

Quasi-induced Exposure Method for Pedestrian Safety at Signalized Intersections

Xuecai Xu^{a,b,*}, S.C. Wong^b, Keechoo Choi^c

^aSchool of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, China

^bDepartment of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China

^cDepartment of Transportation Engineering, TOD-based Sustainable Urban Transportation Center, Ajou University, Korea

Abstract:

To investigate pedestrian exposure when only pedestrian-crash data are available, the quasi-induced exposure method is used to identify the factors that contributed to pedestrian-vehicle crashes in Las Vegas from 2004 to 2008. The results show that overall crash severity, light conditions and weather conditions are potentially risky factors in pedestrian crashes, and the time of day or day of the week are less risky. Multiple-correspondence analysis (MCA) is then used to investigate how pedestrian crash severity is influenced by the interactions among a range of variables. The results indicate that among instances of pedestrian-vehicle crashes, cases of property damage only (PDO) tend to occur during off-peak daytime hours and under dark or dawn light conditions in clear weather. Crashes resulting in injuries usually occur during weekday peak hours in daylight and under rainy conditions. Most fatal crashes occur during off-peak hours, at night on the weekends.

Keywords: Pedestrian Crash; Signalized Intersection; Quasi-induced Exposure Method; Multiple-correspondence Analysis

Introduction

During 2012 in the U.S., pedestrian/motor-vehicle crashes resulted in 4,743 deaths and 76,000 injuries. This casualty rate averaged over 13 deaths per day, and the costs involved amounted to over \$5.2 billion for the year (NHTSA, 2012). A large proportion of those killed had been drinking (40%). Road crashes involving pedestrians are a significant problem, particularly in cities with high volumes of tourists such as Las Vegas, where many people travel on foot. On average, 310 pedestrian/motor-vehicle crashes occur each year in Las Vegas, with around 110 of these at signalized intersections. The number of pedestrian injuries did decrease slightly between 2004 and 2008, but more efforts and resources should be allocated to the prevention of pedestrian crashes.

Various studies have been carried out to investigate pedestrian-related safety at signalized intersections, to evaluate the characteristics of pedestrians and to determine the factors that influence pedestrian/vehicle crashes. However, few researchers have examined pedestrian/vehicle-related factors from the perspective of pedestrian exposure, as insufficient data are available.

In recent years, the quasi-induced exposure method, as developed by Carr (1970) and presented by Haight (1970), has received increasing attention. Two significant advantages of this method are its theoretical simplicity and its independence from the data requirements associated with conventional exposure metrics. This method has been used to compare the characteristics of at-fault and not-at-fault drivers involved in road crashes (Martinez-Ruiz et al., 2013) and to explore the effects of various factors potentially associated with an increased risk of road crashes. Although the quasi-induced exposure method has been widely adopted during the last 20 years (Jiang and Lyles, 2010), it has rarely been used to investigate pedestrian safety at signalized intersections. The first aim of this study is, therefore, to use the quasi-induced exposure method to deduce the pedestrian/vehicle-related factors involved in crashes at signalized intersections (instead of using the commonly applied measure of assessing the incidence of at-fault or not-at-fault drivers and of responsible or non-responsible pedestrians). The study's second aim is to use multiple-correspondence analysis (MCA) to test how the interactions among influencing variables affect the responsible pedestrians. This approach to crash analysis can better inform policies aimed at reducing the pedestrian-related crashes.

Literature Review

Pedestrian exposure (defined as the exposure of pedestrians to collisions with motor vehicles) is a significant risk factor for pedestrian travelers. Data on pedestrian exposure or on pedestrian-volume counts are not readily available. Hence, pedestrian exposure has been previously investigated through the use of various pedestrian crash prediction models.

For general roadways, Qin and Ivan (2001) developed a model for predicting pedestrian exposure in rural areas, and they used the volume of pedestrians crossing the street as the measurement metric. The results showed that pedestrian exposure was significantly affected by the number of lanes, the area type and the local sidewalk system. Similar studies have been undertaken by Cameron (1982), Keall (1995) and Chu (2003). These researchers have used various pedestrian-exposure measures, e.g., population, number of roads crossed, time spent walking, and pedestrian volume multiplied by vehicular volume. However, these measures have not been fully accepted by researchers.

For the pedestrian exposure at signalized intersections, Kennedy (2008) estimated pedestrian volumes and the incidence of crashes at urban signalized intersections. In that study, a scaling technique was developed to estimate pedestrian volume, and the number of crashes per 100 miles traveled by pedestrians was taken as the criterion. Compared with previous research (Zegeer et al., 2005; Molino et al., 2008), the results of Kennedy's approach proved to be more reliable than those obtained using the general linear modeling method. Similar other pedestrian-vehicle crash prediction models have been developed using various count-data and regression models. The results have indicated that pedestrian crashes are significantly influenced by both vehicle volume and pedestrian volume, with the vehicle volume being more influential than the pedestrian volume. However, no linear relationship has been found between pedestrian crashes and pedestrian volumes. Recently, Greene-Roesel et al. (2010) and Schneider et al. (2012) provided an overview of pedestrian crash exposure at signalized intersections. The sketch-plan, network-analysis and the micro-simulation models are the most commonly used models for estimating pedestrian exposure.

The above-mentioned research explained the situations when the exposure data are available. However, if the exposure data are unavailable, the quasi-induced exposure method is also widely used to estimate the degrees of exposure and risk for different groups of drivers and vehicles. Researchers who have used the quasi-induced exposure method and had promising results include Lyles (1994), Stamatidis and Deacon (1997), Lardelli-Claret et al. (2006), Keall and Newstead (2009), Mendez and Izquierdo (2010), Jiang and Lyles (2007, 2010) and Martinez-Ruiz et al. (2013).

The quasi-induced exposure method was first proposed by Lyles (1994), who identified and defined four critical issues associated with the method's use. These issues concerned one-vehicle crashes, the validity of various methods for assigning fault to those involved in an crash, the inclusion of innocent victims in random samples of drivers, and the overall validity of the quasi-induced approach. The results of Lyles' research confirmed that this approach is useful for gaining insight into crash involvement. Stamatiadis and Deacon (1997) continued Lyles' (1994) methodological exploration, and concluded that quasi-induced exposure is a highly effective technique for measuring the relative exposure of drivers or vehicles in situations for which real exposure data are not available.

Based on the fundamental theory of quasi-induced exposure, two or more vehicle crashes were investigated. Lardelli-Claret et al. (2006) compared two quasi-induced exposure methods for investigating the risk factors for road crashes: one method for single-vehicle crashes and the other for two-vehicle collisions. Multinomial and logistic regression models were used to obtain the odds ratios for each driver- and vehicle-related variable. The results showed that the effects of most driver- and vehicle-related characteristics on single-vehicle crashes were greater than their effects on two-vehicle collisions. Jiang and Lyles (2010) extended this approach to crashes involving three or more vehicles, and demonstrated that all of the non-responsible-driver distributions remained similar in both operational and statistical terms. In a novel application of the quasi-induced exposure method, Martinez-Ruiz et al. (2013) extended the risk factors for road crashes into cyclists, with specific attention to collisions between bicycles and other vehicles on Spanish roads. Their results confirmed that the influence of cyclist-related factors differed depending on the type of crash, but the factors related to the other drivers involved in collisions with cyclists were similar to those observed in other types of crashes.

However, there were also criticisms about quasi-induced exposure method about quasi-induced exposure method. Keall and Newstead (2009) determined the crash type that best represented exposure on the road, and identified the situations in which induced-exposure risk estimates were likely to be biased. In a similar study, Mendez and Izquierdo (2010) explored the direction of bias due to speed, crash avoidability, and injury risk for four types of crashes. These investigators found that more research was required to determine the effects of speed on head-on crashes occurring on undivided roads and crashes happening on multi-lane roads.

Besides the methods about pedestrian risks, other methods about pedestrian injury, such as on-site investigation, mathematical modeling and simulation have been used to evaluate the levels of severity. There are two commonly used modeling approaches for evaluating the severity of injuries at signalized intersections. One method is the ordered response framework, and the other method is the unordered response framework (Abay 2013). Previous studies, such as those of Abdel-Aty (2003), Eluru and Bhat (2007), Sze and Wong (2007), Wang and Abdel-Aty (2008), Eluru et al.

(2008), Clifton et al. (2009), Kwigizile et al. (2011), Mohamed et al. (2013), Abay (2013) or Saidharan and Menendez (2014) have used ordered response models for analysis of injury severity, with an emphasis on enabling pedestrian safety. The unordered response models (of the multinomial, nested and mixed logit/probit types) have also been widely used for evaluating injury severity (Carson and Mannering, 2001; Ulfarsson and Mannering, 2004; Kim et al., 2008; Tay et al., 2011; Abay, 2013; Aziz et al., 2013; Saidharan and Menendez, 2014). However, these ordered and unordered response models each involve different assumptions and restrictions in the analysis of injury severity, and these different approaches contradict each other. Therefore, more suitable methods need to be proposed. Table 1 gives a summary of all the studies reviewed. The main focus of this study is to integrate the quasi-induced exposure method with MCA to better analyze the severity of pedestrian injuries.

Table 1 Summary of Literature Review

Methodology

The quasi-induced exposure method was used to analyze a crash database and investigate pedestrian exposure. Generally, researchers taking a quasi-induced exposure approach have used driver-vehicle crash data, and have ensured validity by comparing distributions of the characteristics (usually driver age, gender and vehicle type) of the non-responsible drivers involved in driver-vehicle crashes.

In this study, pedestrian-vehicle crash data replaced driver-vehicle crash data for representing pedestrian exposure. The initial study population was based on the crash database from the Nevada Department of Transportation (NDOT), which provided both driver-vehicle crash data and pedestrian-vehicle crash data, so that there was little need to collect additional data. Arc GIS was used to isolate the pedestrian-vehicle crashes from the geo-database and to identify appropriate samples of responsible and non-responsible pedestrians for use in the quasi-induced exposure framework.

As Lyles (1994) observed, the fundamental assumptions of the quasi-induced exposure method are that vehicle-vehicle crashes involve an at-fault driver and a not-at-fault driver, and that the dataset of not-at-fault drivers consists of a random sample of motorists and vehicles on the road. It is therefore essential that the available crash data distinguish between “at-fault” and “not-at-fault” drivers. Here, the “at-fault” driver is defined as the driver who causes the crash, and the “not-at-fault” driver is the driver who does not cause the crash. Similarly, it is assumed that pedestrian-vehicle crashes involve responsible and non-responsible pedestrians, and that the distribution of the non-responsible pedestrians involved constitutes a random sample of the population of crashes happening in signalized intersections at specific times and places.

The quasi-induced methodology uses the day of the week and a number of other factors to be described below. Table 2 gives the matrix showing day of the week for both the responsible and non-responsible pedestrians involved in signalized intersection crashes in Las Vegas during 2004. Responsibility is assigned in each pedestrian-vehicle crash. A pedestrian-vehicle combination No.1 (or P1) designates a crash in which an injured pedestrian is considered to be at-fault, and a pedestrian-vehicle combination No.2 (or P2) designates an crash where an injured pedestrian is regarded as not at fault (or as non-responsible). It is assumed that P2s are selected randomly from the whole system. The P2s constitute a random sample of pedestrian-vehicle combinations, and are inductively a measure of exposure. Therefore, the involvement ratio (IR) can be calculated from the P1-P2 matrix as the ratio of the marginal proportions of P1s to P2sas follows:

$$IR = \frac{\text{marginal proportion of P1}}{\text{marginal proportion of P2}}$$

The main point here is that P2 characteristics are measures of exposure. Table 2 shows that the ratio of P1 to P2 characteristics gives a measure of the relative involvement of each characteristic in the causes of crashes at signalized intersections. For example, pedestrian crashes on weekdays are slightly over-represented, with the IR on weekdays being 1.01, but pedestrian crashes on weekends are slightly under-represented, with the IR on weekends being 0.98. In other words, there were fewer pedestrian crashes on weekends. The key point is that responsibility is adjusted for exposure or opportunity (i.e., the proportion of weekday pedestrians under a specified set of conditions). If IR (weekday) =1.0, then the pedestrians on weekdays cause crashes in proportion to their presence on the roadway. More details and examples can be found in Lyles (1994).

Table 2P1-P2 Matrix for Signalized Intersection Crashes in Las Vegas (2004)

In terms of our statistical analysis, a significance level of 0.05 is taken as the threshold, and a two-tailed *t*-test is used to examine whether the difference between responsible and non-responsible pedestrians is statistically significant.

To test how the interactions among all the influencing variables affect the responsible pedestrians, a MCA for the multivariate categorical data is applied. Correspondence analysis (CA) and MCA were developed in 1970 from the factor-analysis method of the French statistician. These methods of analysis have been widely used in the social sciences (LeRoux and Rouanet, 2010) and can be performed using SPSS11.0 software. CA is a metric multidimensional scaling method used to ascertain the presence of relationships between categorical variables and to determine how those variables are related. The basic objective of CA is to simplify the structures of data by reducing their dimensionality. The rows and columns in the contingency table are processed to express the relationships between variables in the form of a lower-dimensional graph.

If one category of a variable is longer than another in one direction, then the difference between these categories is significant in this dimension. Otherwise, the difference is not significant.

MCA is an extension of CA that is used to analyze the pattern of relationships between three or more categorical variables. This kind of analysis is characterized by similar graphical displays in which either the categories of the variables or the individual cases are represented as points. Standard corresponding analysis is applied to an indicator or disjunctive matrix that organizes cases as rows, and categories of variables as columns (Abdi and Valentin, 2007). The value of MCA is that the greater the number of categorical variables, the more obvious the advantages from this method. This quality of MCA suits the characteristics of and the interactions between the variables investigated in this study.

Assume that there are K nominal variables, that each nominal variable has J_k levels, and that the sum of the J_k levels is equal to J . Suppose that there are I observations, and the $I \times J$ indicator matrix is denoted \mathbf{X} , so that CA can be performed on the indicator matrix. The CA provides two sets of factor scores: one score for the rows and the other score for the columns. These factor scores are scaled so that their variance is equal to their corresponding eigenvalues (meaning that the row factor scores are normalized to unity).

Let N denote the grand total of elements in the table, so that the first step of the MCA is to calculate the probability matrix $\mathbf{Z} = \mathbf{X}/N$. Here, \mathbf{r} is denoted as the vector of the row totals of \mathbf{Z} , i.e., $\mathbf{r} = \mathbf{Z}\mathbf{1}$, in which $\mathbf{1}$ is a conformable vector of 1's, and \mathbf{c} is denoted as the vector of the column totals. Thus, $\mathbf{c} = \text{disg}\{\mathbf{c}\}$, and $\mathbf{D}_r = \text{diag}\{\mathbf{r}\}$.

The factor scores can be obtained from the following singular value decomposition:

$$\mathbf{D}_r^{-\frac{1}{2}}(\mathbf{Z} - \mathbf{rc}^T)\mathbf{D}_c^{-\frac{1}{2}} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^T \quad (1)$$

Where $\mathbf{\Delta}$ is the diagonal matrix of the singular values, and $\mathbf{\Lambda} = \mathbf{\Delta}^2\mathbf{A}$ is the matrix of the eigenvalues. The row and columns factor scores can be respectively achieved as follows:

$$\mathbf{F} = \mathbf{D}_r^{-\frac{1}{2}}\mathbf{P}\mathbf{\Lambda}, \text{ and } \mathbf{G} = \mathbf{D}_c^{-\frac{1}{2}}\mathbf{Q}\mathbf{\Lambda} \quad (2)$$

The squared χ^2 distances from the rows and columns to their respective barycenters are obtained as

$$d_r = \text{diag}\{\mathbf{F}\mathbf{F}^T\}, \text{ and } d_c = \text{diag}\{\mathbf{G}\mathbf{G}^T\} \quad (3)$$

Then, to help locate the important factors for a given observation or variable, the

squared cosines between row i and factor l , and between column j and factor l can be obtained as follows:

$$o_{i,l} = \frac{r_{i,l}^2}{d_{r,i}^2}, \text{ and } o_{j,l} = \frac{c_{j,l}^2}{d_{c,j}^2} \quad (4)$$

where $d_{r,i}^2$ and $d_{c,j}^2$ are, respectively, the i -th element of \mathbf{d}_r and the j -th element of \mathbf{d}_c .

Similarly, to help locate the important observations or variables for a given factor, the contribution of row i to factor l and of column j to factor l are respectively obtained as

$$t_{i,l} = \frac{r_{i,l} c_l}{\lambda_l}, \text{ and } t_{j,l} = \frac{c_{j,l} c_l}{\lambda_l} \quad (5)$$

where r_i and c_j are elements of \mathbf{r} and \mathbf{c} , respectively.

Supplementary or illustrative elements can then be projected onto the factors.

Specifically, let \mathbf{l}_{sup}^F be an illustrative row and \mathbf{j}_{sup}^F be an illustrative column to be

projected, so that the coordinates f_{sup} and g_{sup} are obtained as follows:

$$f_{sup} = (\mathbf{l}_{sup}^F \mathbf{1})^{-1} \mathbf{l}_{sup}^F \mathbf{G} \Delta^{-1} \text{ and } g_{sup} = (\mathbf{j}_{sup}^F \mathbf{1})^{-1} \mathbf{j}_{sup}^F \mathbf{F} \Delta^{-1} \quad (6)$$

where the terms $(\mathbf{l}_{sup}^F \mathbf{1})^{-1}$ and $(\mathbf{j}_{sup}^F \mathbf{1})^{-1}$ are used to ensure that the sum of the elements of \mathbf{l}_{sup} or \mathbf{j}_{sup} is equal to one. If this is already the case, then these terms are superfluous.

Next, CA can be performed on the indicator matrix, which provides factor scores for the rows and columns, and MCA can then be re-scaled with the factor scores given.

By producing several binary columns for each variable, MCA codes the data with the constraint that one and only one of the columns reaches the value 1. In this way, artificial additional dimensions are created, as one categorical variable is coded with several columns. Consequently, the inertia (i.e., variance) of the solution space is artificially inflated, and therefore the percentage of inertia explained by the first dimension is severely underestimated. In fact, it can be shown that all of the factors having eigenvalues less than or equal to $1/K$ simply code these additional dimensions. Specifically, if λ_l is denoted as the eigenvalues obtained from the analysis of the indicator matrix, then the corrected eigenvalues of c^l can be computed as

$$c^A = \begin{cases} \left[\left(\frac{R}{R-1} \right) \left(\lambda_i - \frac{1}{R} \right) \right]^2 & \text{if } \lambda_i > \frac{1}{R} \\ 0 & \text{if } \lambda_i \leq \frac{1}{R} \end{cases} \quad (7)$$

From this formula, a better estimate of the inertia can be given and extracted by each eigenvalue. An average inertia can be calculated as

$$\bar{\psi} = \frac{R}{R-1} \times \left(\sum_i \lambda_i^2 - \frac{I-R}{R} \right)^2 \quad (8)$$

A detailed mathematical treatment of this approach can be found in Greenacre (1984), Hoffman and Franke (1986) and in Abdi and Valentin (2007).

Data Description

Appropriate crash data for this study (from 2004 to 2008) were obtained from the geo-database maintained by NDOT. In line with the assumptions of the quasi-induced exposure method, 450 signalized intersections (shown in Figure 1) were selected, and 77,250 crashes were surveyed from the Las Vegas metropolitan area, along with information on crash features (including day, hour, month, location, severity and crash types). Data were also included on traffic volume, vehicle features (vehicle types, vehicle direction, vehicle factors, vehicle driver conditions and vehicle action) and on non-behavioral factors (weather, light, or work zones) for each intersection. The reason that the chosen dataset was selected was, first, that pedestrians were harmed. This fact helped to isolate the pedestrian-vehicle crashes and distinguish between data on responsible and non-responsible pedestrians. For the responsible data, there were 550 pedestrian crashes recorded by police reports and maintained by NDOT from 2004 to 2008, which are official and reliable for the analysis. After the pedestrian-related crashes were extracted, the non-responsible data 76,700 crashes could be assumed to be representative.

Figure 1. Selected Signalized Intersections in Las Vegas.

Results

The distributions of all the selected crashes between vehicles and responsible or non-responsible pedestrians are shown in Table 3. The relative exposure to pedestrian crash severities (from PDO, to injury and to fatality) increased from IR 0.77 to IR 1.24 and to IR 17.67, respectively. In terms of the row distribution, only crashes causing injury were relatively similar among the responsible and non-responsible pedestrian-vehicle crashes. Among the crashes involving non-responsible pedestrians, those causing PDO or fatalities had higher and much lower IRs, respectively.

Table 3. Distribution for Signalized Intersection Crashes in Las Vegas (2004-2008)

It seems reasonable to expect that pedestrian-vehicle exposure should remain stable

(IR=0.97) at peak hours, whereas exposure during off-peak hours should vary. According to the data, off-peak daytime hours were over-involved in crashes, and off-peak nighttime hours were under-involved. This pattern was consistent over the five years' worth of data examined.

As the distribution of crashes might be expected to differ according to the day of the week, this variable was also examined. Data analysis showed that the IRs were very close to 1, which suggested that pedestrian exposure on weekdays and weekends was similar in the case of responsible-pedestrian crashes.

A comparison of different lighting conditions over the five years showed that most of the pedestrian crashes occurred during daylight or darkness, and that more than 60% of the pedestrian-related crashes happened in day light. It thus seems reasonable to expect that the exposures at dawn and dusk and for unreported/unknown conditions were close to 1, because they represented small percentages. Exposure was lower during the daylight hours (IR=0.85) and higher in the hours of darkness (IR=1.39).

The distributions were also examined with reference to changing weather conditions. The incidence of non-responsible- and responsible-pedestrian crashes showed considerable similarity across all of the weather conditions, and the differences between the distributions were very close (IR=1). The results for each year were similar.

Statistical tests were run to ensure a full quantitative assessment on the validity of the measures of exposure. For the numerical row percentages, a *t*-test was performed on the differences between the percentages of responsible and non-responsible pedestrians. The mean difference between these groups was 0.001, which was not statistically significant (*t*-value=0.012, *df* =32, two-tailed probability=0.99), implying that there was no significant difference between the percentages of responsible and non-responsible pedestrians.

A one-way analysis of variance was conducted with the actual row proportions (e.g., the percentage of non-responsible pedestrians), because the dependent variables showed that time of day and day of the week did not have any significant influence on pedestrian-related crashes, whereas the effects of overall severity, light, and weather conditions were significant.

After the quasi-induced exposure method confirmed the relative involvement of pedestrian-related variables, MCA was used to test the interactions between the explanatory variables for responsible-pedestrian crashes, so that potential measures could be identified for reducing the pedestrian-related crashes. Table 4 provides the iteration history and shows that the process stopped at the 20th iteration, at which point the convergence test value was reached.

Table 4. Iteration History

Figure 2 displays the interaction results for the principal normalization of all the variables after optimal scaling. The two dimensions together provide an interpretation in terms of distance. If a variable discriminates well, the objects will be close to the categories to which they belong. Ideally, objects in the same category will be close to each other, but categories of different variables will be close only if they belong to the same objects (SPSS Categories 17.0). In Figure 2, dimensions 1 and 2 represent the rows and columns of the MCA, respectively. Severity has three categories: “PDO” and “injury” are close to each other, and “fatal” is very distant. Similarly, for the time of day, the category of peak hours is close to dimension 2; the category of off-peak nighttime hours is close to dimension 1, and only the category of off-peak daytime hours is far from both. Among the five categories of light conditions, “dusk” is near the bottom of the plot; “dawn” is far from the group, and both “daylight” and “dark” are almost the same distance from dimension 1. Similarly, for the weather-condition categories, “rain” is very far from the other categories, and yields a large discrimination measure along dimension 1.

Figure 2. Joint Plot of Category Quantification

In Figure 2, the large discrimination between fatal crashes and PDO/injury crashes is a result of the considerable difference between these categories. Time of day shows a similar difference, as the category of off-peak daytime hours is far from both dimensions, and is substantially different from the other two categories. Among the five categories of light, “dawn” and “dusk,” the greatest difference is between “daylight” and “dark.” In terms of weather conditions, the category of “rain” differs from the others.

The spread of the category quantifications for each variable reflects the variance within that variable, and thus indicates how well each variable is discriminated within each dimension. Among the five variables, almost all of the categories are spread far apart along both dimensions, implying that the five variables discriminate well in both dimensions.

The category-quantification plot provides variable interaction discrimination for determining the dimensions along which a variable discriminates and how that variable discriminates. Figure 2 shows that crashes involving PDO usually occur during off-peak daytime hours, or under dark and dawn light conditions in clear weather. Most crashes causing injury happen during weekday peak hours in daylight or in rainy-weather conditions. Most fatal crashes take place during off-peak nighttime hours on the weekends. These results explain a large number of the phenomena associated with pedestrian-vehicle crashes.

To verify the relationships between the variables, a table of correlations (Table 5) is

constructed. It is clear that an important relationship exists between light conditions and time of day, but the other relationships are less strong.

Table 5. Correlations between Transformed Variables

For each variable in each dimension, the discrimination measure is computed. The measure represents the variance of the quantified variable in that dimension, and has a maximum value of 1. A large discrimination measure corresponds to a large spread in the categories of the variable, which consequently indicates a high degree of discrimination between the categories of the variable in that dimension. The average of the discrimination measures for any dimension equals the percentage of variance calculated in that dimension (SPSS Categories 17.0). The dimensions are ordered according to average discrimination. The first dimension has the largest average discrimination; the second dimension has the second largest average discrimination, and so on.

The discrimination-measures plot in Figure 3 shows that no variable has a large discrimination measure in the first dimension, because almost all of the variables are spread evenly. Severity has a large value in the second dimension, but a small value in the first, and is thus close to the second dimension. Time of day and light conditions have relatively large values in both dimensions, and the values for weather and day of week (although located very close to the origin) discriminate less in the first two dimensions than the values for time of day and light conditions. The weather and day of the week variables are both close to the origin, as obviously the weather in Las Vegas is predominantly clear, and days of the week are counted as either a weekday or as one of the two weekend days.

Figure 3. Discrimination Measures

Discussion

This study presents a quasi-induced method for evaluating pedestrian safety at signalized intersections in a situation where pedestrian exposure data are not available. The study has two primary goals: 1) to describe the methodology and realize the quasi-induced method in pedestrian safety analysis and 2) to examine the significant interactions of variables through multiple correspondence analysis. Both of these goals have been accomplished.

The outcomes of these goals are useful for the improvement of pedestrian safety at signalized intersections. The study demonstrates that the proposed quasi-induced method can be used to evaluate the potential risks of pedestrian-related crashes at signalized intersections. The main advantage of this method is that it can be used to

analyze the influencing variables when pedestrian exposure data are not available.

This method has one weakness that should be considered. The model is appropriate to evaluate pedestrian-related variables when data on the numbers of pedestrian-related crashes are available. However, the lack of data concerning pedestrian attributes and other characteristics of roadways and vehicles (such as geometric design, signal phase or vehicle attributes) may cause under- or over-estimated results, and these potential weaknesses may need to be addressed in further applications.

MCA was selected for this study because this method can handle interactions between three or more categorical variables and produce a simplified matrix displayed in rows and columns. MCA can also generate important risk factor combinations. Crashes involving PDO were found to most commonly occur during off-peak daytime hours, in the dark, or at dawn in clear weather. This finding implies that more attention should be paid to PDO crashes during off-peak hours, and that in poor light the drivers' ability to see may be affected even in clear weather, necessitating the use of headlights to avoid the risk of collisions.

It is understandable that crashes resulting in injuries occur most often during peak weekday hours, when more pedestrians are on the streets. Weekday commuting continues whatever the weather conditions, and injuries resulting from crashes are probably more severe under rainy conditions. Measures can therefore be taken to alert pedestrians to the traffic conditions during peak weekday hours, such as displaying signs or electronic screens, especially during periods of bad weather.

Most fatal crashes take place during off-peak nighttime hours on weekends. People tend to go out to relax or shop until late into the night. There is more risk of crashes at these times, when the pedestrians are less aware of their surroundings and when both pedestrians and drivers can be tired. Basic reminders (for example, a simple message flashing on the pedestrian signal light pole) should be displayed for both pedestrians and drivers, and drivers should be instructed to take particular care regarding pedestrians during nighttime hours on weekends.

The main contribution of MCA in this study is that it can deal with interactions among all the categorical variables easily and reduce the complicated interactions into a two-dimensional graph, which is helpful to explain the relationships among the variables and guide the decision-making direction. However, one limitation of MCA is that it is not easy for practitioners without professional knowledge to discern each variable among the figures in the rows and columns and to interpret them clearly. Practitioners should therefore be aware of concerning the background information on MCA before any general conclusions are drawn about pedestrian safety.

Conclusions

A quasi-induced method was applied to pedestrian-vehicle crash data to identify the influencing factors in situations where only pedestrian crash data are available. The results showed that overall crash severity, light conditions, and weather conditions were potentially risky factors in the pedestrian crashes examined, whereas the time of day and the day of the week were less risky. To investigate the interactions among all of the variables, MCA was conducted. The results showed that for pedestrian-vehicle crashes, those causing PDO tended to occur during off-peak daytime hours or under dark and dawn light conditions in clear weather. Crashes causing injury usually happened during peak weekday hours, and under daylight or rainy weather conditions. Most fatal crashes took place during off-peak nighttime hours on the weekends.

Two key findings can be drawn from the results of the study. First, the quasi-induced exposure method can be used to derive the potential pedestrian/vehicle-related factors affecting safety at signalized intersections. Second, MCA reveals the relationships between pedestrian crash severity and time of day, day of week, light conditions, and weather conditions, all of which should be taken into account in policies that aim to reduce the pedestrian-related crashes. However, one deficiency of this study is that the IR in the quasi-induced exposure method was not utilized fully in the MCA method, and the means of integrating these two approaches organically needs to be investigated further.

The results of this analysis provide potential insights for practitioners and policy makers concerning pedestrian safety at signalized intersections. Planners may need to improve pedestrianization design to accommodate the walking behavior of pedestrians. Safety officials need to provide education programs and take pedestrianization measures for reducing the risks to pedestrians. In terms of empirical concerns, practitioners and policy makers can take various efficient, coherent and suitable measures to improve pedestrian safety at signalized intersections. Such measures can include using different light intensities during different time periods and various weather conditions, or setting up electronic signs during the peak periods and bad weather conditions.

Acknowledgements

The authors are grateful to the Nevada Department of Transportation for providing the database used in this research, and they thank Professors Harry Teng and Mohamed Kaseko, from the University of Nevada at Las Vegas, for their valuable suggestions. This study was jointly supported by the Fundamental Research Fund for the Central Universities (HUST: 2013QN031), the National Natural Science Foundation of China (NSFC) (No: 51208222), the Scientific Research Foundation for Returned Overseas Chinese Scholars, the State Education Ministry of China and the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 717512).

References

- Abay, K. A., 2013. Examining pedestrian-injury severity using alternative disaggregated models. *Research in Transportation Economics* 43(1), 123-136.
- Abdel-Aty, M.A., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research* 34(5), 597-603.
- Abdi, H., Valentin, D., 2007. Multiple correspondence analysis. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage, pp. 657-663.
- Aziz, H.M.A., Ukkusuri, S.V., Hasan, S., 2013. Exploring the determinants of pedestrian-vehicle crash severity in New York City. *Crash Analysis and Prevention* 50, 1298-1309.
- Cameron, M.H., 1982. A method of measuring exposure to pedestrian crash risk. *Crash Analysis and Prevention* 14(5), 397-405.
- Carr, B. R., 1970. A statistical analysis of rural Ontario traffic crashes using induced exposure data. *Proceedings of the Symposium on the Use of Statistical Methods in the Analysis of Road Crashes*. Paris, France: OECD, pp. 86-72.
- Carson, J., Mannering, F., 2001. The effect of ice warning signs on crash frequencies and severities. *Crash Analysis and Prevention* 33(1), 99-109.
- Chu, X., 2003. The fatality risk of walking in America: A time-based comparative approach. *Walk21 Conference: Health, Equity and the Environment*. Portland, Oregon, February, 2003.
- Clifton, K. J., Burnier, C.V., Akar, G., 2009. Severity of injury resulting from pedestrian-vehicle crashes: What can we learn from examining the built environment? *Transportation Research Part D* 14(6), 425-436.
- .
- Eluru, N., Bhat, C., 2007. A joint econometric analysis of seatbelt use and crash-related injury severity. *Crash Analysis and Prevention* 39(5), 1037-1049.
- Eluru, N., Bhat, C., Hensher, D., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Crash Analysis and Prevention* 40(3), 1033-1054.
- Greenacre, M.J., 1984. *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Greene-Roesel, R., Diogenes, M.C., Ragland, D.R., 2010. Estimating pedestrian crash exposure, California PATH Research Report: UCB-ITS-PRR-2010-32. Safe Transportation Education and Research Center, California Partners for Advanced Transit and Highways, Berkeley, CA.
- Haight, F. A., 1970. A crude framework for bypassing exposure. *Journal of Safety Research* 2(1), 26-29.
- Hoffman, D. L., Franke, G. R., 1986. Correspondence analysis: Graphical representation of categorical data in marketing research. *Journal of Marketing Research* 23, 213-227.
- Jiang, X., Lyles, R.W., 2007. Difficulties with quasi-induced exposure when speed varies systematically by vehicle type. *Crash Analysis and Prevention* 39(4), 649-656.

- Jiang, X., Lyles, R.W., 2010. A review of the validity of the underlying assumptions of quasi-induced exposure. *Crash Analysis and Prevention* 42(4), 1352-1358.
- Keall, M.D., 1995. Pedestrian exposure to risk of road crash in New Zealand. *Crash Analysis and Prevention* 27(3), 729-740.
- Keall, M., Newstead, S., 2009. Selection of comparison crash types for quasi-induced exposure risk estimation. *Traffic Injury Prevention* 10(1), 23-29.
- Kennedy, J.F., 2008. Estimating pedestrian volumes and crashes at urban signalized intersections. Thesis: Virginia Polytechnic Institute and State University, Falls Church, VA.
- Kim, J. K., Ulfarsson, G. F., Shankar, V. N., Kim, S., 2008. Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Crash Analysis and Prevention* 40(5), 1695-1702.
- Kwigizile, V., Sando, T. Chimba, D., 2011. Inconsistencies of ordered and unordered probability models for pedestrian injury severity. *Transportation Research Record* 2264, 110-118.
- Lardelli-Claret, P., Jimenez-Moleon, J.J., Luna-del-Castillo, J.D., Garcia-Martin, M., Moreno-Abril, O., Bueno-Cavanillas, A., 2006. Comparison between two quasi-induced exposure methods for studying risk factors for road crashes. *American Journal of Epidemiology* 163, 188-195.
- LeRoux, B., Rouanet, H., 2010. *Multiple Correspondence Analysis: Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage.
- Lyles, R., 1994. Quasi-induced exposure: To use or not to use? CDROM, 74th Annual Meeting of the Transportation Research Board, Washington D.C., January 1994.
- Martinez-Ruiz, V., Lardelli-Claret, P., Jimenez-Mejias, E., Amezcua-Prieto, C., Jimenez-Moleon, J.J., Castillo, J.D.L., 2013. Risk factors for causing road crashes involving cyclists: An application of a quasi-induced exposure method. *Crash Analysis and Prevention* 51(3), 228-237.
- Mendez, A.G., Izquierdo, F.A., 2010. Quasi-induced exposure: The choice of exposure metrics. *Crash Analysis and Prevention* 42(2), 582-588.
- Mohamed, M.G., Saunier, N., Miranda-Moreno, L.F., Ukkusuri, S.V., 2013. A clustering regression approach: A comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US, and Montreal, Canada. *Safety Science* 54, 27-37.
- Molino, J., Kennedy, J., Johnson, P., Beuse, P., Emo, A., Do, A., 2008. Pedestrian and bicyclist exposure to risk: A methodology for estimation in an urban environment. CDROM, 88th Annual Meeting of the Transportation Research Board, Washington D.C., January 2008 .
- National Highway Traffic Safety Administration (NHTSA), 2012. Online website: <http://www.nhtsa.gov/Pedestrians> .Accessed on Nov.24th, 2014.
- Qin, X., Ivan, J.N., 2001. Estimating pedestrian exposure prediction model in rural areas. *Transportation Research Record* 1773, 89-96.
- Sasidharan, L., Menendez, M., 2014. Partial proportional odds model — An alternate choice for analyzing pedestrian crash injury severities. *Crash Analysis and Prevention* 72, 330-340.

- Schneider, R.J., Henry, T., Mitman, M.F., Stonehill, L., Koehler, J., 2012. Development and application of the San Francisco pedestrian intersection volume model. CDROM, 91st Annual Meeting of the Transportation Research Board, Washington D.C., January 2012 .
- Stamatiadis, N., Deacon, J.A., 1997. Quasi-induced exposure: Methodology and insight. *Crash Analysis and Prevention*, 29(1),37-52.
- Sze N.N, Wong S.C., 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Crash Analysis and Prevention* 39(6), 1267-1278.
- Tay, R., Choi, J., Kattan, L., Khan, A., 2011. A multinomial logit model of pedestrian-vehicle crash severity. *International Journal of Sustainable Transportation* 5, 233-249.
- Ulfarsson, G.F., Mannering, F.L., 2004. Difference in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car crashes. *Crash Analysis and Prevention* 36(2), 135-147.
- Wang, X., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Crash Analysis and Prevention* 40(8), 1674-1682.
- Zegeer, C., Stewart, R., Huang, H., Lagerwey, P., Feganes, J., Campbell, B., 2005. Safety Effects of Marked versus Unmarked Crosswalks at Uncontrolled Locations. FHWA-HRT-04-100. FHWA, U.S. Department of Transportation, McLean, VA.

FIGURE CAPTIONS LIST

Figure 1 Selected Signalized Intersections in Las Vegas

Figure 2 Joint Plot of Category Quantification

Figure 3 Discrimination Measures

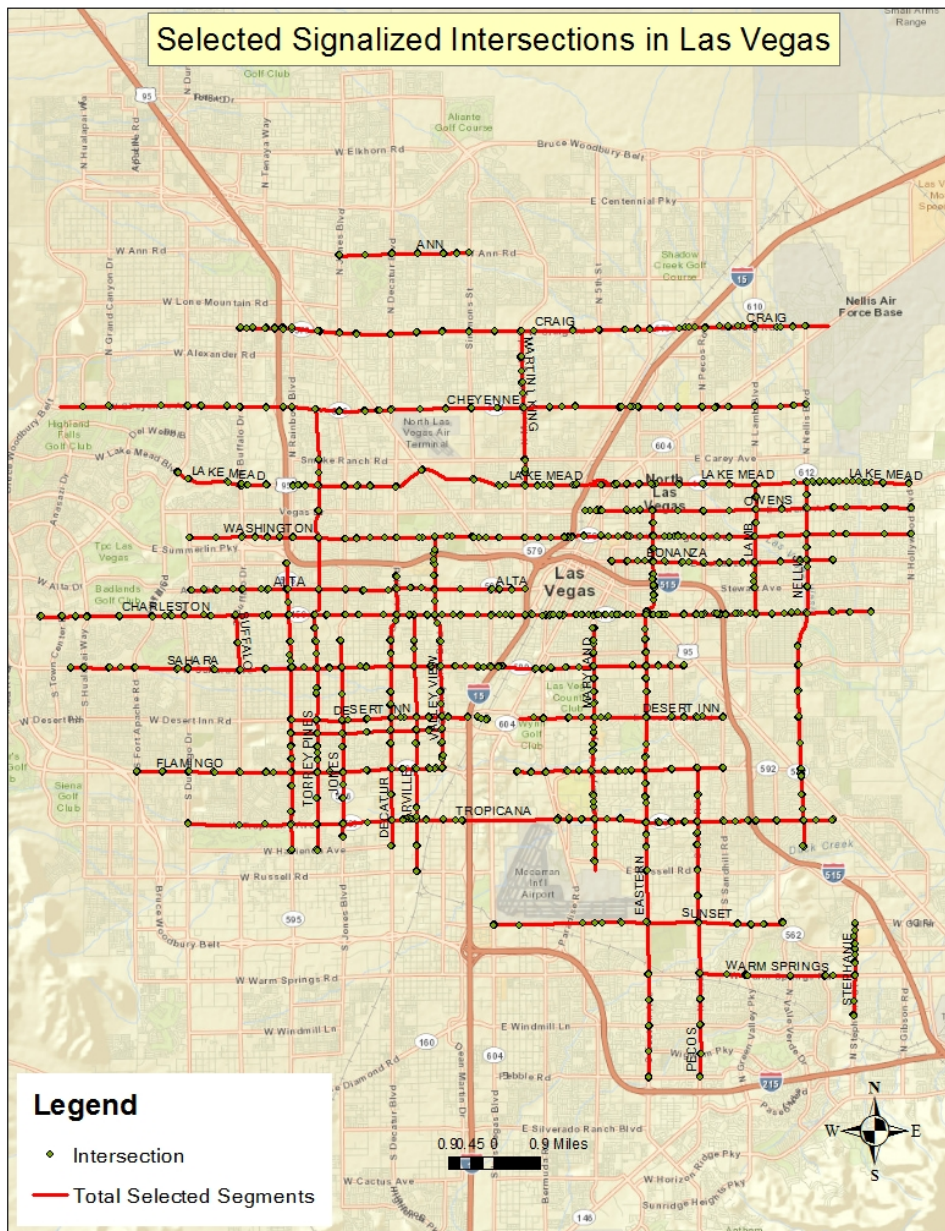


Figure 1. Selected Signalized Intersections in Las Vegas. (450 signalized intersections were selected for analysis, from the City of Las Vegas, the City of North Las Vegas, the City of Henderson, and Clark County.)

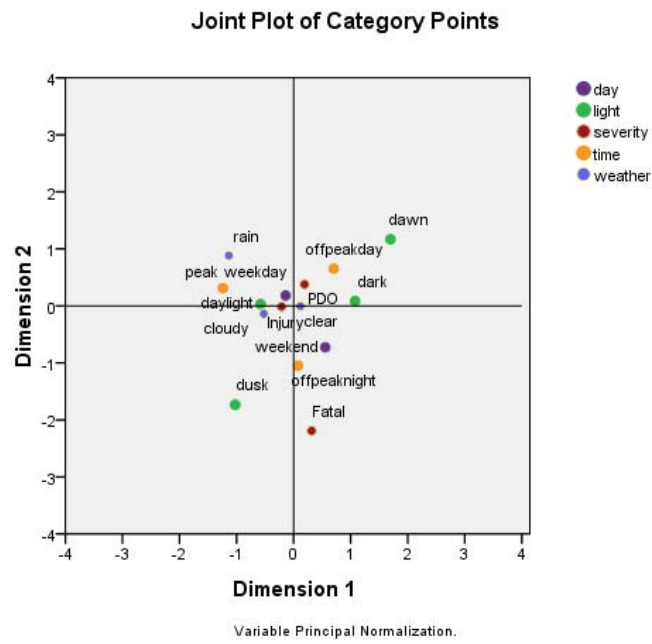
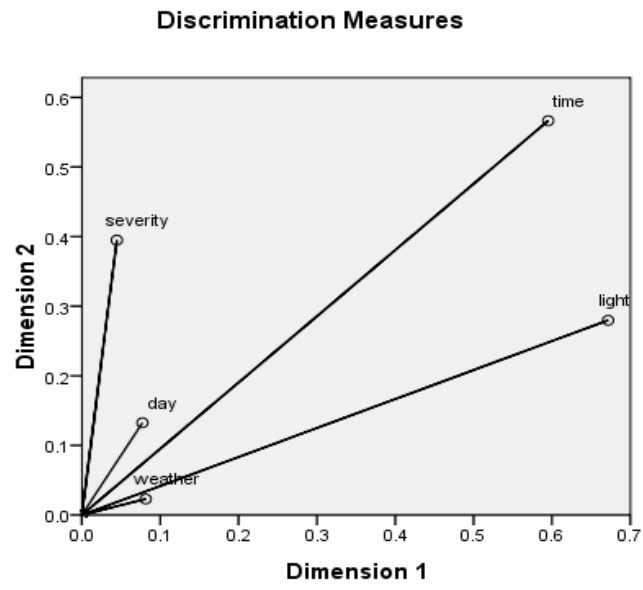


Figure 2. Joint Plot of Category Quantification



Variable Principal Normalization.

Figure 3. Discrimination Measures

TABLE CAPTIONS LIST

Table 1 Summary of Literature Review

Table 2 P1-P2 Matrix for Signalized Intersection Crashes in Las Vegas (2004)

Table 3 Distribution for Signalized Intersection Crashes in Las Vegas (2004-2008)

Table 4 Iteration History

Table 5 Correlations between Transformed Variables

Table 1 Summary of Literature Review

Item		Scholar	Findings/Contribution
Pedestrian exposure	General roadway	Qin & Ivan (2001)	Pedestrian exposure prediction in rural areas
		Cameron (1982), Keall (1995), Chu (2003)	Pedestrian exposure measures
	Signalized intersection	Kennedy (2008)	Pedestrian volumes and crashes estimation at urban signalized intersections
		Greene-Roesel et al. (2010), Schneider et al. (2012)	Overview of pedestrian accident exposure at signalized intersections
Quasi-induced exposure	Basic theory	Lyles (1994)	Quasi-induced exposure was firstly proposed
		Stamatiadis and Deacon (1997)	Induced exposure estimates provided an accurate reflection of exposure to multiple-vehicle accidents
	Extension	Lardelli-Claret et al. (2006)	Two quasi-induced exposure methods were compared, single-vehicle and two-vehicle collisions
		Jiang and Lyles (2010)	Three or more vehicles involved crashes were extended
		Martinez-Ruiz et al. (2013)	Risk factors for road crashes involving cyclists were investigated
	Controversial	Keall and Newstead (2009)	Induced-exposure risk estimates were likely to be biased
		Mendez and Izquierdo (2010)	Explored the direction of bias due to speed
Econometric modeling	Ordered response model	Abdel-Aty (2003), Eluru & Bhat (2007), Sze and Wong (2007), Wang & Abde-Aty (2008), Eluru et al. (2008), Clifton et al. (2009), Kwigizile et al. (2011), Mohamed et al. (2013), Abay (2013), Saidharan & Menendez (2014)	Injury severity was analyzed with emphasis on pedestrian safety
	Unordered response model	Carson and Mannering, 2001; Ulfarsson and Mannering, 2004; Kim et al., 2008, 2010; Tay et al., 2011; Abay, 2013;	Multinomial, nested and mixed logit/probit models were employed to evaluate injury severity

		Aziz et al., 2013; Saidharan & Menendez, 2014	
--	--	---	--

Table 2 P1-P2 Matrix for Signalized Intersection Crashes in Las Vegas (2004)

		P2		P1-Total
		Weekday	Weekend	
P1	Weekday	113 59.8%	76 40.2%	189 63.9%
	Weekend	74 69.2%	33 30.8%	107 36.1%
P2-Total		187 63.2%	109 36.8%	296
IR(weekday) = $63.9/63.2 = 1.01$ IR(weekend) = $36.1/36.8 = 0.98$				

Table 3. Distribution for Signalized Intersection Crashes in Las Vegas (2004-2008)

Variable	Category	Responsible Pedestrians		Non-responsible Pedestrians		Total	
		n	%	n	%	n	%
Severity (Severity)	PDO	259	47.1%	47028	61.3%	47287	61.2%
	Injury	262	47.6%	29476	38.4%	29738	38.5%
	Fatal	29	5.3%	196	0.3%	225	0.3%
Time of day (Time)	Peak hours	128	23.3%	18472	24.1%	18600	24.1%
	Off-peak daytime hours	239	43.5%	40878	53.3%	41117	53.2%
	Off-peak nighttime hours	183	33.3%	17350	22.6%	17533	22.7%
Day of week (Day)	Weekday	434	78.9%	60344	78.7%	60778	78.7%
	Weekend	116	21.1%	16356	21.3%	16472	21.3%
Light condition (Light)	Dawn	11	2.0%	894	1.2%	905	1.2%
	Daylight	331	60.2%	54636	71.2%	54967	71.2%
	Dusk	14	2.5%	1561	2.0%	1575	2.0%
	Dark	191	34.7%	19153	25.0%	19344	25.0%
	Unreported/unknown	3	0.5%	456	0.6%	459	0.6%
Weather condition (Weather)	Clear	461	83.8%	64077	83.5%	64538	83.5%
	Cloudy	78	14.2%	10258	13.4%	10336	13.4%
	Rain	9	1.6%	1813	2.4%	1822	2.4%
	Snow/crosswinds/other	2	0.4%	552	0.7%	554	0.7%
Total		550	100	76700	100	77250	100

Table 4. Iteration History

Iteration Number	Variance Accounted For		Loss
	Total	Increase	
20 ^a	1.433479	.000008	3.566521

a. The iteration process stopped because the convergence test value had been reached.

Table 5. Correlations between Transformed Variables

	Severity	Time	Light	Weather	Day
Severity ^a	1.000	-.026	.112	.046	.040
Time ^a	-.026	1.000	.407	.079	.036
Light ^a	.112	.407	1.000	.065	.101
Weather ^a	.046	.079	.065	1.000	.033
Day ^a	.040	.036	.101	.033	1.000
Dimension	1	2	3	4	5

a. Missing values were input in the mode of the quantified variable.