# A Methodology for Studying Crash Dependence on Demographic and Socioeconomic Data

MICHAEL D. PAWLOVICH, REGINALD R. SOULEYRETTE, AND TIM STRAUSS

Many agencies use traffic crash data to identify problems, establish goals and performance measures, measure progress of specific programs, and support development and evaluation of highway and vehicle safety countermeasures. Traditionally, efforts have considered only crash data and roadway network attributes and have not taken adjacent demographics, socioeconomics, land use, and other non-roadway variables into consideration. The evaluation of non-roadway variables may support two related types of safety management efforts: identification of additional causal factors for roadway crashes and identification of empirical relationships between crashes and non-roadway factors. The second may provide improved estimates of the impact of future changes in land use, demographics, and socioeconomics. Recent efforts use Geographic Information Systems (GIS) or non-spatial relational databases to combine crash and other data to assess correlation and causation. The variety of data available, both within the traditional approach and with the addition of demographic, socioeconomic, and land use data, creates a complex analytical environment. The complexity of these analyses warrants development of a typology to structure an assessment of the best approach in a given situation. This paper presents a concept typology to organize the use of GIS, along with statistical techniques, to explore the relationship between crash incidence and underlying demographic, socioeconomic, and land use data.

## INTRODUCTION

Many agencies use traffic crash data to identify problems, establish goals and performance measures, measure progress of specific programs, and support development and evaluation of highway and vehicle safety countermeasures. Traditionally, efforts have considered only crash data and roadway network attributes and have not taken adjacent demographics, socioeconomics, land use, and other non-roadway variables into consideration.

Engineers have long studied relationships between traffic crashes and potential causal factors, traditionally focusing on roadway geometrics. The studies have generally not considered characteristics of demographics, socioeconomics, and land use in the area proximate to crashes. Efforts on a microscopic (intersection or corridor) level have been made to determine crash causality based on socioeconomic and demographic features, but little has been done to expand this to a macroscopic (network or citywide) level.

Many studies have focused on determining causal factors for traffic crashes (*1,2,3,4*). Other studies utilize traffic crash data to determine cost effectiveness of improvements (*5,6*). Still others focus on factors to reduce crash frequency, fatalities and injuries, and response time. Few sources mention land use, demographics, or socioeconomics in relation to traffic accidents. Of these, two older sources (1965 and 1969) focus on demographics of persons involved in crashes (*7,8*), one mentions land development and traffic influences on road accidents (*9*), and another analyzes a variety of subjects in addition to land use (*10*). No articles found considered demographic, socioeconomic, or land use data in relation to traffic crashes on a macroscopic level.

The evaluation of non-roadway variables may support two related types of safety management efforts. First, it may identify additional causal factors for roadway crashes. Once all such factors have been identified and analyzed, better-informed decisions can be made to remediate existing or potential hazardous locations. Identified non-roadway causal factors will enable engineers and planners to design and plan safer roadways and neighborhoods by providing a clearer picture of contributing factors in certain crashes or crash types. Changes in crash numbers would more clearly be linked to actual causes.

Second, short of causality, the identification of empirical relationships between crashes and non-roadway factors may provide better estimates of the impact of future changes in land use, demographics, and socioeconomics. Such empirical relationships would be useful to guide the allocation of emergency response (e.g., ambulance and police) resources necessary to respond to the potential additional demands presented by new residential and economic developments located in specific locations, and by changing demographic and socioeconomic patterns. However, as few studies have considered the relationship of demographics, socioeconomics, or land use to crashes or crash rates, these variables are not available for design, planning, or analysis.

Recent efforts use Geographic Information Systems (GIS) or non-spatial relational databases to combine crash and other data to assess correlation and causation (*11,12,13*). GISs provide excellent tools to analyze location specific crash data. Multiple layers

Center for Transportation Research and Education, 2625 N. Loop Drive, Suite 2100, Ames, Iowa 50010-8615.

can be viewed and analyzed at once. In addition, GISs enable development of a methodology to consider non-roadway variables.

Currently, the Center for Transportation Research and Education (CTRE) is developing a GIS-based accident location and analysis system for the state of Iowa that facilitates spatial analyses of crash incidence. Iowa is fortunate 1) to have a comprehensive location–based database covering 10 years of all traffic crashes on all road systems, and 2) to have developed one of the better systems for analyses, Personal Computer-based Accident Location and Analysis System (PC-ALAS). Many approaches are available to analyze crash data in a GIS environment. The variety of data available, both within the traditional approach and with the addition of demographic, socioeconomic, and land use data, creates a complex analytical environment. The complexity of these analyses warrants development of a typology to structure an assessment of the best approach in a given situation.

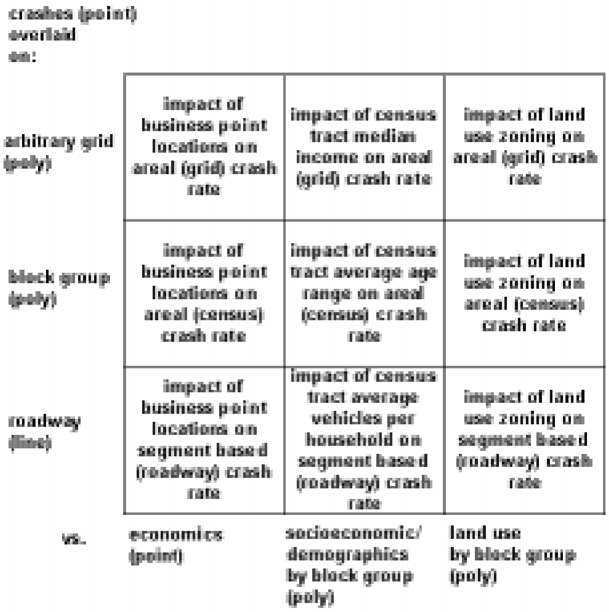
A topological (i.e., based on feature class) division of crash rates includes three types of geographic representations: point, line, and polygon. Points can either represent the location of a single crash or the location of a point where multiple crashes have occurred. Lines can denote a segment or corridor with multiple crashes or a network made up of a series of lines, combining the crashes on each line to develop the crash rate. Polygon representations combine crashes within an area in order to develop an areawide crash rate. Polygons can be further divided, representing regions using an arbitrary grid or block groups, depending on the data available and the desired analysis.

Utilizing the topological representation of crash rates as the dependent (Y) variable, various independent (X) variables, which also can be represented with varying topology, can be used to determine potential causal relationships. As the analysis of these topological representations can become quite complex, a classification scheme (typology) to structure the analyses is helpful. This paper presents a concept typology to organize the use of GIS, along with statistical techniques, to explore the relationship between crash incidence and underlying demographic, socioeconomic, and land use data.

## TYPOLOGY

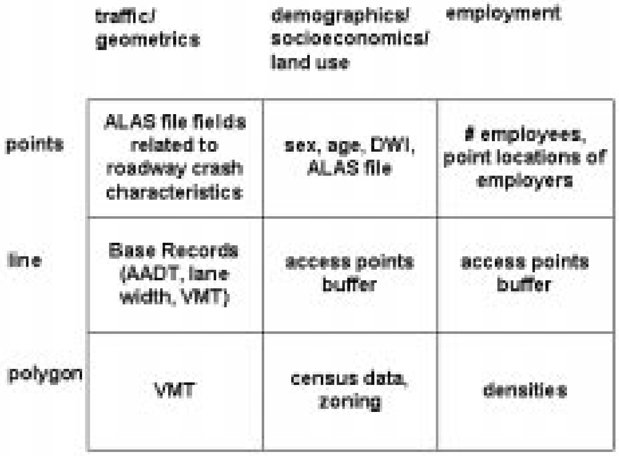
The typology, as shown in Figure 1, consists of two topological dimensions: dependent variable (crash rate) and independent variable (here, demographic, socioeconomic, and land use). Prior to representation with varying topology, the dependent variable must be created from a spatial combination of crash incidence and exposure (traffic levels). In this study, crash locations are represented as points. Three methods are presented here to develop crash rates (point on line and point on polygon, including arbitrary grid and census block), creating three sets of dependent variables for subsequent analyses. In addition, the three sets of dependent variables are statistically related to three independent variables, creating nine or more possible types of analyses.

Each dependent variable/independent variable pair can be utilized to assess the impact of different features of the independent variables on crash rates as shown in Figure 2. For example, the arbitrary grid/economics combination could be utilized to determine impact of business point locations on an areal (grid) crash rate. Employment densities within grids could be related to crash rates within grids. In addition, given an accident location, all businesses within a grid of an accident location could be determined.



| crashes (point) overlaid on: | | | |
|---|---|---|---|
| arbitrary grid (poly) | impact of business point locations on areal (grid) crash rate | impact of census tract median income on areal (grid) crash rate | impact of land use zoning on areal (grid) crash rate |
| block group (poly) | impact of business point locations on areal (census) crash rate | impact of census tract average age range on areal (census) crash rate | impact of land use zoning on areal (census) crash rate |
| roadway (line) | impact of business point locations on segment based (roadway) crash rate | impact of census tract average vehicles per household on segment based (roadway) crash rate | impact of land use zoning on segment based (roadway) crash rate |
| vs. | economics (point) | socioeconomic/ demographics by block group (poly) | land use by block group (poly) |

**FIGURE 1 Crash rate typology.**



| | traffic/ geometrics | demographics/ socioeconomics/ land use | employment |
|---|---|---|---|
| points | ALAS file fields related to roadway crash characteristics | sex, age, DWI, ALAS file | # employees, point locations of employers |
| line | Base Records (AADT, lane width, VMT) | access points buffer | access points buffer |
| polygon | VMT | census data, zoning | densities |

**FIGURE 2 Causal factors matrix.**

## METHODOLOGY

The methodology to develop dependent and independent variables consists of six main steps, as shown in Figure 3. The first three steps are used to develop independent variable data, the fourth to develop dependent variables, the fifth to develop independent variables themselves, and the sixth to apply statistical techniques to determine significant causal relationships. In this paper, the methodology is demonstrated at the block group level.

### Step 1

Data available include census data, crash data, infrastructure data (mainly roadway data), and employment data. An enormous amount of data is available, resulting in computational problems for standard statistical packages; therefore, the variable list was narrowed to those viewed as most promising by the authors. Initially, we arbitrarily selected variables that seemed most likely to affect crash rates.

Census data contains over 3,400 data elements that can all be referenced to geographic regions. These elements are divided into several main headings and subheadings. To provide more manageable census data for the purposes of this paper, data under certain headings, unlikely to relate to crash rates, were discarded from consideration. The remaining data elements were contained in 261 main headings and subheadings. From these we arbitrarily selected a portion for subsequent analyses. The selection left us with a large, but much more manageable number of total variables (225 variables under 25 main headings) to consider.

Infrastructure data obtained from the Iowa Department of Transportation (DOT) include several background GIS coverages such as hydrology, rail lines, secondary roads, primary roads, and municipal roads. The latter three of these were of significant interest for this paper. Included as attributes for the road coverages are AADT and lane length. The AADT and lane length were used to calculate Vehicle Miles Traveled (VMT) for each roadway segment. The VMT, combined with total crashes along each roadway segment, were used to construct the dependent variable, crash rate, for each roadway element.

Crash data also obtained from the Iowa DOT includes many data elements related to the crash, the vehicles and drivers, and the injured persons. These data elements are explained in a recently published report detailing current efforts to expand and improve Iowa's accident location and analysis system (*14*). For the purposes of this paper, relevant information is the incidence and location of crashes. Combining these data with VMT results in crash rate (crashes/100 million VMT). Future efforts may consider specific crash attributes.

Socioeconomic data contains: business name, address, city, state, zipcode, Standard Industrial Code (SIC), and number of employees. The data were obtained from the Iowa Department of Workforce Development (DWD) and are confidential. The SIC code is the most important element, though number of employees may be of interest. The other variables were used to create a point location map of businesses using the geocode function of a GIS.

### Step 2

The second step was to develop or locate polygon coverages of block groups and point coverages of businesses and other land use data. Polygon coverages of block groups, illustrated in Figure 4, were obtained from a commercial product and exported to GIS format. The point coverages of businesses, displayed in Figure 5, were developed from DWD data. Utilizing the address code of the DWD data and a commercial street address database, point coverages were developed using the geocode function of a desktop GIS.

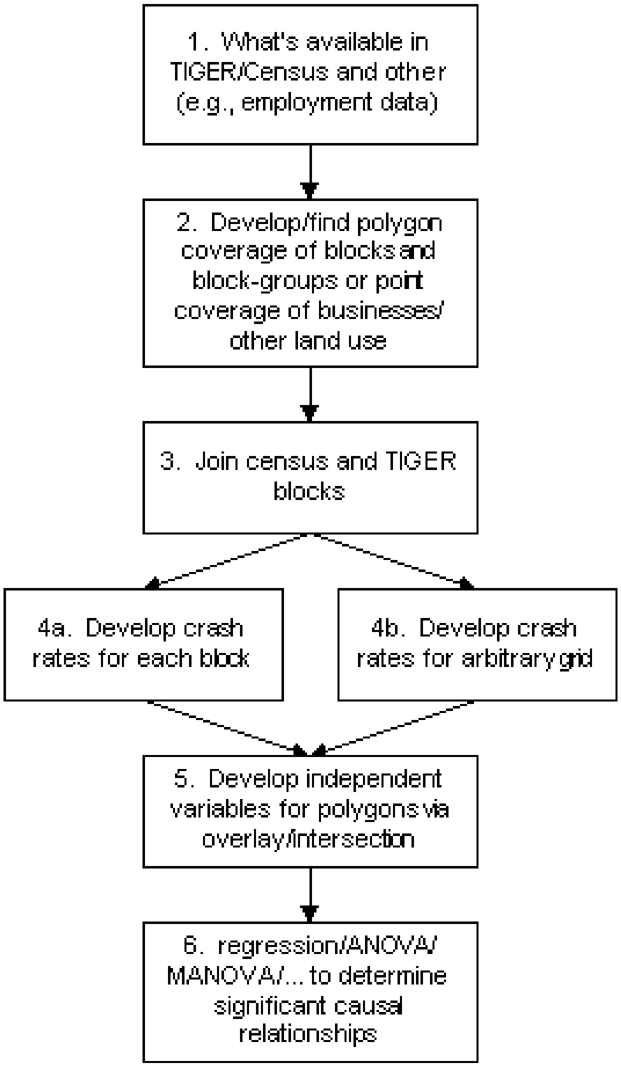

1. What's available in TIGER/Census and other (e.g., employment data)
2. Develop/find polygon coverage of blocks and block-groups or point coverage of businesses/ other land use
3. Join census and TIGER blocks
4a. Develop crash rates for each block
4b. Develop crash rates for arbitrary grid
5. Develop independent variables for polygons via overlay/intersection
6. regression/ANOVA/ MANOVA/... to determine significant causal relationships

**FIGURE 3 Equation development process.**

### Step 3

The third step was to join census data and Topologically Integrated Geographic Encoding and Referencing (TIGER) blocks (i.e., polygons for block groups). However, the census data and TIGER block join was found to be commercially available; therefore, no work was entailed in this step other than querying for desired census information and exporting it into GIS format.

### Step 4

The fourth step was to develop crash rates, the dependent variable, for each independent variable type: block group, arbitrary grid, and linear system. For this paper, only the first, crash rates for block groups, was completed. The latter two independent variable types will be considered in future efforts.

To develop crash rates for block groups, first the Iowa DOT roadway coverages were spatially joined to crash data, as shown in Figure 6. The resultant table was then summarized to produce a table of crashes by using the roadway coverage index fields. The summarization table and roadway coverages were then spatially joined, creating roadway coverages with total crashes along each roadway segment.

The VMT for each block group was calculated by first spatially joining each block group to the road coverages, as shown in Figure 7. The resultant road coverages with the block group table was then exported to dBase format and imported into Microsoft Access. Within Access, the crashes, average annual daily traffic (AADT), and lane length (meters) were grouped by area identifier using a summation for each. The grouping results were saved and imported into GIS and then spatially joined to the block group coverage. Each block group now includes total number of crashes, total AADT, and total lane length as attributes. Additional fields, VMT and crash rate, were created and their values calculated for each block group using the following formulas:

- VMT = total AADT * 365 days/year * lane length (meters) / 1.609 meters/mile; and
- Crash rate = total crashes/(1000000*VMT).

After generation of the grid, development of crash rates for an arbitrary grid would proceed similarly. Linear system crash rates would involve the spatial joining of the roadway and crash data and the subsequent calculation of VMT and crash rate.

### Step 5

The fifth step was the development of the independent variables, for polygons via overlay/intersection. This was accomplished by joining the census data to the crash rate data. An overlay of block groups, crashes, and business locations is shown in Figure 8.

### Step 6

The sixth step uses the independent and dependent variables developed previously to examine the data for significant causal relationships. Various statistical techniques may be utilized, including linear regression, analysis of variance (ANOVA), multiple regression ANOVA (MANOVA), factor analysis, bivariate regression, time series, and spatial regression. Within the selected GIS environment, few rigorous statistical techniques were available; however, a simple bivariate regression script is available for analyzing data. In addition, exporting the database to a spreadsheet allowed for more involved multiple regression. For this paper, various variables were tested against the block group crash rate for a variety of regions to ascertain whether there were any obvious causal factors. This was done using a desktop statistical software package once the data had been exported in delimited text format from the GIS. Future efforts will involve statistical packages allowing the more robust analyses desired.

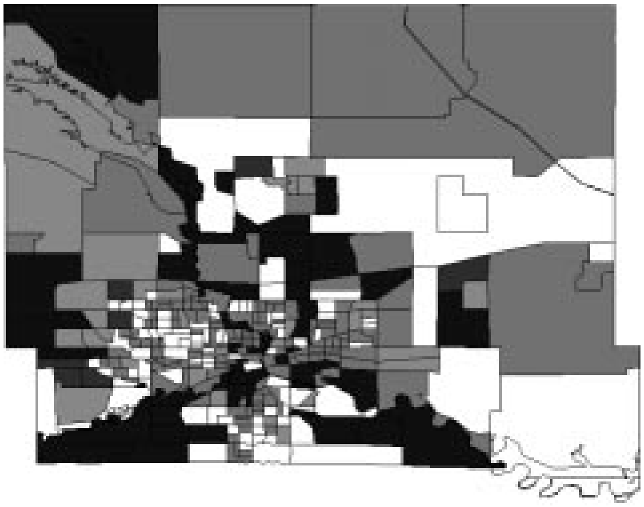


FIGURE 4 Block groups.

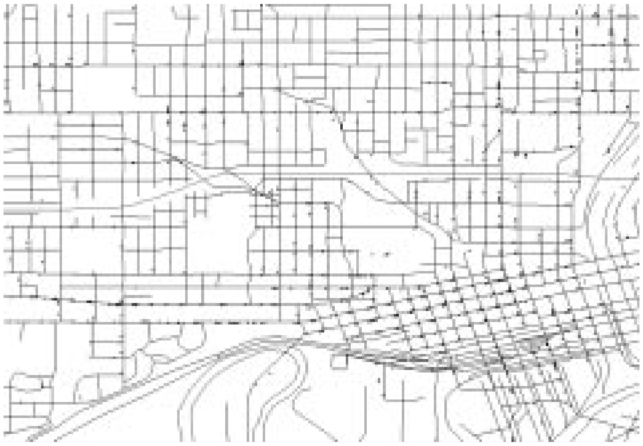


FIGURE 5 Socioeconomic point locations.

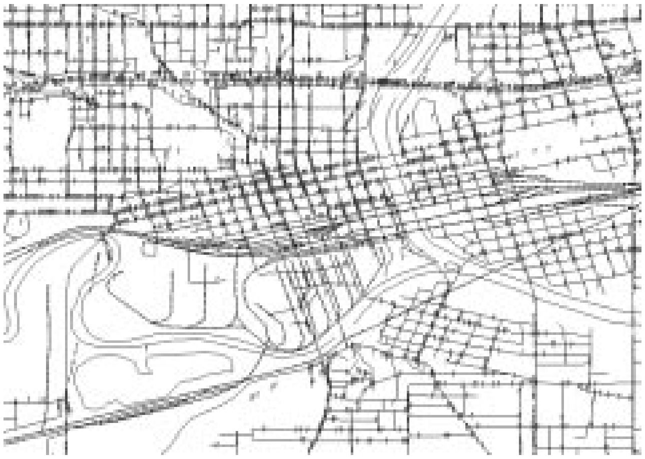


FIGURE 6 Iowa DOT roadway and crash coverages.

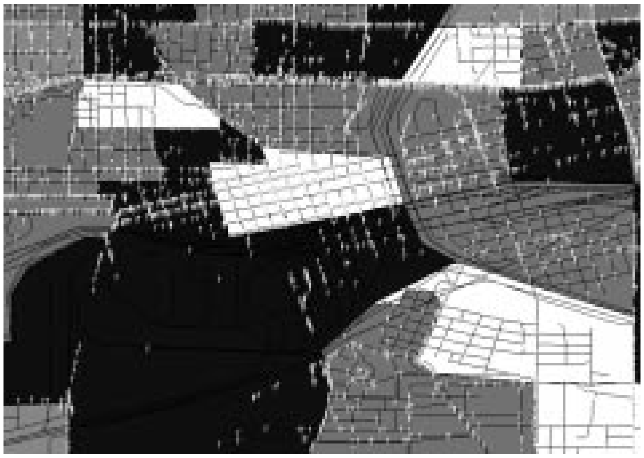


**FIGURE 7 Roadway, crash and block group coverages.**

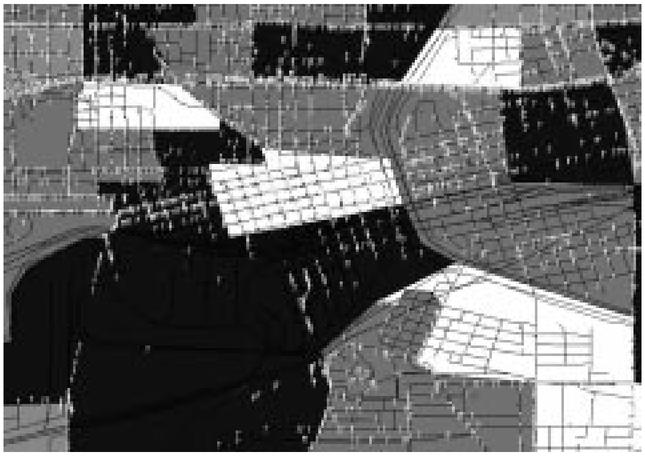
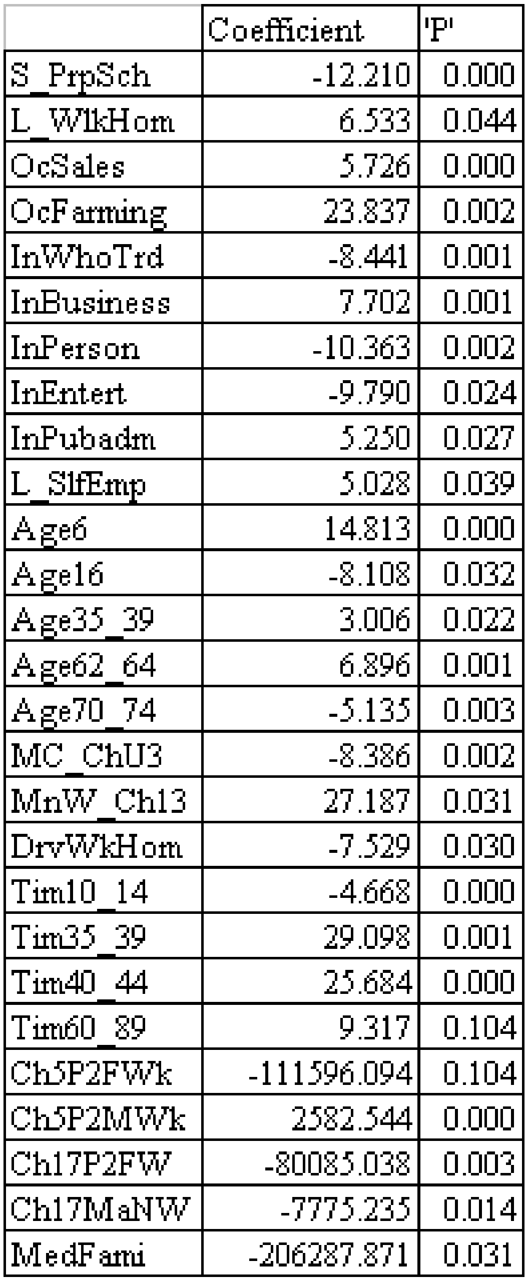


**FIGURE 8 Block groups, crashes, and business location coverages.**



| | Coefficient | 'P' |
|---|---|---|
| S_PrpSch | -12.210 | 0.000 |
| L_WlkHom | 6.533 | 0.044 |
| OcSales | 5.726 | 0.000 |
| OcFarming | 23.837 | 0.002 |
| InWhoTrd | -8.441 | 0.001 |
| InBusiness | 7.702 | 0.001 |
| InPerson | -10.363 | 0.002 |
| InEntert | -9.790 | 0.024 |
| InPubadm | 5.250 | 0.027 |
| L_SlfEmp | 5.028 | 0.039 |
| Age6 | 14.813 | 0.000 |
| Age16 | -8.108 | 0.032 |
| Age35_39 | 3.006 | 0.022 |
| Age62_64 | 6.896 | 0.001 |
| Age70_74 | -5.135 | 0.003 |
| MC_ChU3 | -8.386 | 0.002 |
| MnW_Ch13 | 27.187 | 0.031 |
| DrvWkHom | -7.529 | 0.030 |
| Tim10_14 | -4.668 | 0.000 |
| Tim35_39 | 29.098 | 0.001 |
| Tim40_44 | 25.684 | 0.000 |
| Tim60_89 | 9.317 | 0.104 |
| Ch5P2FWk | -111596.094 | 0.104 |
| Ch5P2MWk | 2582.544 | 0.000 |
| Ch17P2FW | -80085.038 | 0.003 |
| Ch17MaNW | -7775.235 | 0.014 |
| MedFami | -206287.871 | 0.031 |

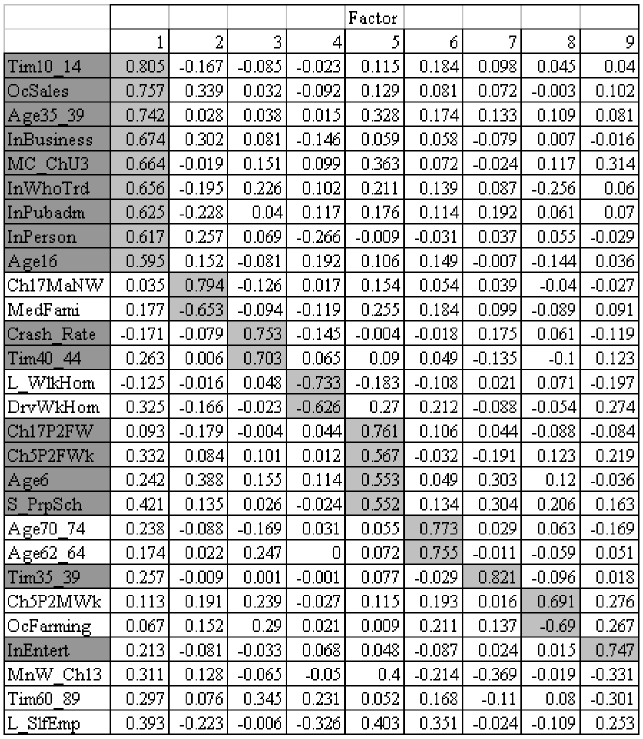
**FIGURE 9 Regression results.**

## ANALYSIS

The initial analysis effort utilized a subset of the available data. Future efforts will analyze the data more comprehensively. However, a variety of factors contributed to the limited analyses performed at this time.

Using a metropolitan region for the analyses, the data fields were pared to an arbitrary set of variables of interest. These variables were then exported to delimited text format from the GIS and imported into the desktop statistical package for analyses. A backward, stepwise, linear regression was performed, using entry and exit probabilities of 0.15. The correlation coefficient (R = 0.677) and the coefficient of determination (R-squared = 0.458) indicate that the equation is moderately successful in predicting crash rates at the block group level (see Figure 9). Several independent variables remained in the estimated equation, 27 in all. Of these, 10 were statistically significant at the 0.001 level, 15 at the 0.01 level, and 25 at the 0.05 level. Independent variables with a calculated statistical significance of "0.000" included:

- Persons 3 years and over enrolled in preprimary school (S_PrpSch)
- Employed persons 16 years and over who are in sales occupations (OcSales)



| | | | | | Factor | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Tim10_14 | 0.805 | -0.167 | -0.085 | -0.023 | 0.115 | 0.184 | 0.098 | 0.045 | 0.04 |
| OcSales | 0.757 | 0.339 | 0.032 | -0.092 | 0.129 | 0.081 | 0.072 | -0.003 | 0.102 |
| Age35_39 | 0.742 | 0.028 | 0.038 | 0.015 | 0.328 | 0.174 | 0.133 | 0.109 | 0.081 |
| InBusiness | 0.674 | 0.302 | 0.081 | -0.146 | 0.059 | 0.058 | -0.079 | 0.007 | -0.016 |
| MC_ChU3 | 0.664 | -0.019 | 0.151 | 0.099 | 0.363 | 0.072 | -0.024 | 0.117 | 0.314 |
| InWhoTrd | 0.656 | -0.195 | 0.226 | 0.102 | 0.211 | 0.139 | 0.087 | -0.256 | 0.06 |
| InPubadm | 0.625 | -0.228 | 0.04 | 0.117 | 0.176 | 0.114 | 0.192 | 0.061 | 0.07 |
| InPerson | 0.617 | 0.257 | 0.069 | -0.266 | -0.009 | -0.031 | 0.037 | 0.055 | -0.029 |
| Age16 | 0.595 | 0.152 | -0.081 | 0.192 | 0.106 | 0.149 | -0.007 | -0.144 | 0.036 |
| Ch17MaNW | 0.035 | 0.794 | -0.126 | 0.017 | 0.154 | 0.054 | 0.039 | -0.04 | -0.027 |
| MedFami | 0.177 | -0.653 | -0.094 | -0.119 | 0.255 | 0.184 | 0.099 | -0.089 | 0.091 |
| Crash_Rate | -0.171 | -0.079 | 0.753 | -0.145 | -0.004 | -0.018 | 0.175 | 0.061 | -0.119 |
| Tim40_44 | 0.263 | 0.006 | 0.703 | 0.065 | 0.09 | 0.049 | -0.135 | -0.1 | 0.123 |
| L_WlkHom | -0.125 | -0.016 | 0.048 | -0.733 | -0.183 | -0.108 | 0.021 | 0.071 | -0.197 |
| DrvWkHom | 0.325 | -0.166 | -0.023 | -0.626 | 0.27 | 0.212 | -0.088 | -0.054 | 0.274 |
| Ch17P2FW | 0.093 | -0.179 | -0.004 | 0.044 | 0.761 | 0.106 | 0.044 | -0.088 | -0.084 |
| Ch5P2FWk | 0.332 | 0.084 | 0.101 | 0.012 | 0.567 | -0.032 | -0.191 | 0.123 | 0.219 |
| Age6 | 0.242 | 0.388 | 0.155 | 0.114 | 0.553 | 0.049 | 0.303 | 0.12 | -0.036 |
| S_PrpSch | 0.421 | 0.135 | 0.026 | -0.024 | 0.552 | 0.134 | 0.304 | 0.206 | 0.163 |
| Age70_74 | 0.238 | -0.088 | -0.169 | 0.031 | 0.055 | 0.773 | 0.029 | 0.063 | -0.169 |
| Age62_64 | 0.174 | 0.022 | 0.247 | 0 | 0.072 | 0.755 | -0.011 | -0.059 | 0.051 |
| Tim35_39 | 0.257 | -0.009 | 0.001 | -0.001 | 0.077 | -0.029 | 0.821 | -0.096 | 0.018 |
| Ch5P2MWk | 0.113 | 0.191 | 0.239 | -0.027 | 0.115 | 0.193 | 0.016 | 0.691 | 0.276 |
| OcFarming | 0.067 | 0.152 | 0.29 | 0.021 | 0.009 | 0.211 | 0.137 | -0.69 | 0.267 |
| InEntert | 0.213 | -0.081 | -0.033 | 0.068 | 0.048 | -0.087 | 0.024 | 0.015 | 0.747 |
| MnW_Ch13 | 0.311 | 0.128 | -0.065 | -0.05 | 0.4 | -0.214 | -0.369 | -0.019 | -0.331 |
| Tim60_89 | 0.297 | 0.076 | 0.345 | 0.231 | 0.052 | 0.168 | -0.11 | 0.08 | -0.301 |
| L_SlfEmp | 0.393 | -0.223 | -0.006 | -0.326 | 0.403 | 0.351 | -0.024 | -0.109 | 0.253 |

**FIGURE 10 Factor analysis results.**

- Persons aged 6 years old (Age6)
- Workers 16 years and over, not working at home, whose travel time to work was 10-14 minutes (Tim10_14)
- Workers 16 years and over, not working at home, whose travel time to work was 40-44 minutes (Tim14_44)
- Families with two parents and children under 6 years with only the mother in the labor force (Ch5P2MWk)

At first glance, the results seem promising. A few caveats are necessary, however. First, one must be careful about assigning causality to the above relationships. The results should not be interpreted as indicating that crashes are caused by 6-year-old children, and their preprimary aged siblings, with commuting salesman fathers and stay-at-home mothers. The results might indicate, however, that locations with families having these characteristics, may have above-average block-group-based crash rates (controlling for a limited set of other factors). Second, regression diagnostics indicated that the above results should be viewed cautiously. In particular, extreme values may be skewing the results. A series of scatterplots with the dependent variable and selected independent variables indicated that this is the case. A few block groups with extremely high values greatly affected the results. Normally, these observations would be discarded, but in crash analysis they are usually the ones of most interest. Third, although a surprising number of interrelated independent variables entered the final equation, a high degree of multicollinearity, typical with socioeconomic data, can make regression parameters unstable and substantive interpretation difficult.

To assess the third point, a principal components analysis (PCA) was performed on the 27 independent variables to identify highly correlated groups of variables (minimum eigenvalue = 1.0, varimax rotation). Nine factors resulted from this and are presented in Figure 10. Analysis of the results requires a bit of interpretation. For instance, variables with high loadings on the first component include:

- Workers 16 years and over, not working at home, whose travel time to work was 10-14 minutes (Tim10_14)
- Employed persons 16 years and over who are in sales occupations (OcSales)
- Persons aged 35-39
- Employed persons 16 years and over who are in public administration industries (InPubadm).

This component can be interpreted as representing neighborhoods with thirtysomething professional service industry workers with average commute times. Other components can be similarly interpreted.

To assess the potential impact of these groups of variables on block-group-based crash rates, the components can, in turn, be used as independent variables in a regression analysis. The results were less useful than the original regression, however, since the R (0.276) and $R^2$ (.076) were lower (worse for empirical prediction), and the difficult interpretation of the components that entered the equation makes it difficult to derive much meaning from the results that we could use to establish causal factors. Additionally, only factor 1 (0.054) and factor 6 (0.000), had p-values of statistical significance.

## CONCLUSIONS

The development of the above typology is a promising approach, but mainly for empirical prediction (e.g., to estimate impact of changes on number of crashes in a given development, city, or emergency response district). Statistical associations and patterns discovered through using this typology can be examined for possible causal significance, which would then be assessed via more detailed studies.

Though regression using individual variables gave good results for prediction, the results are hard to interpret substantively. In addition, use of PCA made the equation worse and did not aid interpretation. The next step would be mapping to find block groups with high values for the two statistically significant components.

Some data issues made the made the analysis more difficult. Extreme values/outliers make it difficult to get meaningful results from regression. Additionally, these extreme values/outliers skew the analysis. However, these values are more interesting in from a safety perspective.

Additionally, the typology might be developed further for two different types of analyses. Examining immediate corridor proximity would facilitate causal analysis and engineering countermeasures. Examining broader areas (e.g., block groups) is better for planning applications such as police and fire response or broader estimates of changes in crash statistics and patterns.

## REFERENCES

1. Zhou, M. and V. P. Sisiopiku. Relationships Between Volume-to-Capacity Ratios and Accident Rates. *Transportation Research Record No. 1581.* National Academy Press, Washington, D.C., 1997, pp. 47-52.
2. Raub, R. A. Occurrence of Secondary Crashes on Urban Arterial Roadways. *Transportation Research Record No. 1581.* National Academy

Press, Washington, D.C., 1997, pp. 53-58.

3. Access Management Slows Incidence of Traffic Accidents. *Public Works*, February 1995, pp. 39-41.
4. Dart, O. K. and L. Mann. Relationship of Rural Highway Geometry to Accident Rates in Louisiana. *Highway Research Record*, Vol. 312, 1970.
5. *Workshop on Development of the Interactive Highway Safety Design Model Accident Analysis Module.* Publication No. FHWA-RD-96-075, U.S. Department of Transportation/Federal Highway Administration/Research and Development, Turner-Fairbank Highway Research Center, McLean, Virginia, February 1997.
6. *Knowledge Acquisition Methods for the IHSDM Diagnostic Review Expert System.* Publication No. FHWA-RD-97-134, U.S. Department of Transportation/Federal Highway Administration/Research and Development, Turner-Fairbank Highway Research Center, McLean, VA, December 1997.
7. Muench, L. O. *The Relationship of Demographic Variables to Fatal Motor Vehicle Accidents.* Master's Thesis, Iowa State University, Ames, Iowa, 1965.
8. Tack, L. R. *Relationship of Demographic Variables to Non-Fatal Motorcycle Accidents, Iowa, 1962-1966.* Master's Thesis, Iowa State University, Ames, Iowa, 1969.
9. Podkowicz, C. The Influence of Land Development and Traffic on Road Accidents. *Transport Quarterly*, Vol. 3, No. 2, 1991.
10. Del Mistro, R. F. and R. Fieldwick. *The Contribution of Traffic Volume, Speed, Congestion, Road Section Block Length, Abutting Land Use and Kerbside Activity to Accidents on Urban Arterial Roads.* National Institute for Transport and Road Research, Pretoria, 1981.
11. Quiroga, C. A. and D. Bullock. Geographic Database for Traffic Operations Data. *Journal of Transportation Engineering.* American Society of Civil Engineers, Vol. 122, No. 3, May/June 1996, pp. 226-234.
12. Saccomanno, F. F., K. C. Chong, and S. A. Nassar. Geographic Information System Platform for Road Accident Risk Modeling. *Transportation Research Record No. 1581.* Transportation Research Board/National Research Council, National Academy Press, Washington, D.C., 1997, pp. 18-26.
13. Pawlovich, M. D. and R. R. Souleyrette. A GIS-based Accident Location and Analysis System (GIS-ALAS). *1996 Semisesquicentennial Transportation Conference Proceedings.* Iowa Department of Transportation/Iowa State University/Center for Transportation Research and Education (CTRE), Ames, Iowa, 1996, pp. 29-34.
14. Souleyrette, R. R., T. Strauss, M. Pawlovich, and B. Estochen. *GIS-based Accident Location and Analysis System (GIS-ALAS) Project Report: Phase 1.* Iowa Department of Transportation/Center for Transportation Research and Education, April 1998.