A SYSTEMIC SAFETY ANALYSIS OF PEDESTRIAN CRASHES: LESSONS LEARNED

Yiyi Wang, Ph.D.
(Corresponding Author)
Assistant Professor
Cobleigh Hall 214
Montana State University
Bozeman, MT 59717
Phone: (406) 994-6051
Fax: (406) 994-6105
yiyi.wang@ce.montana.edu

Shivam Sharda
Graduate Research Assistant
Montana State University
Cobleigh Hall 406
Montana State University
Bozeman, MT 59717
Phone: (406) 539-9267
Fax: (406) 994-6105
shivam.sharda8@gmail.com

Haizhong Wang, Ph.D.
Assistant Professor
101 Kearney Hall
Oregon State University
Corvallis, OR 97331
Oregon State University
Phone: 541-737-8538
Fax: 541-737-3052
Haizhong.Wang@oregonstate.edu

Word Count: 4,490 + 7 figure/tables = 6,240
Submission Date: July 31, 2015
Submitted to the 95th Transportation Research Board Annual Meeting, Washington D.C., 2016
ABSTRACT

The objectives of this study are twofold. First, this study synthesized relevant research and limitations of systemic (pedestrian) safety analysis. Second, the study explored one of the limitations identified relative to ranking roadway facilities based on primary risk factors. An intersection-pedestrian-crash data set from Austin, Texas was used. Comparisons among candidate ranking methods suggest that results were very sensitive to the weights used, and the grouping-based ranking method combining arbitrary weights produced relatively stable results. However, the finding is inconclusive, and more research would be needed to investigate alternative methods, especially the ones suitable for making decisions under multiple criteria (risk factors).

Keyword: pedestrian safety; systemic safety selection; negative binomial; grouping; elasticity; weights
1. INTRODUCTION

Walking has been shown to provide health benefits and, when practiced collectively, can moderate congestion and ease dependence on fossil fuels. However, the benefits of walking often come with a vulnerability to road injury or death. On average, 5,000 pedestrians are killed in roadway crashes every year in the United States, representing 14 percent of all traffic fatalities (1). Decades of research have produced a solid body of knowledge and guidelines to improve motor vehicle safety. By contrast, much less information is available for pedestrian safety. The findings derived from motor vehicle safety studies relate little to pedestrian safety because (a) vehicles and pedestrians operate in strikingly different ways (e.g., travel speed), and (b) traditional engineering measures (e.g., lane width and rumble strips) are concerned with safety and efficiency of vehicular movements, offering little relief or even adverse influence on safety of non-motorized modes.

A systemic selection approach offers a new way to analyze and improve pedestrian safety. This approach focuses on identifying risk factors and countermeasures that can benefit a larger number of locations. A systemic perspective can be especially useful in the pedestrian safety domain because pedestrian crashes are rare, and their spatial patterns are more sporadic than those of vehicular crashes (2). Solely relying on crash counts without accounting for the underlying risk factors (e.g., pedestrian activities, provision of bike/walk facilities, traffic, or built environment) provides limited and even biased information about where safety measures are needed and what measures would best improve pedestrian safety.

The objectives of this study are twofold. First, this study synthesized relevant research on systemic (pedestrian) safety analysis and identified the limitations. Second, the study explored one of the limitations relative to ranking roadway facilities (i.e., intersections) and discussed lessons learned based on the results using a data set of intersection pedestrian crashes. The rest of the paper is structured into five sections. Literature Review assesses the strengths and limitations associated with the systemic selection approach. The Data section summarizes the intersection pedestrian crash data set from Austin, Texas. The Methodology section presents the statistical models used in this study and precedes Results and Analysis. The paper ends with lessons learned and suggestions for future research in Conclusions.

2. LITERATURE REVIEW

The systemic selection approach consists of four steps: identifying risk factors and formulating a set of criteria for the problem of interest (e.g., pedestrian crashes); identifying primary risk factors and ranking facilities based on the risk factors and associated weights; recommending cost-effective measures; and prioritizing locations and countermeasures for implementation. This approach has been applied to study traffic crashes along county roads in Minnesota (3), roadway departure crashes at the bridges in New Jersey (4), and most recently rural unpaved roads in Iowa (5). A study (6) demonstrated an application of the systemic approach to analyze pedestrian safety at signalized intersections in the Twin Cities metro area. The study ranked the intersections using a simple weighted average of scores based on a range of risk factors including speed limit, traffic volume, and an indicator for multi-lane undivided roadways.
Summarized in the following paragraphs are elements crucial to developing a systemic safety approach for studying pedestrian safety, namely exposure term, design parameters, and the ranking/weighting methods used to prioritize locations.

Key to anticipating pedestrian crash risk is pedestrian volume, which serves as the exposure term. A facility with zero pedestrian crashes over a given period does not equal safety for pedestrians; it could be a mere artifact of low pedestrian volume resulting from a poor or dangerous walking environment. Like traffic volume, pedestrian volume tends to vary over space and time. Unlike traffic volume, pedestrian volumes typically are not monitored by transportation agencies. Exceptions include the Puget Sound Regional Council (PSRC) and the Delaware Valley Regional Planning Commission (DVRPC), both of which have extensive count programs. For the regions/areas without reliable pedestrian count data, it would be necessary to impute pedestrian exposures. For that purpose, two fundamental types of models are available: direct-demand models and choice-based models (7). Direct-demand models produce aggregated counts at the area level using land use and network attributes. Choice-based models observe a top-down approach similar to four-step travel demand models and produce link-level pedestrian volumes (see, e.g., MoPeD [8], PedContext [9], and the Portland Pedestrian Model [10]). While the direct-demand models offer quick proxies for pedestrian exposure, the choice-based models are able to reflect the mechanisms of pedestrian behaviors (e.g., mode and route choices).

Contextual factors have been shown to influence pedestrian safety through leveraging vehicle speed and traffic conflicts (11, 12, and 13). While there are no formal rules about what should be counted as contextual factors, they generally portray the built environment in a manner that goes beyond traffic conditions and roadway attributes. Examples include building configurations, street network patterns, and land use details. Google Street Views can facilitate data collection efforts and reduce costs. For instance, the U.S. Road Assessment Program (usRAP) developed an efficient procedure that can collect data on every three-mile section of roadway that contains approximately 40 to 50 variables about the roadway, roadside, and intersections using Google’s Street Views function (14).

At the heart of the systemic selection approach lies the ability to reveal the underlying crash risk and prioritize locations based on a set of criteria. Grouping has been used to formulate the criteria through detecting the ranges (intervals) of the risk factors that carry a disproportionate number of crashes (6 and 15). While quick and simple, this method can overlook the correlations among risk factors (e.g., higher speed often coincides with wider shoulder). Statistical models, on the other hand, allow for correlations among the risk factors, and uncertainty via random-effect error terms (16, 17, 18, and 19). For example, a study (20) used a crash prediction model and found that more pedestrian crashes occurred in areas with educational facilities and a higher percentage of commercial land use. Researchers (17) found that greater connectivity and transit access significantly increased injurious pedestrian crashes via an ordered probit model for severity of injury. Another study (21) concluded that pedestrians at facilities with no sidewalk present were twice as likely to be hit by a car as pedestrians at a site with sidewalks after controlling for crosswalk, on street parking, shoulder width, traffic control, vehicular speed, and traffic volume.

The ranking among candidate intersections or segments is also influenced by the weights assigned to the risk factors. The weights represent various importance levels of the factors (e.g., the safety impact of number of lanes outweighs that of the wait length of pedestrian signal) and
convert disparate metrics/units into a single uniform score. Three weighting methods are
commonly used to rank locations (15): total number of factors meeting the hazard criteria (6); a
product of that number and the actual crash frequency (15); and the weighted risk scores (see,
e.g., [9]). The FHWA tool (6) recommends weights from 0.5 to 1, depending on the confidence
level and crash history, and suggests that these numbers be used as a guide not a standard. A
study (15) investigated the sensitivity of ranking results by shifting the weights from 1 to 2 based
on certain rules. The authors concluded that the systemic selection process tested in their study
tended to yield high sensitivity to changes in the weights.

Having recognized the limitations of the systemic safety analysis, the rest of the paper
investigated the limitation relative to ranking methods. Using an empirical data set of
intersection pedestrian crashes from Austin, Texas, the research compared various ranking
methods when being combined with two weighting schemes and tested the impacts on ranking
results.

3. DATA

The pedestrian crash data set was collected from multiple sources, with summary statistics
shown in Table 1. Land use variables were calculated based on parcel maps available from the
City of Austin’s geographic file clearinghouse. Dissimilarity index (DI) measures the
mixing/interspersing of parcels with different land use (22). It was obtained by dissolving land
parcels into 100-meter by 100-meter grid cells and counting the number of dissimilar cells
among the eight neighboring cells for each hectare cell. Land use entropy measures the balance
of land use, expressed by $\sum_{j=1}^{4} p_j \frac{\ln(p_j)}{\ln(4)}$, where, $P_j$ = proportion of land use category $j$ within a
half mile radius of the hectare grid cell; $j=4$, or the total number of land use categories
(residential, commercial, office, and industrial). Entropy was calculated for a half-mile radius
around each hectare grid cell. Speed limit is available from the City of Austin’s road network
shapefile, and traffic volumes (average daily traffic [ADT]) come from the Texas Department of
Transportation (TxDOT) traffic counts archive in 2011.

To capture the intersection characteristics, researchers used Google Street Views to canvass 36
corridors (excluding access-controlled roadways and interchanges) in or near the center of the
city where the majority of pedestrian crashes occurred (shown in Figure 1), resulting in an array
of design parameters for a total of 1,150 intersections. The design parameters include sidewalk
provisions, bike lane presence, one-way street, median (painted, raised, or missing), crosswalk
(painted, raised, or missing), traffic control (signalized, stop-controlled, or none), number of
lanes for each approach, number of approaches, and whether light poles were present. These
parameters were found to affect pedestrian safety at intersections in significant ways (18, 23, 24,
and 25).

Crash records from 2004 to 2009 are available from the Crash Records Inventory System (CRIS)
of TxDOT. Four types of pedestrian crashes are recorded: intersection, intersection-related, non-
intersection (mid-block), and drive-way crashes. The pedestrian crashes that were coded as
intersection or intersection-related were retrieved from CRIS and joined to the respective
intersection locations. The intersection influence areas were defined based on a buffer with a
radius that was 1/30 of the average block size (the road segment length between neighboring
intersections). The buffer size was selected to avoid overlapping influence areas. For those
crashes that lay outside the buffer areas, new influence areas were computed based on the
stopping sight distance with speed limit being the proxy for travel speed for each intersection
approach (26). The new influence areas were then used to ascribe intersection crashes to their
respective intersections.

FIGURE 1 Visited Intersections overlaid onto the City Boundary of Austin, Texas

TABLE 1 (a) Summary Statistics of Signalized Intersections (No. of Obs. = 399
intersections)

<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Pedestrian Crash Counts</td>
<td>0</td>
<td>5</td>
<td>0.697</td>
<td>0.8241</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissimilarity Index</td>
<td>0</td>
<td>1</td>
<td>0.629</td>
<td>0.2705</td>
<td>0.625</td>
<td>1</td>
</tr>
<tr>
<td>Land Use Entropy</td>
<td>0</td>
<td>0.996</td>
<td>0.658</td>
<td>0.199</td>
<td>0.722</td>
<td>0.722</td>
</tr>
<tr>
<td>No. of Bus Stops within 0.2 mi.</td>
<td>0</td>
<td>25</td>
<td>6.979</td>
<td>5.615</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Sidewalk Presence</td>
<td>0</td>
<td>1</td>
<td>0.983</td>
<td>0.131</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lighting Presence</td>
<td>0</td>
<td>1</td>
<td>0.979</td>
<td>0.140</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pedestrian Crosswalk</td>
<td>0</td>
<td>1</td>
<td>0.972</td>
<td>0.164</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Percent of One-way Street</td>
<td>0</td>
<td>1</td>
<td>0.174</td>
<td>0.322</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bike Lane Presence</td>
<td>0</td>
<td>1</td>
<td>0.308</td>
<td>0.462</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
4. METHODOLOGY

As the application of systemic safety tools on pedestrian safety is still in its infancy, this study generally follows the steps of the current systemic planning process while modifying the process to better meet the challenges associated with pedestrian crash analysis. Respective of those challenges identified earlier in the paper, the methodology section is organized into two parts: (a) estimating pedestrian volumes or the exposure term for predicting crash occurrence; (b) calibrating the pedestrian crash models to assess the effects of different risk factors.

### TABLE 1 (b) Summary Statistic of Stop-Controlled Intersections (No. of Obs. = 751 intersections)

<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Pedestrian Crash Counts</td>
<td>0</td>
<td>3</td>
<td>0.133</td>
<td>0.377</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dissimilarity Index</td>
<td>0</td>
<td>1</td>
<td>0.648</td>
<td>0.258</td>
<td>0.625</td>
<td>0.875</td>
</tr>
<tr>
<td>Land Use Entropy</td>
<td>0</td>
<td>0.996</td>
<td>0.631</td>
<td>0.216</td>
<td>0.658</td>
<td>0.629</td>
</tr>
<tr>
<td>No. of Bus Stops within 0.2 mi.</td>
<td>0</td>
<td>23</td>
<td>4.326</td>
<td>3.612</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Sidewalk Presence</td>
<td>0</td>
<td>1</td>
<td>0.956</td>
<td>0.205</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lighting Presence</td>
<td>0</td>
<td>1</td>
<td>0.671</td>
<td>0.470</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pedestrian Crosswalk</td>
<td>0</td>
<td>1</td>
<td>0.299</td>
<td>0.458</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Percent of One-way Street</td>
<td>0</td>
<td>1</td>
<td>0.037</td>
<td>0.161</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bike Lane Presence</td>
<td>0</td>
<td>1</td>
<td>0.228</td>
<td>0.419</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No. of Approaches</td>
<td>3</td>
<td>4</td>
<td>3.228</td>
<td>0.419</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Total No. of Lanes</td>
<td>6</td>
<td>25</td>
<td>10.74</td>
<td>2.876</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Percent of Painted/Raised median</td>
<td>0</td>
<td>1</td>
<td>0.649</td>
<td>0.252</td>
<td>0.667</td>
<td>0.667</td>
</tr>
<tr>
<td>Average speed limit</td>
<td>25</td>
<td>48.33</td>
<td>38.35</td>
<td>4.468</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>Truck composition (percent single truck + percent combo truck)</td>
<td>0.05</td>
<td>8.2</td>
<td>2.396</td>
<td>1.259</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>Avg. of Approach ADTs</td>
<td>0</td>
<td>55,050</td>
<td>16,439</td>
<td>11,956</td>
<td>16,350</td>
<td>0</td>
</tr>
<tr>
<td>Pedestrian Walk Miles Travelled</td>
<td>0</td>
<td>17,859</td>
<td>1,003</td>
<td>2,704</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
4.1 Pedestrian Volume Estimation

For this research study, pedestrian volume is proxied by walk miles traveled (WMT), a variable that has been computed by a study (16) for the same study region over a similar study duration. They estimated WMT at the traffic analysis zone (TAZ) level based on 569 walk trips from the 2006 Austin Travel Survey (ATS). A weighted log-linear model was used with the following explanatory variables: population; employment counts by type (retail, service, and basic); zone size (land area in square miles); sidewalk lengths; and coded lane miles by functional class. Weights were assigned to each zone based on the sample representation relative to the population to account for heteroscedasticity due to varying sampling rates across TAZs. The final zonal WMTs were imputed by multiplying the average WMT per ATS respondent by the zonal population.

4.2. Pedestrian Crash Models

Negative Binomial (NB) models, along with zero-inflated-type models, were used to establish statistical relationships between pedestrian crash counts and contributing factors. The NB models outperform the other models based on prediction accuracy, Akaike information criterion (AIC), and Bayesian information criterion (BIC). The NB models also capture over-dispersion where the conditional variance of a variable exceeds the conditional mean (27). Over-dispersion results from heterogeneity or uncontrolled variation specific for each intersection due to unobserved influence factors; the NB model can account for heterogeneity thanks to a mix of the Poisson kernel and the gamma-distributed error term. The NB model is written as:

\[ Y_i \sim \text{Poisson}(\lambda_i) \]

\[ \lambda_i = (\text{Pedestrian volume} \times \text{Traffic volume})^\alpha \times \exp(\beta_0 + X_{i1} \times \beta_1 + X_{i2} \times \beta_2 + \cdots + \epsilon_i) \]  

Where,

- Yi = Observed crash counts at intersection i;
- \( \lambda_i \) = Expected number of pedestrian crashes;
- \( \beta_0 \) = A constant term representing the base-line crash risk;
- \( X_{ij} \) = the \( j^{th} \) covariate of intersection \( i \);
- \( \beta_i \) = Coefficient describing the direction and importance of \( X_{ij} \) in explaining the crash risk;
- \( \epsilon_i \) = Error term; and \( \exp(\epsilon_i) \) follows a gamma distribution, Gamma \( (K, \theta) \).

The parameter \( K \) is the shape parameter, and \( \theta \) is the scale parameter measuring the spread of the distribution. The shape parameter is set to equal the inverse of \( \theta \) such that the mean of \( \exp(\epsilon_i) \) is 1 to avoid shifting the mean crash frequency. The Gamma distribution along with the parameter estimates were used to simulate 1,000 error terms for calculating the mean crash risk.

5. RESULTS AND ANALYSIS

This section focuses on comparing the effects of the ranking and weighting methods on the selection results of dangerous locations and suggest methods that perform well. The following paragraphs present the results in regard to prominent risk factors, and a sensitivity analysis to understand how ranking results shift as different weights are used.
5.1 Identification of Risk Factors

Two methods are used to identify risk factors: a model-based approach (which considers cross-correlations among covariates and random effects and provides statistically defensible results) and an empirical approach (a quicker and easier way to formulate the risk criteria but that can overlook confounding effects). For the model-based approach, the NB models (shown in Tables 2 and 3) were selected among a number of candidate models based on goodness-of-fit (e.g., AIC and root-mean-squared errors [RMSE]). The inverse of theta measures the departure between the conditional mean and the conditional variance. Both models signal the presence of over-dispersion, with more severe dispersion found and controlled for the signalized intersections.

### TABLE 2 NB Model Results for Stop-Controlled Intersections

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Estimated Coefficient</th>
<th>t-statistic</th>
<th>Marginal Effect</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept Term</td>
<td>-5.085</td>
<td>-5.622</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>ln(PV) + ln(Avg. ADT)</td>
<td>0.168</td>
<td>3.138</td>
<td>0.021</td>
<td>0.118</td>
</tr>
<tr>
<td>No. of Approaches</td>
<td>0.404</td>
<td>1.706</td>
<td>0.056</td>
<td>0.367</td>
</tr>
<tr>
<td>Percent One-way Street</td>
<td>0.923</td>
<td>2.050</td>
<td>0.113</td>
<td>0.004</td>
</tr>
<tr>
<td>No. of Bus Stops within 0.2 miles</td>
<td>0.054</td>
<td>1.825</td>
<td>0.007</td>
<td>0.051</td>
</tr>
<tr>
<td>Dissimilarity Index</td>
<td>0.683</td>
<td>1.651</td>
<td>0.084</td>
<td>0.102</td>
</tr>
</tbody>
</table>

AIC: 640.72; 2 x log-likelihood: -622.721; Number of observations: 751

Null deviance: 390.56 on 752 degrees of freedom
Residual deviance: 358.37 on 745 degrees of freedom
Theta: 1.135
Std. Err.: 0.629
RMSE: 1.284

### TABLE 3 NB Model Results for Signalized Intersections

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Estimated Coefficient</th>
<th>t-statistic</th>
<th>Marginal Effect</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept Term</td>
<td>-3.369</td>
<td>2.722</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Total No. of Lanes</td>
<td>0.061</td>
<td>3.355</td>
<td>0.048</td>
<td>0.252</td>
</tr>
<tr>
<td>No. of Bus Stops within 0.2 miles</td>
<td>0.075</td>
<td>5.513</td>
<td>0.059</td>
<td>0.121</td>
</tr>
<tr>
<td>Percent One-way Street</td>
<td>-0.648</td>
<td>-2.324</td>
<td>-0.511</td>
<td>-0.029</td>
</tr>
<tr>
<td>ln(PV) + ln(Avg. ADT)</td>
<td>0.096</td>
<td>2.425</td>
<td>0.076</td>
<td>0.137</td>
</tr>
<tr>
<td>Percent Painted/Raised Median</td>
<td>-0.384</td>
<td>-1.956</td>
<td>-0.303</td>
<td>-0.094</td>
</tr>
</tbody>
</table>

AIC: 987.61; 2 x log-likelihood: -967.606; Number of observations: 399

Null deviance: 481.52 on 396 degrees of freedom
Residual deviance: 416.07 on 388 degrees of freedom
Theta: 4.90
Std. Err.: 2.42
RMSE: 3.39
The exposure term is represented by a product of pedestrian volume and traffic volume. Higher traffic volume (or exposure) means higher crash frequency (28, 29, and 30). As confirmed by the NB models, a higher exposure corresponds to an increase in the expected crash frequency, but their relationship appears to be non-linear as reflected by a power/exponent estimated to be 0.096 (t-stats = 2.425). This indicates that the mean crash frequency does not increase in proportion to the increase in exposure, but grows at a much slower rate as exposure increases (or safety in numbers).

For stop-controlled intersections, crash risk increases with land use mix (DI) after controlling for exposure. The DI variable is estimated to be significant under a 90% confidence interval. Land use mix represents the connectivity among different types of land use and the ease of reaching nearby services and activities. Intuitively, a higher level of mix means more activities, traffic, and perhaps more non-motorized trips, which all increase the complexity of travel for all road users. The number of approaches at intersections is positively related to pedestrian crash risk. This is expected, as traffic conflict points grow exponentially with the number of approaches and the associated traffic movements. Adding an intersection approach also means higher traffic volume, a proven hazard for pedestrians (see, e.g., 18 and 31).

The number of bus stops within 0.2 miles of intersections is found to contribute to a higher crash risk, irrespective of intersection types. Bus stops are located either on the far side (a short distance past the intersection) or the near side (a short distance before the intersection). Either location can present a safety hazard for pedestrians crossing the street to catch a bus or exit the bus to access adjacent land use. Bus stops are often located near land use that generates pedestrian activities such as daycare, hospitals, and shopping centers. This can be particularly dangerous for young children, seniors, and physically challenged pedestrians.

The percentage of one-way streets at an intersection was computed as the number of approaches that are one-way divided by the total number of approaches. One-way streets can increase crash risk when drivers traveling along the one-way streets overlook pedestrians trying to cross from the opposing direction (25 and 32). The results suggest that the percentage of one way streets is positively correlated with pedestrian crash risk at stop-controlled intersections, but negative correlated with pedestrian crash risk at signalized intersections. In the study area, one-way streets generally coincide with intersections where traffic along one street should yield to the crossing street. One-way streets present a safety hazard at stop-controlled intersections possibly because pedestrians often need to use their own judgment to cross the street without protection from any type of signals. Drivers traveling along the major approach may not be able to stop the vehicle in time due to high speed. In contrast, a higher share of one-way streets can bring safety benefits to signalized intersections thanks to a shorter crossing distance while traffic signals provide safe crossing time for pedestrians.

Providing a median (raised or painted) is estimated to improve pedestrian safety at signalized intersections in statistically and practically significant ways. Medians may substitute for refuge islands for pedestrians who are unable to cross the street during a signal phase. The study is not able to separate the effects between raised medians and painted medians as the painted median is the dominant type for the study area. The presence of medians appears to be insignificant for stop-controlled intersections, possibly because 92 percent of stop-controlled intersections have a painted or raised median. Monotonic variables like this can prevent the model from detecting significance of the variables.
The total number of lanes serves as a proxy for the total crossing distance at intersections. As expected, longer distance correlates with higher crash risk at signalized intersections. But its effect on crash risk is insignificant at stop-controlled intersections. A plausible reason relates to the smaller range of the variable (total number of lanes), which may diminish the model’s ability to discern statistical significance. Although an increase in traffic-lane width might reduce vehicle collisions (33, 34 and 35), the benefits for motorists often come at the expense of increasing pedestrian crash risk.

### An Empirical Perspective

This method generated bar plots for the same set of risk factors that appeared in the NB models. The plots illustrate the ranges of risk factors where a disproportionate number of pedestrian crashes occurred. Figures 2 and 3 present the bar plots for signalized and stop-controlled intersections. The intervals that bear a disproportionate number of crashes are reported in Table 4.

#### TABLE 4 Ranges of the Risk Factors for Stop-Controlled and Signalized Intersections

<table>
<thead>
<tr>
<th>Risk Factors</th>
<th>Stop-Controlled Intersections</th>
<th>Signalized Intersections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissimilarity Index</td>
<td>0.5 - 1.0</td>
<td>--</td>
</tr>
<tr>
<td>No. of Bus Stops within 0.2 miles</td>
<td>4 – 11</td>
<td>6 – 11</td>
</tr>
<tr>
<td>Percent of One-way Street</td>
<td>0.75 – 1</td>
<td>0.25 - 0.50</td>
</tr>
<tr>
<td>Ln(PV) + Ln(Average ADT)</td>
<td>6 – 9</td>
<td>4.5 - 7.5</td>
</tr>
<tr>
<td>No. of Approaches</td>
<td>&gt; 3</td>
<td>--</td>
</tr>
<tr>
<td>Total No. of Lanes</td>
<td>--</td>
<td>18 – 26</td>
</tr>
<tr>
<td>Percent of Raised/Painted Median</td>
<td>--</td>
<td>0 - 0.25</td>
</tr>
</tbody>
</table>

These intervals were used with the grouping-based method to assess the risk levels of intersections. Details are provided in Section 5.2.
FIGURE 2 (a) Distribution of crashes against dissimilarity index (Stop-Controlled Intersections)

FIGURE 2 (b) Distribution of crashes against number of bus stops within 0.2 miles of intersections (Stop-Controlled Intersections)
FIGURE 2 (c) Distribution of crashes against percentage of one-way streets (Stop-Controlled Intersections)

FIGURE 2 (d) Distribution of crashes against the logarithmic sum of traffic and pedestrian exposures (Stop-Controlled Intersections)
FIGURE 2 (e) Distribution of crashes against number of approaches (Stop-Controlled Intersections)

FIGURE 3 (a) Distribution of crashes against the number of bus stops within 0.2 miles of signalized intersections

FIGURE 3 (b) Distribution of crashes against the percentage of one-way streets (signalized intersections)
FIGURE 3 (c) Distribution of crashes against the number of lanes (signalized intersections)

FIGURE 3 (d) Distribution of crashes against the percentage of raised/painted medians (signalized intersections)
5.2 Ranking and Identifying Dangerous Locations

An important question relative to location ranking is how to weight each risk factor to gauge the overall hazard level for an intersection. Existing methods can skew the underlying crash risk for an intersection by treating each of its factors as equal or assigning arbitrary weights when some factors wield stronger safety impacts than the others.

A remedy is to use weights that vary based on the importance of the risk factors. In addition to the grouping-based method, an alternative method was developed harnessing the results of statistical models. The following sections summarize the techniques and results for the two ranking methods: a grouping-based approach and a weighted-average approach.

5.2.1 A Grouping-Based Approach

The values in Table 4 were used to calculate the hazard scores for each intersection. For each risk factor that meets the criteria, a point is added to that intersection. Hence, the final score of an intersection means the number of risk factors that meet the hazard criteria, similar to the procedure adopted by the researchers (15). If two or more intersections were tied, an additional risk factor was used to differentiate their hazard levels. The logarithmic sum of pedestrian and traffic volume was selected to be that factor because it has been shown to affect crash risk in important ways (18, 24 and 36). In other words, the tied intersections are ranked in decreasing orders of their exposure term in order to produce a distinct sequence of ranks with no ties.

5.2.2 A Weighted-Average Approach

As shown in Equation 2, the hazard score of an intersection is expressed as a weighted average of the mean crash counts estimated by the NB model and the crash counts observed during the same time period,

\[
\text{Hazard Score} = 0.5 \times \text{observed crash counts} + 0.5 \times \text{estimated crash counts}
\]  

(2)

5.2.3 Comparison of the Two Approaches
Having defined the weighting methods, the question becomes whether they produce consistent rankings and which method is better.

The first question is answered by evaluating the correlations across the ranking results produced by the two methods. The Kendall’s rank correlation coefficient is a non-parametric measure to test the statistical association between two vectors of quantities. It is preferred over other alternatives (such as the Spearman’s rho correlation) because it is built on statistical distribution with nicer properties and offers straightforward interpretation of the association between two sets of rankings (37). It is mathematically expressed as (15 and 37):

\[
T_b = \frac{n_e - n_d}{\sqrt{n_0 - T} \times \sqrt{n_0 - U}}
\]

Where, \(n_o = n \times (n-1)/2\); \(n_e = \) number of concordant pairs; \(n_d = \) number of discordant pairs; \(T=\) number of tied values in the first vector; and \(U = \) number of tied values in the second vector. The correlations are summarized in Table 5.

Results suggest that the intersections ranked higher (more dangerous) by the grouping-based approach were also ranked higher by the weighted-average approach, although the precise position of an intersection tends to shift between the two methods. The Kendall’s coefficient for the signalized intersection is 0.43 and 0.61 for stop-controlled intersections.

5.2.4 Sensitivity Analysis

A sensitivity analysis was conducted to determine which method produced relatively stable or consistent rankings irrespective of the weights being used. The Kendall’s tau coefficient was used to detect the changes in the new ranking results after the weights were systematically altered. The coefficients, shown in Table 5, measure the correlation between the baseline rankings and the new rankings for each method. A larger positive coefficient indicates that intersections retain similar positions under the new weights; a negative coefficient implies that locations ranked on the top tend to be ranked at the bottom under the new weights.

The baseline rankings from both methods (grouping-based and weighted average) were obtained by assuming uniform weights (1’s), as indicated in Sections 5.2.1 and 5.2.2. Elasticities were used to weight the risk factors; with a comparative look at arbitrary weights (the new weights double the initial weights of those risk factors that meet the criteria, while everything else stays the same). Results of the sensitivity analysis are summarized in Table 5. The cell values are the Kendall’s tau coefficient between the new ranking (using new weights) and the baseline ranking. Numbers in the parentheses report the number of intersections that stay in the top 30 list, the number of intersections that retain the exact position, and the number of intersections ranked among the top 30 but with no reported crashes.

TABLE 5 Comparisons among Different Ranking Methods and Weight Choices

<table>
<thead>
<tr>
<th></th>
<th>Ranking with Arbitrary Weights</th>
<th>Ranking with Elasticity-based Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grouping-based ranking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stop-controlled intersections</td>
<td>0.87 (26; 22; 16)</td>
<td>0.73 (24; 12; 19)</td>
</tr>
<tr>
<td>Signalized intersections</td>
<td>0.67 (16; 0; 4)</td>
<td>0.76 (25; 1; 5)</td>
</tr>
</tbody>
</table>
Shown in Table 5, the ranking results are highly sensitive to weight choices, as echoed in (15). For the signalized intersections, the weighted-average method when combining arbitrary weights offers the highest stability (i.e., largest Kendall’s tau coefficient). This is also true for the grouping-based method when combining arbitrary weights for the stop-controlled intersections. It is unclear why the two types of intersections responded differently to the ranking methods. A plausible explanation is that the discrepancy can result from the different sample sizes. The number of signalized intersections (No. of Obs. = 399) in the sample is approximately half of the number of stop-controlled intersections (No. of Obs. = 751). The results seem to suggest that grouping-based method performs well when the sample size is fairly big; but a weighted-average method can surpass the grouping-based method when sample size is small, possibly thanks to its superior ability to control for confounding effects than the grouping-based approach.

CONCLUSIONS

This research study identified the limitations associated with the systemic selection approach and investigated alternative ranking methods using a data set of intersection pedestrian crashes from Austin, Texas. Echoed in FHWA (6), Google Street Views were used to collect attributes that were unavailable from archival data sources. Negative binomial (NB) models were used to relate traffic crash frequency to traffic conditions, intersection attributes (e.g., crossing distance, bike lane, one-way street, median type, crosswalk, and sidewalk), and contextual factors (e.g., land use mix), while controlling for pedestrian exposure.

While methodology (e.g., statistical models and grouping) is well-established for identifying risk factors, the methods for ranking network locations based on the risk factors is less understood. Therefore, the study explored different ways to rank locations through analyzing the sensitivity of ranking results to the ranking methods and the weight choices. Aside from the grouping-based method, a weighted-average method was developed combing the NB-based coefficient estimates and crash histories. Comparison between the two methods suggest that both were sensitive to the weights used, but the grouping-based ranking method combining arbitrary weights produced relatively stable results. However, the finding is inconclusive, and more research would be needed to investigate the ranking methods, especially the ones suitable for making decisions under multiple criteria (risk factors). The Analytic Hierarchy Process (AHP) (38) seems promising in that regard.

However, the results are confined to the sample being studied and inconclusive due to relative small sample size and potential biases that result from coarse estimates of pedestrian volume at the area level among others. Future work can further investigate if and how statistical modeling should be incorporated into the systemic selection approach. Statistical models like cross-sectional or time-series models provide more objective estimates of the expected crash frequency, an important input factor for gauging the benefits of countermeasures. Current practices often rely on grouping, which cross tabulates crashes against intervals/ranges of risk

<table>
<thead>
<tr>
<th>Weighted-average ranking</th>
<th>Stop-controlled intersections</th>
<th>Signalized intersections</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.72 (8; 0; 9)</td>
<td>0.75 (11; 0; 19)</td>
</tr>
<tr>
<td></td>
<td>0.79 (20; 0; 3)</td>
<td>0.71 (18; 1; 3)</td>
</tr>
</tbody>
</table>
factors in order to identify the intervals where a disproportionate number of crashes occurred. While quick and easy, this approach can overlook cross-correlations among risk factors (e.g., a wider shoulder might exhibit a spurious association with higher crash counts if speed is not controlled). Key technical questions remain that would benefit from future research, but this study offers new gleanings and suggests new directions.

ACKNOWLEDGEMENTS

The authors would like to thank Jianming Ma, Daniel Young, and Jen Duthie for providing data and Andrew Scott for editorial support.
REFERENCES


(33) Zegeer, C. V., R. C. Deen, and J. G. Mayes. The Effect of Lane and Shoulder Widths on Accident Reductions on Rural, Two-Lane Roads., Kentucky Transportation Centre Research Report, Paper 811, 1980.


