Safety Science 82 (2016) 315–324

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Pedestrian and bicyclist flows in accident modelling at intersections. Influence of the length of observational period

Höskuldur R.G. Kröyer *

Department of Technology and Society, Lund University, Sweden

ARTICLE INFO

Article history:
Received 14 April 2015
Received in revised form 2 September 2015
Accepted 11 September 2015
Available online 22 October 2015

Keywords:
Safety performance function
Pedestrians
Bicyclists
Accidents
Observation period
Safety in numbers

ABSTRACT

Safety performance functions are frequently used to describe the relation of various factors to the number of accidents, often focusing on exposure of vulnerable road users. Those studies require reliable data regarding exposure, but collecting such data is resource demanding, often resulting in short observational periods. This might produce measurement errors, influencing the reliability of the models.

The aim of this work is twofold: to create a safety performance functions for pedestrian and bicyclist accidents at urban intersections, and to analyze the reliability of accident models based on short observational periods, i.e. how does random variation, resulting from short observational periods influence the models? This provides an aid for choosing between increasing the number of sites and the length of the observational period at each site. Accident data were compiled and traffic counting was conducted at 113 urban intersections in Sweden. Multiple samples were created from the counting sessions, facilitating tests of the models' reliability based on lengths of observational periods between 15 and 180 min. Four safety performance functions (accident types) were created. All models showed a safety in numbers effect, including the model for single pedestrian accidents, which might suggest that maintenance and infrastructure quality constitute an important factor for the safety in numbers effect. A sensitivity analysis showed geometric factors, describing the infrastructure quality, influencing the safety in numbers effect, further supporting this hypothesis. The safety performance functions based on short observational periods showed low reliability, indicating that those models are subjected to a considerable measurement error.

1. Introduction

Many cities aim to increase their modal share of pedestrians and bicyclists because of the expected positive influence on the traffic environment, sustainability, and public health. But since exposure is strongly related to the number of accidents, such increases might result in a higher number of accidents involving pedestrians and bicyclists. Therefore, models that focus on these relations are of great importance.

Several known factors influence the risk of an accident and thereby the number of accidents, e.g. road design (Chang, 2005; Elvik and Vaa, 2004), road user behavior, temporal environment factors such as weather (Eisenberg, 2004; Shankar et al., 1995) and speed environment (Nilsson, 2004). Another influential factor is the traffic volume (i.e. exposure); it has been shown to have a strong relation to the number of accidents, for both pedestrians and bicyclists (Elvik, 2009; Geyer et al., 2006; Jonsson, 2005). This relation is often approximated with the mathematical model presented in Eq. (1) (Elvik, 2009), where N is the number of expected accidents, E_i is the exposure (e.g. traffic volume) for different travel modes, β_0 is the intercept of the model, and β_i is a constant. This relation seems to be non-linear, where the estimated β_i parameters are usually between 0 and 1 for traffic volume variables; hence, the accident risk per road user is lower when the traffic volumes, or exposure, of vulnerable road users is higher (Brüde and Larsson, 1993; Elvik, 2009; Geyer et al., 2006; Jacobsen, 2003; Jonsson, 2005, 2013; Turner et al., 2006). This phenomenon is frequently referred to as *safety in numbers* (e.g. Jacobsen, 2003; Jonsson, 2013).

$$N = e^{\beta_0} \prod E_i^{\beta_i} \tag{1}$$

Mathematical models in the form of Eq. (1) are often referred to as safety performance functions. Some studies apply Poisson regression for the modelling process (Geyer et al., 2006; Ye et al., 2013). Poisson distribution assumes equal mean and variance, and since accident data are often over-dispersed, this might influence the significance level of the parameters (Cameron and Trivedi, 1990; Poch and Mannering, 1996; Washington et al., 2013).

* Tel.: +46 46 222 9198; fax: +46 46 222 9100.
E-mail address: hoskuldur.kroyer@tft.lth.se

http://dx.doi.org/10.1016/j.ssci.2015.09.015
0925-7535/

Negative binomial regression (based on negative binomial distribution) is better able to handle this over-dispersion (Poch and Mannering, 1996; Washington et al., 2013).

Accident data frequently demonstrate an unusually high proportion of units (in this case, the unit is sites) with zero accidents compared to the mathematical distributions the model is based on; some authors have therefore suggested using zero inflated Poisson or zero inflated negative binomial models to improve the models (Chin and Quddus, 2003; Jang et al., 2010; Miaou, 1994; Yan et al., 2012). This results in a better fit of the model to the data (Lord et al., 2005). Those models assume, however, that the unit can be in either one of two states: the site is either 'safe' or the number of accidents follows a Poisson or negative binomial distribution (Lord et al., 2005; Washington et al., 2013). This might not be an accurate description, since no traffic site is 'safe' (Lord et al., 2007), and the extreme number of zero accidents likely results from low exposure (Lord et al., 2005). Therefore, the negative binomial regression continues to be frequently used for constructing safety performance functions.

The negative binomial distribution is described by Eq. (2) (Washington et al., 2013), where λ is the mean, κ^{-1} is the overdispersion parameter and the variance is $\lambda + \lambda^2/\kappa$. The resulting safety performance function takes the form shown in Eq. (3), where N is the number of expected accidents, E_1 and E_2 are the exposure variables and β_0, β_1, β_2 and β_k are constants. Additional explanatory variables (E_k) are sometimes included in the model (e.g. geometric variables).

$$P(Y=y) = \left(\frac{\kappa}{\kappa+\lambda}\right)^{\kappa} \frac{\Gamma(\kappa+y)}{\Gamma(y)\Gamma(\kappa)} \left(\frac{\lambda}{\kappa+\lambda}\right)^{y} \tag{2}$$

$$N = e^{\beta_0} \cdot E_1^{\beta_1} \cdot E_2^{\beta_2} \cdot e^{\Sigma\beta_k E_k} \tag{3}$$

Some studies include only geometric variables and exclude exposure variables for pedestrians and bicyclists (Chin and Quddus, 2003; Hosseinpour et al., 2013). A possible explanation for this is that traffic volumes are not always available, and it is demanding to collect them. However, since there seems to be a non-linear relation between exposure and risk, this limits the models' validity, and they can be improved by including traffic volumes (Jonsson, 2005). Some authors apply proxy variables to account for exposure, e.g. bicycle lane kilometers (Wei and Lovegrove, 2013), share of cycling and walking as travel mode (Jacobsen, 2003), the mean distance travelled per day by a typical cyclist or pedestrian (Jacobsen, 2003; Schepers, 2012), and simulated exposure data (Geyer et al., 2006). This is a step in the right direction and provides interesting aspects; however, the relation between those proxy variables, the actual traffic volumes and the accident location is in some cases unknown.

Traffic counts are in some ways preferable; though they have some limitations (see Elvik, 2013a, 2015). Several studies (e.g. Ekman, 1996; Jonsson, 2005, 2013; Wang and Nihan, 2004) have used exposure values from counting as the basis for models that are location specific (i.e. accident data and exposure data are related to certain locations). For practical reasons, those studies usually apply average traffic volumes (e.g. average daily traffic), but it can be argued that this is also a proxy variable, where the exposure variable of choice would be the number of interactions or occurrences that can, and have, given probability of resulting in an accident (Elvik, 2013a), or focus on the exposure at the time of the accident (Mensah and Hauer, 1998).

Anyhow, when creating a safety performance function, the researcher must weigh the benefits of including many sites, which will produce a highly statistically significant model and might allow for the inclusion of several explanatory variables, against having longer observational periods, which will provide a more reliable estimation of the actual traffic volumes. Since collecting exposure data is demanding, short observational periods are often used, as brief as 15–20 min per site (e.g. Jonsson, 2013; Schepers et al., 2011). This might influence the reliability of the exposure estimation, and, consequently, the reliability of the safety performance function itself (Davis, 2000; Maher and Summersgill, 1996).

The aims of this study are (1) to create a safety performance function for accidents involving pedestrians and bicyclists at urban intersections, based on three hours of observations, and (2) to analyze the reliability of accident models based on short observational periods, i.e. explore how random variation, resulting from short observational periods at each site influences the models, and thereby provide an aid for choosing between increasing the number of sites and extending the length of the observational period per site.

2. Method

Creating a safety performance function requires a number of steps: site selection, compilation of accident data, determination of potential explanatory variables, data collection and statistical modelling.

2.1. Process of site selection and collection of accident data and geometric variables

The first step was to decide on sites to be included in the study. For a city to be eligible it had to fulfill three criteria: (a) the municipalities' population was between 50 and 200 thousands, (b) the traffic accident data for the period 2008–2012 had been recorded by the Swedish Traffic Accident Data Acquisition (STRADA) and (c) at least 10 main street intersections existed within city limits, that were eligible for the study. Six cities were chosen: Eskilstuna (population 99,729), Halmstad (94,084), Helsingborg (132,989), Kalmar (63,887), Kristianstad (81,009) and Västerås (142,131; SCB, 2014). The next step was choosing which intersections to include in the study. Eligible intersections fulfilled 4 criteria: (a) two main streets crossed there, (b) they were not roundabouts, (c) they exhibited at least some potential for bicyclist and pedestrian flows, and (d) nothing suggested that any change had occurred in the physical layout between 2008 and 2012 that would have significant influence on the accidents or the bicyclist or pedestrian flows. The final dataset included 113 intersections, 16 in Eskilstuna, 25 in Halmstad, 27 in Helsingborg, 12 in Kalmar, 20 in Kristianstad and 13 in Västerås.

Accident data were compiled from STRADA, which contains police and hospital records for traffic accidents that occur in Sweden. That the accident data includes hospital reports is essential to this work, since single pedestrian and single bicyclist accidents are frequently missing from police reports (Elvik and Mysen, 1999; Jonsson et al., 2011). All accidents registered in STRADA that occurred at each intersection or incoming links between 2008 and 2012 were studied and those who belonged to the intersections were included in the dataset. Basing the analyses on accident data for five years was a compromise. If the data period is too long, it might involve a time trend bias (i.e. the traffic is constantly changing; hence, older accident data might not reflect the current situation), while if it is too short it might be influenced by extreme random variability (Wei and Lovegrove, 2013). Four accident types were included in the dataset: (i) single pedestrian accidents, (ii) single bicyclist accidents, (iii) collisions between a motorized vehicle and a pedestrian, and (iv) collisions between a motorized vehicle and a bicyclist (single accidents involve only one road user – observe that those models exclude accidents between pedestrians and bicyclists; and between two or more bicyclists).