

Can a safety-in-numbers effect and a hazard-in-numbers effect co-exist in the same data?

Rune Elvik

Institute of Transport Economics

Gaustadalleen 21, NO-0349 Oslo, Norway

E-mail: re@toi.no

ABSTRACT

Safety-in-numbers denotes a non-linear relationship between exposure (traffic volume) and the number of accidents, characterised by declining risk as traffic volume increases. There is safety-in-numbers when the number of accidents increases less than proportional to traffic volume, e.g. a doubling of traffic volume is associated with less than a doubling of the number of accidents. Hazard-in-numbers, a less-used concept, refers to the opposite effect: the number of accidents increases more than in proportion to traffic volume, e.g. is more than doubled when traffic volume is doubled. This paper discusses whether a safety-in-numbers effect and a hazard-in-numbers effect can co-exist in the same data. It is concluded that both effects can exist in a given data set. The paper proposes to make a distinction between partial safety-in-numbers and complete safety-in-numbers. Another issue

that has been raised in discussions about the safety-in-numbers effect is whether the effect found in some studies is an artefact created by the way exposure was measured. The paper discusses whether measuring exposure as a rate or a share, e.g. kilometres travelled per inhabitant per year, will generate a safety-in-numbers effect as a statistical artefact. It is concluded that this is the case. The preferred measure of exposure is a count of the number of road users. The count should not be converted to a rate or to the share any group of road user contribute to total traffic volume.

Key words: safety-in-numbers; hazard-in-numbers; statistical artefact; co-existence of effects

1 INTRODUCTION

Safety-in-numbers is a phenomenon that has been the focus of many recent studies of the risks faced by pedestrians and cyclists (see, for example, Pucher and Buehler 2006, Elvik 2009, Vandenbulcke et al. 2009, Nordback et al. 2013). It refers to the tendency for the risk of accident faced by each pedestrian or cyclist to fall as the number of pedestrians or cyclists increase. There is, in other words, a non-linear relationship between the volume of pedestrians and cyclists and the number of accidents involving these groups of road users. The number of accidents increases less than proportional to traffic volume. This has been used as a basis for arguing that measures designed to stimulate people to walk or bike may not necessarily be associated with a large increase in the number of accidents involving pedestrians or cyclists.

It is, however, not clear that such an interpretation of the safety-in-numbers phenomenon is fully justified (Bhatia and Wier 2011). It cannot even be ruled out that the findings of some studies claiming to support a safety-in-numbers effect are pure statistical artefacts (Brindle 1994, Knowles et al. 2009). Finally, the often partial nature of the safety-in-numbers effect has not been fully understood; one could argue that both a safety-in-numbers effect and a hazard-in-numbers effect could be found in the same data set.

It is therefore important to get a deeper understanding of the safety-in-numbers effect, in particular if one wants to use this effect to argue that more walking or cycling can be encouraged without worrying about a large increase in the number of accidents. Such an argument can only be made if: (1) The safety-in-numbers effect is

causal, not just a statistical association that may have other explanations, such as better infrastructure or differences with respect to who walks or cycles; (2) The safety-in-numbers effect is complete, not just partial (see section 2 for an explanation of the difference between a partial and complete safety-in-numbers effect); and (3) The safety-in-numbers effect is real, not simply a statistical artefact.

The objectives of this paper are: (1) to explore whether there could simultaneously be effects that can be interpreted both as a safety-in-numbers effect and as a hazard-in-numbers effect in the same data set; and (2) to explore whether an apparent safety-in-numbers effect could be a statistical artefact. It is stressed that the analyses presented in this paper are exploratory and are only intended to demonstrate that certain effects are logically possible. This is not intended to suggest that these effects are actually common.

2 COMMON FORMS OF ACCIDENT PREDICTION MODELS

The most common form of accident prediction model in studies of the relationship between traffic volume and the number of accidents is:

$$\text{Expected number of accidents} = e^{\beta_0} (PED)^{\beta_1} (MV)^{\beta_2} e^{(\sum_{i=1}^n \beta_i X_i)} \quad (1)$$

PED (alternatively CYC) denotes pedestrian (or cyclist) volume, MV denotes motor vehicle volume (usually in terms of AADT = Annual Average Daily Traffic), e is the exponential function, X_i ($i = 1$ to n) represents risk factors influencing safety, e.g. the mean speed of traffic, the number of travel lanes, the number of legs in junctions, etc. and β_i are coefficients which are normally estimated by means of negative

binomial regression. Note that the following formulations are mathematically identical:

$$X^\beta = e^{(\ln(X) \cdot \beta)} \quad (2)$$

It is therefore common to include variables representing traffic volume as natural logarithms in accident prediction models. This represents no restriction on the values of the estimated coefficients. Another form of model used in studies designed to investigate the safety-in-numbers effect is the following (Jacobsen 2003, equation 2, slightly re-written):

$$\text{Injury rate} = \frac{\text{Injuries}}{\text{Km travelled}} = \alpha \cdot \left(\frac{\text{Km travelled}}{\text{Number of inhabitants}} \right)^{(\beta-1)} \quad (3)$$

This formulation represents the most common type of model in the data sets examined by Jacobsen.

There are five important differences between models of the type shown in equation 1 and models of the type shown in equation 3. In the first place, the first type of model uses the number of accidents as dependent variable, the second uses injury rate (number of injured road users per unit of exposure) as dependent variable. In the second place, the first type of model represents exposure to risk as a count; the second type of model represents it as a rate (kilometres per inhabitant) or a share (percent of journeys to work on foot). In the third place, the first type of model represents the effects of exposure on accidents as (constant) elasticities, i.e. the coefficients in equation 1 show the percentage change in the number of accidents associated with a one percent growth in traffic volume, whereas in the second type of model the effects of exposure are modelled as risk elasticities. In the fourth place, the

first type of model normally includes a count of traffic volume for at least two groups of road users (pedestrians, cyclists, motor vehicles); the second type of model includes traffic volume for just a single group of road users. In the fifth place, the first type of model often includes a number of independent variables in addition to traffic volume; in the second type of model, a measure of exposure to risk tends to be the only independent variable. In sum, these differences are essential and make a direct comparison of the results obtained by the different types of models difficult, if not meaningless.

It is generally regarded as evidence of a safety-in-numbers effect if both of the coefficients referring to traffic volume in the first type of model (equation 1) are less than 1. If a coefficient is less than 1, it means that the number of accidents increases by less than 1 percent when traffic volume increases by 1 percent. This implies that the risk per road user is lower when there are many road users than when there are few. Likewise, in the second type of model (equation 3), a negative risk elasticity is consistent with a safety-in-numbers effect.

In models of the first type, it is important to understand that the coefficient estimated for each variable represents its effect on accidents controlling for all other variables included in the model. Thus, coefficients of, for example 0.5 for motor vehicle volume and 0.7 for pedestrian volume imply that the number of pedestrian accidents increases less than motor vehicle volume, keeping pedestrian volume constant, and less than pedestrian volume, keeping motor vehicle volume constant. These coefficients therefore only show a partial safety-in-numbers effect. When the sum of the coefficients is greater than 1, the number of accidents more than doubles

when the sum of pedestrian volume and motor vehicle volume doubles. If pedestrian or cyclist volume is highly correlated with motor vehicle volume, there will be no overall safety-in-numbers effect with respect to total traffic volume if the sum of the coefficients is greater than 1.

3 EFFECTS FOUND IN A REAL DATA SET

To show how both a safety-in-numbers effect and a hazard-in-numbers effect can occur in the same data, data for 159 marked pedestrian crossings in the city of Oslo will be applied. These data have been analysed by means of negative binomial regression and two accident prediction models were fitted to the data (Elvik, Sørensen and Nævestad 2013):

1. One model used the total number of accidents as dependent variable. There were 316 accidents in total.
2. One model used the number of accidents related to the pedestrian crossings as dependent variable. There were 149 accidents related to the pedestrian crossings.

The total number of accidents includes all types of accidents occurring within a zone of 50 metres to each side of the pedestrian crossing (100 metres in total). Crossing-related accidents include those that are related to use of the crossing, such as pedestrians hit when using the crossing, or rear-end accidents occurring because a car brakes hard to avoid hitting a pedestrian.

The following independent variables were included in both models:

1. The natural logarithm of the total number of road users crossing at pedestrian crossings
2. The natural logarithm of annual average daily traffic (AADT)
3. The product of the number of road users crossing at pedestrian crossings and AADT
4. The number of legs at the crossing location (an indicator of the number of directions from which traffic that may conflict with crossing pedestrians enters)
5. The number of driving lanes at the crossing location (a count variable varying from 1 to 6)
6. The type of traffic control (none or traffic signals; coded as 0 or 1)
7. The percentage of road users crossing outside the marked crossing
8. The mean speed of motor vehicles approaching a marked crossing (km/h)
9. Whether formal warrants for the use of marked pedestrian crossings were satisfied or not (1 if satisfied, 0 otherwise).

In the present context, it is the coefficients referring to pedestrian volume and motor vehicle volume that are of primary interest.

Figure 1 shows that there is very little correlation between motor vehicle volume and pedestrian volume at the 159 marked crossings. Motor vehicle volume (AADT) varied from 500 to 19,500. Pedestrian volume was, in general, much lower, ranging from less than 10 to a little more than 5,000. There was, however, sufficient variation in both motor vehicle traffic volume and pedestrian volume to detect any safety-in-numbers effect.

Figure 1 about here

In the model using the total number of accidents as dependent variable, the coefficient for motor vehicle volume was 0.591; the coefficient for pedestrian volume was 0.312. These coefficients sum to 0.903, suggesting that there will be a safety-in-numbers effect associated with any combination of motor vehicle and pedestrian volume. The number of accidents predicted by these coefficients was estimated for all 159 marked pedestrian crossings. Figure 2 shows the results. The lowest total traffic volume (500 vehicles; 43 pedestrians; total 543) was given the value of 1.0. The number of accidents predicted for this volume was likewise given the value of 1.0.

Figure 2 about here

Figure 2 shows that there is, except for a few borderline cases at very low traffic volumes, a complete safety-in-numbers effect for the whole range of traffic volumes (sum of pedestrians and motor vehicles). In the model using accidents that were judged to be related to the pedestrian crossings as dependent variable, the coefficient estimates were 0.533 for motor vehicle volume and 0.761 for pedestrian volume. The sum of the coefficients is 1.294. The value of the coefficient for motor vehicle volume is close to that found when using the total number of accidents as dependent variable (0.591). However, the value of the coefficient for pedestrian volume (0.761) suggests that accidents related to the pedestrian crossings are considerably more sensitive to pedestrian volume than the total number of accidents (0.312). The coefficients were once more applied to estimate the predicted number of accidents in each pedestrian crossing. The results are shown in Figure 3.

Figure 3 about here

The lowest traffic volume was again given the value of 1.0 and the number of accidents predicted for this volume given the value of 1.0. It is readily apparent that the results are very different from those shown in Figure 2. There is a safety-in-numbers effect for 89 pedestrian crossings, a hazard-in-numbers effect for 69 pedestrian crossings (these number add to 158; the first crossing is used as reference and therefore not counted).

The coefficients are, however, consistent with partial safety-in-numbers effects with respect both to pedestrian and motor vehicle volume. One may wonder, however, if it is appropriate to speak of a safety-in-numbers effect when each pedestrian faces a higher risk of accident when both motor vehicle volume and pedestrian volume increase. As an example, the predicted number of crossing-related accidents increased by a factor of 4.72 when a crossing with 1,800 motor vehicles and 416 crossing pedestrians was compared to a crossing with 900 motor vehicles and 88 crossing pedestrians. Thus, the 416 pedestrians crossing at the busiest crossing each faced a risk more than twice as high as the risk faced by the 88 pedestrians crossing at the less busy crossing.

4 SAFETY-IN-NUMBERS AS A STATISTICAL ARTEFACT

It has been argued that some of the studies claiming to show a safety-in-numbers effect are likely to show a relationship that could be a pure statistical artefact (Brindle 1994, Knowles et al. 2009). Recall that in some studies, pedestrian (or cyclist) risk was measured as the number of injured road users per kilometre walked (or cycled).

Exposure to risk was measured as the number of kilometres walked per inhabitant.

In other words risk equals A/B and exposure equals B/C .

It is obvious that defining exposure and risk this way can generate a spurious negative relationship between exposure and risk that looks like a safety-in-numbers effect. Consider what happens when B increases. All else equal, the value of A/B will decrease, i.e. risk is reduced. When B increases, the value of B/C also increases, i.e. exposure increases at the same time as risk decreases. There will thus, by definition, be a negative relationship between exposure and risk.

To test if this relationship could be a pure statistical artefact, random numbers were generated for motor vehicle volume, pedestrian volume and the number of accidents. Series of 159 random numbers were generated to simulate a sample of the same size as the marked pedestrian crossings in the city of Oslo. Motor vehicle volume was random between 500 and 20,000; pedestrian volume was random between 10 and 5,000 and the number of accidents was random between 0 and 10. These ranges are consistent with those observed for the 159 pedestrian crossings in Oslo.

The risk of accident was measured as the number of accidents per 1,000 pedestrians. If there is safety-in-numbers, this risk should fall as a function of the number of pedestrians. Exposure to risk was measured as the number of pedestrians per motor vehicle. This measure of exposure, although uncommon, is not entirely meaningless. One could, for example, argue that pedestrians will more easily be able to “force” car drivers to comply with their duty to give way at pedestrian crossings the more numerous pedestrians are compared to the number of car drivers. A plot of

pedestrian risk against pedestrian exposure was generated. It is reproduced in Figure 4.

Figure 4 about here

There is a remarkably strong negative relationship between pedestrian exposure and pedestrian risk. This suggests a very strong safety-in-numbers effect. However, as the relationship is based on random numbers, it must be a pure statistical artefact. There is no correlation between the three variables: motor vehicle volume, pedestrian volume and the number of accidents. The negative relationship emerges solely as a result of the way risk and exposure have been defined.

As a further test, the analysis relying on random numbers was repeated, using the number of pedestrians as measure of exposure. Figure 5 shows the results. There is again a negative relationship between pedestrian risk and the number of pedestrians, suggesting a safety-in-numbers effect. The relationship is weaker and more noisy than the one shown in Figure 4. Still, it is clearly discernible.

Figure 5 about here

It is only when pedestrian volume is less than about 500 that the risk curve in Figure 5 starts to rise rapidly. For higher pedestrian volumes, the curve is considerably flatter, although not perfectly horizontal as it ought to be when there is no safety-in-numbers effect. There could be a very simple explanation for the negative relationship (given that it is based on random numbers). When a single accident occurs by chance at a low pedestrian volume, the denominator of the estimate of risk has a low value, so risk is estimated to be high. In real data, characterised by a partly systematic, partly random relationship between pedestrian volume and the number of

accidents, the occurrence of a single accident leading to an inflated estimate of risk at low volumes may be less likely than in the random data that serve as basis for Figure 5.

5 DISCUSSION

Two questions motivated the research presented in this paper: (1) What do we mean by safety-in-numbers?; and: (2) How can we know that we have identified a real safety-in-numbers effect? These questions have not received the attention they deserve in recent studies of the safety-in-numbers phenomenon.

In accident prediction models of the type shown in equation 1 in section 2 of the paper, the usual interpretation is that there is safety-in-numbers if the coefficients for traffic volume are less than one. A coefficient less than 1 implies that the risk of accidents declines as volume increases; thus each road user faces a lower risk.

However, accidents involving pedestrians and cyclists depend both on the number of pedestrians or cyclists and the number of motor vehicles. If the gain in safety for each pedestrian and cyclist as they become more numerous is offset by the added risk posed by an increasing number of motor vehicles, there really is no safety-in-numbers. This will be the case whenever the sum of the coefficients for the two volumes is greater than one.

Table 1 shows coefficients estimated in a number of accident predictions models based on data for pedestrian volume, cyclist volume and motor vehicle volume. In most studies, the sum of the coefficients is greater than one, suggesting that the data

only contains a partial safety-in-numbers effect, i.e. a safety-in-numbers effect observed for pedestrians or cyclists when motor vehicle volume is kept constant.

Table 1 about here

There could be a safety-in-numbers effect in parts of a data set where the coefficients add to more than one. However, such an effect would be contingent on a low correlation between pedestrian or cyclist volume and motor vehicle volume. Thus, for the pedestrian crossings in Oslo, a safety-in-numbers effect was found in 89 out of 159 crossings, even in the model with coefficient estimates of 0.761 for pedestrian volume and 0.533 for motor vehicle volume.

As far as injury prediction models of the type shown in equation 3 in section 2 of the paper are concerned, these models cannot be trusted to reveal a real safety-in-numbers effect. There are three principal reasons for this. First, the models tend to define exposure to risk in a way that entails a risk of creating an artificial relationship between exposure and risk which is consistent with a safety-in-numbers effect.

Obviously, a relationship that could arise as an artefact could also be real. It does not have to be an artefact. This ambiguity makes it almost impossible to interpret the results of models based on equation 3. The relationships revealed by such models could be real, but they could also be pure artefacts. Second, the models include data for a single road user group only, i.e. only pedestrians or only cyclists. However, as shown in the analyses of the data for marked pedestrian crossings in Oslo, the risk of accident is also influenced by motor vehicle volume. Hence, no model can reveal a complete (as opposed to partial) safety-in-numbers effect without including data on

both pedestrian or cyclist volume and motor vehicle volume. Third, models of the type shown in equation 3 do not, in general, control for confounding factors.

For these reasons one should place considerably less trust in models based on equation 3 than in models based on equation 1.

6 CONCLUSIONS

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

1. Two different forms of accident prediction models have been applied in research aiming to determine if there is a safety-in-numbers effect, i.e. a tendency for the risk faced by each road user to decline as the number of road users increases.
2. Only accident prediction models that include data based on counts of traffic volume for all relevant groups of road users (pedestrians, cyclists, motor vehicles) can reveal a true safety-in-numbers effect. Simpler models entail a non-negligible risk of showing relationships that are partly or entirely statistical artefacts.
3. A distinction should be made between partial safety-in-numbers and complete safety-in-numbers. There is partial safety-in-numbers if the risk faced by each road user of type A declines as the number of road users of type A increases, keeping the number of road users of type B constant. There is complete safety-in-numbers if the risk faced by each road user of type A

declines throughout the range of combined volumes of road users of types A and B.

ACKNOWLEDGEMENT

This study was funded by the Norwegian Public Roads Administration (Statens vegvesen)

REFERENCES

- Bhatia, R., Wier, M. 2011. "Safety in Numbers" re-examined: Can we make valid or practical inferences from available evidence? *Accident Analysis and Prevention*, 43, 235-240.
- Brindle, R. E. 1994. Lies, Damned Lies and "Automobile Dependence" – some hyperbolic reflections. *Proceedings of ATRF 1994*, 117-132.
- Elvik, R. 2009. The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis and Prevention*, 41, 849-855.
- 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., Rafoard, N., Pham, T., Ragland, D. R. 2006. Safety in numbers. Data from Oakland, California. *Transportation Research Record*, 1982, 150-154.
- 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.
- Jacobsen, P. L. 2003. Safety in numbers: more walkers and bicyclists, safer walking and cycling. *Injury Prevention*, 9, 205-209.
- Jonsson, T. 2005. Predictive models for accidents on urban links. A focus on vulnerable road users. Ph. D. dissertation. Bulletin 226. Lund Institute of Technology, Department of Technology and Society, Traffic Engineering, Lund.
- Leden, L. 2002. Pedestrian risk decrease with pedestrian flow. A case study based on data from signalized intersections in Hamilton, Ontario. *Accident Analysis and Prevention*, 34, 457-464.
- Knowles, J., Adams, S., Cuerden, R., Savill, T., Reid, S., Tight, M. 2009. Technical Annex to PPR445 Collisions involving pedal cyclists on Britain's roads: establishing the causes. Wokingham, Transport Research Laboratory.
- Lyon, C., Persaud, B. N. 2002. Pedestrian collision prediction models for urban intersections. *Transportation Research Record*, 1818, 102-107.
- Nordback, K., Marshall, W. E., Janson, B. N. 2013. Bicyclist safety performance functions for a US city. Paper presented at 2013 TRB Annual Meeting, Washington D. C.
- Pucher, J., Buehler, R. 2006. Why Canadians cycle more than Americans: A comparative analysis of bicycling trends and policies. *Transport Policy*, 13, 265-279.

Vandenbulcke, G., Thomas, I., Geus, B. De, Degraeuwe, B., Torfs, R., Meeusen, R., Int Panis, L. 2009. Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. *Transport Policy*, 16, 77-87.

Zegeer, C. V., Stewart, J. R., Huang, H. H., Lagerwey, P. A., Feaganes, J., Campbell, B. J. 2005. Safety effects of marked versus unmarked crosswalks at uncontrolled locations: final report and recommended guidelines. Report FHWA-HRT-04-100. Georgetown Pike, Federal Highway Administration.

LIST OF FIGURES AND TABLES

Figure 1:

Volume of motor vehicles and pedestrians at 159 marked pedestrian crossings in the city of Oslo

Figure 2:

Complete safety-in-numbers effect at 159 marked pedestrian crossings in the city of Oslo

Figure 3:

Co-existence of safety-in-numbers and hazard-in-numbers for 159 marked pedestrian crossings in Oslo

Figure 4:

Fictitious safety-in-numbers effect generated by combining random numbers

Figure 5:

Fictitious relationship between pedestrian volume and pedestrian accident rate generated by random numbers

Table 1:

Coefficients estimated in accident predictions model including data on both pedestrian or cyclist volume and motor vehicle volume

Figure 1:

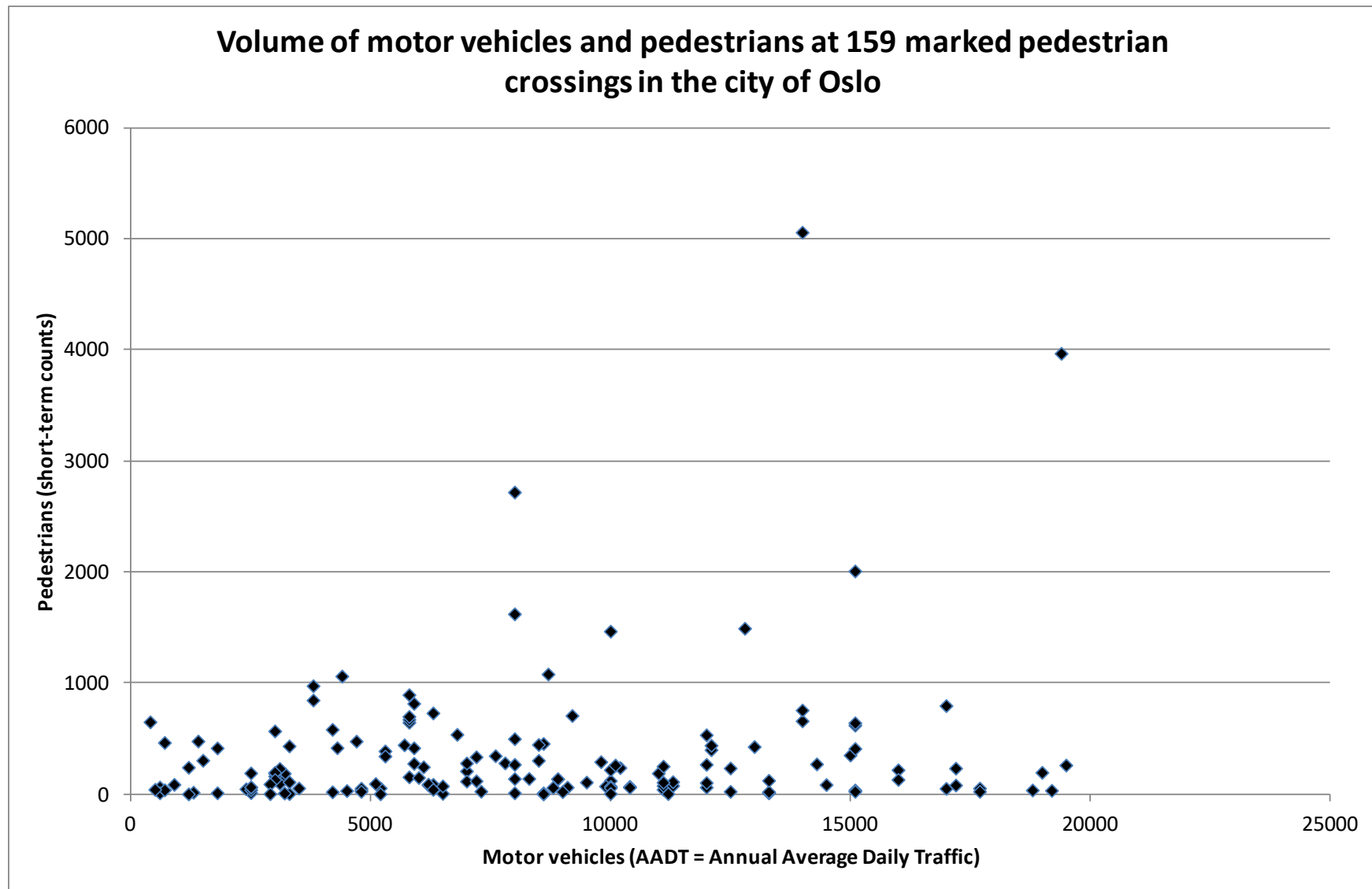


Figure 2:

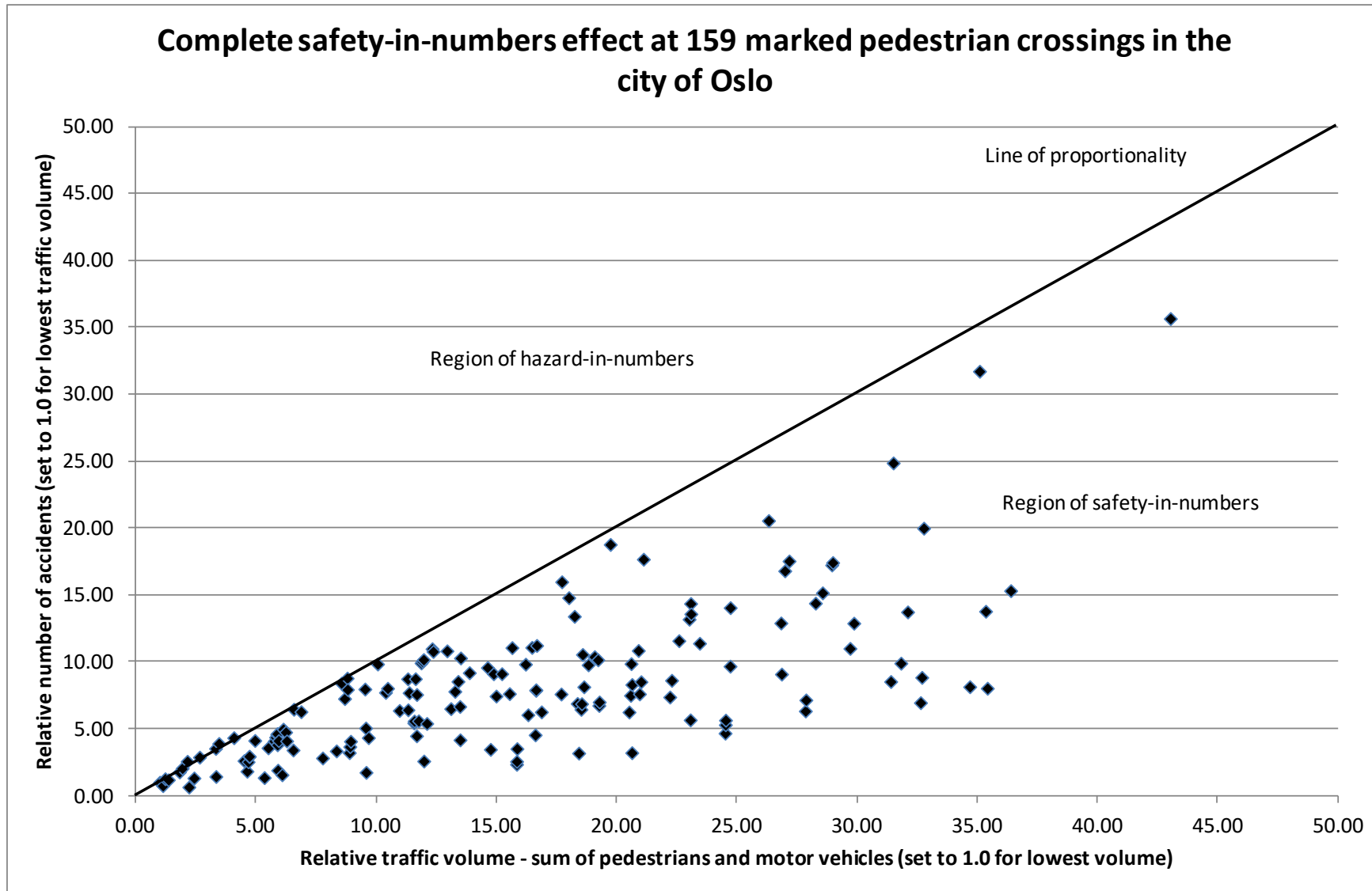


Figure 3:

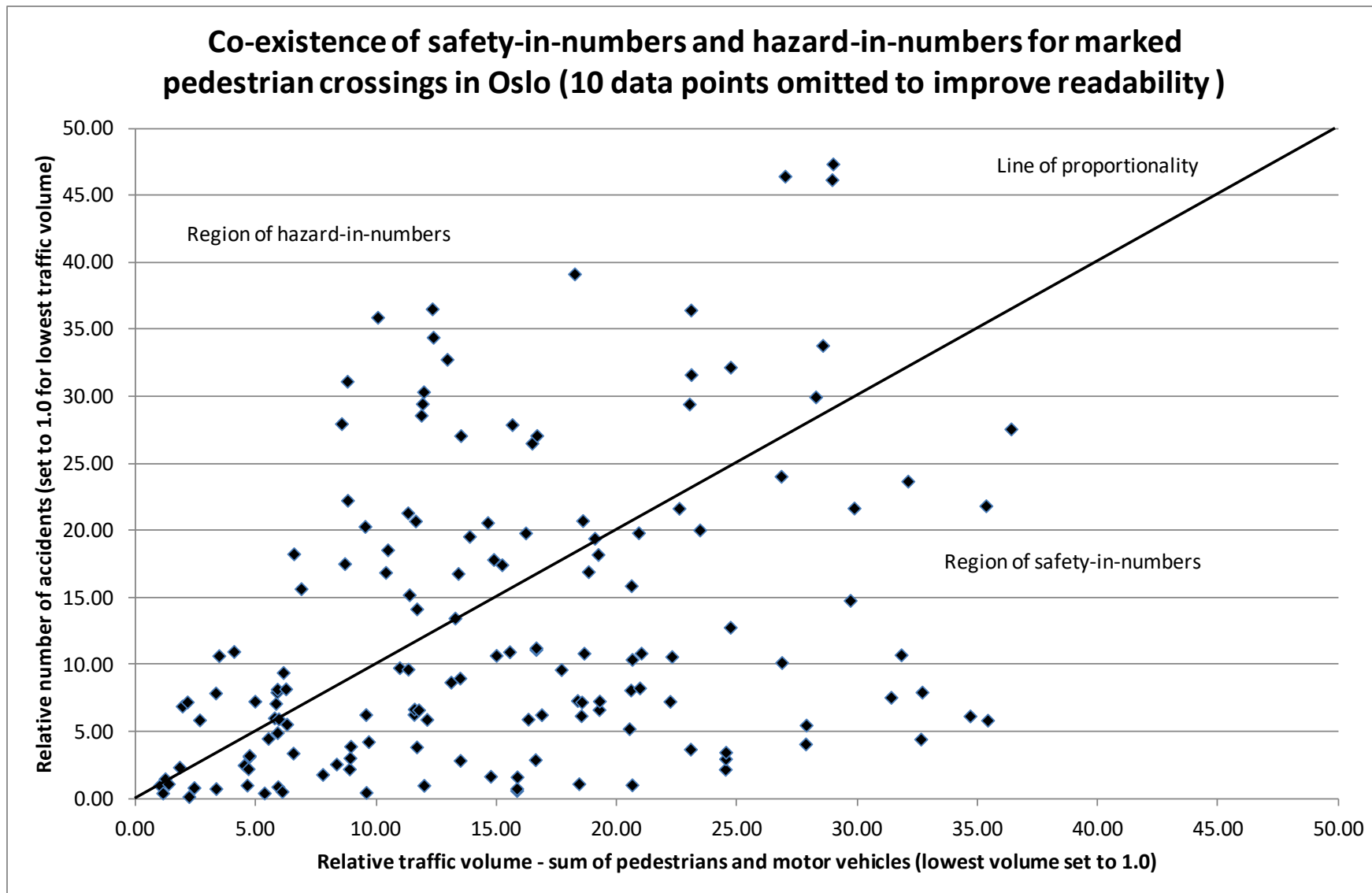


Figure 4:

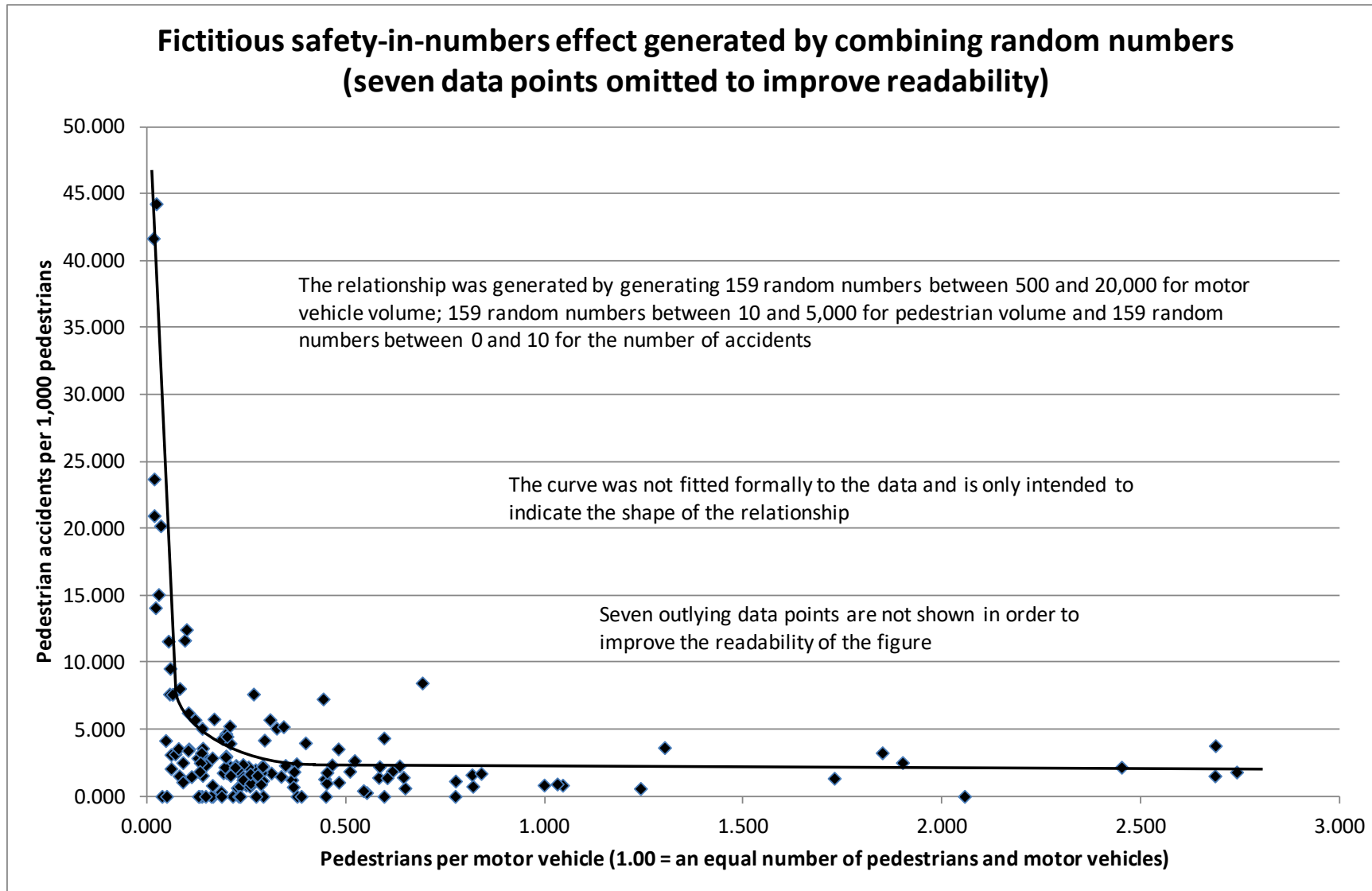


Figure 5:

Fictitious relationship between pedestrian volume and pedestrian accident rate generated by random numbers

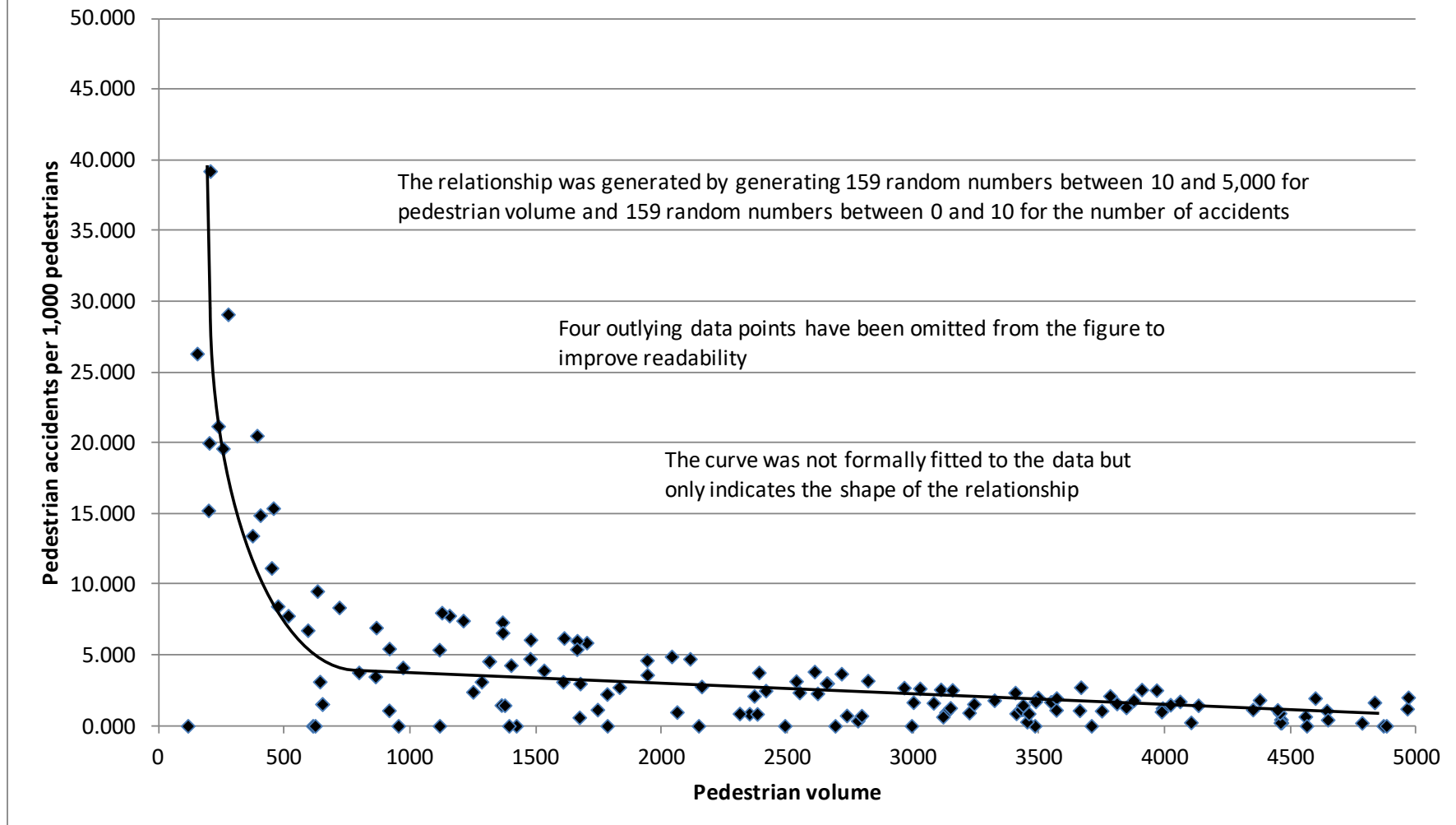


Table 1:

Study	Country	Sample size	Estimate of exponent for traffic volume		
			Motor vehicles	Pedestrians	Cyclists
Brüde and Larsson 1993	Sweden (junctions)	285	0.50	0.72	
Brüde and Larsson 1993	Sweden (junctions)	377	0.52		0.65
Leden 2002	Canada (junctions)	749	0.86	0.48	
Leden 2002	Canada (junctions)	126	1.19	0.33	
Lyon and Persaud 2002	Canada (junctions)	684	0.57	0.74	
Lyon and Persaud 2002	Canada (junctions)	263	0.40	0.41	
Lyon and Persaud 2002	Canada (junctions)	122	0.53	0.66	
Lyon and Persaud 2002	Canada (junctions)	123	0.58	0.71	
Jonsson 2005	Sweden (road sections)	393	0.83	0.38	
Jonsson 2005	Sweden (road sections)	393	0.76		0.35
Zegeer et al. 2005	United States (marked crosswalks)	1000	1.01	0.38	
Zegeer et al. 2005	United States (unmarked crosswalks)	1000	0.30	0.60	
Geyer et al. 2006	United States (junctions)	247	0.16	0.61	
Harwood et al. 2008	United States (junctions)	450	0.05	0.41	
Harwood et al. 2008	United States (junctions)	1433	0.40	0.45	
Elvik 2013	Norway (marked crosswalks)	159	0.59	0.31	
Elvik 2013	Norway (marked crosswalks)	159	0.53	0.76	
Nordback et al. 2013	United States (junctions)	105	0.64		0.53
Nordback et al. 2013	United States (junctions)	106	0.58		0.65
Mean (unweighted)			0.58	0.53	0.55