Evaluating the "Safety In Numbers" Effect With Estimated Pedestrian Activity

Brendan Murphy
Graduate Research Assistant, Accessibility Observatory
University of Minnesota, Department of Civil, Environmental, and Geo- Engineering 500 Pillsbury Drive SE, Minneapolis, MN 55455 USA
murph677@umn.edu

David Levinson
RP Braun-CTS Chair of Transportation Engineering
Director of Network, Economics, and Urban Systems Research Group
University of Minnesota, Department of Civil, Environmental, and Geo- Engineering
500 Pillsbury Drive SE, Minneapolis, MN 55455 USA
dlevinson@umn.edu http://nexus.umn.edu

Andrew Owen
Director of Accessibility Observatory
University of Minnesota, Department of Civil, Environmental, and Geo- Engineering
500 Pillsbury Drive SE, Minneapolis, MN 55455 USA
aowen@umn.edu http://nexus.umn.edu

3008 words + 8 figures + 2 tables
August 1, 2015

ABSTRACT

Pedestrian and bicyclist collision risk assessment offers a powerful and informative tool in urban planning applications, and can greatly serve to inform proper placement of improvements and treatment projects. However, sufficiently detailed data regarding pedestrian and bicycle activity are not readily available for many urban areas, and thus the activity levels and collision risk levels must be estimated. This study builds upon other current work by Murphy et al. (*1*) regarding pedestrian and bicycle activity estimation based on centrality and accessibility metrics, and extends the analysis techniques to estimation of pedestrian collision risk. The Safety In Numbers phenomenon, which refers to the observable effect that pedestrians become safer when there are more pedestrians present in a given area, i.e. that the individual per-pedestrian risk of a collision decreases with additional pedestrians, is a readily observed phenomenon that has been studied previously. The effect is investigated and observed in acquired traffic data, as well as estimated data, in Minneapolis, Minnesota.

INTRODUCTION

Pedestrian collision risk assessment, and its extension to collision risk assessment for bicyclists, offers a powerfully informative tool in urban planning regimes. Sufficiently detailed, local risk assessment estimates for active travel flows may markedly increase pedestrian and bicyclist safety in urban environments, when used to inform the placement and implementation of safety improvements and traffic calming measures. However, a rigorous process for constructing granular risk assessment data sets has not yet been adopted and standardized, for multiple reasons. A model of active transport risk assessment is uninformative if the pedestrian and vehicular flows do not accurately represent corresponding levels *in situ*, and many cities do not have complete data sets of active transport flow levels, instead favoring counts of vehicle traffic. As such, active transport flow levels must be extrapolated from sparse data sets, a technique commonly used in planning applications in Europe and Asia, but not yet in the United States (Raford and Ragland (*2*)).

Additionally, safety levels associated with these modes continue to be a problem, with 1.24 million vulnerable road users (VRUs) being killed in on-road accidents in 2010, and another 20-50 million injured globally. Further, a full 22% of traffic deaths worldwide are pedestrians, which is quite a high figure considering the transportation mode of walking harbors little danger unto itself (World Health Organization (*3*)). Non-motorized transportation as a set of modes tends to be some degree of unsafe in most average developed urban areas, except where specific programs and treatments have been employed to address the safety concerns, such as in Copenhagen, Denmark (Jensen (*4*)).

This investigation aims to evaluate whether the Safety In Numbers phenomenon is observable in both originally collected data and an extrapolation model in the Midwestern, U.S. city of Minneapolis, Minnesota. Safety in Numbers (SIN hereafter) refers to the phenomenon that pedestrians as road users become safer when there are more pedestrians present in a given locale or area, e.g. that the per-pedestrian risk of injurious interaction with motorized vehicles decreases as a function of the increasing flow of pedestrian traffic. SIN is well-supported by pedestrian crash data across a number of studies in various urban environments and reviews (Jacobsen (*5*), Leden (*6*), Bhatia and Wier (*7*)). The concept has seen relatively widespread adoption in urban planning schools of thought, though its temporal causality is not clear-cut (Bhatia and Wier (*7*)), and it is commonly discussed only in the context of pedestrian risk depending on pedestrian flow levels. The USDOT Strategic Plan for Fiscal Years 2012-2016 aims to reduce non-vehicle-occupant fatalities to 0.15 per 100 million vehicle-miles-traveled (VMT) by 2016. However, such a goal does not account for risk dependence on pedestrian flow levels, and thus the federal guidelines ignore the SIN effect.

By necessity, data informing placement of improvements and projects for walking and bicycling safety must be sufficiently granular; travel behavior studies are typically performed at the Transit Analysis Zone (TAZ) level, which is insufficiently fine-grained to allow for analysis of the shorter-distance travel modes of bicycling and walking (Schneider et al. (*8*), Schneider et al. (*9*)) provide a salient starting point for the comprehensive pedestrian risk assessment model, as a granular focus on a specific university campus included factors of pedestrian volume, vehicle volume, and an environmental factor (crosswalk length). Wier et al. (*10*) provide precedent for area-level modeling of pedestrian risk incorporating zoning and land use characteristics. These levels of detail correlate well with the realities of implementation of pedestrian safety investments, which occur not on the city-wide level, but within specific intersections and road segments. Pedestrian traffic, car traffic, and crash data from the city of Minneapolis will be analyzed to test for the SIN effect, as well as extrapolated data on pedestrian traffic levels at intersections where pedestrian counts may not have been available. The estimated data, models, portions of the methodology, and Figure 1 and Figure 2 are reproduced here for clarity from Murphy et al. (*1*).

METHODOLOGY

To predict pedestrian traffic levels, a model was built using census-block level information regarding economic accessibility (access to jobs) by economic sector via both strictly walking and via the net accessibility benefit of public transit, betweenness centrality, and Average Annual Daily Traffic (AADT) (Murphy et al. (*1*)). Then, the existence of the SIN effect was examined within both the collected city data, and the modeled data, for pedestrian activity levels and crash counts.

Data

This section briefly describes the data sources used in the pedestrian activity estimation models, and the data preparation process.

- **Data Sources**
  1. U.S. Census TIGER 2010 datasets: blocks, core-based statistical area (CBSA) for Minneapolis-St. Paul
  2. U.S. Census Longitudinal Employer-Household Dynamics (LEHD) 2011 Origin-Destination Employment Statistics (LODES)
  3. OpenStreetMap (OSM) North America extract, retrieved April 2014
  4. Turning movement counts (TMC) 2000-2013, City of Minneapolis
  5. Average Annual Daily Traffic (AADT) measurements 2000-2013, City of Minneapolis
  6. Traffic crash records 2000-2013, City of Minneapolis
  7. GTFS data from Metro Transit
- **Data Preparation**
  1. Construct pedestrian travel network graph for Minneapolis
  2. Geocode pedestrian Turning Movement Count (TMC) and Average Annual Daily Traffic (AADT) data to spatial locations
- **Accessibility & Centrality Calculation**
  1. For each Census block in Minneapolis, calculate travel time to all other blocks within a 5km radius for a single departure time
  2. Calculate cumulative opportunity accessibility to jobs for each census block, using thresholds of 5, 10, . . . , 30
  3. Calculate net transit accessibility benefit using a threshold of 30 minutes
  4. Calculate betweenness centrality for the Minneapolis OSM road network
- **Model estimation**

1. Construct linear regression of pedestrian behavior on walking accessibility, net transit accessibility, network centrality, and accessibility to job opportunities by sector
2. Assess and validate model on sample of other intersections in Minneapolis

- **Safety analysis**

1. Calculate pedestrian collision risk burden from collected data
2. Calculate pedestrian collision risk burden from extrapolated data
3. Evaluate whether SIN effect present

Intersection locations were determined from OSM road centerline data for the Minneapolis-St. Paul CBSA (Core-Based Statistical Area). The subset of intersections for which count data were available is displayed in Figure 1; these intersections were used to construct the predictive models. Accessibility calculations were performed using OpenTripPlanner (OTP) open-source routing software; GIS work performed in QGIS and PostGIS; network centrality measures computed in ArcMap GIS with the Urban Network Analysis Tools toolbox; statistical work done in SQL, Python, and R. Figure 2 displays the locations of intersections in Minneapolis used to estimate pedestrian activity and collision risk burden.

Accessibility

The first type of explanatory variable used in the model of Minneapolis pedestrian count data is cumulative opportunity accessibility. Using OTP, walking travel times along the network are calculated from each Census block centroid in Minneapolis, to each other block centroid within the travel-time thresholds of 5, 10, . . . , 30 minutes. Job opportunities are summed from each block centroid reachable within a given time threshold, yielding an X-minute accessibility measure. Job opportunities are broken down by economic sector, as defined by the North American Industry Classification System. There are two accessibility calculations used in this study:

1. Accessibility to jobs from Census block centroids by walking
2. Accessibility to jobs from Census block centroids by transit & walking

Pedestrian counts are often taken at intersections in either gross counts, or divided by turning movement type. This study uses Turning Movement Count (TMC) data from approximately 750 intersections in Minneapolis; intersection counts were calculated by adding the various TMC types for each intersection in the analysis group, to yield a gross figure of pedestrian activity within an intersection. Two-hour counts for pedestrian activity were used for morning peak (7-9AM), midday (11am-1pm), and evening peak (4-6PM). Accessibility calculations were performed using the following formulation of a gravity-based model:

$$A_i = \sum_j O_j f\left(C_{ij}\right) \tag{1}$$

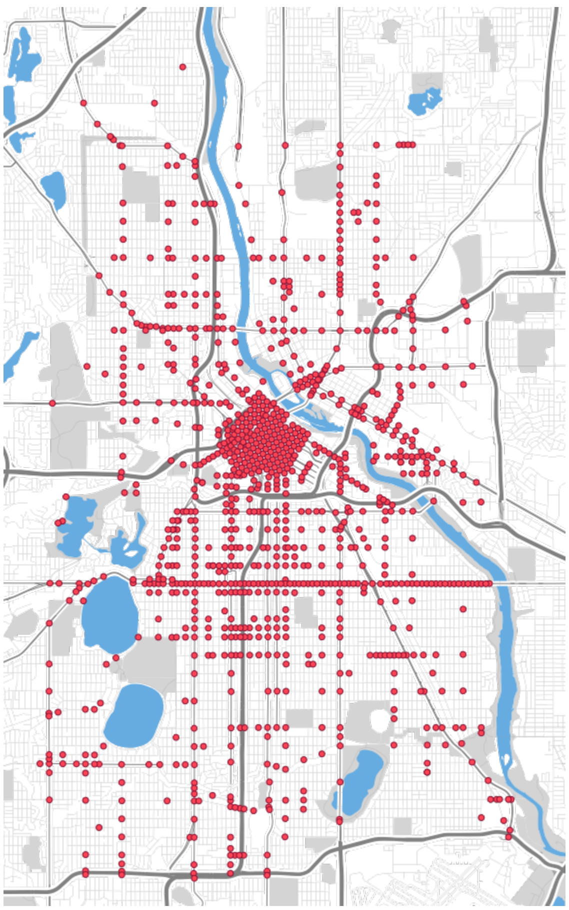


FIGURE 1 Locations of intersections in Minneapolis with raw pedestrian count data.

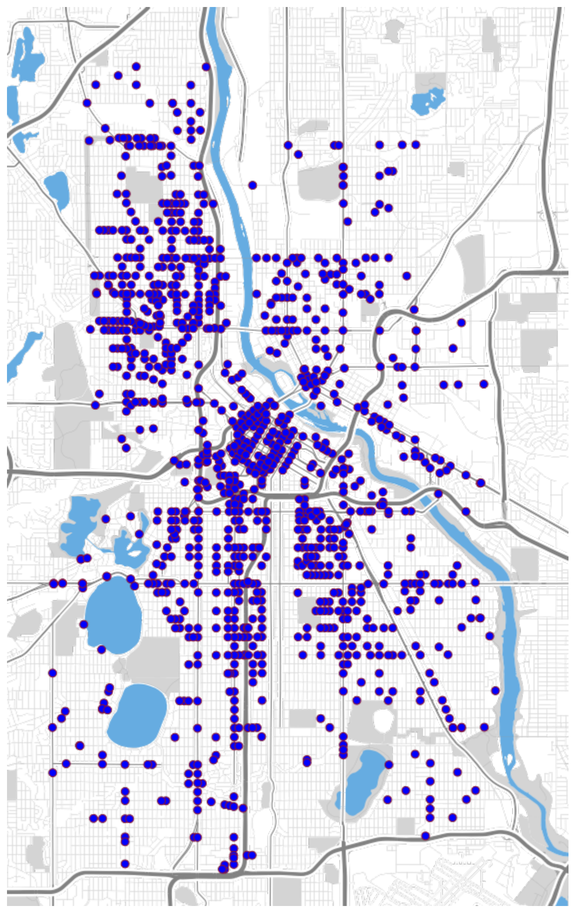


FIGURE 2 Locations of sampled intersections in Minneapolis, used in estimated pedestrian activity analysis.

$$A_i = \text{accessibility for location } i \tag{2}$$
$$O_j = \text{number of opportunities at location } j \tag{3}$$
$$C_{ij} = \text{time cost of travel from } i \text{ to } j \tag{4}$$
$$f\left(C_{ij}\right) = \text{weighting function} \tag{5}$$
$$\tag{6}$$

The choice of weighting function has a large impact on the resulting Accessibility calculations; however, one of the simplest interpretations of cumulative opportunities is an integer count, using the following weighting function:

$$f\left(C_{ij}\right) = \begin{cases} 1 & \text{if } C_{ij} \leq t \\ 0 & \text{if } C_{ij} > t \end{cases} \tag{7}$$

$$t = \text{travel time threshold}$$

Accessibility using this type of weighting function yields an intuitive metric - discrete counts of the numbers of opportunities (jobs, restaurants, other destinations) reachable within a certain time frame. We predicted that locations with higher accessibility values would see greater pedestrian activity, and thus a decreased pedestrian risk burden of collisions, throughout the day. Accessibility for both walking, and walking + transit modes, are used in the activity estimation model; net transit benefit is calculated by subtracting the walking accessibility at each Census block from the walking + transit accessibility value, and including both variables separately in the model allows for targeted analysis of how significant transit access is to pedestrian activity and safety.

Centrality

The underlying network topology, including features such as block size and degrees of connectivity, can influence the walkability of a place. Betweenness centrality was computed in ArcGIS with the Urban Network Analysis Toolbox, to include the influences of network topology in estimating pedestrian activity and safety levels. Various types of network measures of centrality have been proposed in their applicability to estimation of non-motorized activity levels (McDaniel et al. (*11*), Anciães (*12*), Do et al. (*13*)), and safety and collision rates (Zhang et al. (*14*), Dai et al. (*15*)). One of the most common measures of centrality is "betweenness" centrality, or how "between" other nodes or links a given node or link is. When considering route choice and estimating modal traffic flows, link betweenness centrality is often considered, and consists of the proportion of shortest paths between all node pairs that pass through a link or node (McCahil and Garrick (*16*)). Specifically, the metric of stress centrality was used, which consists of counting the number of times each link in a given network is utilized when enumerating the set of shortest baths between all node pairs; this metric is given by:

$$C_s(k) = \sum_{i,j \in V} \sigma_{ij}(k) \tag{8}$$

where σ_{ij} is either 1 if link k is used in shortest path σ_{ij}, and 0 otherwise. This form of stress centrality has been used to spatially assess transportation systems (Derrible (*17*)). Here, simple

TABLE 1 Dataset Summary Statistics

Description	Value
Intersections with evening ped counts	741
Crashes at intersections with evening ped counts	1064 (1052 injuries, 12 deaths)
Crashes at intersections with evening ped counts per year	76
Intersections included in estimation modeling	1123
Crashes at all intersections in estimation modeling	2513 (2478 injuries, 35 deaths)
Crashes at all intersections in estimation modeling per year	179.5
Intersection-μ crashes per year with evening ped counts (incl. zero-crash ints.)	0.1518
Intersection-μ crashes per year with evening ped counts (w/o zero-crash ints.)	0.2647
Intersection-μ crashes per year in estimation modeling	0.1597
Intersection-μ total ped activity per day	633.66, $\sigma = 2023.20$
Intersection-μ morning ped activity per day	194.70, $\sigma = 570.34$
Intersection-μ midday ped activity per day	270.74, $\sigma = 994.79$
Intersection-μ evening ped activity per day	264.52, $\sigma = 733.49$

Note: Summary statistics for datasets used in pedestrian activity analysis: pedestrian turning movements between 2000 and 2013, and aggregate crash reports 2000-2013, for the City of Minneapolis.

stress centrality was calculated with a 5km limiting radius on shortest path length, corresponding to an hour of walking at a typical human pace, as it is not reasonable to include the entire set of road network intersections as possible destinations for a given intersection-origin when walking. Each O/D pair was weighted equally in the centrality calculation.

Pedestrian Safety Estimation

Pedestrian risk-burden for collisions with automobiles was first calculated for the raw data, and then for the estimated data based upon the modeled pedestrian activity. The pedestrian activity model was derived via multiple regression in R using iterative stepwise regression to determine the most highly predictive explanatory variables; pedestrian activity was then estimated by applying the derived model to a subset of intersections, many of which did not have pedestrian count data. Pedestrian collision risk-burden is defined as the number of crashes occurring at an intersection during the 14-year measurement period, per pedestrian walking through that intersection on a given day. If the per-pedestrian rates of crashes are lower at intersections with more pedestrian activity, then a ”safety in numbers” effect is observed. In both safety models, the number of car-pedestrian crashes at intersections is not altered, and is taken from aggregated crash report data.

RESULTS

A parsimonious model for walking activity, in terms of the strongest explanatory variables, is reported in Table 2. Table 1 lists summary statistics for the datasets used in the safety analysis: automobile-pedestrian crashes between 2000 and 2013; pedestrian turning movement counts (TMC) between 2000 and 2013; and automobile AADT figures between 2000 and 2013.

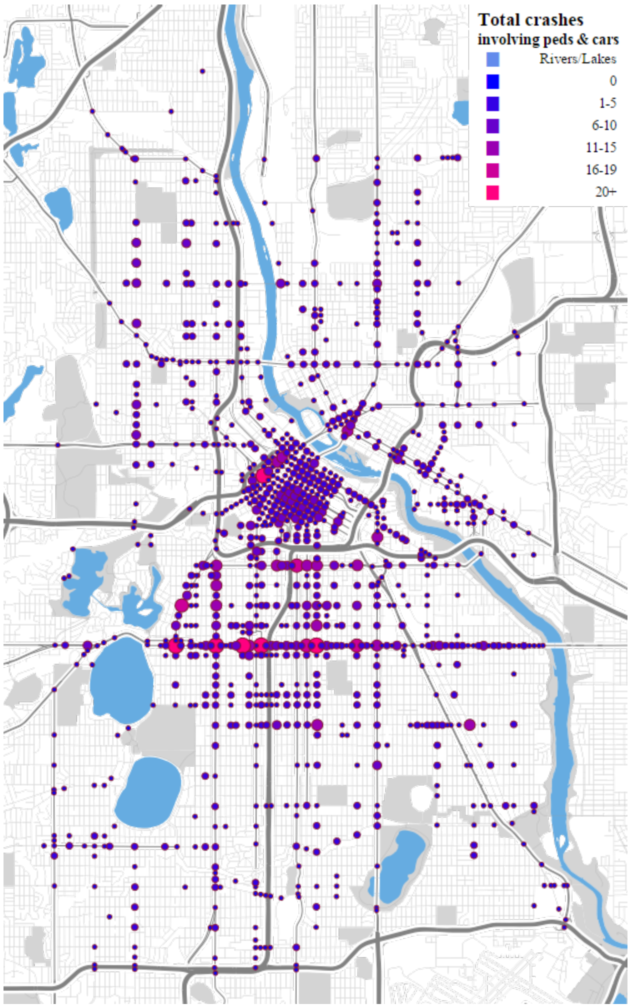
Regression results for the two parsimonious models for walking activity, with and without AADT included, are in Table 2. Accessibility by walking, net transit benefit to accessibility,

TABLE 2 Parsimonious Model Regression Results: With & Without AADT

	Dependent variable:	
	Average PM pedestrians	
	(1)	(2)
Walking accessibility (15-minute)	0.410**	0.649***
	(0.173)	(0.112)
Net transit accessibility (30-minute)	0.320***	0.129**
	(0.093)	(0.053)
Betweenness	0.029	0.487***
	(0.371)	(0.186)
AADT	1.312*	
	(0.679)	
Management jobs 5min	−0.114***	−0.109***
	(0.033)	(0.017)
Education jobs 5min	0.922***	0.700***
	(0.086)	(0.058)
Finance jobs 10min	0.071***	0.054***
	(0.009)	(0.006)
Utilities jobs 15min	−0.968***	−0.729***
	(0.104)	(0.071)
Constant	−15.208	−1.698
	(9.874)	(4.795)
Observations	486	1,016
R^2	0.287	0.226
Adjusted R^2	0.275	0.221
Residual Std. Error	83.830 (df = 477)	72.773 (df = 1008)
F Statistic	23.970*** (df = 8; 477)	42.139*** (df = 7; 1008)
Note:	*p<0.1; **p<0.05; ***p<0.01; (standard error)	

AADT, and accessibility to Finance and Education jobs were all found to be significant predictors of increased pedestrian activity. Accessibility to Management and Utilities jobs were found to be significant predictors of decreased pedestrian activity, relative to other variables. Betweenness centrality was not found to be a significant predictor of pedestrian traffic, but showed weakly positive correlation.

Safety analysis was also performed on the raw data, to attempt to verify the existence of the Safety in Numbers effect. Figure 3 and Figure 4 display the unweighted and weighted levels of pedestrian-auto crashes in Minneapolis between 2000 and 2013, respectively. The pedestrian-weighted data displayed in Figure 4 are plotted in Figure 7, which shows the relationship between per-pedestrian crash risk and the average daily pedestrian use level of an intersection. Such an effect, wherein intersections characterized by greater daily levels of pedestrian activity show lower per-pedestrian crash rates than less-active intersections, appears to be present in the raw data. Figure 8 shows the same relationship, but for estimated pedestrian count data based on the explanatory variables enumerated in Table 2. Exponential models are fitted to both the raw and estimated data, and both datasets appear to show significant trends towards exhibiting the "safety in numbers" effect.



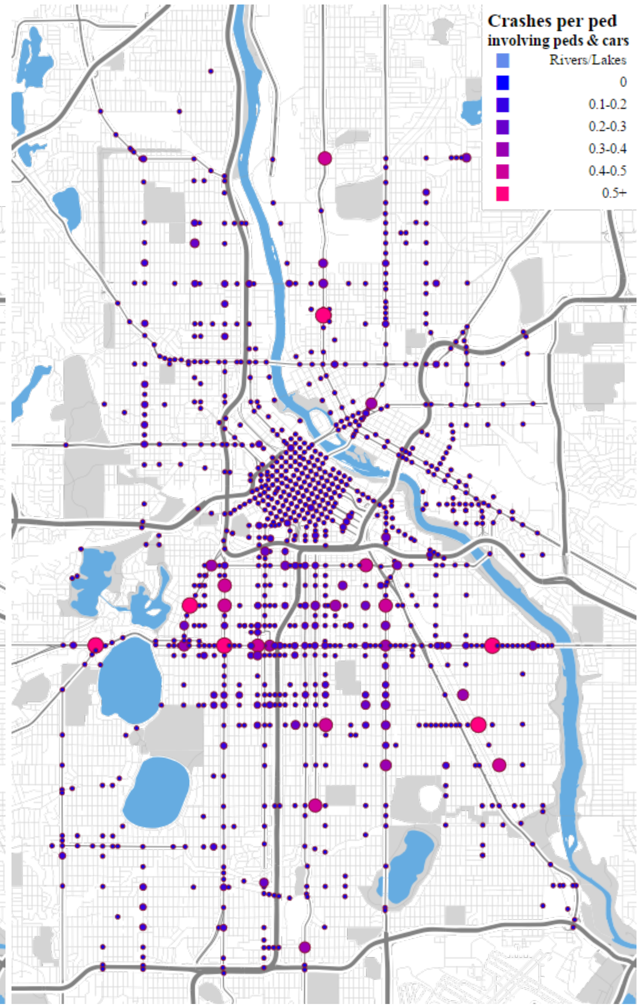


FIGURE 3 Raw levels of pedestrian-auto crashes in Minneapolis, 2000-2013.

FIGURE 4 Pedestrian-weighted levels of ped-auto crashes in Minneapolis, 2000-2013. Pedestrian risk is defined as the number of crashes in 2000-2013, per pedestrian in daily PM peak period.

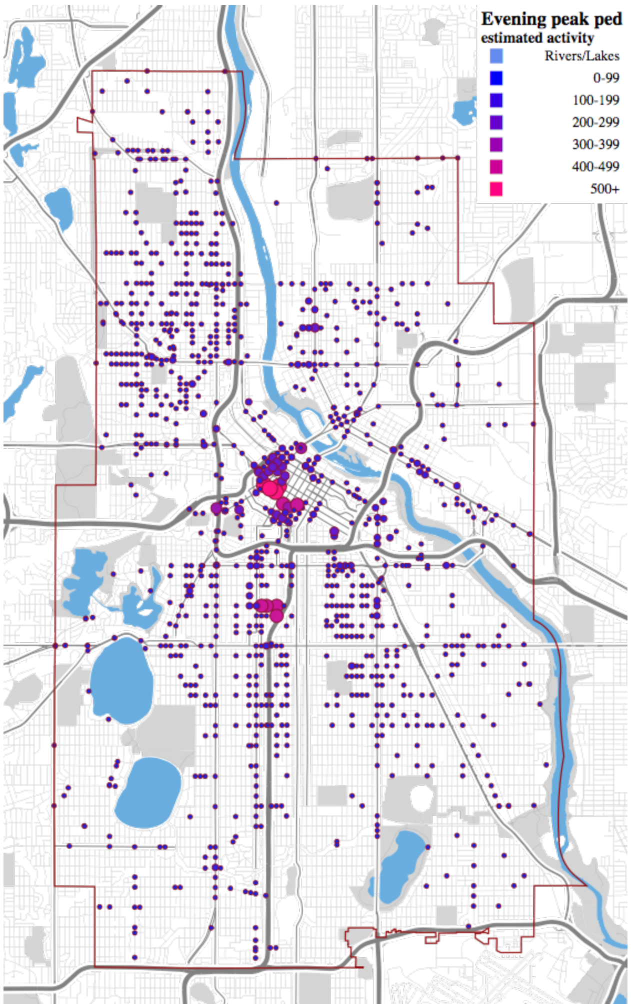


FIGURE 5 Estimated levels of evening peak pedestrian activity in Minneapolis.

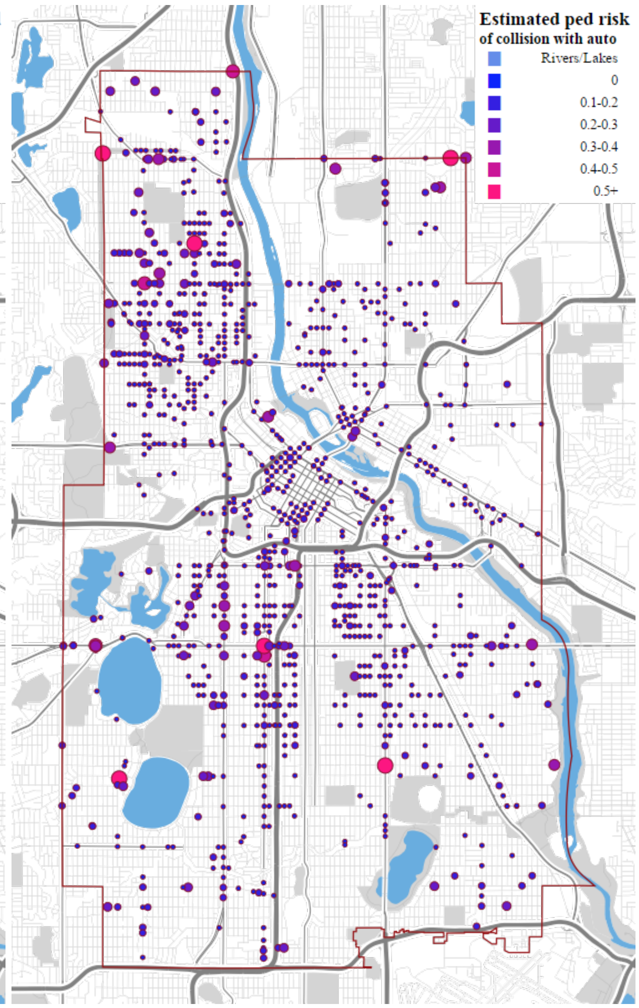


FIGURE 6 Estimated weighted pedestrian risk of crash. Pedestrian risk is defined as the number of crashes in 2000-2013, per pedestrian in daily PM peak period.

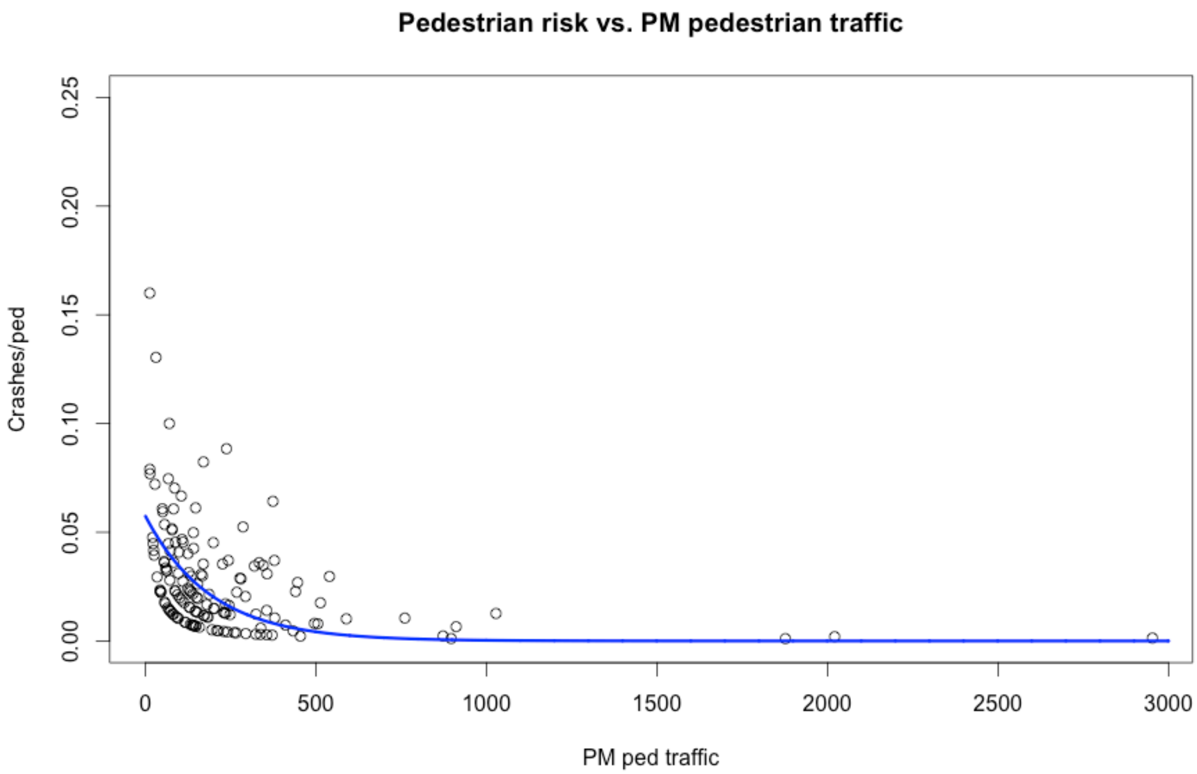


FIGURE 7 Pedestrian risk burden vs. pedestrian traffic levels, raw data; exponential fit, $b = -0.0516$, $RSE = 0.1018$, $p << 0.05$. Pedestrian risk is defined as the number of crashes in 2000-2013, per pedestrian in daily PM peak period.

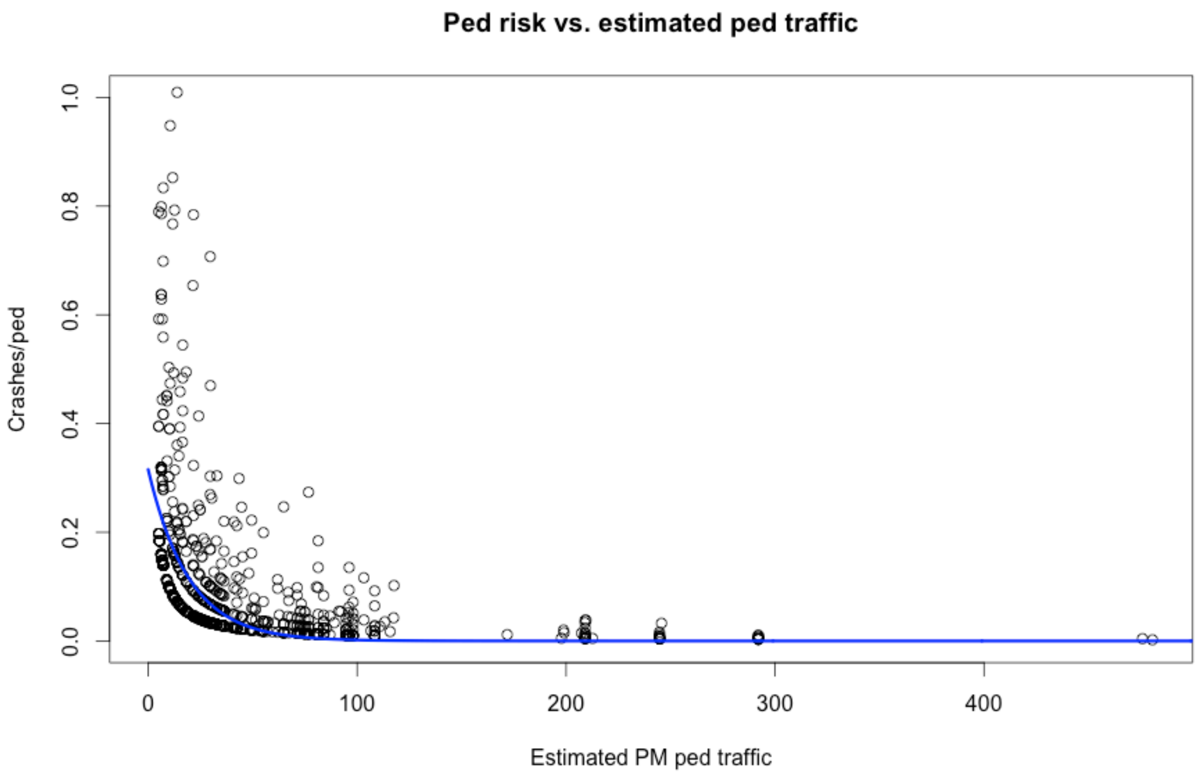


FIGURE 8 Pedestrian risk burden vs. pedestrian traffic levels, estimated data; exponential fit, $b = -0.0248$, $RSE = 0.0664$, $p << 0.05$. Pedestrian risk is defined as the number of crashes in 2000-2013, per pedestrian in daily PM peak period.

DISCUSSION & CONCLUSION

The "safety in numbers" effect was indeed observed in both the raw Minneapolis pedestrian and crash data, as well as the modeled data at the broader sample of intersections (visible in Figure 3 and Figure 4). Intersections characterized by higher per-day pedestrian traffic exhibited lower per-pedestrian crash rates, a phenomenon that has been observed and described previously (see Jacobsen (*5*), Leden (*6*), Bhatia and Wier (*7*)). The precise reasons behind this effect are not definitively known; however, the aforementioned studies have hypothesized psychological effects on drivers, in that when driving in environments characterized by greater average levels of pedestrians, drivers may tend to act with more caution. Additionally, spatial geometric probability of crashes for a given pedestrian necessarily varies with additional pedestrians present within an intersection, due to physical constraints of the built environment.

An ongoing challenge with activity estimation and safety analysis dependent on count and crash data is the issue of data quality and availability. Data practices vary from city to city and state to state, with implications to investigations intending to collate and aggregate safety data for cross-jurisdiction comparison. Additionally, a large amount of city data collection pertaining to street utilization is still performed manually, and such processes are error-prone and inconsistent between jurisdictions. This study used a combination of national (Census, LEHD) datasets and local (Minneapolis traffic) data. Some cities, such as Boston, do not have robust pedestrian and bicycle counting programs throughout the city; others, such as Philadelphia, may have varying data release and non-disclosure agreements between MPOs, cities, and police departments; still other cities may have inconsistent data tracking and release practices, such as Washington, D.C. The collection and processing of pedestrian and bicycle spatial safety data on an aggregate scale becomes exceedingly difficult. Better standards of practice in data collection, management, and distribution are needed.

Visualizing unsafe intersections, or groups of intersections, within an urban area is an important angle of analysis to undertake with the types of datasets used in this investigation. Problematic areas within the city environment become readily apparent; when multiple intersections with relatively high pedestrian injury risk-burden lie in the same corridor, such as Lake Street in Minneapolis, a discussion of pedestrian safety and the surrounding built environment should occur. Some of these problematic areas are visible in Figure 4 with the raw original data, as well as in Figure 6 for the estimated model data. The entire Lake Street corridor stands out as an area with elevated pedestrian risk burdens given the number of pedestrians walking there, compared to downtown. Further, if the sample data were to only contain a few intersections within that corridor, the predictive models would enable planners and engineers to construct a more complete picture of pedestrian safety and activity throughout the entire corridor. Through the pedestrian risk-burden analysis, it is also possible to see intersections with a disproportionately high rate of crashes for its level of pedestrian activity, which serves as salient information for urban planners and engineers wishing to alter the built environment to increase pedestrian and bicyclist safety factors.

Future Directions

As mentioned in Murphy et al. (*1*), the analysis framework outlined in this report constitutes Phase I of the overall investigation. Phase II will extend the framework to the bicycling mode, and look at estimating the collision risk variability for bicycling throughout an urban area. Activity levels at the granularity level of intersections will be estimated for bicycles in much the same fashion as for walking, using explanatory variables of a time-threshold value of bicycle accessibility to jobs by sector, betweenness centrality, and net transit accessibility benefit. However, we hypothesize that adapting the betweenness measure to use spatial work trip distributions given by LEHD data will more closely reflect actual pedestrian use-cases than all-to-all O/D pair analysis. OpenTripPlanner software will again be utilized for accessibility calculations, and the modeling and analysis framework for safety levels will be analogous to the above.

REFERENCES

[1] Murphy, B., D. Levinson, and A. Owen, Accessibility and Centrality Based Estimation of Urban Pedestrian Activity, 2015.

[2] Raford, N. and D. Ragland, Space Syntax: Innovative Pedestrian Volume Modeling Tool for Pedestrian Safety. *Transportation Research Record*, Vol. 1878, No. 1, 2004, pp. 66–74.

[3] World Health Organization, Global status report on road safety. *Injury prevention*, 2013, p. 318.

[4] Jensen, S. r. U., Safety effects of blue cycle crossings: A before-after study. *Accident Analysis and Prevention*, Vol. 40, No. 2, 2008, pp. 742–750.

[5] Jacobsen, P., Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention*, Vol. 9, No. 3, 2003, pp. 205–209.

[6] Leden, L., Pedestrian risk decrease with pedestrian flow. A case study based on data from signalized intersections in Hamilton, Ontario. *Accident Analysis and Prevention*, Vol. 34, No. 4, 2002, pp. 457–464.

[7] Bhatia, R. and M. Wier, "Safety in Numbers" re-examined: Can we make valid or practical inferences from available evidence? *Accident Analysis and Prevention*, Vol. 43, No. 1, 2011, pp. 235–240.

[8] Schneider, R. J., L. S. Arnold, and D. R. Ragland, *Pilot Model for Estimating Pedestrian Intersection Crossing Volumes*, 2010.

[9] Schneider, R. J., R. M. Ryznar, and A. J. Khattak, An accident waiting to happen: A spatial approach to proactive pedestrian planning. *Accident Analysis and Prevention*, Vol. 36, No. 2, 2003, pp. 193–211.

[10] Wier, M., J. Weintraub, E. H. Humphreys, E. Seto, and R. Bhatia, An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident; analysis and prevention*, Vol. 41, No. 1, 2009, pp. 137–45.

[11] McDaniel, S., M. B. Lowry, and M. Dixon, Using Origin-Destination Centrality to Estimate Directional Bicycle Volumes, 2013, p. 15.

[12] Anciães, P. R., Urban transport , pedestrian mobility and social justice A GIS analysis of the case of the Lisbon Metropolitan Area. , No. July, 2011.

[13] Do, M. T., O. Grembek, D. Ragland, and C.-Y. Chan, Weighting integration by block heterogeneity to evaluate pedestrian activity, 2013, pp. 14p, schmas,\rtabl., ill., bibliogr.

[14] Zhang, Y., J. Bigham, D. Ragland, and X. Chen, Investigating the associations between road network structure and non-motorist accidents. *Journal of Transport Geography*, Vol. 42, 2015, pp. 34–47.

[15] Dai, D., E. Taquechel, J. Steward, and S. Strasser, The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban university campus. *The western journal of emergency medicine*, Vol. 11, No. 3, 2010, pp. 294–301.

[16] McCahil, C. and N. W. Garrick, The Applicability of Space Syntax to Bicycle Facility Planning. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2074, No. -1, 2008, pp. 46–51.

[17] Derrible, S., No Title. In *Using Network Centrality to Determine Key Transfer Stations in Public Transportation Systems*, 2012.