On the causal effect of proximity to school on pedestrian safety at signalized intersections: A heterogeneous endogenous econometric model

Shahram Heydari, PhD (Corresponding Author)
Lecturer
Department of Civil, Maritime and Environmental Engineering
University of Southampton, Southampton, UK
s.heydari@soton.ac.uk

Luis Miranda-Moreno, PhD
Associate Professor
Department of Civil Engineering and Applied Mechanics
McGill University, Montreal, Canada
luis.miranda-moreno@mcgill.ca

Adrian J. Hickford
Research Assistant
Department of Civil, Maritime and Environmental Engineering
University of Southampton, Southampton, UK
A.J.Hickford@soton.ac.uk

ABSTRACT

Pedestrian safety in proximity to schools is a major concern of transportation authorities, local governments, and residents. In fact, several countermeasures (e.g., school-zone speed limits) are usually in place around schools to provide a safer environment, especially for school-age children. Two questions arise here: (i) are transportation facilities in proximity to schools truly safer than other facilities given a variety of implemented road safety interventions around schools? and (ii) how can we answer the previous question properly using a reliable approach that accounts for possible confounding? While previous literature has mixed results and does not provide clear methodological/empirical guidelines in this regard, we propose an approach that answers the above questions. We illustrate our method on a sample of intersections in Montreal, Canada. Specifically, to underpin a causal interpretation, for the first time in the extent of transportation literature, we develop a heterogeneous endogenous econometric model that estimates the causal effect of proximity to school on pedestrian safety, addressing a complex endogenous relationship between the two. Various built environment, traffic exposure, and road geometric/operational characteristics are considered. The results indicate that if endogeneity is not accounted for, the effect of proximity to school is underestimated, while not being significant at a 5% level of significance. However, after accounting for confounding factors, the proposed endogenous model indicates that proximity to school deteriorates pedestrian safety. Therefore, traffic safety countermeasures and policies in place (if any) during the study period have not been sufficient and/or effective in improving pedestrian safety at intersections near schools. Our heterogeneity in mean and variance formulation provided more insights. For example, we found that, interestingly, as pedestrian volume increases at intersections around schools, the adverse effect of proximity to school on pedestrian safety decreases, a possibility not previously explored in the extent of road safety literature, confirming a strong safety-in-numbers effect.

Keywords: pedestrian safety, proximity to school, endogeneity, safety-in-numbers, built environment

1. INTRODUCTION

Pedestrian safety in proximity to schools is a major concern of transportation authorities, local governments, and residents. This could be mainly attributed to a relatively large proportion of child and adolescent pedestrians around schools. In fact, local authorities, by designing and implementing safety improvement programs, often aim to enhance traffic safety in proximity to schools, particularly for pedestrians, which in turn helps to promote commuting to school by walking. An exemplar of such programs is the federal safe route to school initiative in the US, which aimed to improve safety around schools and promote active commute to school (Yu and Zhu, 2016). This initiative was found to be effective in improving pedestrian and cyclist safety as discussed, for example, by DiMaggio and Li (2013) and Ragland et al. (2014).

There is a whole body of literature discussing various issues (urban planning, health and safety, etc.) relating to school siting (Ewing and Green, 2003; McDonald, 2010; Guliani et al., 2015; Yu and Zhu, 2016; Vitale et al., 2019). With respect to traffic safety, a relatively limited number of studies, as part of their larger analyses, have examined the association between pedestrian safety and proximity to school (or the relationship between safety and the number of schools in an area) at both micro and macro levels (Clifton and Kreamer-Fults, 2007; Clifton et al., 2009; Zahabi et al., 2011; Ukkusuri et al., 2012; Xin et al., 2017; Bhat et al., 2017). Previous studies have identified some common attributes associated with school siting that increase or decrease crash risk.

For instance, exposure to traffic increases the likelihood of crashes involving pedestrians, so reducing the interaction between pedestrians and vehicles through higher levels of more connected sidewalks contributes to a reduction in pedestrian crash risks (Yu and Zhu, 2016; Hwang et al., 2017). Schools located within a network of more local roads are safer than those near highways where the volume and speed of traffic is higher (Dumbaugh and Li, 2011; Yu, 2015; Yu and Zhu, 2016). Commercial land use tends to increase crash risks, as not only are there higher levels of traffic (Clifton and Kreamer-Fults, 2007), but entryways tend to be set back, resulting in more potential conflicts (Dumbaugh and Li, 2011). The presence of transit stops locally, which are likely to involve higher pedestrian activity and to be sited along major arterial highways for greater accessibility, also increase crash risk near schools (Miranda-Moreno et al., 2011; Ukkusuri et al., 2012; Yu and Zhu, 2016; Briz-Redón et al., 2017; Heydari et al., 2017).

However, these and other studies have mixed results (Merlin et al., 2020). In fact, there are inconsistencies in previous research regarding whether pedestrian crash and injury risk propensities increase or decrease in proximity to schools or by the number of schools, say, in a geographic area. For example, Zahabi et al. (2011) found that proximity to school decreases the likelihood of fatality or serious injury for pedestrians while Clifton et al. (2009) found the contrary. Similarly, there are inconsistencies in whether a higher proportion of children or adolescents in a specific area (which is often the case around schools) decreases pedestrian risk propensity for injuries. For instance, Bhat et al. (2017) found that a higher proportion of young individuals decreases pedestrian risk propensity for serious injuries. In contrast, a study conducted by Amoh-Gyimah et al. (2016) is not in accordance with Bhat et al.'s finding.

We speculate that such mixed results may be mostly because of the fact that proximity to school is often endogenous, as we describe below. Endogeneity issues are shown to play an important role in delineating inferences in the context of traffic safety analysis (Shankar and Mannering, 1998; Carson and Mannering, 2001; Kim and Washington, 2006; Winston et al., 2006; Eluru and Bhat, 2007; Bhat et al., 2014; Roesel, 2017; Sarwar et al., 2017; Xu et al., 2017; Afghari et al., 2018). For example, Kim and Washington (2006) addressed selection bias in estimating the effect of left-turning lanes on angle crashes. In their study, a simple negative binomial model indicated that the presence of left-turning lanes at intersections has an increasing effect on angle crashes; however, the authors found the contrary after controlling for confounding. Endogeneity issues could easily arise in the context of traffic safety around schools since safety treatments are not assigned randomly, and several interrelated confounding factors may affect crash mechanisms.

Some examples of "conflicting" factors affecting safety or the perception of safety (perceived safety importance) around schools are as follows: (i) speed limits are generally lower in proximity to schools, which is expected to reduce the risk of crash and injury severity sustained by pedestrians; (ii) land-use characteristics such as the area of commercial land-use could be different in proximity to schools increasing traffic exposure and conflicts between pedestrians and vehicles (Heydari et al., 2017; Bhat et al., 2017); (iii) school zones are supposed to be safer, for instance, due to an increased driver awareness because of the presence of traffic signs and school-age children, but other risk factors would be in place as well.

For example, school zones could be more dangerous due to a high proportion of children and teenagers who, due to their natural propensity for inattentiveness, could be more likely to be involved in risk-taking behaviors (e.g., unsafe crossing, disobeying signs, etc.) compared to adults. In fact, due to their limited cognitive abilities, very young individuals are more likely to be involved in traffic conflicts (Johnson et al., 2004; Stocker et al., 2015; Amoh-Gyameh et al., 2016; Bhat et al., 2017). In this regard, a study conducted by Rothman et al. (2016), considering child pedestrian safety around one hundred elementary schools in Toronto, indicated that dangerous pedestrian behaviors and/or drop-off were pervasive around schools; and finally (iv) perception of safety and risk compensation (offset hypothesis) could also affect safety mechanisms around schools. For instance, pedestrians could trade off enhanced safety in proximity to schools due to lower speed limits for less attentive behaviors resulting in an increased risk. A detailed discussion related to risk compensation is provided in Winston et al. (2006), who estimated the effect of airbags and antilock brakes on safety. Another exemplar of risk compensation can be found in Lv et al. (2015).

To summarize, safety mechanisms at intersections located in proximity to schools may be different from those away from schools in three major aspects: type of road users, their behavior, and safety features at nearby transportation facilities. Different safety improvement programs may be in place around schools so that transportation facilities (e.g., intersections or road segments) in proximity to schools are often subject to certain safety treatments. Since data are usually scarce in terms of details relating to type of interventions and their effects, one could focus on proximity to school to estimate the joint effect of various countermeasures that might be in place around schools in a jurisdiction. This could be justified due to the fact that safety interventions and policies, regardless of treatment types and their individual effectiveness, are collectively expected to enhance the safety of transportation facilities around schools.

Despite the general consensus about the importance of traffic safety around schools, studies that examine the causal effect of proximity to school on safety are surprisingly rare. While previous literature does not provide clear methodological or empirical guidelines, this paper introduces an econometric framework that quantifies the causal effect of proximity to school on pedestrian safety. This in turn allows the analyst to, indirectly, assess whether implemented safety policies and countermeasures around schools, if any, are truly effective in enhancing traffic safety.

1.1.The current paper

Based on the above discussion, the goal of this paper is to estimate the causal effect of proximity to school on pedestrian safety at signalized intersections. To this end, we answer the following questions:

(i) Are intersections in proximity to schools truly safer than those away from schools given a variety of implemented safety improvement programs and policies?
(ii) How can we answer the previous question properly using a reliable approach that accounts for possible confounding?

We therefore develop a framework to answer the above questions, focusing on intersection safety. However, our approach can be readily applied to examine safety of other transportation facilities around schools or where self-selection bias exists. Our model is an extension of the approach proposed initially by Terza (1998) in econometrics literature, and its later treatment by Kozumi (2002) from a Bayesian perspective. Specifically, a heterogeneous (random parameter) Bayesian endogenous econometric model is developed to identify safety correlates of pedestrian safety in urban areas using a host of variables, including built environment characteristics and traffic exposure measures. In accordance with real-world scenarios, the proposed approach allows the causal effect of proximity to school to vary across intersections (capturing its heterogeneous effect on pedestrian safety) by employing a heterogeneous endogenous model. We then extend to a heterogeneous endogenous model with heterogeneity in mean and variance. This helps us to provide further interesting insights with the same set of variables available in the data.

Due to computational complexities involved in the presence of endogeneity, most endogenous modeling applications, especially in traffic safety research, employed an instrumental variables approach (see Washington et al. (2011) for a discussion on this approach). However, as discussed by Washington et al. (2011), a system-equation method is more valuable since it explicitly accounts for the contemporaneous correlation between equations (and error terms). In this regard, a limited number of traffic safety studies, providing valuable methodological and empirical insights, have employed system-equation methods in the extent of the crash literature mostly during the past five to ten years (Eluru et al., 2007; Eluru et al., 2010; Paleti et al., 2010; Abay et al., 2013; Bhat et al., 2014; Lavieri et al., 2016). The estimation of system-equations is often computationally expensive and requires simulation-based approaches. An important study conducted by Bhat et al. (2014), however, estimates model parameters analytically. Of course, analytical approaches are often superior to simulation-based techniques. In this paper, similar to the latter examples, we directly estimate a system-equation model, explicitly accommodating the correlation between error terms. Note that in contrast to the propensity score method, which is also employed in traffic safety literature to address issues relating to confounding and selection bias (e.g., Li et al., 2013), the system-equation approach does not require a control group, avoiding sensitivity of inferences to the choice of a control group, thereby reducing data requirements.

While most instances of traffic safety studies that deal with endogeneity issues use classical statistics, we adopt a Bayesian approach — as a viable alternative — in estimating our endogenous specification. Alternative densities such as lognormal can be assumed for varying regression coefficients and can be easily estimated in our setting. More importantly, we employ a heterogeneity in mean and variance approach the use of which is rare if non-existent in investigating the (varying) effect of treatments, especially while considering a system-equation approach. Our model formulation therefore allows us to explain the varying effect of proximity to school in our data, revealing risk factors causing heterogeneity in the effect of proximity to school on pedestrian safety. A limited number of previous traffic safety studies have implemented and discussed different versions of such models, highlighting their advantages (Venkataraman et al., 2014; Seraneeprakarn et al., 2017; Heydari et al., 2018). These models are particularly appealing in assessing treatments and in the presence of endogeneity as they provide further empirical insights using the same set of variables available in the data. For example, we would be able to identify factors that minimize or maximize the beneficial effect of a safety intervention. This is while previous traffic safety research rarely allows the effect of treatments to vary across the sample. This is the first instance of such an endogenous model in the extent of road safety literature.

This is also the first study that examines safety-in-numbers (Elvik and Bjørnskau, 2017; Murphy et al., 2017) within a causal modeling framework, explicitly exploring how pedestrian safety in proximity to school varies according to pedestrian volume. Such an investigation, while being different from examining safety-in-numbers across the entire sample, has not been explored in the extent of road safety research previously and is of major empirical importance. That is, in this research, we investigate safety-in-numbers at two different levels. Further empirical contribution relates to the fact that, while many studies use proxy exposure measures (e.g., population density) for non-motorized and/or motorized flows in modeling pedestrian safety, we use direct pedestrian counts and disaggregated motorized (right-turning, non-turning, and left-turning) volume, providing more accurate and detailed information for safety policy design and analysis.

2. Methodological approach

Our approach is based on simultaneous equation modeling in which two sets of interrelated equations are specified to jointly model two correlated outcomes: proximity to school and pedestrian injury frequency. Various factors (e.g., built environment characteristics) that affect the presence of school in an area could affect pedestrian safety as well. At the same time, proximity to school could affect safety at intersections for various reasons discussed in Section 1. Therefore, proximity to school may be endogenous. Note that one may consider the presence of school in proximity of a transport facility (here intersections) as a treatment implemented non-randomly, leading to selection bias.

We model the presence of a school near an intersection as a binary outcome S that is associated with a latent variable $s*$ being a function of exogenous explanatory variables $\boldsymbol{Z}$ (observables that influence the presence of school in proximity to transportation facility of a specific type) and an error term ε.

$$S = g(\boldsymbol{Z}, \varepsilon) \tag{1}$$

If $s*$ has a positive value, the binary outcome S is equal to one; otherwise, it is equal to zero. The presence of school near a transportation infrastructure can thus be modeled as in (2), where we observe S; however, $s*$ is unobservable leading to a binary probit model. In our application, the latent variable $s*$ measures the propensity of the presence of a school in proximity to a transportation facility.

$$S = \begin{cases} 1 & if \quad s^* > 0 \\ 0 & if \quad s^* \leq 0 \end{cases} \tag{2}$$

Similarly, traffic safety (here, pedestrian injury counts) y is a function of exogenous variables $\boldsymbol{X}$ (various site characteristics) including built environment variables, proximity to school S (endogenous variable), and an error term ξ capturing unobservable factors that affect safety:

$$y = f(\boldsymbol{X}, S, \xi) \tag{3}$$

As discussed earlier, transportation facilities such as intersections in proximity to schools are more likely to be subject to safety treatments. Moreover, multiple factors may affect both y and S at the same time. Therefore, it is plausible to hypothesize that unobservable factors ξ not only affect y, but also affect S, leading to misleading inferences if only (3) is considered because S and ξ could be correlated. To avoid the endogeneity bias, we model y and S simultaneously by allowing for correlation between ε and ξ as described in the subsequent sections. We first discuss our proposed endogenous approach, followed by a heterogeneous Bayesian endogenous econometric model that assumes the effect of proximity to school varies across sites. We then extend the model to a heterogeneous endogenous econometric model with heterogeneity in mean and variance, which allows us to find an explanation for the varying effect of proximity to school on pedestrian safety.

2.1. Endogenous econometric model

Let y_i and λ_i, respectively, denote the observed and the expected injury frequency for intersections ($i = 1,2,\ldots, n$), $\boldsymbol{\beta} = (\beta_1, \beta_2,\ldots, \beta_m)$, including a constant term, be the vector of regression coefficients corresponding to $\boldsymbol{X}$, and γ be a fixed parameter associated with S. Since we account for endogeneity of proximity to school, γ implies causality. Given the above notation, crash frequencies (here, pedestrian injury counts) can be modeled as in (4).

$$\begin{aligned} &y_i \sim Poisson\ (\lambda_i) \\ &log(\lambda_i) = \boldsymbol{\beta X_i} + \gamma S_i + \xi_i \end{aligned} \tag{4}$$

Note that the above specification accounts for over-dispersion as it includes an error term. The next step is to define an equation for the latent variable s^*. Let $\boldsymbol{\delta} = (\delta_1, \delta_2,\ldots, \delta_k)$, including a constant term, denote the vector of regression coefficients corresponding to $\boldsymbol{Z}$. We can write

$$s_i^* = \boldsymbol{\delta Z_i} + \varepsilon_i \tag{5}$$

Thus, the larger the value of $s*$, the higher the probability of a school existing in proximity to a transport facility (here, an intersection). As in Terza (1998) and Kozumi (2002), the error terms (ε and ξ) are modeled jointly, assuming a bivariate normal (BVN) density with mean zero and covariance matrix Σ defined as

$$\Sigma = BVN \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \tag{6}$$

where the variance of ε_i (relating to the latent variable) is set to unity as the identification restriction; σ^2 is the variance of ξ_i and ρ is the correlation coefficient. This model formulation allows us to account for potential endogeneity of S. Note that in the above model formulation $\boldsymbol{X}$ and $\boldsymbol{Z}$ may be overlapping.

2.2. Heterogeneous endogenous econometric model

To account for unobserved heterogeneity (Mannering et al., 2016; Heydari, 2018) more fully, it is reasonable to consider that the causal effect of proximity to school on pedestrian safety may vary from one intersection to another. Therefore, one can allow γ to vary across the sample by modeling this coefficient as a random parameter. In random parameters models, the typical approach is to assume a parametric density (e.g., normal, lognormal, etc.) for random parameters.

Assuming a normal density for γ_i, we can summarize the above formulation in (7), where the effect of proximity to school varies across sites through γ_i.

$$\begin{aligned} &y_i \sim Poisson\ (\lambda_i) \\ &log(\lambda_i) = \boldsymbol{\beta X_i} + \gamma_i S_i + \xi_i \\ &\gamma_i \sim normal(\mu, v) \\ &s_i^* = \boldsymbol{\delta Z_i} + \varepsilon_i \\ &(\xi_i, \varepsilon_i) \sim BVN\left(0, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right) \end{aligned} \tag{7}$$

where μ and v are the mean and the variance of the random parameter associated with proximity to school.

2.3.Heterogeneous endogenous econometric model with heterogeneity in mean and variance

A more useful and informative approach, compared to the one discussed in Section 2.2, is to model the mean and the variance of the random parameter associated with proximity to school as a function of explanatory variables available in the data. In traffic safety research, such models are usually referred to as random parameters models with heterogeneity in mean and/or variance (Venkataraman et al., 2014; Seraneeprakarn et al., 2017; Heydari et al., 2018). As discussed in Section 1, we believe the latter approach is particularly useful in the context of causal inference such as the one discussed in this paper since it would explain the heterogeneous effect of an endogenous risk factor (or treatment) of interest; here, proximity to school. That is, the model allows the analyst to explain variations in both mean and variance of the varying effect as a function of explanatory variables available in the data.

In this paper, we model the mean and the variance of the random parameters γ_i as a function of explanatory variables $\boldsymbol{W_\mu}$ and $\boldsymbol{W_v}$ as in (8), where η is a constant term and $\boldsymbol{\alpha}$ is the vector of coefficients associated with the vector of explanatory variables $\boldsymbol{W}$.

$$\begin{aligned} &y_i \sim Poisson\ (\lambda_i) \\ &log(\lambda_i) = \boldsymbol{\beta X_i} + \gamma_i S_i + \xi_i \\ &\gamma_i \sim normal(\mu_i, v_i) \\ &\mu_i = \eta_\mu + \boldsymbol{\alpha}_\mu \boldsymbol{W}_{\mu_i} \\ &v_i = \eta_v + \boldsymbol{\alpha}_v \boldsymbol{W}_{v_i} \\ &s_i^* = \boldsymbol{\delta Z_i} + \varepsilon_i \\ &(\xi_i, \varepsilon_i) \sim BVN\left(0, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right) \end{aligned} \tag{8}$$

Note that model formulations defined in (7) and (8) account for both endogeneity and heterogeneity of proximity to school. However, the model specified in (8) relaxes the homoscedasticity assumption in the effect of proximity to school. That is, while the formulation in (7) assumes that the variance of the effect is fixed, heteroscedasticity is explicitly accounted for in (8).

2.4.Specification of priors and model estimation

We specified non-informative normal priors normal(0, 100) for regression coefficients ($\boldsymbol{\beta}$, $\boldsymbol{\delta}$, η, and $\boldsymbol{\alpha}$) and the mean parameters (μ and m) associated with random parameters. We took advantage of the possibility of specifying prior beliefs through prior specification in the Bayesian approach (see, e.g., Heydari et al., 2013; Miranda-Moreno et al., 2013; Heydari et al., 2014; Nowakowska, 2017; Farid et al., 2017). We specified truncated normal priors (truncated at zero) for the regression coefficients η_v and $\boldsymbol{\alpha_v}$ in (8) to ensure the positivity of the variance term v_i. Similarly, we specified a uniform(-1, +1) for correlation ρ. The choice of the latter prior was to ensure that the value of correlation falls in the desired interval [-1,1]. Note that, for example, the non-linear weighted least squares, suggested by Terza (1998), may lead to unrealistic values of correlation that are beyond the desired interval when modeling (1) and (3) jointly (Kozumi, 2002). Relatively vague uniform priors uniform(0, 10) were specified for standard deviation σ_v corresponding to variance parameters v. For parameters in the covariance matrix defined in (6), we specified relatively vague uniform priors uniform(0, 10) for σ.

In this paper, even though more complex endogenous models are formulated in (7) and (8), the estimation can be easily handled within the Bayesian framework through standard Markov chain Monte Carlo (MCMC) algorithms. Using a data augmentation approach (Albert and Chib, 1993; Kozumi, 2002), the presence of school can be modeled as in (2) in which we observe S; however, the vector of latent variables $s^* = (s_1^*,\ldots, s_n^*)$ is unobservable and is jointly drawn at each iteration of our MCMC simulations with other unknown model parameters. Using freely available software WinBUGS (Lunn et al., 2000), we ran 40,000 iterations with two chains, discarding the first 15,000 iterations for convergence requirements. The posterior densities are thus obtained based on 50,000 samples. Gelman-Rubin statistic (Gelman and Rubin, 1992), history plots, and Monte Carlo errors were used to ensure convergence and the accuracy of estimates.

3. Data and variables considered

The data used in this paper contain pedestrian injury counts for 647 signalized intersections in Montreal from 2003 to 2008. Summary statistics of the data are provided in Table 1. The spatial distribution of the intersections is displayed in Fig. 1. While details on various data sources and the way these were cleaned and processed are provided in Strauss et al. (2014), here we discuss the major features of the final data set.

The pedestrian injury frequencies for these sites were provided by Montreal ambulance services for the 6-year study period. Accidents are considered as having occurred at an intersection if they are within 15m of the center point of the intersection, and were provided at the level of the individual rather than at the level of the crash. Differing injury-severity levels (e.g., minor, major, fatal) were not available in the data so that conducting an injury-severity analysis was not possible. Other data related to road environment were made available:

(i) Exposure measures including pedestrian and vehicle counts were obtained following a data collection campaign undertaken at signalized intersections in Montreal in 2008 and 2009 for an 8-hour period (6:00am–9:00am, 11:00am–1:00pm and 3:30pm–6:30pm) (Strauss et al., 2014);
(ii) Intersection geometric and control characteristics collected from site visits (by a data collection team from McGill University) in 2010-2012, including characteristics such as intersection form (e.g., three-leg), crosswalk length, the number of approaches and lanes, presence and width of medians, crosswalk length, approach speed limit, signal type, phasing; and
(iii) Built environment and socio-economic variables using geographic information systems data from a variety of sources (Statistics Canada, DMTI Spatial Inc., Société de transport de Montréal (STM) and Agence métropolitaine de transport), including the length of cycling facilities, employment, area of commercial land use, location of schools, presence of metro (subway) stations, bus stops, and schools in the proximity of intersections (within a range of 50 m, 400 m or 800 m).

Vehicle counts were provided at a disaggregate level (left-turning, right-turning, non-turning flows), being free from ecological bias. This allowed us to capture variation in the effect of each disaggregate traffic volume on pedestrian safety; and eventually, identify the one that has the largest impact on pedestrian injury frequencies. Relating to pedestrian safety and traffic control, dedicated pedestrian signal was another important variable available in the data. Where these traffic lights are installed, a full phase or part of a phase is given to pedestrian crossing, keeping other traffic lights red. This is expected to improve the safety of pedestrians at signalized intersections as found, for example, by Heydari et al. (2017). Among 647 signalized intersections, 160 (24.7%) had traffic lights with a dedicated pedestrian signal.

To define the presence of school in the vicinity of intersections, a buffer of 400 m around intersections was considered. The choice of this buffer is in accordance with earlier literature (Clifton et al., 2009; Zahabi et al., 2011). Among 647 signalized intersections, 364 (56.26%) intersections were in a distance of less than 400 m from a school. Lastly, we created a new variable being the ratio of non-motorized volume, based on the idea that this ratio may affect injury counts among pedestrians. In fact, a number of previous studies indicate that the safety of pedestrians improves as this ratio increases. In this regard, interested readers are referred to Leden (2002), Pucher and Buehler (2008), Jacobsen (2015), and Stoker et al. (2015).

4. Results and discussion

Tables 2, 3 and 4 provide a summary of the results. As reported in Table 2, proximity to school is not statistically important at a 5% level of significance when using a simple non-endogenous count model. After addressing endogeneity, however, proximity to school becomes statistically important in explaining pedestrian safety. The results indicate that neglecting the covariance between the propensity of proximity to school and pedestrian crash risk propensity, results in an under-estimated effect of proximity to school on pedestrian injury counts. The results of the endogenous (fixed-effects) model and both heterogeneous econometric models are relatively similar, except in the effect of proximity to school that is found to vary across the sample, so we do not discuss the former further. The varying effect of proximity to school on pedestrian safety can be explained, for example, due to variations in intersection geometric and operational characteristics—reflecting the reality of the data. The heterogeneity in mean and variance specification allowed us to identify the sources of variability in the varying effect of proximity to school, providing useful empirical information (see Section 4.2.3).

The estimated correlation parameters (and their standard deviations) under the heterogeneous model and the one with heterogeneity in mean/variance are, respectively, -0.23 and -0.3. When allowing for heterogeneity in mean and variance of the variable proximity to school, the model fit improves; the deviance information criterion (DIC) (Spiegelhalter et al., 2002) reduces from 1579.47 to 1572.95. Note also that in this research we did not examine temporal instability (Mannering, 2018), which may affect our results. Obviously, addressing temporal instability in the presence of endogeneity would add to the complexity of the problem. Also, note that we modeled other independent variables as random parameters; however, this was unnecessary as variances were negligible and/or model fit did not improve. Finally, we considered other distributions, such as lognormal, to model the varying effect of proximity to school as an alternative to a normal density; however, the model fit deteriorated as DIC increased from 1579.47, in the normally distributed random parameter specification, to 1586, in the lognormally distributed specification.

4.1. Posterior estimation summary of regression coefficients

Both heterogonous models provide similar estimations of regression coefficients in general although the differences become more obvious when considering other criteria and features of the models, as discussed in Section 4.2. Based on the binary probit model in our system-equation approach, we found that employment, the presence of bus stop, the ratio of non-motorized traffic to motorized traffic, and land use mix increase the propensity of the presence of school in the vicinity of intersections. Based on the Poisson count model in our system-equation approach, we found that pedestrian counts, left-turning AADT, the area of commercial land use, dedicated traffic lights for pedestrians, and proximity to school can explain pedestrian injury frequencies at signalized intersection in Montreal during the study period. It is important to highlight that based on our analysis, right-turning AADT and non-turning AADT did not have any statistically important effects on pedestrian safety.

With respect to comparison with previous research, it is important to consider that, since we conducted an intersection-level study (and used a system-equation approach), a direct comparison with mostly macro-level studies (or those that focus on injury-severity analysis) is not straightforward. For example, we used "exact" exposure measures (i.e., pedestrian and vehicle counts) while most previous studies used proxy exposure measures such as population, vehicle kilometers traveled, total transit commuters, and the number of commuters commuting to work (see e.g., Ukkusuri et al. (2012), Cai et al. (2016), and Yasmin and Eluru (2016)).

Due to the use of "exact" exposure measures, many proxy measures appear to be less useful in our model. We also use disaggregated vehicle volumes (left turning, right turning, and non-turning flows) that increases the accuracy of the model, allowing us to isolate the association between varying turning movements and pedestrian safety. This in turn enables us to make a stronger empirical contribution compared to many previous studies when it comes to understanding the effect of exposure measures, and designing safety interventions more effectively. As discussed by Washington et al. (2010), in general, a micro-level analysis provides more detailed insights (compared to an area-level analysis) and would be more useful in terms of designing and implementing safety treatments.

Overall, our results are in accordance with Strauss et al. (2014) and Heydari et al. (2017) who analyzed the same data. Similar to the latter studies, for example, we found that dedicated pedestrian crossing signal improves pedestrian safety as indicated above. The main difference is that the presence of school was not found to be statistically important in the model developed by the above-mentioned authors since issues of confounding relating to the presence of schools were not accounted for. Another difference is that Strauss et al. (2014) considered only total vehicle volume as the motorized exposure measure. Note that Heydari et al. (2017) modeled pedestrian and cyclist injury frequencies simultaneously while Strauss at al. (2014) developed separate models for pedestrians and cyclists.

Our study is in accordance with previous research with respect to the effect of exposure and commercial land use on pedestrian safety (e.g., see Ukkusuri et al., 2012; Bhat et al., 2017). Similar to previous research (e.g., Strauss et al., 2014; Yu and Zhu, 2016), we found that the presence of bus stop in the proximity of intersections deteriorates pedestrian safety. For example, Yu and Zhu (2016) found that higher transit stop densities (in an area) decrease pedestrian safety. As discussed in previous studies the deteriorating effect of commercial land use and the presence of bus stop on pedestrian safety can be explained due to the fact that these factors increase conflicts between motorized traffic and pedestrians (Yu and Zhu, 2016). Similarly, a higher number of driveways on commercial zones is associated with higher potential for conflicts between various road users, and increase crash risk propensity (Dumbaugh and Li 2011; Strauss et al., 2014; Yu and Zhu, 2016).

As Zahabi et al. (2011), we found that speed limit does not have an important effect on pedestrian safety; however, some previous studies such as Ewing and Dumbaugh (2009) and Abdel-Aty et al. (2007) indicated that speed is associated with higher pedestrian crash injury/fatality risks. In this regard, as discussed by Yu and Zhu (2016), the key issue is speed differential and the fact that vehicles traveling at higher speeds on arterials entering local roads often increase conflicts with pedestrians. An explanation for our finding perhaps is that speed differential is not considerable in our data; and therefore, speed limit does not appear to have a significant effect on safety. Also, another explanation could be that in dense urban areas (such as Montreal) traffic speed is often low and speed limit may not necessarily represent operating speed (Ewing and Dumbaugh, 2009).

Other factors such as employment and land use mix that have been shown to have a bearing on pedestrian safety (e.g., see Cai et al. (2016); Heydari et al. (2017)) appear in the binary probit component of our system-equation. The sign of employment is positive in previous research (Miranda-Moreno et al., 2011; Cai et al., 2016) and in our binary probity component; therefore, our finding is in accordance with most previous research. This is because in our study employment is positively associated with the presence of school, and the latter is positively associated with pedestrian safety. Similarly, our finding with respect to land use mix is in accordance with previous studies such as Clifton and Kreamer-Fults (2007), Zahabi et al. (2011).

Although pedestrian volume and injury frequency are positively correlated, given their non-linear relationship and the values of coefficients that are smaller than one, this increase is less than in proportion to traffic exposure, indicating a safety-in-numbers effect (Elvik and Bjørnskau, 2017). This finding is in accordance with earlier literature that suggests higher pedestrian activity may improve pedestrian safety (Jacobsen, 2015; Stoker et al., 2015; Murphy et al., 2017).

4.2. Policy analysis

We discuss policy implications from different perspectives: marginal effects, the covariance in our system-equation approach, and the sources of variability in the effect of proximity to school.

4.2.1. Marginal effects

To clearly understand the effect of the explanatory variables on pedestrian injury frequency, we estimated the average marginal effects. These can be obtained, for example, for the *m*th continuous variable and proximity to school (a categorical variable) from (9) and (10), respectively.

$$\frac{\partial E(y|S_i, \boldsymbol{X_i}, \boldsymbol{\beta}, \gamma_i)}{\partial (X_m)} = \frac{1}{N}\sum_{i=1}^{N} \beta_m exp(\boldsymbol{\beta X_i} + \gamma_i S_i) \tag{9}$$

$$\frac{\partial E(y|S_i, \boldsymbol{X_i}, \boldsymbol{\beta}, \gamma_i)}{\partial (S)} = \frac{1}{N}\sum_{i=1}^{N} [exp(\boldsymbol{\beta X_i} + \gamma_i S_i) - exp(\boldsymbol{\beta X_i})] \tag{10}$$

The estimated average marginal effects (and their standard deviations) are reported in Table 5. The marginal effects of exposure variables (pedestrian flow and motorized traffic volume) are similar under both models although the heterogeneous endogenous model slightly underestimates the effect of exposure on pedestrian safety (see Table 5). Under both models, pedestrian volume has a greater impact on pedestrian safety compared to motorized traffic volume. We found that as the exposure variables increase, pedestrian injury frequency increases. Based on the heterogeneous endogenous model with heterogeneity in mean and variance, we infer from the results that intersections that are in proximity to schools have an average expected pedestrian injury frequency that is 0.239 higher than other intersections. Similarly, intersections that are in proximity of bus stops have an average expected pedestrian injury counts that is 0.375 higher than other intersections. Intersections that are equipped with dedicated traffic lights for pedestrian crossing have an average expected pedestrian injury frequency that is 0.124 lower than other intersections. Marginal effects also indicate that one unit (here, 10000 m^2) increase in the area of commercial land use, on average, leads to 0.103 increase in the expected pedestrian injury frequency.

4.2.2. Implications of the estimated covariance in our system-equation

The covariance between the proximity-to-school propensity and the pedestrian crash risk propensity was found to be negative, -0.25. This is an important finding in that the sign of the covariance indicates that unknown factors that increase the likelihood of the presence of school in the vicinity of intersections, decrease the pedestrian crash risk propensity. Interestingly, this implies that intersections in proximity to schools are in general low crash risk propensity locations. At the same time, the positive sign of the parameter associated with proximity to school indicates that this variable increases pedestrian injury frequencies. Therefore, one may conclude that this is not due to the fact that intersections in the vicinity of schools are inherently high crash propensity locations, but other factors may play a key role here. Certainly, this interesting finding requires further in-depth investigations (including more detailed data). Firstly, it is imperative to understand the mechanism in which safety interventions are selected and applied around schools, which may vary from one jurisdiction to another. This said, we hypothesize that perhaps implemented countermeasures in school zones and those (if any) at intersections in the vicinity of schools but out of the boundary of school zones may be in contrast somehow. Such issues may cause confusion among road users; the overall effect of which makes the intersections in proximity to schools less safe. A similar discussion with respect to the effect of flashing lights on intersection safety is provided by Bhat et al. (2014). One may also consider the phenomenon of road crash migration and the so-called “Kangaroo” effect, similar to what happens in proximity to locations where speed cameras are installed. For a discussion in this regard, see Mountain et al. (2005) and Li et al. (2013). Around schools, for example, drivers may trade off lower speeds in school zones for higher speeds and/or abrupt deceleration/acceleration before or after school zones, decreasing traffic safety.

4.2.3. Sources of variability in the effect of proximity to school on safety

The sources of variability (and the magnitude of their impact) in the effect of proximity to school were identified by employing a heterogeneity in mean and variance specification (discussed in Section 2.3). We found that variation in the mean of the random regression coefficient associated with proximity to school can be explained by pedestrian volume. Specifically, the results showed that the mean effect of proximity to school on pedestrian injury frequency decreases as pedestrian volume increases, suggesting a strong safety-in-numbers effect at intersections around schools. This means that an increased pedestrian volume contains the identified adverse effect of proximity to school on pedestrian safety. In the extent of road safety research, to our knowledge, this is the first instance of investigating safety-in-numbers effect with respect to the effect of a risk factor (e.g., proximity to school) on safety, using a rigorous model formulation. Note that previous research, which discusses safety-in-number issues, mostly studied the direct effect of pedestrian volume on safety not its impact on the effect of other factors on safety. This is why as discussed in Section 4.1, the model indicates that pedestrian volume is positively (while non-linearly) associated with pedestrian safety. Given the equation for the mean of the random parameter associated with proximity to school (see Table 4), for a 10% increase in pedestrian volume, the difference in the mean effect of proximity to school on pedestrian safety will be smaller than zero: -0.014; i.e., ln(1.10)x(-0.152).

The results also indicated that the variance of the varying effect of proximity to school is not independent across intersections. Specifically, we found that this variance can be explained by the variable dedicated traffic lights for pedestrians. This implies that the net effect of uncertainty is systematic across the sample (for a general discussion on issues relating to heteroscedasticity, see Washington et al., 2011). In fact, the variance of the effect of proximity to school increases at intersections equipped with dedicated pedestrian traffic lights, implying that the effect of proximity to school is more dispersed in the presence of this specific control type. Given the equation for the variance of the effect of proximity to school (see Table 4), the variance increases significantly (almost 13 times) from 0.09 to 1.16 in the presence of dedicated traffic lights. A detailed investigation (perhaps, including field visits and additional data collection) is therefore required to reveal reasons behind this noteworthy finding. For example, certain risk factors might play a role in hampering the beneficial effect of dedicated pedestrian lights in the vicinity of schools.

5. Summary and conclusions

With respect to the association between proximity to (or presence of) school and traffic safety, previous studies have mixed results perhaps due to ignoring issues of confounding. This article contributes to the crash literature in developing a Bayesian heterogeneous endogenous econometric model that estimates the causal effect of proximity to school on pedestrian safety. The model allows the effect of proximity to school to vary across the sample, capturing unobserved heterogeneity more fully. Within a causal framework, we employ a heterogeneity in mean and variance model providing more insights, with the same set of variables available in the data. Our proposed model allows us to explain variation in both mean and variance of the random parameter associated with proximity to school. Also, in spite of most causal settings developed in traffic safety research, we assign a density to the correlation parameter to better capture uncertainty around this parameter.

We used disaggregate motorized exposure measures disentangling different motorized movements. Doing so allowed us to reveal their individual effects on pedestrian safety, which has valuable policy implications. For example, this is useful in implementing more cost-effective countermeasures according to each disaggregate motorized volume compared to the overall motorized volume. We found that pedestrian volume, left-turning volume, the presence of bus stop, and the area of commercial land use are positively associated with pedestrian injury counts, decreasing pedestrian safety. However, dedicated traffic lights for pedestrians is negatively associated with pedestrian injury counts, improving pedestrian safety. After adjusting for endogeneity, we found that pedestrian injury counts increase at intersections located in proximity to schools. We found that proximity to school increases pedestrian injury frequencies at nearby intersections. This indicates that intersections in proximity to schools were not safer than other intersections during the study period. Consequently, it can be implied that the collection of safety interventions that were in place, if any, were not effective in improving pedestrian safety at intersections around schools. Obviously, this confirms the need for further in-depth investigations to be able to plan more effective safety improvement programs.

This work also investigated safety-in-numbers at two different levels simultaneously for the first time in the extent of road safety literature. First, it examined safety-in-numbers by considering the overall effect of pedestrian volume on safety across the sample (the usual approach considered by previous research). Second, for the first time, this study investigated safety-in-numbers within a causal modeling framework, explicitly exploring how pedestrian safety in proximity to schools varies according to pedestrian volume. Based on the heterogeneity in mean and variance model, we found that as pedestrian volume increases at intersections around schools, the effect of proximity to school on pedestrian safety decreases. This important finding implies that a higher level of pedestrian activity would enhance pedestrian safety around schools.

In the first case, the sign of pedestrian volume was found to be positive although it supports a safety-in-numbers effect as the coefficient is smaller than 1. In the second case, the sign of pedestrian volume is negative, indicating a strong safety-in-numbers effect. This interesting finding reveals that safety-in-numbers effects may vary even across the same data set, for example, according to location or the effect of risk factors on safety. The practical implication of our finding with respect to safety-in-numbers is that while most previous research discusses that improving pedestrian safety helps promote walking commute to schools, our results imply that local authorities should also consider increasing the level of walking commute as a viable alternative to improve pedestrian safety at nearby intersections.

Also, it is imperative to monitor pedestrian injury frequencies and driver behavior (e.g., compliance with speed limits) at intersections that are in the vicinity of schools to assess the possibility of crash migration. Police enforcement and publicity campaigns to raise awareness among road users would be needed to enhance safety in such locations. We found that proximity to school decreases safety although nearby intersections are intrinsically safe. As discussed by Bhat et al. (2014) with respect to flashing lights, local authorities should perhaps consider implementing safety treatments based on concrete evidence (e.g., high crash risk location) rather than obvious features such as geometry. Our inference in this regard remains, however, inconclusive as further detailed studies are needed to understand the underlying mechanism of crashes around schools.

With respect to engineering interventions, signalized intersections around schools should be equipped with dedicated pedestrian traffic lights to provide further protection to pedestrians. A particular attention should be given to implementing countermeasures at intersections where left-turning movement is allowed since left-turning vehicles have a deteriorating influence on pedestrian safety. Using traffic-calming strategies, as discussed by Yu and Zhu (2016), should be considered in locations where bus stops are present. Also, education and publicity campaigns should be considered to raise awareness among drivers and pedestrians, especially school-age children, about the high crash risk at transit stops. Finally, it is important to limit the area of commercial land use around schools as evidence shows commercial activities increase conflicts and decrease safety.

The data set employed in this research allowed us to conduct a micro-level (intersection-level) analysis, which is more suitable (compared to macro-level analysis) to provide detailed insights into designing safety improvement programs. However, more valuable information would be obtained by including other variables such as school enrolment and the presence of universities and colleges in the vicinity of intersections. Note that our findings are based on the case study used in this article during a specific study period; and therefore, caution should be taken in generalizing the findings to other cities or time periods. Other factors related to safety such as traffic calming measures may be endogenous as well; we should highlight that the modeling approach proposed in this research is suitable for evaluating the effectiveness of such treatments as well.

ACKNOWLEDGEMENTS

The authors would like to thank the City of Montreal, the Department for Public Health, Urgence-Sante' Montreal, and the data collection and preparation team at McGill University.

REFERENCES

Albert, J., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669–679.

Afghari, A., Washington, S., Haque, M., Li, Z., 2018. A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk. Analytic Methods in Accident Research 18, 1-14.

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: a cross-comparison of estimation methods. Accident Analysis and Prevention 93, 147-159.

Bhat, C., Astroza, S., Lavieri, P., 2017. A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. Analytic Methods in Accident Research 16, 1–22.

Bhat, C., Born, K., Sidharthan, R., Bhat, P., 2014. A count data model with endogenous covariates: formulation and application to roadway crash frequency at intersections. Analytic Methods in Accident Research 1, 53-71.

Briz-Redón, Á. Martínez-Ruiz, F. Montes, F., 2019. Estimating the occurrence of traffic accidents near school locations: A case study from Valencia (Spain) including several approaches. Accident Analysis and Prevention, 132.

Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models. Accident Analysis and Prevention 93, 14–22.

Carson, J., Mannering, F., 2001. The effect of ice warning signs on ice-accident frequencies and severities. Accid. Anal. Prev. 33 (1), 99–109.

Clifton, K.J., Burnier, C., Akar, G., 2009. Severity of injury resulting from pedestrian–vehicle crashes: what can we learn from examining the built environment? Transportation Research Part D 14 (6), 425-436

Clifton, K.J., Kreamer-Fults, K., 2007. An examination of the environmental attributes associated with pedestrian–vehicular crashes near public schools. Accident Analysis and Prevention 39 (4), 708-715.

DiMaggio, C., Li, G., 2013. Effectiveness of a safe routes to school program in preventing school-aged pedestrian injury. Pediatrics 131 (2), 290–96.

Dumbaugh, E., Li, W., 2011. Designing for the safety of pedestrians, cyclists, and motorists in urban environments. Journal of the American Planning Association 77 (1), 69–88.

Eluru, N., Bhat, C., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention 39 (5), 1037-1049.

Elvik, R., Bjørnskau, T., 2017. Safety-in-numbers: a systematic review and meta-analysis of evidence. Safety Science 92, 274–282.

Ewing, R., Dumbaugh, E., 2009. The built environment and traffic safety: a review of empirical evidence. Journal of Planning Literature 23 (4), 347–67.

Ewing, R., Greene, W., 2003. Travel and Environmental Implications of School Siting. Washington, DC: US Environmental Protection Agency.

Farid, A., Abdel-Aty, M., Lee, J., Eluru, N., 2017. Application of Bayesian informative priors to enhance the transferability of safety performance functions. Journal of Safety Research 62, 155–161.

Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4), 457–472.

Guliani, A., Mitra, R., Buliung, R., Larsen, K., Faulkner, G., 2015. Gender-based differences in school travel mode choice behaviour: examining the relationship between the neighbourhood environment and perceived traffic safety. Journal Transport and Health 2 (4), 502–511.

Heydari, S., 2018. A flexible discrete density random parameters model for count data: embracing unobserved heterogeneity in highway safety analysis. Analytic Methods in Accident Research 20, 68–80.

Heydari, S., Fu, L., Miranda-Moreno, L., Joseph, L., 2017. Using a flexible multivariate latent class approach to model correlated outcomes: a joint analysis of pedestrian and cyclist injuries. Analytic Methods in Accident Research 13, 16–27.

Heydari, S., Fu, L., Thakali, L., Joseph, L., 2018. Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: applications to grade crossing safety analysis. Analytic Methods in Accident Research 19, 33–48.

Heydari, S., Miranda-Moreno, L., Amador, L., 2013. Does prior specification matter in hotspot identification and before-after studies? Transportation Research Record 2392, 31-39.

Heydari, S., Miranda-Moreno, L., Lord, D., Fu, L., 2014. Bayesian methodology to estimate and update safety performance functions under limited data conditions: a sensitivity analysis. Accident Analysis and Prevention 64, 41-51.

Hwang, J. Joh, K. Woo, A., 2017. Social inequalities in child pedestrian traffic injuries: differences in neighborhood built environments near schools in Austin, TX, USA. Journal of Transport and Health 6, 40–49.

Jacobsen, P., 2015. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Injury Prevention 21 (4), 271–275.

Kim, D., Washington, S., 2006. The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. Accident Analysis and Prevention 38 (6), 1094-1100.

Kozumi, H., 2002. A Bayesian analysis of endogenous switching models for count data. Journal of the Japan Statistical Society 32 (2), 141-154.

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(4), 457–464.

Li, H., Graham, D., Majumdar, A., 2013. The impacts of speed cameras on road accidents: an application of propensity score matching methods. Accident Analysis and Prevention 60, 148-157.

Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10 (4), 325–337.

Lv, J., Lord, D., Zhang, Y., Chen, Z., 2015. Investigating Peltzman effects in adopting mandatory seat belt laws in the US: Evidence from non-occupant fatalities. Transport Policy 44, 58-64.

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analytic Methods in Accident Research 17, 1–13.

Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1-16.

McDonald, N., 2010. School siting: contested visions of the community school. Journal of the American Planning Association 76 (2), 184–98

Merlin, L., Guerra, E., Dumbaugh, E., 2020. Crash risk, crash exposure, and the built environment: A conceptual review. Accident Analysis and Prevention 134, 105244.

Miranda-Moreno, L., Heydari, S., Lord, D., Fu, L., 2013. Bayesian road safety analysis: incorporation of past evidence and effect of hyper-prior choice. Safety Research 46, 31-40.

Miranda-Moreno, L., Morency, P., El-Geneidy, A.M., 2011. The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. Accident Analysis and Prevention 43 (5), 1624–1634.

Mountain, L., Hirst, W., Maher, M., 2005. Are speed enforcement cameras more effective than other speed management measures?: The impact of speed management schemes on 30mph roads. Accident Analysis & Prevention 37 (4), 742-754.

Murphy, B., Levinson, D., Owen, A., 2017. Evaluating the safety in numbers effect for pedestrians at urban intersections. Accident Analysis and Prevention 106, 181-190.

Nowakowska, M., 2017. Selected aspects of prior and likelihood information for a Bayesian classifier in a road safety analysis. Accident Analysis and Prevention 101, 97-106.

Pucher, J., Buehler, R., 2008. Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. Transport Reviews 28 (4), 495–528.

Ragland, D., Pande, S., Bigham, J., Cooper, J., 2014. Ten years later: examining the long-term impact of the california safe routes to school program. Paper presented at the Transportation Research Board 93rd Annual Meeting, Washington, DC.

Roesel, F., 2017. The causal effect of wrong-hand drive vehicles on road safety, CEPIE Working Paper, No. 15/17, Technische Universität Dresden, Center of Public and International Economics (CEPIE), Dresden.

Sarwar, M., Fountas, G., Anastasopoulos, P., 2017. Simultaneous estimation of discrete outcome and continuous dependent variable equations: a bivariate random effects modeling approach with unrestricted instruments. Analytic Methods in Accident Research 16, 23-34.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., Milton, J., 2017. Occupant injury severities in hybrid-vehicle involved crashes: a random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research 15, 41–55.

Shankar, V., Mannering, F., 1998. Modeling the endogeneity of lane-mean speeds and lane-speed deviations: a structural equations approach. Transportation Research Part A: Policy and Practice 32 (5), 311-322.

Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2002. Bayesian measures of complexity and fit (with discussion). Journal of the Royal Statistics Society, Series B 64 (4), 1–34.

Stoker, P., Garfinkel-Castro, A., Khayesi, M., Odero, W., Mwangi, M., Peden, M., Ewing, R., 2015. Pedestrian safety and the built environment: a review of the risk factors. Journal of Planning Literature 30 (4), 377–392.

Strauss, J., Miranda-Moreno, L., Morency, P., 2014. Multimodal injury risk analysis of road users at signalised and non-signalised intersections. Accident Analysis and Prevention 71, 201-209.

Terza, J., 1998. Estimating count data models with endogenous switching: sample selection and endogenous treatment effects. Journal of Econometrics 84 (1), 129-154.

Ukkusuri, S., Miranda-Moreno, L., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Safety Science 50 (4), 1141-1151.

Venkataraman, N., Ulfarsson, G., Shankar, V., Deptuch, D., 2014. A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. Analytic Methods in Accident Research 2, 12–20.

Vitale, M., Millward, H., Spinney, J., 2019. School siting and mode choices for school travel: Rural–urban contrasts in Halifax, Nova Scotia, Canada. Case Studies on Transport Policy 7 (1), 64-72.

Washington, S.P., Van Schalkwyk, I., You, D., Shin, K., Samuelson, J.P., 2010. Forecasting the safety impacts of socio-demographic changes and safety countermeasures. NCHRP 8–44(2): Plansafe User Manual. Transportation Research Board, Washington, DC.

Winston, C., Maheshri, V., Mannering, F., 2006. An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes. Journal of Risk and Uncertainty 32 (2), 83–99.

Xin, C., Guo, R., Wang, Z., Lu, Q., Lin, P., 2017. The effects of neighborhood characteristics and the built environment on pedestrian injury severity: a random parameters generalized ordered probability model with heterogeneity in means and variances, Analytic Methods in Accident Research 16, 117-132.

Xu, X., Wong, S., Zhu, F., Pei, X., Huang, H., Liu, Y., 2017. A Heckman selection model for the safety analysis of signalised intersections. PLOS ONE 12(7): e0181544. https://doi.org/10.1371/journal.pone.0181544

Yu, C. Y., 2015. How differences in roadways affect school travel safety. Journal of the American Planning Association 81 (3), 203–220.

Yu, C.Y., Zhu, X., 2016. Planning for safe schools: Impacts of school siting and surrounding environments on traffic safety. Journal of Planning Education and Research 36 (4), 476–486.

Zahabi, A., Strauss, J., Manaugh, K., Miranda-Moreno, L., 2011. Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes. Transportation Research Record 2247 (1), 81–90.

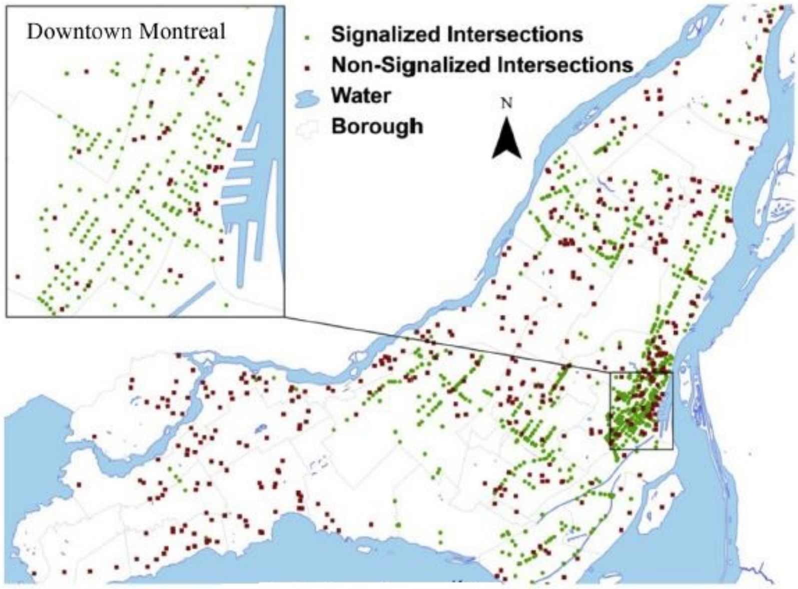

Fig. 1. Spatial distribution of intersections in the data (from Strauss et al. (2014)).

Table 1. Summary statistics of the data

Variable types	Variables	Mean	Std. Dev.	Min	Max
Crash type	Pedestrian injury counts	1.15	1.88	0.00	16.00
Exposure measure	Pedestrian counts	1578.07	3531.82	1.00	40958.00
	Total non-motorized volume	2022.99	3792.45	2.96	41541.00
	Left-turning motorized volume	2602.72	2641.86	0.00	23843.00
	Right-turning motorized volume	2668.01	2697.45	0.00	23792.00
	Non-turning motorized volume	19467.96	11084.39	1790.00	76525.00
	Total motorized volume (AADT)	24738.65	12526.06	3751.00	84386.00
	Ratio of non-motorized to motorized	0.13	0.30	0.00	4.01
Built environment	Employment[1] (in 0000s)	0.58	0.30	0.03	1.49
	Commercial area[1] (in 0000s m^2)	1.21	1.34	0.00	8.70
	Land use mix[1]	0.67	0.15	0.00	0.92
	Length of cycling facilities[2] (km)	0.54	0.61	0.00	2.96
	Number of schools[2]	1.05	1.24	0.00	6.00
	Number of subway stations[2]	0.44	0.70	0.00	4.00
	Presence of bus stop[3]	0.71	0.46	0.00	1.00
	Presence of School[2]	0.56	0.50	0.00	1.00
	Presence of subway stations[2]	0.34	0.48	0.00	1.00
Geometric/operational	Maximum speed (km/h)	61.82	9.88	50.00	100.00
	Dedicated traffic light for pedestrians	0.25	0.43	0.00	1.00
	Three-leg intersection	0.22	0.42	0.00	1.00
	Presence of raised median	0.48	0.50	0.00	1.00
	Total number of lanes	6.87	2.63	3.00	16.00

[(1)] buffer of 800 meters; [(2)] buffer of 400 meters; [(3)]buffer of 50 meters

Table 2. Posterior estimation summary – Simple Poisson gamma (negative binomial) model

	Posterior Mean	Std. Dev.	95% Bayesian interval 2.50%	97.50%
Intercept	-6.086	0.440	-6.997	-5.286
ln(pedestrian counts)	0.354	0.033	0.291	0.427
ln(left-turning AADT)	0.246	0.052	0.150	0.353
Presence of bus stop	0.754	0.148	0.456	1.047
Area of commercial land use	0.111	0.045	0.022	0.197
Dedicated traffic lights for pedestrians	-0.410	0.150	-0.708	-0.118
Proximity to school[(1)]	0.247	0.137	-0.016	0.478

[(1)] Non significant at a 5% level of significance.

Table 3. Posterior estimation summary – Heterogeneous endogenous model

	Posterior Mean	Std. Dev.	95% Bayesian interval 2.50%	97.50%
Count model (modelling injury frequency)				
Intercept	-5.451	0.462	-6.339	-4.495
ln(pedestrian counts)	0.322	0.038	0.251	0.401
ln(left-turning AADT)	0.284	0.052	0.173	0.380
Presence of bus stop	0.719	0.159	0.415	1.031
Area of commercial land use	0.117	0.045	0.029	0.205
Dedicated traffic lights for pedestrians	-0.393	0.152	-0.693	-0.099
Proximity to school (Mean)	0.536	0.237	0.056	0.958
Proximity to school (Std. Dev.)	0.190	0.135	0.014	0.486
Binary probit model (modelling presence of school)				
Intercept	-1.755	0.201	-2.150	-1.374
Employment	2.104	0.222	1.671	2.541
Presence of bus stop	0.392	0.123	0.153	0.635
Ratio of non-motorized to motorized	1.524	0.425	0.767	2.424
Land use mix	0.632	0.295	0.067	1.204

Table 4. Posterior estimation summary - Heterogeneous endogenous model with heterogeneity in mean and variance

			95% Bayesian interval	
	Posterior Mean	Std. Dev.	2.50%	97.50%
Count model (modelling injury frequency)				
Intercept	-6.062	0.603	-7.226	-4.873
ln(pedestrian counts)	0.417	0.058	0.306	0.539
ln(left-turning AADT)	0.295	0.056	0.191	0.402
Presence of bus stop	0.707	0.154	0.408	1.001
Area of commercial land use	0.118	0.045	0.028	0.206
Dedicated traffic lights for pedestrians	-0.577	0.181	-0.938	-0.228
Proximity to school (Mean)				
Intercept	1.493	0.565	0.421	2.623
ln(pedestrian counts)	-0.152	0.070	-0.294	-0.018
Proximity to school (variance)				
Intercept	0.090	0.068	0.008	0.210
Dedicated traffic lights for pedestrians	1.066	0.693	0.077	2.661
Binary probit model (modelling presence of school)				
Intercept	-1.722	0.203	-2.132	-1.335
Employment	2.080	0.226	1.649	2.529
Presence of bus stop	0.390	0.124	0.149	0.637
Ratio of non-motorized to motorized	1.526	0.416	0.766	2.379
Land use mix	0.589	0.285	0.033	1.148

Table 5. Posterior summary of average marginal effects.

	Posterior Mean	Std. Dev.
Heterogeneous endogenous model		
ln(Pedestrian volume)	0.267	0.038
ln(Left-turning AADT)	0.236	0.048
Presence of bus stop	0.362	0.071
Dedicated traffic lights for pedestrians	-0.064	0.023
Proximity to school	0.268	0.106
Area of commercial land use	0.097	0.039
Heterogeneous endogenous model with heterogeneity in mean and variance		
ln(Pedestrian volume)	0.364	0.057
ln(Left-turning AADT)	0.258	0.055
Presence of bus stop	0.375	0.071
Dedicated traffic lights for pedestrians	-0.124	0.049
Proximity to school	0.239	0.140
Area of commercial land use	0.103	0.040