Using GIS to Determine Risk Factors of Urban Bicyclists

Buffalo, NY Case Study

Elizabeth Cahill, Jean-Claude Thill, Eric Delmelle

Abstract

As issues related to oil dependencies, rising gas prices and global warming come to the forefront of topics of concern for Americans, the need for alternative modes of transportation has become critical. Urban settings are seemingly ideal for bicycling to become a predominant mode given the compactness of destinations. However, in the United States, bicycling is both scarcely used and very dangerous as bicyclists are 12 times more likely to be killed than automobile drivers.

The purpose of this study is to use GIS to model bicycle accidents in order to determine and compare risk factors of both child and adult bicyclists. Physical road characteristics such as intersection composition (arc-node connectivity), and road classification in addition to social variables and potential trip attractors will be examined. Given the spatial nature of these variables, a spatially weighted regression model will be incorporated to limit potential dependencies amongst the attributes.

Keywords: Bicycling, Accident Analysis

1. Introduction

Bicycling as a mode of transportation has numerous benefits in an urban environment ranging from cost efficiency to the increased health of bicyclists in a time when health related problems in this country due to a lack of physical activity are flourishing. Additionally, it reduces the number of automobiles on the road thereby causing less congestion; it requires no fossil fuel consumption and as a result emits no air or noise pollution.

Despite these positive facts, only a very small portion of the population in the United States utilizes this transport mode (Pucher et al. 1999). In 1995, bicycling’s modal share of all trips was only 0.9 percent. Of these bicycling trips, the majority were for social or recreational purposes, as commuting to work composed only 9 percent of all bicycling journeys. Pucher et al. (1999) identified eight deterrents to cycling in the United States: public attitude, public image, city size and density, cost of car use and public transportation, income, climate, danger, and infrastructure. The danger aspect recognized by Pucher will be the focus of this research.

In the US, bicyclists are 12 times more likely to be killed than automobile drivers and in New York State, of all reported accidents, bicyclists and pedestrians compose 30 percent of all traffic fatalities and 16 percent of all serious traffic accidents (Brustman 1999). The high risks associated with bicycling have proven to be a deterrent to the number of people commuting by this mode. A study by Noland (1995) on perceived risk
and bicycle modal choice showed that improvements in bicycle safety resulted in a proportionally higher number of people commuting by bicycle. It has also been shown that an increase in the number of people walking or bicycling yields a lower accident rate for these two transportation modes as drivers are more aware of the presence of these people on the roads (Jacobsen 2003). This appears to create a cyclical problem as a greater number of cyclists will make the activity safer, yet people don’t want to utilize the mode until it is safe. It is clear that measures must be taken to amend some of the risks associated with biking which could then ignite an increase in riders that would further enhance the safety of the activity.

Bicycle accidents

For the United States as a whole, according to the US Department of Transportation’s Fatality Analysis Reporting System and the National Center for Statistics and Analysis, 725 bicyclists were killed and 41,000 injured in 2004. These fatalities accounted for 2 percent of all traffic fatalities annually. Nearly 20 percent of all cyclists killed in crashes were between the ages of 5 and 15. The fatality rate of this group was 3.1 per million people, 24 percent higher than the rate for all cyclists (2.5 per million). The injury rate for the ages 5 to 15 group was 286 per million compared to 140 per million for bicyclists of all ages. Eighty percent killed and 76 percent injured were male. Sixty six percent of fatalities were in an urban environment and 30 percent occurred at non-intersection locations. Incidents were most frequent in the summer months of June, July and August (36 percent) and between the hours of 5 and 9pm (30 percent) (U.S.DOT 2004).

The cause of bicycle-automobile collisions has been shown to most often be a result of the bicyclist not obeying traffic laws (Garder 1994). A study of accidents at intersections by Wachtel and Lewiston (1994) found that bicyclists riding against the flow of traffic were on average 3.6 times more likely to be involved in a collision. In an investigation of demographic and socio-economic characteristics of bicycle crashes, Epperson (1995) found the majority of the bicyclists involved in Dade County, Florida were of a lower economic class. A significant socio-economic gradient in the accident location and neighborhoods of children involved in traffic accidents has been discovered (Braddock et al. 1991, Christie 1995, Roberts et al. 1994). According to Laflamme et al. (2000), traffic injuries are not only the most common cause of childhood death in the industrialized world, but are one of the causes of mortality with the greatest social class gradient. Petch et al. (2000) found that the majority of accidents occur close to home, and the intensity would decline with separating distance. LaScala et al. (2004) found that annual child traffic injuries were more likely to occur in places with a high youth population density, high unemployment rates, lower incomes, and higher traffic flow.

From a physical road perspective, Hoque and Andreassen (1986) studied pedestrian accidents in relation to their street segment and intersection classification. The authors determined that the highest accident rates for intersection collisions occurred at the higher-order intersection classes of primary/primary, primary/secondary, and secondary/secondary types while the highest rate for street segment accidents occurred on secondary arterial roads. Bicycle accidents at intersections were examined for both spatial and physical predictors (Mitra et al. 2007) revealing a significant positive correlation with average daily traffic rates at major and minor-road intersections.
Additionally, the presence of bars and pubs, colleges and universities, and middle schools close the intersections had a positive effect on bicycle collisions.

2. Purpose /problem statement

The purpose of this study is to address bicycle accidents from a geographic perspective. We concentrate on statistical and spatial analysis methods as well as Geographic Information System techniques to determine variation amongst accidents. The study will focus on both physical and social variables in an urban environment and specifically highlight the different risk factors between child and adult bicyclists. A comparison will also be made between the location of the accident and the location of the bicyclists’ residence to gain insight on the actual characteristics of the riders as well as the location of crash sites.

3. Methodology

Descriptive Statistics and Visualization

General descriptive statistics are computed for accidents in SPSS including histograms, graphs, and numerical summaries. Accidents are visualized within ArcGIS by geocoding addresses to the New York State Accident Location Information System (ALIS) street network. Kernel density mapping is used as a means for visually identifying clustered areas.

Distance Calculation

To determine the distance traveled from the home address to the accident site, ArcView’s Network Analyst can compute the shortest path between the two locations. An Avenue script loops through the database of origins and accident locations, calculates the network distance, and writes the result to a new field in the table. A limitation of this method is the assumption that bicyclists began their journey from the provided home address. While this assumption will be the cause to some error in these calculations, the values should still give a general overview to how far bicyclists are riding.

Regression Analysis

A regression analysis is used to identify potential risk factors and predictors of bicycle accidents based on accident location as well as characteristics of city bicyclists. Given the difference between child and adult bicyclists in terms of traffic experience and bicycling purpose, separate models are created for both groups. Four models are therefore created for the different dependent variables: adult home addresses, child home addresses, adult accident locations, and child accident locations. These are all normalized by census tract population and all variables are aggregated to the census tract.

A step-wise multiple regression analysis is run in SPSS to identify the significant independent variables for each model. A 0.05 significance level is set for entry into the model and 0.10 level for removal. To determine and account for potential spatial effects in these models, diagnostics are run in GeoDa to check for spatial dependencies in the residuals. The Lagrange Multiplier test statistics pertaining to both the spatial lag (LM lag) and spatial error models (LM error) are calculated. The following guidelines on
selecting a proper regression model, suggested by Anselin (2005), are used: if neither of the LM test statistics are significant then the OLS results are kept, however if the results of either of the tests are significant, then the corresponding spatial model is used as the alternative regression. If both statistics are significant, then the Robust forms of the statistic are examined and the significant one is chosen. In the case where both tests reject the null hypothesis, the one that is most significant is used. Queen contiguity is selected for the weights matrix in the spatial models. All models are checked for collinearity and reported only in the case of a problem.

**GIS Data Description and Development**

As Census Tracts are selected as the aggregation unit for the analysis, all variables, both independent and dependent, must be assigned to these units.

**Crash Data**

Data for this study were collected for the years 2003-2004 from the city of Buffalo Police Department accident records. Information regarding both the driver and the bicyclist are recorded in these records as well as location information and other factors surrounding the accident.

Both the accident points and the addresses of the bicyclists are geocoded to the street network. Aggregating these points to the tract level provides a problem in the case where the points are located on a street serving as a tract boundary. Several options exist in this situation. The first is to use the address assigned to the accident by the reporting officer as recorded on the Motor Vehicle Police Accident Report. In this case, points with addresses on the right side of the road are assigned to the tract on the right and vice versa for accidents on the left-hand side. This operation is easily performed when accidents are geocoded with a specified offset and the street and boundary layers are perfectly aligned enabling a spatial join of the two layers and therefore providing a count for each tract. A significant drawback of this method stems from the accuracy of the recorded address. It is very likely that the reporting police officer will look to find the closest visible address to the scene regardless of where on the road the accident actually occurred. This is especially true for accidents occurring close to the road’s centerline. Verifying the addresses provided on the reports is very difficult and so using this method could result in inaccurate counts.

Another option is recording boundary accidents twice – once in each of the tracts it shares or up to 5 times for accidents occurring at the convergence of 5 different census tracts. This method also introduces error by creating a false number of total accidents and also gives more weight to the boundary accidents over others as they are counted more frequently.

A third alternative is to apportion the accident, giving a count of one half to each tract if it shares a border between two units, one third if it is at an intersection of three units, etcetera. This is a more time consuming process, as no automated process is readily available within a GIS software package. However, it should produce the most accurate representation of the number of accidents per tract of the options presented and is the approach taken for this analysis.
Street Network
The New York State Accident Location Information System (ALIS) street network is used in this analysis. This dataset contains address ranges thereby allowing accident data to be geocoded to the streets. Additionally, improvements have been made to this network in street geometry, accuracy, and attributes compared to the TIGER Line files developed by the US Census Bureau.

As the classification of both intersections and street segments are considered in the regression analysis, these variables need to be aggregated as well. The Framework Classification Code (FCC) is used to categorize both the segments and intersections. Four classes are used for the segments: A2, A3, A4, A6. Class A2 roads are non-limited access highways, typically with a speed limit of 45 miles per hour. Class A3 and A4 roads serve as local city streets; A3 with a speed limit of 35mph and A4 with a speed limit of 25mph. Finally, A6 class roads have a speed limit of 20mph and are typically collector roads, round-abouts, and highway ramps. All limited-access highways restricting bicycle use are removed from the study. The percentage of each class of road is compiled and assigned to their corresponding tract.

Intersections are divided into groups according to the street segments forming the node. After network topology is built and from-to-nodes are assigned to the segments, a Python script is run to get the number of intersecting segments of each node as well as their FCC class. This information is written to the node’s attribute table and classified into the following groups: A2/A2, A3/A3, A4/A4, A6/A6, A2/A3, A2/A4, A2/A6, A3/A4, A3/A6, A4/A6. Percentages of each intersection classification are then compiled for the tracts. Intersections on the boundary of more than one tract are assigned to each of the tracts it intersects to more accurately assess the influence of each type of intersection within each boundary.

Socio-Demographic Variables
Socio-demographic and average income statistics at the tract level are downloaded from the U.S. Census Bureau’s website and joined to the ALIS boundary layer. From the 2000 Census, the following variables are selected or derived to be used in the regression model: average household size, house density, average income, percent under poverty level, percent male, and percent of individuals in the age groups 1-17, 18-29, 30-64, and 64 and above.

Non-motorized Transportation Attractors
The 2000 Business Census dataset is used to provide information on the number of employees and businesses in each zone. It is hypothesized that dense areas of either places of work or businesses would be associated with larger numbers of bicyclists. As this dataset is available only at the zip code level for the City of Buffalo, areal interpolation is used to derive estimates of these concentrations at the tract level. As in all cases when dealing with aggregated data, a source of error lays in the assumption that the values are evenly distributed throughout the aggregation unit especially as the size of the unit increases.

Other potential trip attractors include schools and universities. These layers are obtained from the New York State GIS Clearinghouse and a count of each per census tract is added.
4. Results from the GIS analysis and discussion

Dataset Overview

The general accident characteristics gathered from the police accident records are very consistent with findings from previous studies noted in the literature review. A total of 280 bicycle accidents are collected for the years 2003 and 2004 for the city of Buffalo’s five police districts with no fatalities recorded.

From the distribution of accident cases through the year it can be seen there is a clear relationship with warmer weather and the amount of bicycling. The graph in Figure 4.1 illustrates a resounding trend towards an increase in accidents in the warmer weathered summer months and a noticeable decrease in cases as the weather becomes less favorable towards bicycling. This is not indicative of more dangerous riding conditions in the summer (exposure), rather of a greater number of cyclists (population at risk) and therefore an overall increase in accident frequency.

![Graph showing distribution of accidents by month.](image)

Figure 4.1. Distribution of accidents by month.

There are discernible differences in the gender traits of bicyclists involved in the crashes as males overwhelm females, composing 80 percent of the victims. Younger riders are also dominant in the findings with a modal age of 13, a minimum of 5 and a maximum of 77. Bicyclists aged 11, 12, 13, and 14 totaled 20 percent of all cases (Figure 4.2).
Accidents occur 66 percent of the time at intersections and 34 percent at mid-block locations. An intersection is defined as the crossing of two or more roadways not categorized as driveways; it is information directly collected from the MV-104 accident report. When the cases are split into two groups on the basis of age (aged 15 and younger and 15 and older), clear differences are observed. For the young age group, 57 percent of the accidents occur at intersections while 43 percent take place at mid-block locations; the older group however has a greater difference between the two location types: 73 percent take place at intersections while only 27 percent are at mid-block locations. Behavioral differences of these two groups can be an explanation for this statistic. Younger aged riders tend to use bicycling more as a means of recreation than for transportation. If they are riding around in their own residential neighborhood, they are more likely to ride into the middle of the road from a sidewalk or driveway without looking, and consequently increasing their chance of being hit by a car. On-street parking creates an additional risk to children as vehicle presence can inhibit a child’s ability to notice oncoming cars and also obstructs the driver’s view of the situation. Adults, on the other hand, are more likely to be bicycling as a means for transportation, following the road network from origin to destination with a more thorough knowledge of traffic patterns and laws. For these types of bicyclists, intersections are a greater source of confusion and danger. Often bicyclists will ignore traffic signals and ride through intersections causing collisions with turning vehicles. Separate bicycling lanes are typically situated to the extreme right so if a right-hand turning lane is present, and a bicyclist is continuing straight accidents can easily occur. At intersections, the majority of collisions occur when the vehicle is going straight ahead or making a right turn.
**Data Visualization**

A kernel density map is created with ArcGIS Spatial Analyst to visualize the concentration of points and to find where the highest densities occur in the city. When the accidents are divided into the two age groups used for comparison in this study, 15 and younger and 16 and older, the geographic distribution of the respective accident events can be compared through density maps. The age groups are selected on the basis that in New York State, driving becomes legal at the age of 16 and so education about traffic rules and safe driving courses are taught which can potentially affect bicycling behavior in relation to these laws. The maps in Figure 4.3 show the two kernel density maps for these groups with a 400-meter search bandwidth. The adult accidents seem to follow a much more linear pattern along arterials while the child accidents have less of a detectable pattern characterized by tiny pockets scattered throughout the metropolitan area. There is a concentration of accidents around the neighborhoods of South Buffalo for both groups, but the accidents involving adults follow the main road of South Park Ave., while the youth accidents are more dispersed within the local streets.

![Figure 4.3. Kernel density maps with a search radius of 400m for adults (a) and children (b).](image)

**Distance Analysis**

Using ArcView’s Network Analyst combined with an Avenue script, the network distances between the home address and accident address are computed. Out of all 280 accidents, the distance is calculated for 231 cases. The omitted incidents are attributed to
a lack of data entered for the bicyclist on the police report, or an address that was out of the range of the city study area. The average distance traveled for all riders is 1,501.5 meters with a minimum of 3.2 and a maximum of 8,770.9 meters. When separated by age, the average distance for bicyclists aged 15 and under is 933.4 meters while the average distance to those older than 15 is twice this amount, 1,850.3 meters. These results show that the younger riders that are very prominent in the number of accidents are unlikely to travel far from their own neighborhoods. A correlation between the bicyclist’s age and the distance traveled is computed and Figure 4.4 demonstrates the positive relationship.

Figure 4.4. Relationship between age of bicyclist and distance from home address to accident site.

**Regression Model Results**

The following independent variables are tested for each of the 4 models:

- Average household size
- Road density
- House density
- Average income
- Percent under poverty level
- Percent male
- Density of employees
- Number of schools
- Number of universities
- Percent of individuals ages 1-17
- Percent of individuals ages 18-29
- Percent of individuals ages 30-64
- Percent of individuals ages 65 and over
- Percentage of A2, A3 and A4 roads.
Four regression models of accident rates are presented here. Accidents rates are for adults or children bicycling victims, and based on their home location or the site of the accident. A stepwise approach to the linear regression analysis is chosen to select significant variables and minimize multicollinearity problems.

**Adults – Accident Location**

The location of adult accident locations revealed a large influence of physical road characteristics. Percentage of A3/A3 intersections is the most significant variable ($t$-value = 5.661). While the percentage of class A3 roads has a negative coefficient in the model (-0.580), it has a positive Pearson’s Correlation with the dependent variable (0.531). This mismatch with the model can be explained by the expected strong correlation between A3 roads and A3/A3 intersections (0.875). The percentages of both A3/A4 and A6/A6 intersections have a negative significant value ($t$-value = -1.899, -0.314) and are also negatively correlated with the rate of adult accidents.

From the socio-demographic variables, only percent male ($t$-value = 3.387) and percent ages 65 and up ($t$-value = 0.265) are significant. The overall model has an $R$-squared value of 0.771. Results of the spatial diagnostics tests in GeoDa indicate no problem of spatial dependencies with a Lagrange Multiplier (LM) significance of 0.169 for the spatial lag model and 0.484 for the spatial error model. Based on these findings, the Ordinary Least Squares (OLS) model can be kept.

**Children – Accident Location**

The regression model for child accidents is a much poorer fit than the adult model signifying a difficulty in fitting children’s behavior to a mathematical model. The only significant variable is the percentage of A3/A3 intersections ($t$-value = 3.586) giving the overall model an $R$-squared value of 0.226. This model has no significant values in the spatial dependency tests (LM-Lag = 0.812, LM-Error = 0.673), so the OLS model remains.

**Adults – Home Location**

The location of adults involved in bicycle accidents is best explained by the percentage of class A2/A2 intersections ($t$-value = 7.993) and the percentage of individuals ages 30 to 64 ($t$-value = 5.307). The percentage of people ages 65 and older has a negative coefficient ($t$-value = -2.393) and is significantly negatively correlated with the dependent variable. This is a change from the model of adult accident rates based on accident locations, where this age group was positively associated. The strong presence of the 30-64 age groups describes the adult population most likely to use bicycling as a transportation mode. The location close to larger intersections may indicate a home within a closer physical location of amenities such as employment or shopping thereby
encouraging more bicycle use. Overall, the model has an $R$-squared value of 0.829, explaining a large portion of the adult bicyclists’ characteristics.

Spatial dependency diagnostics reveal no significant values with a LM-lag significance of 0.47 and a LM-error value of 0.15.

**Children – Home Location**

The literature reviewed suggests a strong relationship with socio-economic variables and children involved in traffic accidents. This model identifies housing density as a positive predictor ($t$-value = 0.534) which can be interpreted as a measure of lower economic status, especially in a city such as Buffalo where there are no dense locations of very expensive or upper-class residences. This variable is a reflection of neighborhoods in which children are bicycling as they are more likely to ride close to their own home. The presence of schools is also a significant positive variable in this model ($t$-value = 0.338). As schools are primarily located in the proximity of children’s home, this statistic is not surprising; it mirrors the presence of the population at risk. Road Density is the third significant variable in the model showing a negative relationship ($t$-value = -0.430). The correlation matrix reveals a slightly negative and insignificant correlation between road density and the dependent variable (-0.013) and a strong positive correlation with housing density (0.408); therefore this coefficient may be partly a result of multicollinearity. The $R$-squared value for this mode is 0.334 and no spatial dependencies are found (LM-lag = 0.015, LM-error = 0.515).

**Discussion and Findings**

This study determined evident differences between child and adult bicycle accidents. The under 16 year old class of bicyclists comprised a significant portion of the total events and followed a much less predictable pattern of accidents compared to the adults. While adult accidents illustrated a linear spatial pattern along major arterials in the city, child accidents showed smaller clusters or spots throughout the city and were concentrated inside neighborhoods on local roads.

An analysis of distance between the address of the bicyclist involved in the accident and the crash site demonstrated the hypothesis that younger riders are most likely to be involved in a crash in their own neighborhood rather than en route to a specific destination as is more likely the case with adults. This point is particularly important for any analysis of pedestrian or bicycling accidents; the purpose of the travel activity of these two age categories is so different, it would be inappropriate to group them in a single study.

Four regression models investigated both the physical street and intersection properties as well as social-demographic variables in determining potential risk factors for urban bicyclists. Both child and adult bicycle accident locations were positively associated with the presence of class A3/A3 intersections – or local streets with a 35 mile per hour speed limit. Few socio-demographic variables proved to be influential in the location of these accidents – none for children and just percent male and the age category of 65 and older were positively associated with adult accidents. As males overwhelmed females in the accident data, encompassing 80 percent of the cases, this variable was not
a surprise. The presence of the older age group category has no intuitive explanation. Surprisingly, none of the variables expected to serve as non-motorized trip attractors resulted in a significant value, nor did any of the socio-economic variables. Because the A3 street and intersection variables were so strong, it leads to the conclusion that road infrastructure plays the most important role in determining the physical location of bicycle accidents. Cyclists are at the greatest risk when riding on local city roads where there is the potential for either the bicyclist to take for granted regular traffic laws because the perceived risk may not be great, or drivers may not be prepared for the emergence of a bicyclist on the road. These local roads also tend to be narrower and allow on street parking, two hazardous situations for bicyclists. Further research should focus on ways to improve the safety of these local roads and intersections.

Characteristics of where bicyclists originate have age-specific tendencies. Children tend to come from locations with dense housing and close to schools while adults originate from areas near larger intersections (class A2/A2) and a high concentration of 30-64 year olds. For the children, this information can be used to target traffic safety programs in these neighborhoods.
5. Bibliography


Elizabeth Cahill  
Lecturer  
Department of Geography  
University of Idaho  
McClure Hall, Moscow, ID 83844  
Ph: 208/885-6091; Fax: 208/885-2855; E-Mail: ecahill@uidaho.edu

Jean-Claude Thill  
Knight Distinguished Professor  
Department of Geography & Earth Sciences  
The University of North Carolina at Charlotte  
9201 University City Blvd, Charlotte, NC 28223  
Ph.: 704/687-5909; Fax: 704/687-5966; E-Mail: jfthill@uncc.edu

Eric Delmelle  
Assistant Professor  
Department of Geography  
University of Idaho  
McClure Hall, Moscow, ID 83844  
Ph: 208/885-6452; Fax: 208/885-2855; E-Mail delmelle@uidaho.edu