Spatial Analysis of Child Pedestrian and Bicycle Crashes: Development of a Safety Performance Function for Areas Adjacent to Schools

by

Adam McArthur, E.I.T.
Department of Civil and Environmental Engineering
Wayne State University
5050 Anthony Wayne Drive
Detroit, MI 48202
E-mail: az9034@wayne.edu

Peter T. Savolainen, Ph.D., P.E.
Department of Civil and Environmental Engineering
Wayne State University
5050 Anthony Wayne Drive
Detroit, MI 48202
Telephone: 313/577-9950; Fax: 313/577-8126
E-mail: savolainen@wayne.edu
(corresponding author)

Timothy J. Gates, Ph.D., P.E., P.T.O.E.
Department of Civil and Environmental Engineering
Wayne State University
5050 Anthony Wayne Drive
Detroit, MI 48202
Telephone: 313/577-2086; Fax: 313/577-8126
E-mail: tjgates@wayne.edu

Words excluding Tables = 4,817
Figures: 1 x 250 = 250
Tables: 2 x 250 = 500
Total Words = 5,567

Date Submitted: November 15, 2013
ABSTRACT
Pedestrian and bicycle safety are particular concerns among school aged children as traffic crashes continue to be a leading cause of death for children ages 5 to 14. To address this concern, Safe Routes to School (SRTS) programs have been implemented in communities across the country, with one of the primary objectives of such programs being to provide for safe and convenient routes for children to walk or bike to school. Unfortunately, SRTS programs allow for the allocation of a very limited pool of funds for such projects. Consequently, it is imperative that programs are implemented at locations where they are likely to result in the greatest impact. The primary focus of this study was to develop a safety performance function (SPF) for use in prioritizing candidate schools for SRTS programs. Traffic crashes over a five-year analysis period were examined and linked to data that included school enrollment, socioeconomic and demographic data for each school district, and the functional class of the roadway on which each school was located. Schools on local roads were found to experience more crashes than those located on other, higher class road facilities. Crashes also varied with average family size, the number of parents per household, population density, and median family income. Interestingly, crashes were less frequent in school districts that exhibited greater ethnic diversity. The SPF developed as a part of this research can be used for prioritization of candidate schools, as well as to assess the efficacy of SRTS programs on a longitudinal basis, providing a valuable tool for school administrators and others involved in child pedestrian and bicycle safety.
INTRODUCTION
Each year, more than 4,000 pedestrians and nearly 700 bicyclists are fatally injured in traffic crashes throughout the United States [1,2]. While pedestrian and bicycle traffic comprise a very small proportion of all travel nationwide, these fatalities account for 15 percent of the national total. Pedestrian and bicycle safety are particular concerns among school aged children as pedestrian fatalities account for 19 to 23 percent of all traffic fatalities among children ages 0 to 15 while 10 percent of all bicyclist fatalities involved children in this youngest age group [1,2]. Child pedestrian and bicyclist activity tends to be highest during the periods before and after school, periods that also tend to coincide with high levels of traffic volume, increasing the potential risk for child pedestrian crashes [3].

Safety concerns present one of several factors that explain the relatively low rate of children who currently walk or bike to school [4,5]. Prior research has shown various demographic, attitudinal, psychosocial, and environmental factors to influence parents’ decisions as to whether to allow children to walk or bike to school. Among the most influential factors are land use and geographic factors, income, parental attitudes and schedules, and traffic/crime related fears [6,7]. Children with less independent mobility, fewer traveling companions, and those less likely to engage in conversations about safety have been found more likely to be driven to school [8]. Parental intervention and influence may increase rates of active-travel-to-school (ATS) for both pedestrian and bicyclists [9,10]. In support of increased ATS, a national study of state laws showed that legislation requiring minimum bussing distances, hazardous route exemptions, crossing guards, speed zones, and other traffic control measures all influenced ATS policies [11]. Within the scope of planning and engineering, the proximity of homes to a school, the level of traffic exposure in the surrounding neighborhood, and the connectivity of the roadway network around the school have all been shown to be significant determinants of children walking or bicycling to school [12-14].

To address some of the aforementioned issues, Safe Routes to School (SRTS) programs have been implemented in communities across the country, with one of the primary objectives of such programs being to provide for safe and convenient routes for children to walk or bike to school. The SRTS program funds a variety of projects, including infrastructure improvements (e.g., sidewalk and bike lane installation), educational programs, and on-site safety audits.

Unfortunately, SRTS programs allow for the allocation of a very limited pool of funds for such projects. Consequently, it is imperative that programs are implemented at locations where they are likely to result in the greatest impact. To this end, the National Center for Safe Routes to School presented a practical process for identifying and prioritizing schools for SRTS programs [15]. This methodology provided a balance between what is known about preventing child pedestrian injuries and what can realistically be accomplished without having to perform comprehensive field reviews of each site. Primary emphases for prioritization are centered on three major factors: crash history; public and school officials’ concerns; and current or potential pedestrian use. While this framework presents a promising starting point, its application is somewhat limited due to the relative infrequency of child pedestrian and bicycle crashes near schools and the inherent subjectivity involved in the more qualitative assessments.

The objective of this research was to develop a safety performance function (SPF) for K-8 schools that could be used as part of a data-driven approach to assist communities in identifying or prioritizing candidate schools for SRTS funding. Furthermore, this SPF can also be used to conduct longitudinal assessments of the effectiveness of SRTS programs in reducing child pedestrian and bicycle crashes, filling an important gap in the research literature [16].
LITERATURE REVIEW
The research literature includes a few studies that have examined crashes involving school aged pedestrian or bicyclists, as well as a more extensive volume of recent research that has examined pedestrian- and bicycle-involved crashes more broadly.

Studies of Pedestrian and Bicycle Crashes Involving Children
A 1997 study in England involved the development of a procedure to identify the number of crashes occurring during the commute to and from school [17]. Schools were ranked on the basis of the number of crashes occurring within a 250-meter radius from 8:00 AM to 9:00 AM and from 3:30 PM to 4:30 PM. Crash statistics were subsequently linked to self-reported travel data. The resultant crash rates were compared between two school locations, though no significant differences were identified.

Exploratory spatial analysis was used to map the distribution of crashes involving child pedestrians less than 10 years of age in Lille, France [18]. The results showed that crashes tended to be concentrated in a select number of areas rather than randomly localized across space. Furthermore, the proportion of crashes tended to decrease at distances further from schools.

A more recent study in Baltimore, Maryland examined pedestrian-involved crashes in the vicinity of public schools [19]. Separate models were estimated to examine crash frequency and severity. The results showed that both measures tended to decrease at locations where driveways or turning bays were present near the school entrance. Conversely, both frequency and severity increased when recreational facilities were present on the site. Each of these factors is likely to be capturing the effects of varying exposure, as are several of the other significant variables identified in the study, which included transit access, commercial access, and population density.

Spatial and temporal analyses of child pedestrian crashes in Santiago, Chile were conducted to identify critical areas that exhibited high crash concentrations and frequencies for periods of five or more years during the period from 2000 to 2008 [20]. Seven critical areas were identified and these areas tended to be located in districts with lower middle socioeconomic statuses. Crashes tended to be clustered by time of day and no trends were evident with respect to the children’s ages, which ranged from 5 to 18 years old.

General Spatial Analyses of Pedestrian and Bicycle-Involved Crashes
Beyond studies focused on schools and young children, there has been considerable recent research to assess the crash risks of pedestrians and bicyclists. A study in Hamilton, Canada mapped commuter cyclist collision risk per distance travelled [21]. The authors recommended this method be used as a basis for transportation policy as opposed to more traditional analyses focused on rates per capita.

A spatial epidemiologic approach was used to study the relative risk factors of bicycle and pedestrian crashes at the neighborhood level in Buffalo, New York [22]. Among the factors considered in the evaluation were roadway functional classes, as well as socio-economic and demographic variables. Results show neighborhood-level characteristics, such as ethnicity, education level, and land use type to significantly influence the frequency of pedestrian and bicycle crashes.

A study in four Florida counties examined trends in crash frequency with respect to trip productions and attractions within traffic analysis zones while controlling for the effects of road characteristics [23]. The results showed that pedestrians and bicyclists tended to be overrepresented in intersection crashes. Crashes involving pedestrians and bicyclists also tended...
to increase with trip productions and trip attractions on a zonal level while crashes were underrepresented on roads with a 35-mph speed limit.

A New York City study involved the development of a random parameter negative binomial model to examine factors influencing the frequency of pedestrian crashes at the census tract level [24]. Crashes tended to be higher in African-American and Hispanic neighborhoods, as well as in areas with lower levels of education. Areas with more schools or more concentrated commercial or industrial development were also prone to more crashes.

A study in Ciudad Juárez, Mexico used linear regression to examine the density of pedestrian crash data at the census tract level [25]. Pedestrian crashes were more likely in areas with: greater population and employment density; higher levels of commercial and retail land use; and larger proportions of persons age 65 and above.

Collectively, the extant research literature summarized here shows that a variety of analytical techniques can be utilized to assess pedestrian and bicycle crashes in a spatial context. Regardless of the approach, several common factors emerge as important, particularly information about the transportation infrastructure, as well as socioeconomic and demographic characteristics.

METHODOLOGY

Available safety resources are optimally invested at those locations that present the greatest opportunity to reduce the frequency of traffic crashes, injuries, and fatalities. Traditionally, such locations were identified on the basis of historical crash counts. However, given the random and rare nature of traffic crashes, such data may not necessarily lead to the identification of the locations with the highest potential for crashes in the future. This is particularly true for pedestrian- and bicycle-involved crashes, which tend to occur much less frequently on a site-specific basis in comparison to motor vehicle crashes.

In light of these facts, the Highway Safety Manual suggests that historical crash counts be supplemented with crash estimates provided by safety performance functions (SPFs). SPFs allow for the prediction of the expected number of crashes at a given location based upon site-specific information such as exposure (e.g., traffic volumes), roadway geometry, and other relevant factors that are correlated with crash counts. The primary focus of this study was to develop an SPF for use in the Michigan SRTS program.

Data Collection and Summary

In order to develop the safety performance functions (SPFs), details of those schools eligible for Safe Routes to School (SRTS) funding were obtained from the State of Michigan’s Center for Educational Performance and Information (CEPI), which is responsible for the collection and analysis of performance data for Michigan’s K-12 public schools and students [26]. Data were collected for a total of 2,404 schools that included any students in grades kindergarten through eighth grade. Each school was geocoded in a geographic information system (GIS) database for use in the subsequent spatial analysis.

The spatial analysis involved using the coordinates for each school to create a one-mile buffer. Data were then obtained for all crashes involving child pedestrians and bicycles ages 5 to 14 (kindergarten through eighth grade) within the buffer surrounding each of the schools. These data were obtained from the Michigan State Police (MSP) crash database, a condensed version of which can be accessed from the MSP Office of Highway Safety Planning’s Michigan Traffic Crash Facts (MTCF) Data Query Tool [27]. Given the limited number of crashes experienced on an annual basis at even the schools with the highest crash frequencies, five years of data were
obtained (from 2007 to 2011). A total of 7,781 crashes occurred at the eligible schools over the five-year analysis period. Figure 1 provides both a probability density function and cumulative distribution function, which illustrate the percentage of schools that experienced specific crash frequencies during the analysis period. The data appear to be well approximated by a Poisson or negative binomial distribution, with roughly 50 percent of the schools experiencing zero crashes during the analysis period.

![Probability Density and Cumulative Distribution Functions](image)

**FIGURE 1** Probability Density and Cumulative Distribution Functions of Child Pedestrian and Bicyclist Involved Crashes within One Mile of a School, 2007 to 2011

The school and crash data were then combined with demographic and socioeconomic information obtained from the United States Census Bureau [28]. These data were obtained at the school district level and matched with each of the K-8 schools included in the study. The data included as a part of the analysis were: average family size (number of persons in the household related to the householder); average number of parents in each household; total district population by age, race, and sex; and median income. Various combinations of these and other socioeconomic characteristics were examined as part of the modeling process.

Lastly, the National Functional Class (NFC) code [29] was obtained for the primary roadway on which each school was located based upon its official address. This information was obtained from the Michigan Department of Transportation (MDOT) roadway database [30]. Table 1 provides summary statistics for those factors that were found to be significant as a part of the analysis.
TABLE 1 Summary Statistics

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crashes from 2007 to 2011</td>
<td>1.31</td>
<td>2.17</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Census Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Family Size</td>
<td>3.05</td>
<td>0.18</td>
<td>2.34</td>
<td>3.98</td>
</tr>
<tr>
<td>Children Ages 5 to 14</td>
<td>10,543</td>
<td>23,349</td>
<td>25</td>
<td>102,261</td>
</tr>
<tr>
<td>Average Parents per Household</td>
<td>1.72</td>
<td>0.13</td>
<td>1.25</td>
<td>1.89</td>
</tr>
<tr>
<td>Median Family Income ($)</td>
<td>54,626</td>
<td>19,420</td>
<td>9,906</td>
<td>163,366</td>
</tr>
<tr>
<td>Population Density (1,000 per sq. mi.)</td>
<td>1.60</td>
<td>1.86</td>
<td>0.02</td>
<td>10.76</td>
</tr>
<tr>
<td>Proportion of Non-White Households</td>
<td>0.19</td>
<td>0.23</td>
<td>0.02</td>
<td>0.97</td>
</tr>
<tr>
<td>Roadway Functional Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Roads (proportion)</td>
<td>0.51</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Collectors (proportion)</td>
<td>0.23</td>
<td>0.42</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Arterials (proportion)</td>
<td>0.26</td>
<td>0.44</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>School Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students Enrolled</td>
<td>397</td>
<td>203</td>
<td>10</td>
<td>1,298</td>
</tr>
</tbody>
</table>

Statistical Methods

Once the analysis dataset was created, a safety performance function (SPF) was estimated in order to examine the relationships between the factors listed in Table 1 and the frequency of child pedestrian and bicycle crashes near each school during the five-year analysis period. Consistent with contemporary research [31] and the recommended methodology from the Highway Safety Manual [32], the SPF takes the form of a negative binomial regression model. The negative binomial specification provides the probability $P(n_i)$ of school $i$ experiencing a total of $n_i$ child pedestrian- or bicycle-involved crashes over the five-year analysis period:

$$P(n_i) = \frac{\Gamma(\theta+n_i)}{\Gamma(\theta)\Gamma(n_i)} u_i^\theta (1-u_i)^{ni} \quad (1)$$

where $u_i = \frac{\theta}{(\theta+\lambda_i)}$, $\theta = \frac{1}{\alpha}$, $\Gamma(\cdot)$ is a gamma function, and the mean number of crashes at a given location, $\lambda_i$, is a linear function of the covariates:

$$\lambda_i = EXP(\beta X_i + \epsilon_i) \quad (2)$$

where $X_i$ is a vector of data describing the traffic, socioeconomic, demographic, and enrollment characteristics of school $i$, $\beta$ is a vector of estimable parameters; and $exp(\epsilon_i)$ is a gamma-distributed error term with mean one and variance $\alpha$. The resulting mean variance relationship is given by

$$Var[n_i] = E(n_i)[1 + \alpha E(n_i)] \quad (3)$$
If $\alpha$ is significantly different from zero, the data are overdispersed or underdispersed. If $\alpha$ is equal to zero, then the negative binomial reduces to the Poisson distribution. Estimation of $\lambda_i$ can be conducted through standard maximum likelihood procedures [31].

Given the nature by which crashes occur, as well as the available data, it is reasonable to expect unobserved heterogeneity due to the presence of unobserved factors (e.g., pedestrian and vehicular volumes, detailed roadway geometry, etc.). To account for this heterogeneity, the constant term is allowed to vary as

$$\beta_i = \beta + \omega_i$$

(4)

where $\omega_i$ is a randomly distributed term. For the constant, this term is assumed to be normally distributed with mean zero and variance $\sigma^2$. The probabilities are now conditional based upon the probability distribution of $\omega_i$. A simulation-based maximum likelihood method is used to estimate the model parameters, further details of which can be found elsewhere [33,34]. In the research literature, this model formulation is also referred to as a “random effects” model.

RESULTS AND DISCUSSION
This random effects negative binomial model is presented in Table 2, which includes parameter estimates, standard errors, p-values, and marginal effects. Marginal effects represent the effect on the expected frequency, $\lambda_i$, that is associated with a one unit change in each independent variable. Marginal effects (ME) are calculated as follows:

$$ME_{\lambda_i} = \frac{\partial \lambda_i}{\partial x_{ik}} = \beta_k EXP(\beta X_i)$$

(5)

where $x_{ik}$ is the value of the $k$th independent variable for school $i$, $\beta_k$ is the estimated parameter for the $k$th independent variable, and $\lambda_i$ is the expected crash count for school $i$. Marginal effects are calculated for each observation and then averaged across the entire sample.

In examining the model results, they are generally found to be consistent with a priori expectations. First, the standard deviation of the constant term was found to be significantly different from zero, which is evidence of unobserved heterogeneity (and “random effects”) that is due to some important factors for which data were not available. Secondly, the overdispersion parameter was also significantly larger than zero, indicating that the negative binomial model outperformed the more restrictive Poisson model (a finding that was confirmed by comparing the log-likelihood values for the two models).

With respect to the variables of interest, crashes were found to increase with the population of children ages 5 to 14 (corresponding to grades kindergarten through eighth grade), as well as with the number of students enrolled in each school. In lieu of actual count data that indicate the number of children who walk or bike to school, these variables are likely capturing the effects of such exposure. This is consistent with prior work that has shown similar trends with respect to student enrollment and population [19].

Also consistent with prior research [8], the location of the school also impacted the expected number of crashes. Schools that were located on local roads (as per the NFC code) experienced more crashes than schools that were located on higher class collector or arterial roads. This finding could be reflective of less regulated travel behavior by both student commuters and motor vehicle drivers in the area adjacent to the school. Field observation conducted as a part of prior research in Michigan [35,36] showed erratic behavior by parents.
who were picking up or dropping off their children at school, as well as by the children themselves who would frequently dart in and out of traffic, particularly on lower volume local roads. This could also be due to underdevelopment of pedestrian and bicycle facilities along local roadways that may discourage safe crossing behavior [13]. Finally, this factor could also be capturing the effects of increased exposure as schools located on local roads may be located within a subdivision. As a result, the proportion of children walking or biking to and from school may be much higher than that of a school located on a major roadway, where it may be more likely that children would be picked up by parents or bused to and from school.

### TABLE 2 – Random Effects Negative Binomial Model Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>P-value</th>
<th>Marginal Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.933</td>
<td>0.654</td>
<td>0.153</td>
<td>N/A</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.635</td>
<td>0.021</td>
<td>&lt;0.001</td>
<td>N/A</td>
</tr>
<tr>
<td>log (Child Population Ages 5-14)</td>
<td>0.228</td>
<td>0.020</td>
<td>&lt;0.001</td>
<td>0.168</td>
</tr>
<tr>
<td>log (K-8 Enrollment)</td>
<td>0.066</td>
<td>0.014</td>
<td>&lt;0.001</td>
<td>0.049</td>
</tr>
<tr>
<td>School Located on Local Roadway</td>
<td>0.172</td>
<td>0.038</td>
<td>&lt;0.001</td>
<td>0.126</td>
</tr>
<tr>
<td>Average Family Size</td>
<td>0.326</td>
<td>0.127</td>
<td>0.010</td>
<td>0.239</td>
</tr>
<tr>
<td>Population Density (persons per sq. mi.)</td>
<td>0.292</td>
<td>0.013</td>
<td>&lt;0.001</td>
<td>0.215</td>
</tr>
<tr>
<td>Median Family Income ($1,000)</td>
<td>-0.017</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>-0.012</td>
</tr>
<tr>
<td>Average Parents per Household</td>
<td>-2.214</td>
<td>0.403</td>
<td>&lt;0.001</td>
<td>-1.630</td>
</tr>
<tr>
<td>Proportion of Non-White Households</td>
<td>-1.644</td>
<td>0.217</td>
<td>&lt;0.001</td>
<td>-1.210</td>
</tr>
<tr>
<td>$\alpha$ (overdispersion parameter)</td>
<td>20.183</td>
<td>7.365</td>
<td>0.006</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Turning to the socioeconomic and demographic factors, crashes were found to increase in school districts with larger average family sizes. Children in larger households may be more apt to walk or bike to school due to limitations in the number of available vehicles. Similar dynamics may explain why crashes decrease in areas with more two-parent households. Two-parent households are more easily able to facilitate automobile transportation for children as shown by previous studies [8,12]. A two-parent household would also make it easier for one of the parents to walk or bike to school with their children. This finding may also capture some effects of greater parental interaction and oversight within such households.

Pedestrian and bicycle crashes were also found to increase in school districts with greater population density, which is likely to reflect differences in travel patterns as denser, more urbanized areas are likely to have higher proportions of children walking or biking to school. Such areas are also likely to be subject to higher traffic volumes, further elevating the risks of crash involvement. Crash frequency was negatively correlated with median family income. Higher income has generally been associated with higher levels of education and higher income areas are also more likely to have greater availability of automobiles for use in transportation children to and from school.

Interestingly, more diverse school districts (as measured by the proportion of non-White families) were found to experience fewer crashes than less diverse districts. However, the reasons for this correlation are not easily explained and further research is warranted to better understand the underlying causes of this result.
CONCLUSIONS
This study provides an important contribution to the state-of-the-art in the development of a safety performance function (SPF) for use in SRTS programs. The SPF could be calibrated to other jurisdictions using the procedures outlined in the Highway Safety Manual [31]. The method can also be easily applied in other jurisdictions using data that are available from the United States Census Bureau and state transportation agencies. Such SPFs can be used for various functions, including two that are of particular relevance for SRTS:

• Prioritizing candidate locations for subsequent investment decisions – Locations that are in need of safety improvements can be identified by comparing observed and expected safety performance. Schools that have experienced significantly more pedestrian or bicycle crashes than expected may be indicative of underlying safety issues. Alternately, schools that have experienced fewer crashes than expected may reflect either safe campuses or “luck” due to the random and rare nature of crashes.

• Conducting longitudinal evaluations of the effectiveness of prior investments – Historically, evaluations of SRTS programs have focused on whether the number of children walking and bicycling to school has increased [16]. Safety performance functions will allow for a quantitative assessment of how well existing programs meet the complementary objective of enhancing the safety for children who walk and bike to school.

As a result of this research, decision-makers will be able to more efficiently prioritize schools for safety improvements or other SRTS programs. In addition to providing this tool, the study results support prior research, which has associated a variety of socioeconomic, demographic, and other school-specific factors with child pedestrian and bicycle crashes [12-14]. Specific findings include:

• Crashes increased with the number of children enrolled at a given school and with the population of children ages 5 to 14 in the accompanying school district;
• Schools located on local roadways experience higher crash frequencies than schools located on other, higher class facilities;
• Crashes also varied with average family size, the number of parents per household, population density, and median family income. Interestingly, crashes were less frequent in school districts that exhibited greater ethnic diversity.

This research can be expanded by conducting more in-depth investigations that consider additional factors, such as pedestrian and motor vehicle volumes, travel speeds, and infrastructure characteristics.

DISCLAIMER
Funding for this research was provided by the Michigan Fitness Foundation through the Michigan Department of Transportation Safe Routes to School Program. The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Michigan Fitness Foundation, the Michigan State Transportation Commission, the Michigan Department of Transportation, or the Federal Highway Administration.
REFERENCES

