Scientific Research Publishing

Current Urban Studies, 2022, 10, 275-292
https://www.scirp.org/journal/cus
ISSN Online: 2328-4919
ISSN Print: 2328-4900

Analysis of Global Road Traffic Death Data Using a Clustering Approach

Utpal Dutta[1]*, Xiaohui Zhong[2], Nouri Gsouda[1]

[1]Department of Civil, Architectural & Environmental Engineering, Detroit, USA
[2]Department of Mathematics, University of Detroit Mercy, Detroit, USA
Email: *Duttau@udmercy.edu

How to cite this paper: Dutta, U., Zhong, X. H., & Gsouda, N. (2022). Analysis of Global Road Traffic Death Data Using a Clustering Approach. *Current Urban Studies, 10,* 275-292.
https://doi.org/10.4236/cus.2022.102017

Received: March 16, 2022
Accepted: June 13, 2022
Published: June 16, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Three years of global traffic death data were analyzed to identify significant socio-economic determinants of global traffic death. Due to wide variation of data elements, the collected data were standardized and then four clusters namely Medium, Low, High and Medium-High were formed. The stepwise regression analysis method was used to find significant determinants. Proportion of urban population, alcohol, number of registered vehicles per million of population and human development index were appeared as significant determinants. However, income was noticed as a significant variable in the case of only cluster Low. The involvement of older drivers in traffic death incident was very low.

Keywords

Global Traffic Death, Clusters, Determinants, Regression, Socio-Economic Factors

1. Introduction

Traffic crashes have been and continue to be one of the primary causes of mortality worldwide, both in developing and developed countries [WHO, 2018]. Road traffic crashes have a widespread and devastating effect on public health, global economy, and poverty. As shown in **Figure 1**, the World Health Organization (WHO) estimates that every year over 1.35 million people have died from traffic crashes worldwide, and 20 to 50 million suffer non-fatal injuries. Traffic crash injuries are considered the eighth-leading cause of death for people of all ages and are the leading cause of death for young people aged 5 - 29 years [WHO, 2018]. Globally, an average of 186,300 children die every year as a result of road traffic crashes, indicating a humanitarian loss of more than 500 children every day [WHO, 2021].

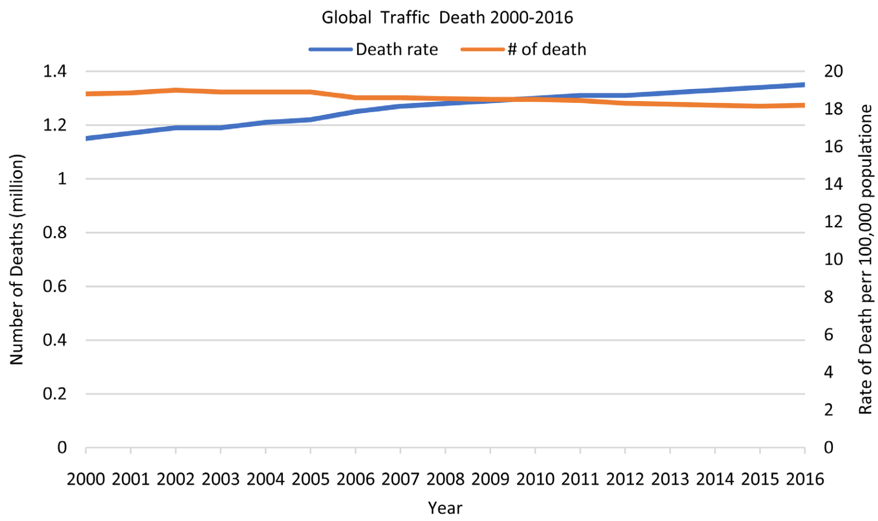

Figure 1. Numbers and rate of road traffic death per 100,000 population: 2000-2016 [WHO, 2018].

World traffic death-related data collected by WHO in three categories based on a country's per capita gross income level (GNI). The countries with a GNI per capita of up to $1005, between $1006 to $12,235 and higher than $12,236, are designated as low-, medium-, and high-income level countries, respectively [WHO, 2018]. It has been observed that developing countries are much more impacted by traffic crashes than developed countries because those countries have not put sufficient investment in institutional capacity to improve road infrastructure, enforce traffic laws (which saves vulnerable road users), and maintain vehicle standards [WHO, 2018]. In one estimate, low- and middle-income countries have 60% of the world's registered vehicles, but experienced over 93% of the world's road traffic deaths [WHO, 2018]. Road traffic death rates in low- and middle-income countries are more than twice of those in high-income countries [WHO, 2018]. Children in low-income countries have road traffic death rates nearly three times as high as those in high-income countries [WHO, 2015b].

In addition to human loss, traffic crashes induce many socio-economic costs. Socio-economic costs including long-term care, hospitalization, material damage, and welfare loss. A significant financial strain can be put on families as a result of road traffic injuries. Many families are driven into poverty because of the cost of traffic crash-related long-term medical care [WHO, 2015a]. In some cases, families may lose the primary breadwinner of the household, or they may need extra funds to care for impaired individuals resulting from traffic crashes. The economically active age groups in low- and middle-income countries are particularly affected by road traffic crashes, which can kill the most productive segments of the population, or those set to contribute to family, society, and the general workforce [WHO, 2015a]. Globally, road crash fatalities and serious injuries are estimated to cost 7 trillion in US dollars for low- and middle-income countries. Overall, road traffic crash deaths and injuries in low and middle-income countries are causing economic losses of up to 6.5% of a country's gross national product (GDP), while in high-income countries are 2% of the GDP [World Bank, 2020; Peden et al., 2004].

Statistical data issued from the WHO shows that all regions of the world are affected by the epidemic of road traffic deaths and injuries, and there has been no decrease in the number of people dying as a result of road crashes in any low-income countries from 2013-2016 [WHO, 2018]. However, substantial improvement has been made in middle and high-income countries to reduce traffic-related deaths [WHO, 2018]. As shown in **Figure 2**, the highest road traffic crash fatality rates in the world were registered in African countries, with an annual rate of 267 per million population in 2016, followed by South-East Asia countries, at 207 per million population, both of which exceed the global average rate of 182 per million population. West Pacific and Eastern Mediterranean countries' rates were close to the global rates with 180 and 169 deaths per million population, respectively. The lowest rates were in Europe and the Americas, with rates of 93 and 156 deaths per million population, respectively [WHO, 2018]. Factors contributing to road traffic crashes vary; they are not simple and straightforward. Some studies found that differences in socio-economic/demographic factors, including income, education, population, education, occupation, culture, residential density, etc., led to the huge variation in fatality risk [Kang, 2001; Nanjunda, 2021]. Other studies have also shown that people with lower social and economic/demographic profiles are often more likely than those with higher socio-economic status to be involved in fatal or non-fatal road accidents [Kang, 2001; Nanjunda, 2021; WHO, 2007].

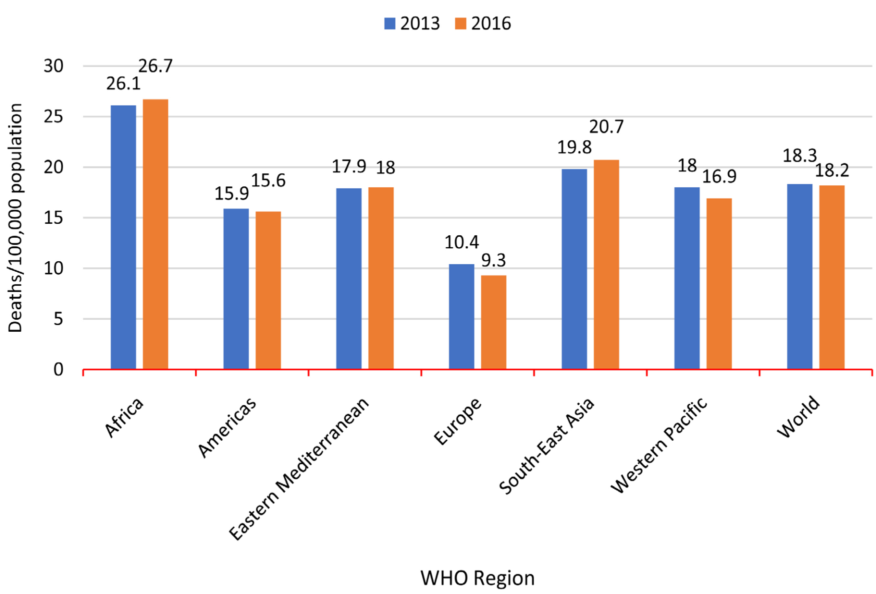

Figure 2. Rates of road traffic death per 100,000 population by WHO regions: 2013, 2016 [WHO, 2018].

Driver-related risk factors, such as drug/alcohol-impaired driving, driver age, and driver health, are often considered as contributing factors in fatal road traffic collisions [WHO, 2007]. Drinking and driving is one of the leading causes of road accidents worldwide, according to WHO's report, "Drinking and Driving: A Road Safety Manual" [WHO, 2007]. Excess alcohol in the blood is found in around 20% of fatally injured drivers in high-income countries, but up to 69% of fatally injured drivers in low-and middle-income countries [WHO, 2007]. Driver age has been described as another significant factor that influences driving behavior and, as a result, is linked to the risk of causing or being involved in accidents. Many studies have shown that the probability of fatal and non-fatal crashes is far higher among the youngest and oldest drivers than among middle-aged drivers [Nanjunda, 2021; WHO, 2007]. All these studies conducted in the last decades concentrated on factors like driver's speeding, behavior, and road/vehicle characteristics; however, the effects of social disparities in traffic crash risks have not been thoroughly researched to date by public health experts yet.

The purpose of this paper is to identify the contributing factors mainly from the social economic aspects of the global traffic death at the macro level by examining three years of traffic death and other related data.

The primary objectives of this effort are:

1) Identify possible relationships between socio-economic/demographic factors and traffic crashes globally and identify the most significant factors.

2) Understand the relationship among these socio-economic/demographic factors.

3) Derive models to relate road traffic crashes and the socio-economic/demographic factors.

4) Elucidate on the problem of traffic crashes and their challenges worldwide.

2. Literature Review

Socio-economic factors related to Traffic Death

Extensive number of studies were done to explore the link between income and road traffic fatalities at a cross-country level [Cociu, 2020; Wintemute, 1985; Jacobs & Cutting, 1986; Söderland & Zwi, 1995; Van Beeck et al., 2000; Kopits & Cropper, 2005; Bishai et al., 2006; Anbarci et al., 2009; Paulozzi et al., 2007]. The general conclusion was that at levels of income below certain threshold, road traffic fatality rate initially increases with income due to increased motorization. After reaching this threshold, traffic fatality rate tends to taper for a range of incomes, and then starts decreasing due to investment in safety measures such as safer cars, preventive measures, better quality roads etc. This inverted U-shaped pattern known as Environmental Kuznets Curve (EKC) was first explicitly pointed out by Van Beeck et al. within the context of traffic fatalities [Van Beeck et al., 2000]. The Kuznets curve (EKC) hypothesized that the relationship between traffic death and economic growth exhibits an inverted U-shape [Grossman & Krueger, 1996]. That is, traffic fatalities increase as economy grows in countries with low income; however, once the economy reaches a certain threshold, fatalities begin to decline due to greater development.

Van Beeck et al. used data from 21 industrialized countries from 1962 to 1990 to examine the impact of Gross Domestic Product (GDP) on traffic fatalities [Van Beeck et al., 2000]. It was observed that the traffic fatality rate rises until per capita GDP reaches to a threshold of around $3000 after which the rate begins to decrease [Van Beeck et al., 2000]. Kopits and Cropper (2005) determined this income threshold to be $8,600 by investigating panel data from 1966 to 1999 of 88 high, medium, and low-income countries [Kopits & Cropper, 2005]. Bishai et al. stated that this income threshold is around $1500 to $8000 [Bishai et al., 2006]. Anbarci et al. concluded that the threshold is $11,454 [Anbarci et al., 2009].

Using province-level panel data in China, Iwata examined the relationship between traffic accidents and economic growth [Iwata, 2010]. Similar to other studies, data exhibited an inverted U-shaped pattern; traffic fatality and injury increase as provincial GDP per capita rise up to about $1500 and $4000 respectively and decline thereafter. However, there is no inverted U-shaped relationship between traffic injury rate and growth in the GDP [Bishai et al., 2006].

Traffic death and other factors:

Factor et al. emphasized the importance of linkage between traffic death and social and cultural characteristics [Factor et al., 2008]. Israeli census data was used with road traffic accident records as a part of this study. While determining the probability of drivers from different social groups to be involved in a fatal or severe road accident, several traffic death-related contributing factors were identified. The factors include religion, lifestyle, gender, age, skill level, and education level. The authors concluded that traffic accidents may be linked to the different habits, skills, and lifestyles of each sub-group.

Anbarci et al. found that corruption, as measured by the international country risk guide, significantly increases the occurrence of fatalities in relatively low-income countries. The authors concluded that this happens via forged driving licenses, low enforcement of rules and regulations, and low vehicle maintenance and security [Anbarci et al., 2009].

Traynor analyzed the relationship between income, population density, the incidence of alcohol abuse and the share of teenage drivers with fatalities across counties in the state of Ohio, U.S. [Traynor, 2008]. He concluded that that the county population density, the presence of interstate highways in rural counties, the prevalence of severe alcohol abuse, the proportion of teen drivers, and the presence of a large college student population all have statistically significant relationships with county fatality rates. On the other hand, La Torre et al. determined that the employment rate and alcohol consumption are important determinants of road traffic fatalities [La Torre et al., 2007].

Söderland and Zwi used data from 83 countries for the year 1990 [Söderland

and Zwi, 1995]. They examined factors such as traffic-related death rate per 100,000 persons per year, traffic related deaths per 1000 registered four-wheeled vehicles per year, the ratio of mid-age to total population mortality, the ratio of the male/female mortality rate, and fatal injuries as a proportion of total injuries. The authors found that GDP per capita is positively correlated with traffic-related deaths per 100,000 population, but negatively correlated with traffic deaths per 1000 registered cars. They also observed that the number of road traffic accident related deaths among youth and elderly people is directly correlated to population density.

Jacobs and Cutting found that the variables having the greatest impact on the number of fatalities are vehicle ownership, per capita GDP, vehicle density, and population per hospital bed [Jacobs & Cutting, 1986]. Bishai et al. showed that increase in per capita GDP growth were associated with an increase in traffic crashes, injuries, and deaths in low-income countries. However, in case of high-income countries an increase in per capita GPD lowers the number of traffic fatalities but does not alter the number of crashes and injuries [Bishai et al., 2006].

Dhibi analyzed data for 120 countries for 38 years period (1975-2012) to identify determinants of the road safety in low- and middle-income countries [Dhibi, 2018]. These data comprised information on traffic fatalities, accidents with injuries, road network quality, traffic exposure and socio-economic variables. He developed relating dependent variables, total traffic fatalities (DEATHS) and total traffic related injuries (INJURED), to independent variables such as income, GDP per capita, and corruption. The main findings concluded the following:

- The income plays an important role in influencing traffic crashes and fatalities.
- The population structure, growth of GDP per capita, population density and the urbanization rate appear to be significant in the DEATHS model.
- The infrastructure variables, especially highways and paved roads are determinants of the number of road traffic fatalities and injuries.
- The behavior related factors such as lack of awareness, a high discount rate on future returns, corruption and lack of democracy, independent of income, play an important role in low and middle income countries.
- Lower rates of fatalities and accidents in high income countries are associated with democracies.
- Many countermeasures and experiences that have proven to be successful in high income countries should be also effective in low- and middle-income countries.

Grimm and Treibich examined five road traffic databases for a prolong time period to establish a linkage between traffic death and other social factors [Grimm & Treibich, 2010]. The databases are the International Road Traffic and Accident Database (IRTAD), the International Road Federation World Road Statistics (IRF), the United Nations Economic Commission for Europe Database

(UNECE), the World Health Organization Database (WHO), and the Community Road Accident Database (CARE). However, they stated that the reliable and systematic data on road traffic injuries and fatalities is scarce.

From the models developed, the authors concluded that:

- In general, an increase in median income by $5000 almost halves the number of road fatalities per 100,000 population.
- In case of medium and high-income countries, there seems to be no strong association between income and fatalities. This seems plausible as here factors such as enforcement of rules and regulations, driving behavior, and traffic density should explain most of the variance, whereas income per capita should only play a minor role.
- Regarding all countries, the main determinants of road traffic deaths per 100,000 population are the number of nurses per 1000 persons, the urbanization rate, the voice and accountability governance index, the number of alcohol related deaths per 100,000 population, the male life expectancy, and the adult literacy rate.
- Population density caused more traffic fatalities and was statistically significant.

It is noted that the past studies used panel cross-country data and were assumed to have homogeneous economic growth effects on the fatality/injury rate across countries.

Highlights of literature review include:

- Significant traffic fatalities impediment factors include population density, gender, income, corruption, alcohol consumption rate, education level, social and cultural characteristics, and number medical facilities per population.
- The relationship between traffic death and income of the countries follows EKG hypothesis, i.e.; traffic fatalities increase as the economy grows in countries with low income; once the economy reaches a certain threshold, then start declining with further economic development.
- The income threshold value for the EKG hypothesis varies from $1500 - $8000.
- There is no inverted U-shaped relationship between traffic injury rate and growth in the per capita income.
- It is very challenging to find good quality global traffic safety data.
- The statistical characteristics of the developed traffic safety models were very weak.
- Many interventions and experiences that have proven to be successful in high income countries are necessarily effective in low- and middle-income countries.
- No studies were done by examining low, medium and high-income counties at the same time. Factor such as education, cultural and social were not considered in any of the studies. Outlier countries such as China, India, Singapore and USA (too many people, too many cars per capita, too many crashes)

did not receive any special attention, even though they may have influenced outcomes. It is noted that within an income category, income level, education level and other varies widely. For this reason, the clustering approach will minimize theses variation because clustering can provide homogeneousness within a data set.

3. Data Collection and Processing

Three years (2008, 2010 and 2013) of traffic death related data of high, medium and low income countries were collected. The sources of data are World Health Organization (WHO), The World Bank, the United Nations (UN), The United Nations Educational, Scientific and Cultural Organization (UNESCO), KOF Swiss Economic Institute, and other international organizations.

Data elements consist of number of road traffic deaths, total population, total area, urban population, alcohol consumption, number of registered vehicles, Education Index, Human Development Index, and GNI. Analysis variables are derived from these elements. The list of countries by GNI categories is presented in Table 1.

Table 1. List of countries under low, medium and high income categories.

Gross National Income (GNI) category by WHO	Name of Countries
Low	Bangladesh, Burkina Faso, Congo, Cambodia, Ghana, Kenya, Kyrgyzstan, Mali, Myanmar, Malawi, Mozambique, Niger, Sierra Leone Tajikistan, Uganda, Zimbabwe
Medium	Albania, Argentina, Angola, Azerbaijan, Barbados, Belarus, Botswana Bulgaria, Brazil, Bhutan, Bolivia, Chile, Costa Rica, Cote d'Ivoire, Cuba, Colombia, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Fiji, Georgia, Guyana, Guatemala, Honduras, Hungary, Iran, Indonesia, Jamaica, Jordan, Kiribati, Laos, Libya, Lebanon, Latvia, Maldives, Malaysia, Mauritius, Morocco, Mongolia, Namibia, Nicaragua Saint Lucia, Sudan, Romania, Serbia, Poland, Lithuania, Sri Lanka, South Africa, Samoa, Suriname Solomon Islands, Seychelles, Thailand, Turkey, Tunisia, Turkmenistan, Tonga, Paraguay, Panama, Peru, West Bank and Gaza Strip
High	Andorra, Australia, Austria Bahrain, Bahamas, Barbados, Belgium, Canada, Cyprus, Croatia, Czech Republic, Denmark, Estonia, France, Finland, Germany, Greece, Iceland, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Malta, Netherlands, New Zealand, Norway, Oman, Portugal, Qatar, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United Arab Emirates,

Three sets of spatial data representing the three years of data are finally prepared for the use of ArcGIS. Variables selected for this study are presented in Table 2. Summary of data is presented in Table 2. Kolmogorov-Smirnov test and *t*-test show that the distributions of the traffic death are significantly different (Table 3) among the three groups of countries with different means. Thus all related variables are standardized using mean and standard deviation. Since China, India, United States, and Singapore represent outliers in population, number of crashes, and vehicle per population, they were removed from the database for this study. They can be studied separately if more detailed information is available.

4. Data Analysis and Modeling

After removing the outliers, the data elements still vary widely within each

Table 2. Variables considered for this study.

Variables	*Definition*
Death/Million Urban Population	Traffic Death/Urban Population in Millions
Vehicles/Million Population	Number of Registered Vehicles/Population in Millions
Vehicles/Million Urban population	Number of Registered Vehicles/Urban population in Millions
Income per Capita ($)	Amount of money earned per person in a nation
Education Index (0 - 1)	Average of Mean Years of Schooling defined by the United Nations Development Program (UNDP)
Human Development Index (0 - 1)	The Human Development Index is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living. The HDI is the geometric mean of normalized indices for each of the three dimensions.
Population density per Sq·km	Population/land area in km^2
Proportion of Urban population %	Urban population/total population
Social Globalization Index (0 - 100)	Developed by World bank considering three factors: Telephone subscriptions, Freedom to visit, International airports
Cultural Globalization Index (0 - 100)	Developed by UNESCO by considering three factors: Gender parity, expenditure on education and Civil freedom
Alcohol Consumption (Lit/capita)	Alcohol Consumption measured by Liter/capita
Average Drivers Ages (years)	Average Drivers Ages

Table 3. Mean and standard deviation of variables by low, medium and high income countries defined by WHO.

Variables	Low Income Countries		Middle Income Countries		High Income Countries	
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Death/Million Urban Population	3210	2650	1370	2990	167	104
Vehicles/Million Population	36,000	38,300	208,000	146,000	610,000	182,000
Vehicles/Million Urban population	147,000	171,000	367,000	277,000	784,000	284,400
Income per Capita ($)	548	277	5243	3006	36457	19379
Education Index (0 - 1)	0.44	0.12	0.63	0.11	0.80	0.07
Human Development Index (0 - 1)	0.48	0.086	0.70	0.081	0.87	0.048
Population Density per Sq·km	187	337	129	201	202	323
Proportion of Urban Population %	27	9.15	58	17.13	78	13.01
Social Globalization Index (0 - 100)	47.04	10.74	67.94	9.30	81.93	8.04
Cultural Globalization Index (0 - 100)	45.93	15.35	61.04	14.37	79.39	15.79
Alcohol Consumption (Liter/capita)	2.51	2.92	4.73	3.47	9.05	3.49
Average of Drivers Ages (years)	36.2	1.31	40.21	3.11	45.21	3.66

GNI category predefined by WHO. In order to address this issue as well as finding finer predictive models, cluster analysis was performed to categorize data set into clusters with more homogeneous characteristics. Models for traffic fatalities were then created within each cluster instead of GNI base.

Cluster analysis can be defined as the art of finding groups within the data [Kaufman & Rousseeuw, 2005]. Cluster analysis is one of the methods of statistical multivariate analysis and it is an exploratory method as well as one of the most important methods of data classification. It refers to procedures aimed at classifying a group of cases or variables in specific ways and arranging them within clusters so that the cases classified within one cluster are relatively similar among them with respect to specific characteristics. The data sets are comparatively less similar between clusters. Accordingly, it can be concluded that the basis of cluster analysis is the arrangement of cases or variables in the form of clusters, which works to reduce the variance within one cluster and to maximize the difference between the different clusters [Grubesic & Murray, 2001; Kumar & Toshniwal, 2015].

The role of cluster analysis is to uncover a certain kind of neutral structure in the dataset, and this tool is not only an important cognitive tool, but can also be used to reduce a large set of data because this tool allows for the replacement of a group of data by its compact characterization [Caruso et al., 2018].

Among many clustering methods in the literature [Wierzchoń & Kłopotek, 2018; Aggarwal & Reddy, 2014], one of the most popular partitioning clustering techniques which has been used extensively in data mining is the k-mean clustering method. This method is used due to the nature of our data sets that is distinct or well separated from each other. To form sensible groups instead of just simply using income, the following variables were used in the cluster analysis: Vehicles/Million Urban population Income/capita, Education Index, Human Development Index, Social Globalization Index, Cultural Globalization Index, Population density per Sq·KM, Proportion of Urban population, Alcohol Consumption, and Average of Drivers Ages.

The K-mean method was used to form 3, 4 or 5 clusters in the initial exploratory. When four (4) clusters were formed, each cluster shows the most homogeneous characteristics. Thus, four clusters were used with various attributes of each clusters presented in Table 4. Distinct characteristics of each cluster are summarized as follows.

Cluster 1—Medium

This cluster consists of 120 data points (representing 1111 and 8 data points of Low, Medium, High income countries respectively) The average age of drivers is about 40 years, which is about 5 years older than cluster High's drivers and 5 years younger than cluster Low and Medium-High's drivers. The median income per capita of this group varies from $840 to $86,790 with an average of $7822, the second lowest among the four clusters. More than 80 percent population of this cluster has an income less than $21,000. For this reason, it can be considered as representation of Medium income countries defined by WHO. It was given a name Medium.

This population consumes an average of 3.5 liters of alcohol per year which is

Table 4. Characteristics of various clusters (*variables were not considered in forming clusters).

Cluster	Medium		Low		High		Medium-High	
Cluster Size	120		45		73		59	
Variables	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
*Death/Million Urban Population	270	200	300	200	80	40	200	90
*Vehicles/Million Population	194, 500	133,400	694,200	53,700	694,000	117,700	392,000	128,400
Vehicles/Million Urban population	330,400	239,300	165,200	17,700	887,000	24,000	603,500	291,000
Income per Capita ($)	$7822	$12700	$1157	$1080	$43038	$17352	$11961	$6481
Education Index (0 - 1)	0.63	0.07	0.43	0.11	0.83	0.06	0.79	0.07
Human Development Index (0 - 1)	0.71	0.07	0.51	0.09	0.89	0.03	0.80	0.004
Population density per Sq. KM	121	143	116	231	136	127	104	128
Proportion of Urban Population %	60.83	17.46	31.34	11.61	77.72	11.30	68.24	15.59
Social Globalization (0 - 100)	67.02	7.32	49.58	10.69	67.02	4.62	78.24	6.17
Cultural Globalization (0 - 100)	59.51	13.24	44.64	15.30	85.64	8.0	73.92	9.14
Alcohol Consumption (Liter/capita)	3.55	2.48	2.61	2.76	9.97	2.19	9.57	3.11
Average Drivers Ages (years)	39.14	2.19	36.47	1.32	46.68	1.61	44.76	2.25

significantly lower than those of clusters High and Medium-High. The average education index is 0.63 on a scale of 0 - 1, compare to 0.43, 0.76 and 0.83 for clusters High, Medium-High, and Low respectively. Sixty percent of population lives in an urban area. One third of the urban population has car ownership, which is twice that of cluster Low. The population density per square mile varies widely from 1.66 to 692 with an average of 120. The average social index is 0.67, which is the second lowest among four clusters (0.67 vs 0.49, 0.77 and 0.78 for clusters Low, High and Medium-High 4).

Cluster 2—Low

This cluster consists of 45 data points (L, M, H: 28, 17, 0), with a maximum income of only $5170. The number of registered vehicles per population and urban population, human development, social and cultural indices are at the lowest among all four clusters. The level of education is low (below 50th percentile). Average drivers are the youngest among four clusters. Alcohol consumption is the lowest among the four clusters. Below one-third of the population resides in an urban setting. More than 62 percent of the data points in this cluster are from low-income countries, with the remaining 38 percent of countries in the middle-income category as defined by WHO. However, as these middle-income countries are at the lower end of income, this cluster can now be considered a representation of the low-income countries as defined by WHO. Thus, this cluster is given the name Low. The number of traffic deaths per million urban population is the highest among the four clusters. However, this variable was not considered in forming clusters. However, this variable was not considered in forming clusters the average social index is 0.67, which is the second lowest among four clusters (0.67 vs 0.49, 0.77 and 0.78 for clusters Low, High and Medium-High).

Cluster 3—High

This cluster consists of 73 data points (L, M, H: 0, 0, 73), with the highest average per capita median income among all four clusters. The minimum income of this cluster population is $13,240 compared to $840 for cluster 1, $131 for cluster Low, and $3,570 for cluster Medium-High. The population of this cluster has the highest education index. Within the four clusters, the highest population density per square km as well percentage of people living in urban area are found in this cluster. The number of registered vehicles per million urban population are the highest among 4 clusters. The average drivers' age is close to 47 years, which is the highest among all four clusters. The average consumption of alcohol is almost 10 liters (significantly higher than clusters Medium and Low). This cluster can be considered a representation of high-income countries. So, it is named High. This cluster experienced the lowest rate of death per million of urban population. It is to be noted that this factor was not considered in forming cluster.

Cluster 4—Medium-High

This cluster consists of 59 data points (L, M, H: 0, 34, 25). This cluster sits between cluster Medium and High. Cluster Medium-High is similar in income to cluster Medium, but they consume as much as alcohol as cluster High. Car ownership is much higher in cluster 4 than in clusters Medium and Low. Like other clusters, the urban population has higher car ownership. However, the population density is the lowest among the four clusters. The percentage of the urban population is closer to clusters Medium and High. The level of education as well as average age of drivers is close to cluster High. The population density is the lowest. About fifty-eight percent of the data points in this cluster consist of data points from the middle-income countries and about forty-two percent of high-income countries. In terms of income, this cluster can be considered somewhere in the middle compared to the other clusters and is named Medium-High. The factor death per million of urban population which was not considered in forming clusters appeared as the second lowest at 200.

Once forming the clusters, correlation analysis was performed to investigate the association of each factor and the variable Death/Million Urban Population within each cluster. This was a preliminary step to explore direction and strength of each factor. The results are presented in Table 5 with significant levels.

Based on the table above, the findings can be listed as follows:

The number of registered vehicles per million urban population is positively correlated to the traffic death. This is true for the all four clusters.

The income per capita is positively correlated with cluster Low. It is noted in past studies that for low income countries, traffic death increases with income. Most of the data point components that represent this cluster are low-income countries and indicate that any increase in the income will lead to an increase in vehicle ownership, thus increase the number of crashes. However, the opposite

Table 5. Correlation analysis considering death/million urban population as dependent variable.

Cluster	**Medium**	**Low**	**High**	**Medium High**
Vehicle per Million Population	0.192†	0.276†	0.040	−0.038
Vehicle per Million Urban Population	0.460†	0.478†	0.480†	0.477†
Income Per Capita	−0.139*	0.152	−0.476†	−0.50†
Education Index	0.067	0.253‡	−0.123	−0.457†
Human Development Index	−0.027	0.266‡	−0.463†	−0.607†
Population Density	−0.078	−0.206	−0.213‡	0.085
Proportion of Urban population	−0.475†	−0.483†	−0.595†	−0.507†
Social Globalization Index	−0.153‡	0.281†	−0.091	−0.283†
Cultural Globalization Index	−0.064	0.175	−0.198‡	−0.369†
Alcohol Consumption	0.216†	0.512†	0.278†	0.022
Average Drivers Ages	0.0478	−0.066	−0.133	−0.289†

Approx. sig levels † $p < 0.001$, ‡ $p < 0.01$, * $p < 0.05$.

effect has been observed in other clusters, especially with cluster High and Medium High, as the correlation was negative, and the strength was moderate. This is compatible with the past studies about an inverted U pattern. Note that in case of cluster High, where 42% of the data points represent the rich countries. The people in those countries can buy high quality vehicles and have more advanced safety tools that will help to reduce the number of traffic crashes.

Alcohol consumption is positively correlated with the traffic death per urban population among the four clusters, with the highest strength seen with cluster Low followed by cluster High. Cluster Low can be related to the weakness of law enforcement, and cluster High's results to the amount of the alcohol consumption in those countries.

For drivers ages, the correlation was negative for clusters Low, High and Medium-High but positive only for cluster Medium.

Human development and social and cultural globalization indexes are negatively correlated with the traffic death per urban population, and cluster two was the only cluster with a positive correlation to the dependent variable because of the magnitude of those indexes which demonstrate values less than those of other clusters.

To further quantify the impacts of different factors on the traffic death rates, linear models are developed. After experimenting with traffic death per million of population, traffic death per million of registered vehicles and traffic death per million of urban population as dependent variable using step-wise regression analysis, it was concluded that traffic death per million of urban population is the most effective dependent variable.

To identify and quantify the significant socio-economic factors, linear models were developed considering Death/Million Urban Population as dependent variables and Registered Vehicles/Population, Registered Vehicles/Proportion of Urban population, Income, Education Index, Human Development Index, Population Density, Proportion of Urban population, Social Globalization, Cultural Globalization, Alcohol Consumption, and Average Drivers Ages variables as independent variables with stepwise regression. Models were derived for each cluster. Only the coefficients with 0.05 significance were kept. The details are listed in **Table 6**.

These models built on the clusters indicate that:

Income impacts the death rate in cluster Low only. It seems that the rate increases as the income increases. Higher proportion of urban population is beneficial to all four clusters in reducing traffic death, especially for countries in cluster Medium.

Alcohol consumption significantly resulted in higher traffic death for the countries in Clusters Low and Medium.

Social and human development index are negatively associated traffic death except for the countries in Cluster Low.

Higher number of register vehicles per million of population positively associated

Table 6. Models of death/million urban population with significant socio-economic determinants by clusters.

Death/Million Urban Population				
Cluster	Medium	Low	High	Medium High
Variables				
Constant	1273	426.6	789	1657
Human Development Index			−853.0	−573
Vehicles/Million Urban Population		473		91.23
Social Globalization Index	−48.0			−4.75
Average Driver Age	−12.5			−12.76
Proportion of Urban population	−5.77	−10.16	−1.73	−1.87
Education Index			228	
Alcohol Consumption	17.7	22.7		
Income Per Capita		0.064		
Vehicles/million population	491.0			
Adjusted R-square	0.469	0.599	0.453	0.653
RMSE	112	128	31.9	47.6

with traffic death for countries in Clusters Low and Medium only. The Proportion of urban population is appeared as a significant variable with negative impact.

Number of registered vehicles per million population appeared as significant variables in Cluster Low and Medium.

Human development index is emerged as a significant determinant cluster of High and Medium-High.

Alcohol consumption is significantly positively related to traffic death.

5. Conclusion

In this study, three years of global traffic death data were analyzed to identify significant socio-economic determinants of global traffic death using cluster based data. The findings are:

Income appears only one time in case of cluster Low as a significant determinant.

Proportion of urban population deters number of traffic death. Urban population may increase pedestrian movement thus reduce travel speed.

It is a well-established fact that alcohol is strongly related to traffic death. This has been also observed in this study.

Number of registered vehicles per million of population has significant influence on the traffic death. However, the number of registered vehicles per urban population is not a significant factor.

Human Development index has appeared as a significant determinant.

It is also observed that older drivers are involved in traffic death to lesser extent.

The proportion of the urban population is negatively correlated to traffic deaths per urban population. Increasing the percentage of the urban population will lead to increased use of other public transportation tools such as buses, trains, trams, bicycles, etc. and will lead to a decreased use of vehicles, resulting in a decrease in the number of traffic deaths; but any increase in the number of registered vehicles per urban population without having the right strategy, such as developing the infrastructures and enforce the traffic safety legislation, will lead to an increase in the number of traffic deaths.

This study was done using country wide data (macrolevel). The quality traffic death data are very scarce. If data are available at city or regional level, the result should more accurate and practical.

Any future studies should examine the trend of traffic death during COVID19 peak period. It has been observed that during 2020 and 2021, in spite of traffic volume as well as number crashes went down significantly, but traffic death increased more than 10 percent in the United States [https://www.michigan.gov/msp, 2021]. Also, the impact of cell phone use on the traffic fatalities as well injuries should be explored.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Aggarwal, C. C., & Reddy, C. K. (2014). *Data Clustering: Algorithms and Applications*. Routledge.

Anbarci, N., Escaleras, M., & Register, C. A. (2009). Traffic Fatalities: Does Income Inequality Create an Externality? *Canadian Journal of Economics, 42,* 244-266. https://doi.org/10.1111/j.1540-5982.2008.01507.x

Bishai, D., Quresh, A., James, P., & Ghaffar, A. (2006). National Road Casualties and Economic Development. *Health Economics, 15,* 65-81. https://doi.org/10.1002/hec.1020

Caruso, G., Gattone, S. A., Fortuna, F., & Di Battista, T. (2018). Cluster Analysis as a Decision-Making Tool: A Methodological Review. In *Proceedings of the International Symposium on Distributed Computing and Artificial Intelligence*. Springer.

Cociu, S. (2020). Environmental Risk Factors Related to Road Traffic Crashes. *Arta Medica, 77,* 93-96.

Dhibi, M. (2018). Road Safety Determinants in Low and Middle Income Countries. *International Journal of Injury Control and Safety Promotion, 26,* 99-107. https://doi.org/10.1080/17457300.2018.1482926

Factor, R., Mahalel, D., & Yair, G. (2008). Inter-Group Differences in Road Traffic Crash Involvement. *Accident Analysis and Prevention, 40,* 2000-2007. https://doi.org/10.1016/j.aap.2008.08.022

Grimm, M., & Treibich, C. (2010). *Socio-Economic Determinants of Road Traffic Accident Fatalities in Low and Middle Income Countries.* International Institute of Social Studies, Working Paper No. 504.

Grossman, G. M., & Krueger, A. B. (1996). The Inverted-U: What Does It Mean? *Environment and Development Economics, 1,* 119-122. https://doi.org/10.1017/S1355770X00000450

Grubesic, T. H., & Murray, A. T. (2001). Detecting Hot Spots Using Cluster Analysis and GIS. In *Proceedings from the Fifth Annual International Crime Mapping Research Conference* (Vol. 26). Springer.

Iwata, K. (2010). The Relationship between Traffic Accidents and Economic Growth in China. *Economics Bulletin, 30,* 3306-3314.

Jacobs, G. D., & Cutting, C. A. (1986). Further Research on Accident Rates in Developing Countries. *Accidents Analysis and Prevention, 18,* 119-127. https://doi.org/10.1016/0001-4575(86)90056-4

Kang, K. (2001). Socioeconomic Characteristics of Speeding Behavior. *Driving Assessment Conference, 1,* 320-324. https://doi.org/10.17077/drivingassessment.1066 https://pubs.lib.uiowa.edu/driving/article/id/28005/

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.

Kopits, E., & Cropper, M. (2005). Traffic Fatalities and Economic Growth. *Accident Analysis and Prevention, 37,* 169-178. https://doi.org/10.1016/j.aap.2004.04.006

Kumar, S., & Toshniwal, D. (2015). A Data Mining Framework to Analyze Road Accident Data. *Journal of Big Data, 2,* Article No. 26. https://doi.org/10.1186/s40537-015-0035-y

La Torre, G., Van Beeck, E., Quaranta, G., Mannocci, A., & Ricciardi, W. (2007). Determinants of Within-Country Variation in Traffic Accident Mortality: A Geographical Analysis. *International Journal of Health Geographic, 6,* Article No. 49. https://doi.org/10.1186/1476-072X-6-49

Nanjunda, D. C. (2021). Impact of Socio-Economic Profiles on Public Health Crisis of Road Traffic Accidents: A Qualitative Study from South India. *Clinical Epidemiology and Global Health, 9,* 7-11. https://doi.org/10.1016/j.cegh.2020.06.002

Paulozzi, L. J., Ryan, W. R., Espitia-Hardeman, V. E., & Xi, Y. (2007). Economic Development's Effect on Road Transport Related Mortality among Different Types of Road Users: A Cross-Sectional International Study. *Accident Analysis and Prevention, 39,* 606-617. https://doi.org/10.1016/j.aap.2006.10.007

Peden, M. et al. (2004). *World Report on Road Traffic Injury Prevention.* World Health Organization.

Söderland, N., & Zwi, A. B. (1995). Traffic-Related Mortality in Industrialized and Less Developed Countries. *Bulletin of World Health Organization, 73,* 175-182.

Traynor, T. L. (2008). Regional Economic Conditions and Crash Fatality Rates—A Cross-County Analysis. *Journal of Safety Research, 39,* 33-39. https://doi.org/10.1016/j.jsr.2007.10.008

Van Beeck, E. F., Borsboom, G. J., & Mackenbach, J. P. (2000). Economic Development and Traffic Accident Mortality in the Industrialized World, 1962-1990. *International Journal of Epidemiology, 29,* 503-509. https://doi.org/10.1093/intjepid/29.3.503

Wierzchoń, S. T., & Kłopotek, M. A. (2018). *Modern Algorithms of Cluster Analysis.* Springer. https://doi.org/10.1007/978-3-319-69308-8

Wintemute, G. J. (1985). Is Motor Vehicle-Related a Disease of Development? *Accident*

Analysis and Prevention, 17, 223-237. https://doi.org/10.1016/0001-4575(85)90055-7

World Bank (2020). *Guide for Road Safety Opportunities and Challenges: Low and Middle Income Country Profiles.*

World Health Organization (WHO) (2007). *Drinking and Driving: A Road Safety Manual for Decision-Makers and Practitioners.*

World Health Organization (WHO) (2015a). *Global Status Report on Road Safety 2015.*

World Health Organization (WHO) (2015b). *Ten Strategies for Keeping Children Safe on the Road.*

World Health Organization (WHO) (2018). *Global Status Report on Road Safety 2018.*

World Health Organization (WHO) (2021). *Road Safety.* https://www.who.int/health-topics/road-safety#tab=tab_1