Are there higher pedestrian fatalities in larger cities?: A scaling analysis of 115 to 161 largest cities in the United States

Yu Sang Chang a,b, Won Jae Lee b,c, and Jae Hee Lee b,c

a College of Business Administration, Department of Global Business Administration, Gachon University, Seong Nam, Republic of Korea; b Gachon Center for Convergence Research, Gachon University, Seong Nam, Republic of Korea; c College of Social Science, Department of Health Care Management, Gachon University, Seong Nam, Republic of Korea

ABSTRACT

Objective: In 2012, 4,743 pedestrians were killed in the United States, representing 14% of total traffic fatalities. The number of pedestrians injured was higher at 76,000. Therefore, 36 out of 52 of the largest cities in the United States have adopted a citywide target of reducing pedestrian fatalities. The number of cities adopting the reduction goal during 2011 and 2012 increased rapidly with 8 more cities. We examined the scaling relationship of pedestrian fatality counts as a function of the population size of 115 to 161 large U.S. cities during the period of 1994 to 2011. We also examined the scaling relationship of nonpedestrian and total traffic fatality counts as a function of the population size.

Methods: For the data source of fatality measures we used Traffic Safety Facts Fatality Analysis Reporting System/General Estimates System annual reports published each year from 1994 to 2011 by the NHTSA. Using the data source we conducted both annual cross-sectional and panel data bivariate and multivariate regression models. In the construction of the estimated functional relationship between traffic fatality measures and various factors, we used the simple power function for urban scaling used by Bettencourt et al. (2007, 2010) and the refined STIRPAT (stochastic impacts by regression on population, affluence, and technology) model used in Dietz and Rosa (1994, 1997) and Yorke et al. (2003).

Results: We found that the scaling relationship display diseconomies of scale or sublinear for pedestrian fatalities. However, the relationship displays a superlinear relationship in case of nonpedestrian fatalities. The scaling relationship for total traffic fatality counts display a nearly linear pattern. When the relationship was examined by the 4 subgroups of cities with different population sizes, the most pronounced sublinear scaling relationships for all 3 types of fatality counts was discovered for the subgroup of megacities with a population of more than 1 million.

Conclusions: The scaling patterns of traffic fatalities of subgroups of cities depend on population sizes of the cities in subgroups. In particular, 9 megacities with populations of more than 1 million are significantly different from the remaining cities and should be viewed as a totally separate group. Thus, analysis of the patterns of traffic fatalities needs to be conducted within the group of megacities separately from the other cities with smaller population sizes for devising prevention policies to reduce traffic fatalities in both megacities and smaller cities.

Introduction

According to the Alliance benchmarking report (Alliance for Biking and Walking 2014), 36 out of 52 largest cities in the United States have adopted their city-wide target of reducing pedestrian fatalities. The number of cities adopting the reduction goal during 2011 and 2012 increased rapidly to 8 cities ranging from Houston, Texas; Cleveland, Ohio; and Tucson, Arizona. As many as 44 states in the United States have also adopted the reduction goal as well. Many of these reduction goals seek to decrease pedestrian fatalities by half within the next 5 years or more. However, some of the most ambitious “Vision Zero” adopted by New York, Los Angeles, San Francisco, and Chicago in recent years seeks to eliminate all pedestrian deaths within 10 years or by 2025.

In 2012, 4,743 pedestrians were killed in the United States, representing 14% of total traffic fatalities (NHTSA 2014). The number of pedestrians injured was higher at 76,000. A pedestrian is defined as “any person on foot, walking, running, jogging, hiking, sitting or lying down who is involved in a motor vehicle traffic crash” (NHTSA 2014, p. 1). A traffic crash is defined as “an incident that involves one or more vehicles where at least one vehicle is in transport” (NHTSA 2014, p. 1). The same report states that 73% of pedestrians were killed in urban areas and 70% of pedestrians were killed at nonintersections in 2012.

In 1994, the number of pedestrian deaths in 115 of the largest U.S. cities was 1,583, representing 29.5% of the total number of 5,374 persons (NHTSA 1995). In 2011, the number of pedestrian deaths in 161 of the largest U.S. cities was 1,321 persons,
representing about 28% of the total 4,726 persons (NHTSA 2014). These averaged percentages of pedestrian deaths in large cities are much higher in comparison to the U.S. national average of 14% in 2011, demonstrating that a larger number of pedestrian deaths took place in urban areas.

However, there is a great degree of variation in pedestrian deaths among individual cities. For example, the number of pedestrian deaths in 2011 ranged from a high of 138 persons in New York City to a low of 0 persons killed in 10 cities, including Laredo, Texas; Durham, North Carolina; Gilbert, Arizona; Boise City, Idaho; Tacoma, Washington; Oxnard, Moreno Valley, and Santa Clara, California; Overland Park, Kansas; and Salem, Oregon.

Translating these numbers of pedestrian deaths into the conventional pedestrian fatality rate per 100,000 population, we have a high of 1.67 for New York City to a low of 0 for each of the 10 cities with 0 fatality counts. Most cities have been using the same population size–adjusted pedestrian fatality rate to compare their progress with respect to other peer cities. For example, in the Pedestrian Crash Analysis, published by the City of Chicago (2011), claims that “the average pedestrian fatality rate in Chicago of 1.77 from 2005 through 2009 was 16% lower than the average rate of the 14 other cities.” Furthermore, “Chicago has the lowest pedestrian fatality rate of large, densely populated cities” (City of Chicago 2011, p. 5).

However, such comparison can be misleading if the pedestrian fatality counts to population size of the city is subjected to a nonlinear scaling relationship. For example, if the scaling relation follows a sublinear relation with the scale exponent of 0.5, then larger population sizes will generate a disproportionately smaller number of pedestrian deaths. For example, City A has a population of 1 million and City B has a population of 2 million. City A recorded an average of 10 pedestrian deaths during the last 5 years. If the scale exponent is 0.5, then City B should expect 14 pedestrian deaths ($2^{0.5} = 1.41$). The pedestrian fatality rate for City A is 1 per 100,000, and for City B, the comparable rate will be 1.4 per 200,000 or 0.7 per 100,000. If City C has a population of 4 million, the comparable rate should be 0.5 per 100,000 ($4^{0.5}$). This simple example explains why it is important to examine the existence of scaling relationships and determine whether such a relationship follows a sublinear, superlinear, or linear pattern.

Therefore, the basic question in this article is to determine the scaling relationship of pedestrian fatality counts as a function of the population size of large individual cities in the United States. More specifically, we plan to examine the following 3 primary questions. First, do larger cities generate disproportionately fewer pedestrian fatalities? Second, does the scaling relationship vary among the subgroups of more populous cities versus less populous cities? Third, what are other variables besides population size that influence this scaling relationship? In order to provide comparable background information, we also plan to examine the same 3 questions for total traffic fatality counts as well as for nonpedestrian fatality counts.

This article is organized into 4 additional sections. A brief literature survey on pedestrian fatality in the large cities is presented in the following section. Next, we describe our method of analysis and the data used. The next section explains the results of our analysis. Finally, limitations of this study are presented.

### A brief literature survey

What is the scaling relationship between the population size of cities and the number of pedestrian deaths? Do other types of traffic fatalities also display a scaling relationship to the population size of cities? There is a large body of literature on the urban scaling theory, which was initially developed by an interdisciplinary team of physicists, economists, and urban specialists at Santa Fe Institute (Bettencourt 2013; Bettencourt et al. 2006, 2007, 2010; Fragkias et al. 2013; Loaf and Barthelemy 2014; Strumsky and Lobo 2008). When cities are viewed as living organisms or ecosystems (Decker et al. 2007; Florida 2004; Fujita et al. 2001; Glaeser and Gottlieb 2009; Samaniego and Moses 2008), cities change shape as the population size increases. For example, economic output and creative ideas grow faster than linearly due to the fact that the number of human interactions in larger cities may increase exponentially.

Bettencourt et al. (2007) have presented empirical evidence on measures such as patents, inventors, R&D employment, and gross domestic product that increase disproportionately faster than the increase in population size. They have also noted measures such as crime, infectious diseases, congestion, and poverty that display the similar scaling pattern. In other words, both positive and negative impacts from rapidly increasing social interactions in populous cities are expected to generate a superlinear scaling relationship. On the other hand, they have also collected other measures dealing with material infrastructure and network of cities that grow slower than the increase in population size. For example, the length of roads, the number of gasoline stations, or the lengths of electric cables display a sublinear scaling relation. Finally, another group of measures dealing primarily with needs of individuals in cities have been found to increase at the same rate as the city’s population size. Measures on household electricity and water consumption or a city’s rate of employment display a linear relation.

We have not been able to discover any systematic research done on the scaling relationship between pedestrian fatalities and population size of cities. However, there is a related topic dealing with pedestrian fatalities as a function of the number of people walking in cities and countries. This literature, known as the “safety in number concept,” was pioneered by Jacobsen (2003). Jacobsen presented empirical data that showed that the relationships between the number of pedestrian injuries or fatalities to the number of people walking in cities or countries follow a sublinear scaling relation. The typical scaling exponent estimated is roughly 0.4, which indicates that doubling the amount of walking can show a 32% increase in injuries ($2^{0.4} = 1.32$). In other words, individual risk while walking in a community with twice as much walking will be reduced to 66% ($2^{-0.6} = 0.66$). Jacobsen (2003) suggested that “the most plausible explanation is behavior modification by motorists when they expect or experience people walking and bicycling” (p. 4).

Recently, the Alliance benchmarking report (Alliance for Biking and Walking 2014) showed that the safety in number concept may have operated in 52 of the largest U.S cities. The report shows that pedestrian fatality rates in individual cities have declined as the percentage of people walking to work increased. However, safety in numbers is still a controversial concept and has attracted a number of articles, both for and
against, on theoretical as well as empirical grounds (L. F. Beck et al. 2007; S. F. Beck et al. 2007; Elvik 2009, 2013; Jacobsen et al. 2009; Paulozzi 2006; Schepers et al. 2014; Thompson et al. 2015).

In summary, the urban scaling theory provides us with a simple power function model to estimate the scaling relationship in our research. The safety in number concept appears to suggest that the scaling relationship governing pedestrian fatalities may be more likely to display a sublinear rather than superlinear scaling pattern. On the other hand, the scaling relationship of pedestrian fatalities in cities is a complex issue subject to influence by multiple factors. Our research aims to discover some of these influencing factors.

Method and data

Following the earlier works on urban scaling by Bettencourt et al. (2007, 2010), we used the same simple power function model to determine the scaling relationship of the 3 traffic fatality measures as a function of the population size of an individual city for a given year. The formula we used for the power function model is

\[ Y_i = a P^b_i, \]  

where \( Y \) is the number of total traffic fatalities including driver, passenger, pedestrian, cyclist, and others. \( Y \) can also represent pedestrian fatalities only or nonpedestrian fatalities. \( P \) is the population size of cities, \( a \) is a constant, and \( i \) indexes the individual city. The exponent \( b \) determines the scaling relationship between \( Y \) and \( P \). When the relationship is superlinear, \( b > 1 \); when the relationship is sublinear, \( b < 1 \); and when the relationship is linear, \( b = 1 \). For example, if city A with 2 million people experiences a proportionately large number of 3 traffic fatalities per 100,000 over city B with 1 million people with only 1 traffic fatality per 100,000, the scaling relationship is superlinear with the value of exponent \( b \) being 1.58. On the other hand, if city A experiences a proportionately lower number of 1.5 fatalities per 100,000 over city B with 1 traffic fatality per 100,000, the scaling relationship is sublinear with the value of exponent \( b \) being 0.58. If city A experiences 2 fatalities per 100,000, then the scaling relationship is linear with the value of exponent \( b \) being 1.0.

Taking the natural logarithm of Eq. (1), we have our estimation equation:

\[ \ln Y_i = \ln a + b \ln P_i + \epsilon_i, \]  

We used Eq. (2) to run a yearly cross-sectional ordinary least squares estimates corrected for heteroskedasticity for the total group of all cities with more than 150,000 people. The number of cities in the total group ranges from 115 cities in 1994 to 161 cities in 2011, reflecting an increase in population. We then ran the same cross-sectional ordinary least squares estimation each year for the 4 subgroups of cities with different population sizes. In order to run the panel data analysis, we expanded Eq. (2) into Eq. (3) as follows:

\[ \ln Y_t = \ln a + b \ln P_t + \epsilon_t, \]  

where \( t \) represents time in years.

To estimate a panel for Eq. (3) for each subgroup, we used a Prais-Winsten regression model with panel-corrected standard errors. The method used the generalized least squares framework that corrects for AR(1) autocorrelation within panels and cross-sectional correlation and heteroskedasticity across panels (N. Beck and Katz 1995).

Next, we expanded Eq. (2) to include other characteristics relating to individual cities. We chose to follow the well-known environmental principle of \( I = PAT \), where \( I \) stands for the environmental impact from population (\( P \)), affluence (\( A \)), and technology (\( T \); Ehrlich and Holdren 1972; Holdren and Ehrlich 1974). More specifically, we used a more refined STIRPAT model known as stochastic impacts by regression on population, affluence, and technology (Dietz and Rosa 1994, 1997; York et al. 2003). The STIRPAT model has been used to examine the relationship between population size and \( \text{CO}_2 \) emissions (Cole and Neumayer 2004; Martinez-Zarzoso et al. 2007; Poumanyvong and Kaneko 2010; Martinez-Zarzoso and Maruitti 2011; Shi 2003). In addition, the STIRPAT model has been used to examine the impact of population, income, and technology in other areas such as material footprint, human ecological footprint, and environmental efficiency of well-being (Dietz et al. 2007, 2009; Fisher-Kowalski et al. 2011; Steinberger et al. 2010).

Although the STIRPAT model had not been used in the analysis of traffic fatality counts in the past, we believe that the conceptual use of the STIRPAT model may be appropriate because traffic fatality counts viewed at a macroperspective similar to \( \text{CO}_2 \) emissions or other environmental and ecological measures may be influenced by such underlying general elements as population size, income level, and technology. Another reason for the use of STIRPAT model is the ready availability of necessary data.

We are proposing that cities with higher income rates may generate a larger number of pedestrian and other traffic fatalities due to more frequent trips made by a larger number of vehicles. On the other hand, cities with lower population densities may generate a larger number of pedestrian fatalities due to fact that road network and infrastructures are designed primarily for vehicle uses, as are many newer cities in the southern part of the United States. Representing personal income per capita (\( I \)) for affluence and population density (\( PD \)) for technology, we expanded Eq. (2) into Eq. (4) and Eq. (3) into Eq. (5) as follows:

\[ \ln Y_i = \ln a + b (\ln N_i) + c (\ln I_i) + d (\ln PD_i) + \epsilon_i \]  

and

\[ \ln Y_t = \ln a + b (\ln N_t) + c (\ln I_t) + d (\ln PD_t) + \epsilon_t. \]  

For the estimation of Eq. (4), we use an ordinary least squares method of cross-sectional multivariate regression corrected for heteroskedasticity. We applied the methodology for the year of 2011 for each subgroup. For the panel data estimation of Eq. (5), we used the same Prais-Winsten regression method.

For our data source, we used Traffic Safety Facts Fatality Analysis Reporting System/General Estimates System annual reports published each year from 1994 to 2011 by the NHTSA. Each annual report contains a table on number of persons killed, population, and fatality rates by city. This table lists the total number persons killed, number of pedestrian deaths, and the U.S. cities with populations of more than 150,000 each year. For
example, the 2011 table lists a total of 161 cities with New York City as the largest with a population of 8,244,910, with 271 being the total number of persons killed and 138 pedestrian killed. Pomona, California, is listed as the smallest with a population of 150,119, with 16 persons killed and 3 pedestrian deaths. There are 161 cities listed in the 2011 report, whereas the 1994 report lists a total of 115 cities. By subtracting the number of pedestrian deaths from the total number killed, we obtained the number of nonpedestrian deaths for each year as well.

In order to obtain income per capita, we used per capita personal income by metropolitan areas from the Bureau of Economic Analysis (2015). Personal income is the sum of net earnings based on place of residence, property income, and personal current transfer receipts (U.S. Bureau of Economic Analysis 2016). Per capita personal income is calculated as the personal income of residents in a given area divided by the resident population. In computing per capita personal income, the Bureau of Economic Analysis uses the Census Bureau’s annual midyear population estimates. Moreover, personal income is measured before the deduction of personal income taxes and other personal taxes and is reported in current dollars (no adjustment is made for price changes).

For the measure of population per unit area, we computed population density by dividing population size by the land area of individual metropolitan areas during the same period (1994 to 2011). The land area data were extracted from the World Atlas. Although there are several methods to calculate population density, we used arithmetic density, which is the total number of people divided by the area of land (measured in square kilometers).

The income per capita and population density data are matched to the individual cities listed in the Traffic Safety Facts reports. The result is that the number of cities with complete data sets ready for multivariate analysis becomes somewhat smaller in comparison to the number of cities used for bivariate analysis, due to missing data. For example, for the year 2011, the total number of cities with complete data sets for multivariate analysis was 97 cities versus 161 cities for bivariate analysis. Similarly, the number of cities with more than 200,000 people included 37 cities for multivariate analysis versus 109 cities for bivariate analysis; 37 cities versus 62 cities with more than 300,000 people, 19 cities versus 27 cities with a population of more than 600,000, and 6 cities versus 9 cities with a population of more than 1 million.

Analysis of results

Results from the cross-sectional and panel data bivariate regression

For all of the cities, ranging from 115 in 1994 to 161 in 2011, the annual cross-sectional bivariate regressions generated nearly linear relations of the population coefficient for pedestrian fatalities. The range of population coefficient is from a high of 1.05 in 2010 to a low of 0.805 in 2000 and the result from the panel data analysis is 0.994. The annual cross-sectional bivariate regressions generated a moderate superlinear relation of the population coefficient for total traffic fatalities. The range of the coefficient is from a high of 1.503 in 1996 to a low of 0.873 in 2001. The panel data analysis of the bivariate regression during 1994 to 2011 generated the population coefficient of 1.052 for total fatalities. As expected, the panel data analysis of nonpedestrian fatalities generated a more pronounced superlinear relation of 1.179 as the population coefficient. All of the population coefficients derived are statistically significant at less than 0.1% level without exception. These results are displayed in Figure 1 and listed in Table A1 (see online supplement). Figure 1 clearly shows that the number of nonpedestrian fatalities increases more than proportionately with population increase, whereas pedestrian fatalities increase sublinearly or nearly linearly with population increase. Figure 1 also shows that the distributions of the annual population coefficients are reasonably consistent throughout the time period within the respective fatality measures.

Next, all of the cities were divided into 4 subgroups of populations with more than 1 million, 600,000, 300,000, and 200,000. Then we ran annual cross-sectional bivariate regressions for the number of pedestrian, total, and nonpedestrian fatalities as a function of population size of cities. We also ran the panel data analysis for each subgroup.

The results of annual cross-sectional and panel data analyses of bivariate regression on pedestrian fatalities are shown in Figure 2 and Table A2 (see online supplement). For the subgroups of cities with populations greater than 1 million, the population coefficients ranged from a high of 0.556 in 1994 to a low of 0.223 in 2004. The coefficient from the panel data analysis is the most significant sublinear relation at 0.354. For the next subgroup with populations greater than 600,000, the population coefficients ranged from a high of 0.976 in 2001 to a low...
of 0.646 in 2004. The coefficient from the panel data analysis is 0.807 with another strong sublinear relation. For the subgroup with populations greater than 300,000, the range of population coefficients is from a high of 0.949 in 1999 to a low of 0.77 in 2005 and the coefficient from the panel data analysis is 0.862. Finally, for the subgroup with populations greater than 200,000, the range is from a high of 1.057 in 2008 to a low of 0.709 in 2000. The coefficient from the panel data analysis is closer to linear at 0.975.

Figure 2 displays the historical distribution of annual population coefficients of pedestrian fatalities by subgroup. Once again, a striking difference exists between sublinear scaling of the largest cities with a population of more than 1 million in comparison to other subgroups of cities. Another important finding is that in all of these city subgroups, population coefficients of pedestrian fatalities display sublinear relations.

Next, the results of population coefficients of total traffic fatalities are shown in Table A3 (see online supplement). For the 8 to 10 cities that have populations greater than 1 million, the population coefficients follow more clear-cut sublinear relations with a high of 0.685 in 1994 to a low of 0.405 in 2004. The coefficient from the panel data analysis is estimated at 0.501. For the 18 to 27 cities that have populations greater than 600,000, the annual population coefficients from the bivariate regressions range from a high of 1.063 in 2000 to a low of 0.740 in 2007. The coefficient from the panel data analysis shows another sublinear relation of 0.82. For the 51 to 62 cities that have populations greater than 300,000, the range is from a high of 1.016 in 1995 to a low of 0.846 in 2005. The coefficient from the panel data analysis is again a sublinear relation at 0.921. Finally, for the 77 to 109 cities that have populations greater than 200,000, the population coefficients range from a high of 1.134 in 2008 to a low of 0.806 in 2001. The coefficient from the panel data analysis is nearly linear at 1.04. A large majority of population coefficients are statistically significant at less than 0.1% level, and all of the coefficients are significant at less than 5% level. Figure 3 displays the historical distribution of annual population coefficients of total fatalities by these subgroups. The most striking difference exists between sublinear scaling of the largest cities with a population of more than 1 million compared to the other subgroups of cities. The implication is that the number of total traffic fatalities is less than proportionate to the population size among these larger cities.

We again use the population size of 600,000 as the dividing line to contrast those cities with superlinear relationships versus those with sublinear relationships in Figure 4. All of the cities with less than 600,000 people combined display a more clear-cut superlinear relation with the panel data result of 1.167.

Lastly, we show the results of annual cross-sectional and panel data analysis of bivariate regression on nonpedestrian fatalities in Figure A1 and Table A4 (see online supplement). For the subgroup of cities with populations greater than 1 million, the population coefficients range from a high of 0.996 in 2008 to a low of 0.739 in 2001. The coefficient from the panel data analysis remains at a sublinear relation at 0.841. For the subgroup with populations greater than 600,000, the range is a high of 1.284 in 2000 and a low of 0.933 in 2007. The coefficient from the panel data analysis is a moderated superlinear relation at 1.101. For the subgroup with populations greater than 300,000, the range is from a high of 1.223 in 1999 to a low of 1.044 in 2007, with a coefficient from the panel data analysis of 1.124. Finally, the subgroup of cities with populations greater than 200,000, the range is from a high of 1.253 in 2008 to a low of 1.069 in 2001. The coefficient from the panel data analysis is 1.217.

Figure A1 displays the historical distribution of annual population coefficients of nonpedestrian fatalities by subgroup. Only the subgroup of cities with populations greater than 1 million display sublinear relations, whereas all other subgroups display superlinear relations.

An alternative approach of subgroup analysis is to create subgroups that are mutually exclusive. For example, we analyzed subgroups of more than 600,000 people versus less than 600,000 people. For pedestrian fatalities, we obtained the panel data population coefficient of 0.807 for the greater than 600,000 subgroup versus 1.186 for the less than 600,000 subgroup. For the total fatalities, the panel data results are 0.882 for the greater than 600,000 subgroup versus 1.167 for the less than 600,000 subgroup.

2011 Cross-sectional and panel data multivariate regressions

As noted earlier, the availability of data on the 2 additional independent variables, income per capita and population density, are limited to a smaller number of cities. In comparison to 161 cities in the group of all cities in 2011 for the bivariate analysis, we have
relevant data for only 97 cities for the multivariate analysis. Similarly, for the subgroup of cities with populations greater than 200,000, the multivariate analysis was based on 63 versus 109 from the bivariate analysis. For the subgroups of cities with populations greater than 300,000, 600,000, and 1 million, the multivariate analysis was based on 37, 19, and 6 cities versus 62 cities, 27 cities, and 9 cities, respectively, from the bivariate analysis.

As shown in Table 1, the 2011 cross-sectional multivariate regression on total fatality generated a population coefficient of 1.11, whereas the panel data multivariate regression generated a coefficient of 1.02. In comparison, the 2011 bivariate regression and the panel data bivariate regression estimated 1.059 and 1.052, respectively, as the population coefficients as shown in Table A1. In other words, the population coefficients from the multivariate models showed a nearly linear relationship between the number of total fatalities and population size of cities, supporting the earlier findings from the bivariate analysis.

To determine the effect of missing cities in the multivariate analysis, we ran a sensitivity analysis by repeating the 2011 bivariate regression with 97 cities only instead of our previous analysis of the total group containing all cities. This nonlinear transformation process applies not only to pedestrian fatalities but to nonpedestrian as well as to total fatalities.

The results of pedestrian fatality shown in Table A5 (see online supplement) from the 2011 multivariate regression and the panel data analysis are also quite similar to the earlier results from the bivariate analysis. The most important finding from pedestrian fatalities is that the same pattern of transformation from a nearly linear population coefficient from the group of all 97 cities moves toward sublinear coefficients for the subgroups of larger cities. In the panel data analysis, the population coefficient of 0.977 for all 97 cities moves to 0.957, 0.832, and 0.768 and eventually to 0.374 for the subgroups of cities with populations of more than 200,000, 300,000, 600,000, and 1 million, respectively. All of these coefficients are again statistically significant at less than 0.1% level. As for income per capita coefficients, all show positive signs and are statistically significant. On the other hand, all of the population density coefficients are statistically not significant with one exception for the group of all 97 cities. In other words, the impact from more densely populated cities versus more dispersed cities in the Southern region is not statistically different as far as pedestrian fatalities are concerned.

Finally, the results of nonpedestrian fatalities shown in the table in Appendix 7 (see online supplement) again display the similar movement of changing population coefficients from the group containing all 97 cities toward the subgroups of larger cities. More specifically, the superlinear population coefficient of 1.17 for the group of 97 cities became 1.201, 1.144, 1.095, and 0.917 for the subgroups with populations of more than 200,000, 300,000, 600,000, and 1 million, respectively. In other words, superlinear population coefficients change to sublinear relations for the subgroup containing the largest cities. Once again, all of these population coefficients are statistically significant at less than 0.1% level. As for income per capita, only 3 subgroups show statistically significant coefficients. Once again, population density coefficients are not statistically significant with one exception for the group containing all 97 cities.

In summary, the overall results from the multivariate analysis confirm population size to be the dominant independent variable in relation to the number of traffic fatalities. Income per capita and

| Subgroup | 6 Cities with more than 1 million people | 19 Cities with more than 600,000 people | 37 Cities with more than 300,000 people | 63 Cities with more than 200,000 people | All 97 cities |
|----------|----------------------------------------|----------------------------------------|----------------------------------------|----------------------------------------|-------------|
| Time \ Variables | 2011 | Panel data | 2011 | Panel data | 2011 | Panel data | 2011 | Panel data | 2011 | Panel data |
| Population | 0.848* | 0.574*** | 0.695*** | 0.872*** | 0.828*** | 0.916*** | 1.024*** | 1.03*** | 1.1*** | 1.0*** |
| (0.198) | (0.081) | (0.123) | (0.074) | (0.095) | (0.035) | (0.096) | (0.031) | (0.068) | (0.028) |
| Population | -0.135 | -0.021 | 0.03 | -0.018 | -0.054 | -0.001 | -0.058 | 0.008 | -0.035 | -1.05*** |
| (0.041) | (0.035) | (0.083) | (0.053) | (0.097) | (0.015) | (0.067) | (0.014) | (0.0049) | (0.029) |
| Density | 0.343 | 0.026** | -0.363 | 0.020** | -0.286 | 0.021** | -0.438 | 0.025** | -1.13*** | 0.01** |
| (0.819) | (0.009) | (0.351) | (0.009) | (0.386) | (0.007) | (0.325) | (0.009) | (0.316) | (0.005) |
| Income | -7.65 | -3.26*** | -4.326 | -7.687*** | -5.800*** | -8.465*** | -7.937*** | -10.138*** | -5.93*** | -9.047*** |
| (4.92) | (1.017) | (1.63) | (0.96) | (1.499) | (0.444) | (1.393) | (0.383) | (1.381) | (0.385) |
| Constant | 0.9204 | 0.967 | 0.6864 | 0.8789 | 0.6848 | 0.8784 | 0.7285 | 0.8249 | 0.7004 | 0.679 |
| R² | 0.02 | 1.0 | 0.02 | 1.0 | 0.02 | 1.0 | 0.02 | 1.0 | 0.02 | 1.0 |

*p<0.05 **p<0.01 ***p<0.001, Standard errors of coefficient estimates appear in parentheses.

Number of cities are based on 2011 data.
Figure 5. Population coefficients of pedestrian, total, and nonpedestrian fatalities by subgroups of cities from panel data multivariate analyses.

capita influences the traffic fatalities in a negative fashion, meaning that more well-to-do cities will experience some increase in traffic fatalities, all other things being equal. On the other hand, population density does not appear to influence the outcome of traffic fatalities in a significant way. The most important finding from the multivariate analysis is to confirm the changing pattern of population coefficients toward sublinear relations as the population size of cities increases.

The most critical findings from our analysis are summarized in Figures 4 and 5. Figures 4 and 5 compare the population scale coefficients of pedestrian, total, and nonpedestrian fatalities by the 5 subgroups of cities. Figure 4 presents the results from the panel data analysis of the bivariate model, and Figure 5 presents the results from the panel data analysis of the multivariate model. Both Figures 4 and 5 show the transformation of population coefficients for the subgroup of megacities and the total group containing all cities in the sample.

In the case of pedestrian fatalities, the most pronounced sublinear relation with coefficients of 0.354 (Figure 4) and 0.374 (Figure 5) for the subgroup of megacities transforms to a moderate sublinear relation for the group of all cities with coefficients of 0.994 (Figure 4) and 0.977 (Figure 5). In general, the relationships between pedestrian fatalities and population size of cities are characterized as sublinear.

In case of nonpedestrian fatalities, the sublinear relation estimated for the group of megacities with the respective population coefficients of 0.841 (Figure 4) and 0.917 (Figure 5) transforms to a superlinear relation for the group containing all cities with population coefficients of 1.179 (Figure 4) and 1.17 (Figure 5). In general, the relationships between nonpedestrian fatalities and population size of cities can be represented as superlinear with the exception of the subgroup of megacities.

In case of total fatalities, the clear-cut sublinear relation with population coefficients of 0.501 (Figure 4) and 0.574 (Figure 5) for the subgroup of megacities again transforms to a linear relation for the group with coefficients of 1.052 (Figure 4) and 1.02 (Figure 5). In general, the relationships between total fatalities and population size of cities can be described as sublinear for larger cities and linear for a majority of all cities in the sample. Another important finding is that this transformation process undergoes a rapid change between the subgroup of megacities and the next subgroup of cities with populations greater than 600,000 for each of the 3 fatality measures. Then the transformation displays a more gradual change to the next subgroups. The transformation process then reaches a plateau, showing a minimal difference in coefficients between the subgroups of cities with populations greater than 200,000 and the group of all cities. This nonlinear transformation process applies not only to pedestrian fatalities but to nonpedestrian as well as total fatalities.

In summary, we have discovered the existence of 3 groups of cities with different population sizes displaying distinctly different scaling patterns. For pedestrian fatalities, the first group of megacities with populations greater than 1 million displays a very pronounced sublinear scaling pattern. The second group of cities with populations greater than 300,000 or 600,000 exhibits a clear-cut sublinear scaling pattern. The third group of cities with populations greater than 150,000 or 200,000 displays nearly the same moderate sublinear scaling pattern. In other words, the number of pedestrian fatalities shows diseconomies of scale as the population size of cities increases.

We have discovered that the same 3 groups of cities with different scaling patterns also exist for nonpedestrian and total fatalities. For nonpedestrian fatalities, the first group of cities displays a moderate sublinear scaling relation, whereas the second group of cities displays a moderate superlinear scaling pattern. The third group of cities displays a more clear-cut superlinear scaling pattern. In other words, the number of nonpedestrian fatalities increases more than proportionately as the population size of cities increases, with the exception of megacities.

As for total fatalities, the first group of megacities shows a pronounced sublinear scaling pattern, whereas the second group of cities displays a more moderate sublinear scaling relation. The third group of cities, on the other hand, displays a linear scaling relation. In other words, the number of total fatalities increases proportionately to the increase of population size of cities when the sample includes a large majority of cities. However, when the sample includes cities with larger population sizes only, the number of total fatalities is influenced by diseconomies of scale.

What implications can we draw from this study? One of the most important implications is that the 9 megacities with populations greater than 1 million are significantly different from the remaining cities and should be viewed as a totally separate group. Then, having a more meaningful and useful comparison of what countermeasures against pedestrian fatalities have or have not worked may be conducted within this group of megacities. In view of the pronounced sublinear scale of 0.354 and 0.374 estimated for pedestrian fatalities within this group, cities with a population size 4 times as large should have about 50% lower fatality rates per 100,000 population as the comparable pedestrian fatality rate. A similar implication applies to the cases of nonpedestrian and total traffic fatalities for this group of megacities, although the degree of sublinear scaling patterns is more moderate.

As for the scaling pattern of all 115 to 161 cities included in this study, what appears to be either linear or nearly linear relations may easily lead one to draw wrong conclusion. Once again, the results of our subgroup analysis indicate that it would be more realistic to separate those cities with populations greater than 300,000 or 600,000 from cities with smaller population sizes. More specifically, we suggest that a group containing the top 27 to 62 cities, excluding megacities, would provide more meaningful comparison. The remaining cities with smaller population sizes will provide another sample group that can be used for more meaningful comparisons.
The major contributions of this article are as follows: first, our findings provide qualified support to the safety in numbers concept. However, the support is limited to cities with population sizes of 600,000 or more. Second, our findings provide partial support to the superlinear urban scaling theory, as applied to pedestrian fatalities and population size of cities. However, the support is limited to cities with populations of 600,000 or less. For cities with more than 600,000 people, the scaling relation is reversed to a sublinear pattern. In short, our findings provide a framework that may be employed for further empirical studies dealing with multiple socioeconomic and environmental factors that influence the outcome of pedestrian and total traffic fatalities in an urban setting.

There are several limitations to our study that should provide future topics for research. In the method of analysis, instead of the power function model we used to analyze scaling relationship, alternative models such as Poisson, negative binomial, or logistic regression can be employed for further analysis.

An equally important limitation of our research is the fact that we have not fully explored several other factors that may significantly influence pedestrian as well as other types of fatalities. For example, pedestrian density in large versus small cities may vary. Furthermore, the relative ratio of female versus male and young versus older pedestrians may vary by population size. A similar question may arise in terms of vehicle and driver density. Once again, the relative ratio of female versus male and young versus older drivers may vary by population size. All of these variations in pedestrian, driver, and vehicle density can influence the scaling relationship of fatalities.

There are several additional local factors such as vehicle type; vehicle speed; traffic control devices; driver age; pedestrian age; distribution of trips by modes such as walking, biking, transit, and car; availability of safe infrastructure for walking, including traffic enforcement and safe intersection design; and several others. The relative variation in these factors may also exist in different population sizes. Many of these local factors are possible candidates to be incorporated in multivariate analysis in future research. Another limitation that needs to be overcome is the lack of usable data that are available for individual cities. Which of these factors is more critical for which type of cities is the empirical question that needs to be explored further again in the context of subgroups of cities with different population sizes. This article represents a modest beginning toward explaining the complex relationship between pedestrian fatalities and the population size of cities.

Acknowledgments

The authors are grateful for insightful comments from 3 anonymous referees that have improved our article. We also acknowledge research assistance provided by our research assistants Sun Joo Park and Eun Sun Lee at the Gachon Center of Convergence Research.

References

Alliance for Biking and Walking. Bicycling and walking in the United States: 2014 Benchmarking report. 2014. Available at: http://www.bikewalkalliance.org/resources/benchmarking. Accessed March 1, 2014.

Beck LF, Paulozzi LJ, Davidson SC. Pedestrian fatalities, Atlanta metropolitan statistical area and United States, 2000–2004. J Safety Res. 2007;38:613–616.

Beck N, Katz JN. What to do (and not to do) with time-series cross-section data. Am Polit Sci Rev. 1995;89:634–647.

Beck SF, Dellinger AM, O’Neil ME. Motor vehicle crash injury rates by mode of travel, United States: using exposure-based methods to quantify differences. Am J Epidemiol. 2007;66:212–218.

Bettencourt LMA. The origins of scaling in cities. Science. 2013;340:1438–1441.

Bettencourt LMA. Lobo J, Helbing D, Kühnert C, West GB. Growth, innovation, scaling and the pace of life in cities. Proc Natl Acad Sci USA. 2007;104:7301–7306.

Bettencourt LMA, Lobo J, Strumsky D. Invention in the city: increasing returns to scale in metropolitan patenting. Res Policy. 2006;36:107–120.

Bettencourt LMA, Lobo J, Strumsky D, West GB. Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. PLoS ONE. 2010;5:e13541.

Bureau of Economics Analysis. Quarterly state personal Income. 2014. Available at http://www.bea.gov/itable/itable.cfm?ReqID=70&step=1&isuri=1&HSID=1 Access May 1, 2015.

City of Chicago. Pedestrian Crash Analysis (2005–2009 Crash Data). Chicago, IL: Chicago Department of Transportation; 2011.

Cole MA, Neumayer E. Examining the impact of demographic factors on air pollution. Popul Environ. 2004;26:5–21.

Decker EH, Kerkhoff A, Moses ME. Global patterns of city size distribution and their fundamental drivers. PLoS ONE. 2007;2(9):e934.

Dietz T, Rosa EA. Rethinking the environmental impacts of population, affluence and technology. Human Ecology Review. 1994;1:277–300.

Dietz T, Rosa EA. Effects of population and affluence on CO2 emissions. Proc Natl Acad Sci USA. 1997;94:175–179.

Dietz T, Rosa EA, York R. Driving the human ecological footprint. Front Ecol Environ. 2007;5:13–18.

Dietz T, Rosa EA, York R. Environmentally efficient well-being: rethinking sustainability as the relationship between human well-being and environmental impacts. Human Ecology Review. 2009;16:114–123.

Ehrlich P, Holdren J. A bulletin dialogue on the ‘closing circle’: critique: one-dimensional ecology. Bull At Sci. 1972;28(5):16–27.

Elvik R. The non-linearity of risk and the promotion of environmentally sustainable transport. Accid Anal Prev. 2009;41:849–855.

Elvik R. Can a safety-in-numbers effect and a hazard-in-numbers effect coexist in the same data? Accid Anal Prev. 2013;60:57–63.

Fischer-Kowalski M, Kraussmann F, Giljum S, et al. Methodology and indicators of economy-wide material flow accounting. J Ind Ecol. 2011;15:855–876.

Florida R. Cities and the Creative Class. New York, NY: Routledge; 2004.

Fragkias M, Lobo I, Strumsky D, Seto KC. Does size matter? Scaling of CO2 emissions and US urban areas. PLoS ONE. 2013;8:e64727.

Fujita M, Krugman P, Venables AJ. The Spatial Economy: Cities, Regions, and International Trade. Cambridge, MA: MIT Press; 2001.

Glaser EL, Gottlieb JD. The wealth of cities: agglomeration economies and spatial equilibrium to the US. J Econ Lit. 2007;45:983–1028.

Holdren JP, Ehrlich PR. Human population and the global environment: population growth, rising per capita material consumption and disruptive technologies have made civilization a global ecological force. Am Sci. 1974;62:282–292.

Jacobsen PL. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj Prev. 2003;9:205–209.

Jacobsen PL, Racioppi F, Rutter H. Who owns the roads? How motorized traffic discourages walking and bicycling. Inj Prev. 2009;15:369–373.

Loaf R, Barthelemy M. How congestion shapes cities: from mobility patterns to scaling. Sci Rep. 2014;4:15561.

Martínez-Zarzoso I, Bengoecha-Moranco A, Morales-Lage R. The impact of population on CO2 emissions: evidence from European countries. Environ Res Econ. 2007;38:497–512.

Martínez-Zarzoso I, Maruotti A. The impact of urbanization on CO2 emissions: evidence from developing countries. Ecol Econ. 2011;70:1344–1353.

NHTSA. Traffic Safety Facts 1994. Washington, DC: Author; 1995. DOT HS 808 292.
NHTSA. *Traffic Safety Facts 2012 Data: Pedestrian.* Washington, DC: NHTSA; 2014. DOT HS 811 888.

Paulozzi LJ. Is it safe to walk in the Sunbelt? Geographic variation among pedestrian fatalities in the United States, 1999–2003. *J Safety Res.* 2006;37:453–459.

Poumanyvong P, Kaneko S. Does urbanization lead to less energy use and lower CO₂ emissions? A cross-country analysis. *Ecol Econ.* 2010;70:434–444.

Samaniego H, Moses M. Cities as organisms: algometric scaling of urban road networks. *J Transp Land Use.* 2008;1:21–39.

Schepers P, Hagenzieker M, Methorst R, Wee BV. A conceptual framework for road safety and mobility applied to cycling safety. *Accid Anal Prev.* 2014;62:331–340.

Shi A. The impact of population pressure on global carbon dioxide emissions, 1975–1996: evidence from pooled cross-country data. *Ecol Econ.* 2003;44:24–42.

Shi A. The impact of population pressure on global carbon dioxide emissions, 1975–1996: evidence from pooled cross-country data. *Ecol Econ.* 2003;44:24–42.

Steinberger JK, Krausmann F, Eisenmenger N. Global patterns of materials use: a socioeconomic and geophysical analysis. *Ecol Econ.* 2010;69:1148–1158.

Strumsky D, Lobo J. Metropolitan patenting, inventor agglomeration and social networks: a tale of two effects. *J Urban Econ.* 2008;63:871–884.

Thompson J, Savino G, Stevenson M. Reconsidering the safety in numbers effect for vulnerable road users: an application of agent-based modeling. *Traffic Inj Prev.* 2015;16:147–153.

US Bureau of Economic Analysis. State personal income. 2014. Available at: http://www.bea.gov/newsreleases/regional/spi/2015/spi0315.htm. Accessed March 1, 2016.

York R, Rosa EA, Dietz T. STIRPAT, IPAT and IMPACT: analytic tools for unpacking the driving forces of environmental impacts. *Ecol Econ.* 2003;46:351–365.