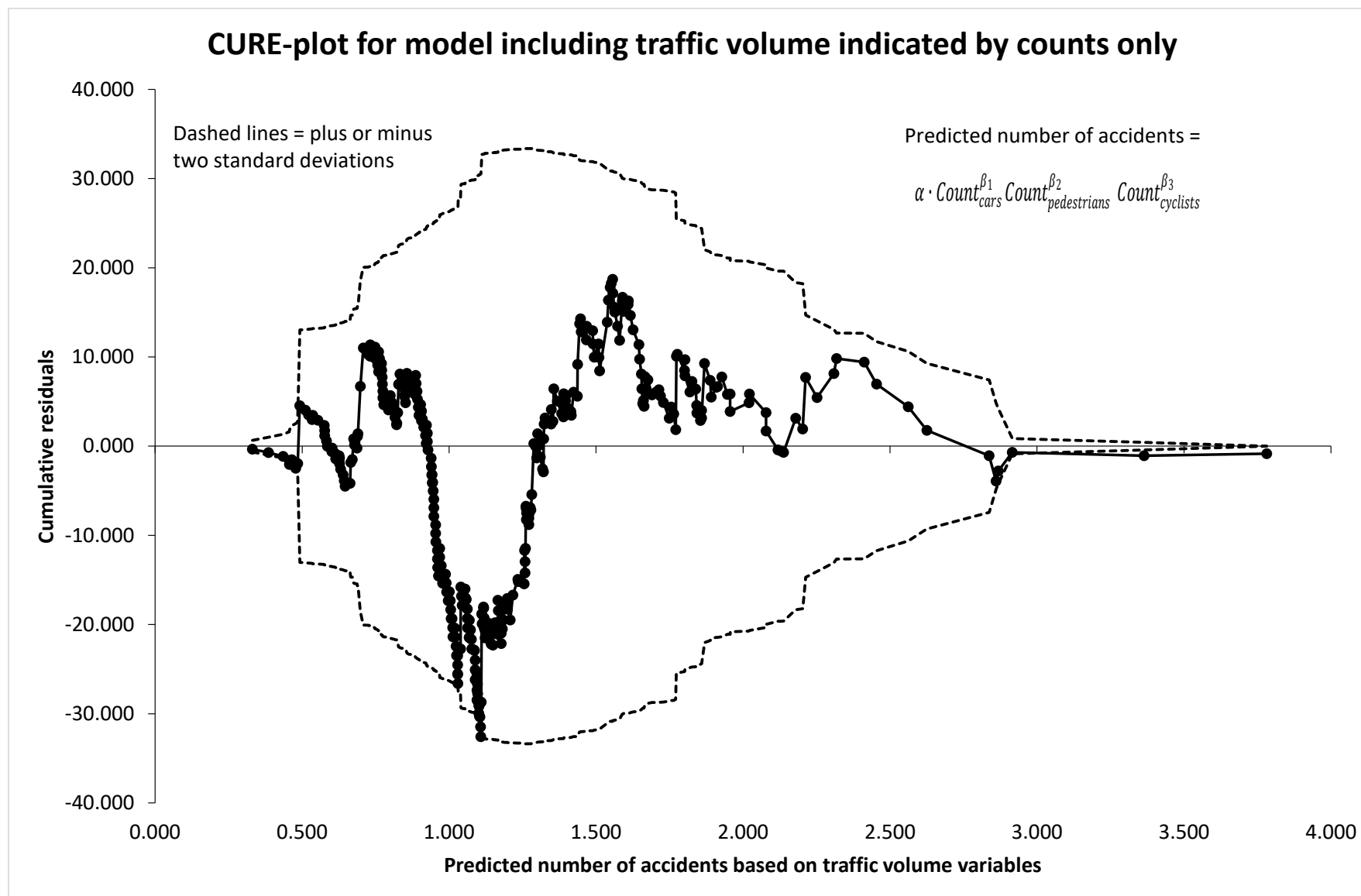


Figure 2:

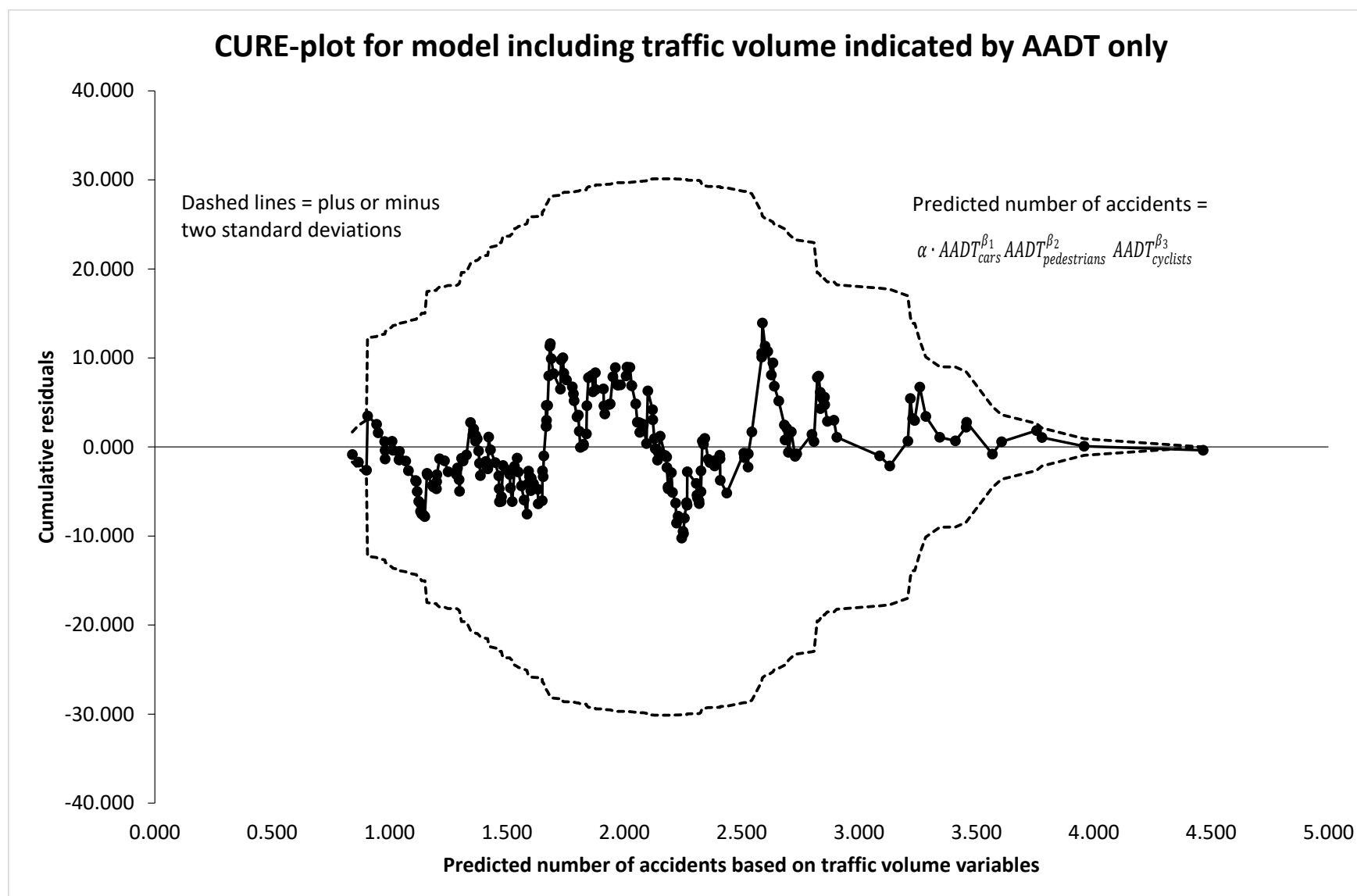
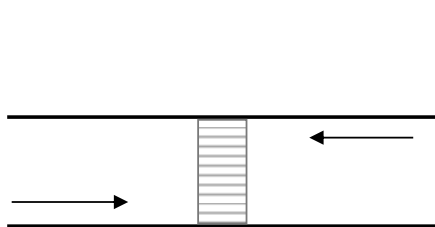
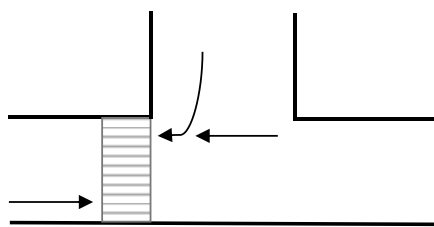


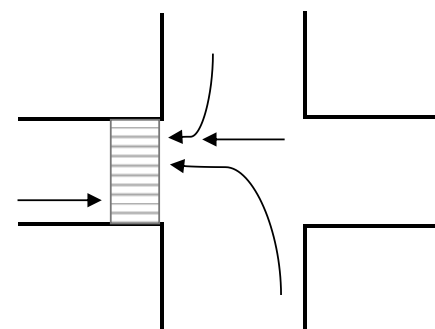
Figure 3:



Entering traffic from two directions:
0.869 accidents per crossing



Entering traffic from three directions:
1.439 accidents per crossing



Entering traffic from four directions:
2.279 accidents per crossing

Figure 4:

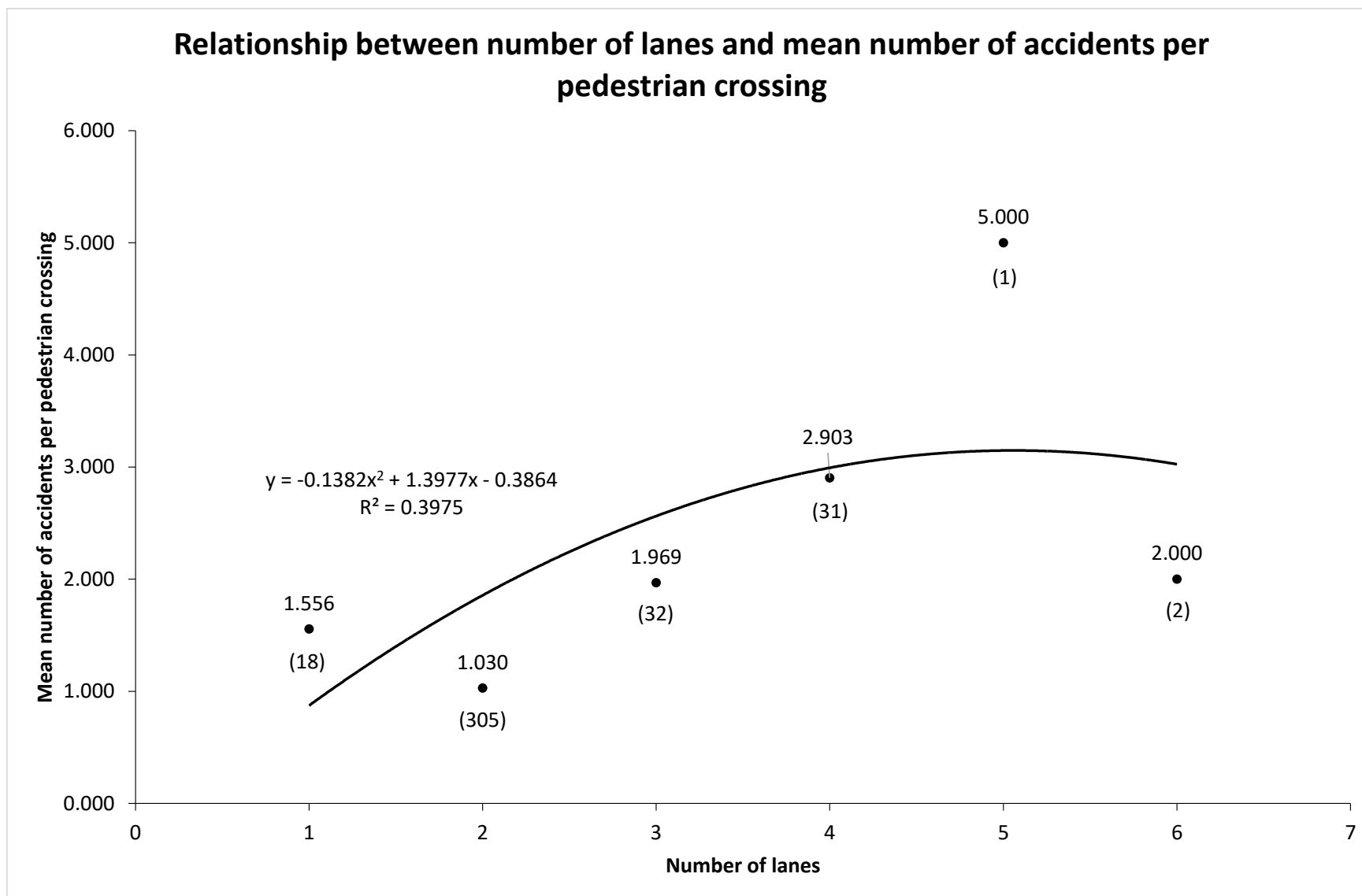


Figure 5:

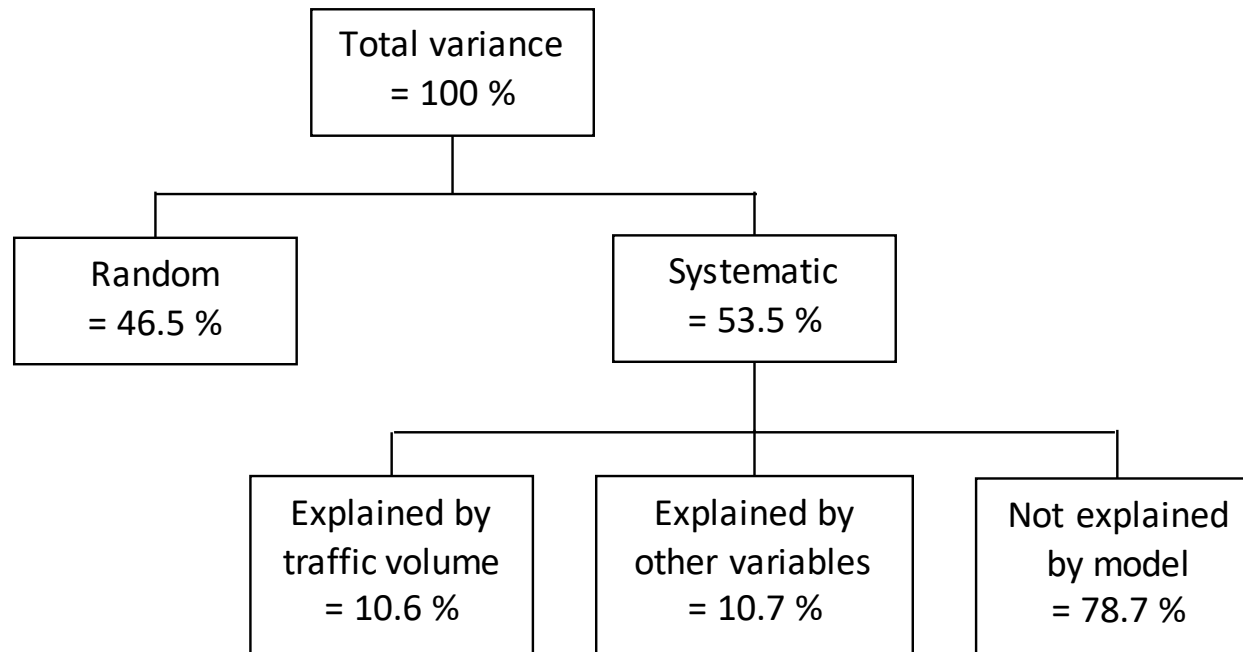


Figure 6:

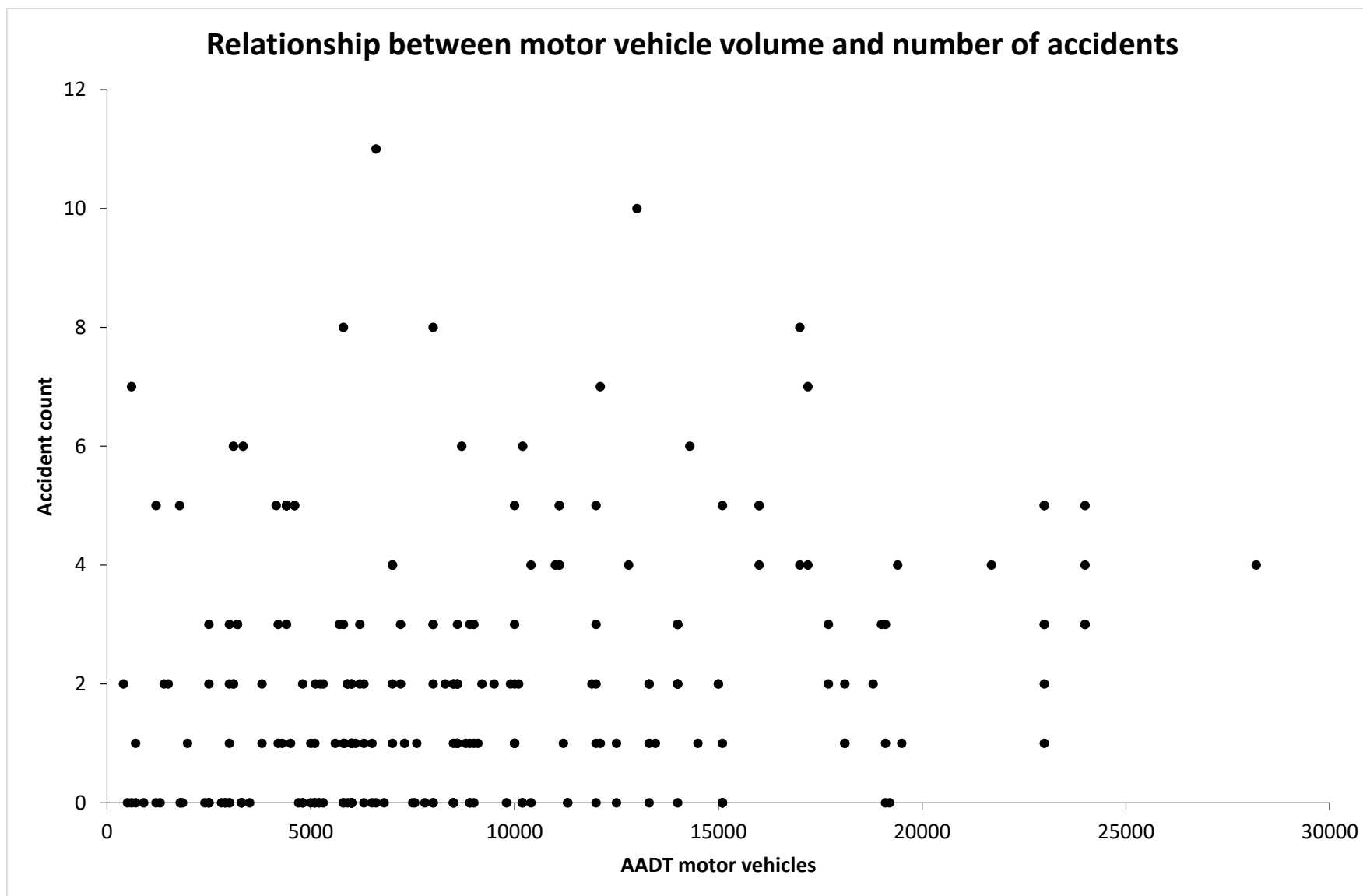


Table 1:

Variable	Definition and explanation	Mean	Minimum	Maximum
<i>Group 1: Dependent variables</i>				
Accidents	Count of police reported injury accidents during 5 years	1.296	0	11
Slightly injured road users	Count of slightly injured road users during 5 years	1.527	0	11
Seriously injured road users	Count of seriously injured road users during 5 years	0.082	0	3
Fatally injured road users	Count of fatally injured road users during 5 years	0.008	0	1
<i>Group 2: Traffic volume variables</i>				
Motor vehicle volume	Annual average daily number of motor vehicles (AADT)	8181	145	28200
Pedestrian volume	Estimated annual average daily number of pedestrians crossing the road	233	0	5000
Cyclist volume	Estimated annual average daily number of cyclists crossing the road	35	0	589
Count of cars	Count of cars made during daytime when data were collected about each pedestrian crossing	7339	33	25863
Pedestrians in maximum hour	Count in pedestrians in the hour with the largest number (short-term count)	62	0	1571
Cyclists in maximum hour	Count of cyclists in the hour with the maximum number (short-term count)	8	0	252
<i>Group 3: Other independent variables</i>				
Arms to be observed	The number of approaches (directions) from which vehicles enter a pedestrian crossing	2.666	2	4
Number of lanes	The number of lanes for vehicles	2.224	1	6
Presence of refuge	If there is a refuge for pedestrians or not (dichotomous variable; 1 if refuge, 0 otherwise)	0.550	0	1
Signal control	Presence of traffic signals (1 if present, 0 otherwise)	0.111	0	1
Speed limit	Speed limit in kilometres per hour (30, 40, 50 or 60)	52.237	30	60
85-percentile speed	The speed below which 85 percent of motor vehicles travel (km/h)	44.759	16	80

Table 2:

	Arms	Lanes	Refuge	Signals	Speed limit	85 th % speed	AADT cars	Count of cars	AADT pedestrians	Count of pedestrians	AADT cyclists	Count of cyclists	Count of accidents
Number of arms	1	.063	.045	.080	-.100	-.447	.002	-.159	.090	.101	.156	.215	.279
Number of lanes		1	.256	.559	.165	.097	.501	.283	.274	.207	.276	.109	.262
Presence of refuge			1	.154	.173	-.076	.202	.199	.047	.143	.054	.063	.047
Traffic signals				1	.112	.091	.423	.319	.174	.184	.141	.105	.299
Speed limit					1	.461	.232	.303	-.278	-.076	-.170	-.084	.093
85th % speed						1	.236	.377	-.232	-.151	-.232	-.195	-.065
AADT cars							1	.739	.131	.118	.123	-.002	.274
Count of cars								1	-.082	-.021	-.050	-.098	.143
AADT pedestrians									1	.952	.322	.314	.166
Count of pedestrians										1	.335	.296	.146
AADT cyclists											1	.890	.153
Count of cyclists												1	.191
Count of accidents													1

Table 3:

Regression coefficients – negative binomial regression – standard errors in parentheses							
Variables	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7
Constant term	-1.671 (0.773)	-1.476 (0.790)	-1.669 (0.871)	-1.342 (0.875)	-1.588 (0.909)	-1.193 (0.927)	-1.197 (0.927)
Ln(AADTcars)	0.190 (0.088)	0.150 (0.095)	0.143 (0.096)	0.078 (0.101)	0.082 (0.100)	0.050 (0.101)	0.048 (0.101)
Ln(AADTpedestrians)	0.062 (0.050)	0.054 (0.050)	0.054 (0.050)	0.075 (0.058)	0.071 (0.058)	0.061 (0.057)	0.066 (0.058)
Ln(AADTcyclists)	0.138 (0.052)	0.132 (0.053)	0.135 (0.053)	0.136 (0.053)	0.126 (0.054)	0.127 (0.054)	0.120 (0.055)
Number of lanes		0.089 (0.084)	0.279 (0.369)	0.262 (0.366)	0.263 (0.364)	0.243 (0.360)	0.225 (0.361)
Lanes squared			-0.032 (0.060)	-0.027 (0.060)	-0.027 (0.059)	-0.036 (0.059)	-0.035 (0.059)
Speed limit 60 km/h				0.280 (0.162)	0.294 (0.162)	0.287 (0.161)	0.274 (0.163)
Speed limit 40 km/h				0.506 (0.267)	0.507 (0.266)	0.529 (0.264)	0.549 (0.266)
Speed limit 30 km/h				0.285 (0.552)	0.218 (0.554)	0.281 (0.550)	0.286 (0.550)
Number of arms					0.085 (0.088)	0.074 (0.087)	0.077 (0.087)
Traffic signals						0.362 (0.203)	0.358 (0.203)
Refuge							0.085 (0.145)
Overdispersion	0.520 (0.111)	0.512 (0.110)	0.510 (0.110)	0.483 (0.107)	0.475 (0.106)	0.457 (0.104)	0.458 (0.104)