Statistical modeling of number of human deaths per road traffic accident in Oromia region, Ethiopia

Merga Abdissa Aga (mergabdisa3@gmail.com)
Salale University https://orcid.org/0000-0003-4291-6247
Berhanu Teshome Woldeamanuel
Salale University
Mekonnen Tadesse
Addis Ababa University

Research article

Keywords: Count Regression models, Hurdle Poisson Model, number of human deaths per accident, Oromia regional state, Ethiopia

Posted Date: May 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-19202/v2

Version of Record: A version of this preprint was published at PLOS ONE on May 19th, 2021. See the published version at https://doi.org/10.1371/journal.pone.0251492.

Abstract

Background: Globally, road traffic accidents are leading causes of death among young people in general, and the main cause of death among young people aged 15–29 years. Recently, in Ethiopia, the average number of road traffic accidents has been increasing, particularly the Oromia regional state is experiencing a higher road traffic fatalities. This study was conducted to identify the major factors associated with the number of human deaths by road traffic accident in the Oromia Regional State, Ethiopia.

Methods: We used data obtained from the Oromia Police Commission Bureau that have been recorded on daily basis road traffic accidents from July, 2016 - July, 2017. Count regression models were was used to assess the factors associated with the number of human deaths from road traffic accidents.

Results: Of the total of 3900 road traffic accidents included in the 1188 (30.5%) were giving rise to fatal. The Hurdle models were of better fit than zero inflated Poisson and zero inflated negative binomial model. Thus, the Hurdle Poisson is recommended in this study. Age of the driver 31-50 years (AOR = 0.289, 95%CI: 0.175, 0.479) and higher than 50 years of age (AOR = 0.311, 95%CI: 0.129, 0.751), driver's years of experience 5-10 years (AOR = 0.014, 95%CI:0.007, 0.027), and more than 10 years (AOR = 0.101, 95%CI:0.057, 0.176), vehicle type automobile (AOR = 8.642, 95%CI:2.7644, 27.023), vehicle years of service 5-10 years (AOR = 2.484, 95%CI:1.194, 5.169), and more than 10 years (AOR = 2.639,95%CI:1.268, 5.497), type of accident, vehicle upside down (AOR = 5.560,95%CI: 2.506, 12.336), causes of accident, turning illegal position (AOR = 0.454, 95%CI:0.226, 0.913), area of accident, residential place (AOR = 108.506, 95%CI: 13.725, 857.798), working areas (AOR = 129.606, 95%CI: 16.448, 1021.263), near hospitals (AOR =23.789, 95%CI: 3.038, 186.298), geographical locations, Western zones (AOR = 0.275, 95%CI: 0.167, 0.455), and South east zones (AOR = 0.624, 95%CI: 0.410, 0.950) were significant associated number of human deaths per road traffic accident factors in the study area.

Conclusion:

In this study,30.5% of accidents were giving rise to at least one human death per road traffic accident and different associated numbers of human deaths per road traffic accident factors have been identified. Thus, interventions by the bodies concerned with introduction educational programs that will create awareness about road traffic accidents and the associated human deaths, especially targeting road users, young drivers, passengers and pedestrians.

Background

Road traffic accidents (RTA) constitute a major public health and development crisis globally. Over 1.2 million people die each year worldwide due to road traffic accidents [1], with millions suffering serious injuries and living with long-term adverse health consequences. Road traffic accidents are a leading cause of death particularly among young people, aged 15–29 years [2]. Even though small numbers of vehicles the vast majority of road traffic deaths happened in low income countries. More than 90% of road traffic deaths occur in low-and middle-income countries, yet these countries have just 54% of the world's vehicles [3].

Road traffic accidents place a heavy burden on national economies as well as on households. In low- and middle income countries they particularly affect the economically active age group, or those set to contribute to the family and the workforce in general. Many families are driven deeper into poverty by the loss of a breadwinner, or by the expenses of prolonged medical care, or the added burden of caring for a family member who is disabled as a result of road traffic injury. The economic costs also strike hard at a national level, imposing a significant burden on health, insurance and legal systems. This is particularly true in countries struggling with other development needs, where investment in road safety is not commensurate with the scale of the problem. Data suggest that road traffic deaths and injuries cause economic losses up to 3% of gross domestic product (GDP) globally, whereas in low- and middle-income countries they are estimated at 5% of GDP [3].

The African Region continues to have the highest road traffic death rates. The situation of road traffic accidents is most severe in sub Saharan Africa, where the lives of millions are lost and significant amount of property is damaged. In Ethiopia, the situation has been worsened as the number of vehicles has increased and consequently due to increased traffic flow and encounters between vehicles and pedestrians. Despite government efforts in the road development, road accidents remain to be one of the critical problems of the road transport sector in Ethiopia [4]. Every year, many lives are lost and much property is destroyed due to road traffic accidents in the country. The country has experienced average annual road accidents of 8115 during 2000-2010 [5].

Among many low-income countries in Africa; Ethiopia has a high rate of road traffic injury and death [6].Road accident in Ethiopia is one of the worst accident records in the world, as expressed per 10,000 vehicles. Moreover, road accidents are concentrated in Addis Ababa, which is the capital city of Ethiopia and Oromia regional State, accounting for 58% of all fatalities and two-third of all injuries [7]. During 2006-2015, the average number of road traffic accidents increased in Ethiopia, where Oromia regional state accounts for the majority of total fatalities occurred and Addis Ababa city administration accounts for the major serious and slight injuries as well as property damages [8].

According to [8], Road traffic accident in Ethiopia is showing a rising trend over the last decades. Moreover, higher road accidents are concentrated in Oromia regional State and Addis Ababa. Not only the traffic accidents in Ethiopia are concentrated in Oromia region and Addis Ababa city, but also the volume of motorized traffic is very high as compared to the other regions of the country.

Previous studies conducted on road traffic accident focused on an Addis Ababa city survey of road traffic accidents. Further, these studies employed descriptive analysis, logistic regression, and Poisson regression models to analyze data [9-13]. In the study of the number of human deaths, count regression models are more appropriate than other methods. Many prior researches categorized the count variable as binary, but this may underestimate the total numbers of deaths due to road traffic accident since multiple human deaths are collapsed into a single unit.

Ethiopia is one of the developing countries with low level of income coupled with a high rate of population growth, where Oromia is the largest region in the country with a total population estimated more than 37 Million [14].To expand our understanding about the most common and consistent factors on the risk of human death due to road traffic accident, this study has identified the human, environmental, road and vehicle characteristics associated with human death by road traffic accident in Oromia regional state of Ethiopia. Moreover, to the best of our knowledge, none of these researchers use zero inflated models and other higher count regression models. Therefore, this study aims to identify the major factors determining the number of human deaths per road traffic accident in the Oromia Regional State, Ethiopia via the appropriate count regression models.

Methods

Study setting and design

A retrospective study was conducted based on data obtained from Oromia regional State police commission, Bureau traffic control and crime investigation department. The study included reported road traffic accidents from July, 2016 to July, 2017. The study was conducted in Oromia regional state, which the largest region of the country. The major towns of the region include Adama, Ambo, Asella, Bale Robe, Bishoftu, Fiche, Goba, Jimma, Metu, Nekemte, Sebeta, Sululta, Shashemene, and Wolliso among many others [15]. The Oromia region has the highest traffic movement in the country, next to Addis Ababa. The major road from Addis Ababa to Djibouti passes through this region. The all major roads between Addis Ababa city and other regions capital cities will pass this region.

Study population, sample size and sampling techniques

All registered road traffic accidents from July 2016 to July 2017 that were documented in the Oromia regional state police commission Bureau; traffic control and crime investigation department were included in the sample.

Variables of the study

The response variable of this study was the number of human deaths per road traffic accident in Oromia region measured as 0, 1, 2, 3…A human death includes pedestrians, road users, passengers or other peoples dead due to road traffic accident. The potential predictors related to the number of human deaths per road traffic accident adopted from various literature [16-22]were demographic characteristics (sex of driver, age of driver), environmental characteristics (accident time, road condition, environment of accident, weather condition, location of accident, day of accident), vehicle related characteristics (vehicle type, vehicle length of service), road related characteristics (road pavement), driving experience, education level of driver, driver-vehicle relationship, accident cause, and accident type (collision between vehicles, collision between vehicle and pedestrian, collision between vehicle and animal, vehicle upside down).

Data collection process

Data were collected from Oromia regional state police commission Bureau; traffic control and crime investigation department, using a checklist that was prepared based on the road traffic control and registry format.

Statistical data analysis

We used STATA version 13 [23]to analyze data. We illustrated the distribution of data using tables, percentages, mean variance, skewness and kurtosis. Count regression models were used as a method to model the discrete nature of response variable (number of human deaths per road traffic accidents) measured as 0, 1, 2, 3… [24].

Count regression models, unlike linear regression, have counts as the response variable that can take only nonnegative integer values, 0, 1, 2, 3… measured in natural units on a fixed scale representing the number of times an event (number of human deaths per road traffic accident) occurs in a fixed domain. For count data, the standard framework for explaining the relationship between the outcome variable and a set of predictors includes the standard count regression models (Poisson Regression and Negative Binomial (NB) Regression), Zero-inflated models (Zero-inflated Poisson (ZIP) and Zero-inflated Negative Binomial (ZINB)) and Hurdle models (Hurdle Poisson (HP) and Hurdle Negative-binomial (HNB)) were used.

The conventional Poisson regression model for count data is often of limited use because empirical count data set typically exhibit over-dispersion and/or have an excess number of zeros. This problem can be addressed by extending the ordinary Poisson regression model by including the dispersion term in the model. The family of generalized linear models (GLMs) negative binomial regression model was the solution [25, 26]. However, although these models typically can capture over-dispersion rather well, they are in many applications not sufficient for modeling excess zeros. Zero-augmented models was address this issue by capturing zero counts [27,28]. Conversely, Zero-inflated models [27], are mixture models that combine a count component and a point mass at zero. A comprehensive and up-to-date account of count models and methods as well as the interpretations of fitted count models are provided by [29]. Hurdle models [28] combine a left-truncated count component with a right-censored hurdle component.

Poisson Regression Model

Poisson regression model is the most common technique employed to model count data. Consider a group of p predictors denoted by the vector x' = [$x_1, x_2,\ldots,x_p$] and let y_i represent counts of events (number of deaths) occurring of the random variable Y in a given time or exposure periods with rate μ. Then the probability mass function for a Poison random variable Y is given by;

See formula 1 in the supplementary files.

where y_i = 0, 1, 2, 3… are discrete counts (the number of human death per accident) and μ is the rate parameter [29].

Then the relationship between the predictors the non-negative mean parameter μ_i is the exponential specification given by;

E(y_i) = μ_i = exp ($x_i'\beta$)

wherex_i' = (1, x_{i1}, … , x_{ip}), is a vector of explanatory variables and β = (β_0,β_1, β_2, … ,β_p)' is the corresponding (p + 1) dimensionalcolumn vector of unknown parameters to be estimated.

The unknown parameters of the model were estimated using the maximum likelihood estimation of the log-likelihood function.

Negative Binomial Regression Model

Given that over-dispersion is the norm, thenegative binomial model has more generalitythan the Poisson model. Over-dispersion is most often caused by highly skewed response/dependent variables or often due to variables with high numbers of zeros [30, 31].The probability mass function of a negative binomial distribution random variable Y is given by;

See formula 2 in the supplementary files.

The mean and variance of NB distribution areE($y|\mu$,δ) = μ, and var($y|\mu$,δ) = μ(1 + $\delta\mu$).Where δ is the dispersion parameter [29].The predictor variables related to the parameterμ through the log-link function defined aslogμ = $x_i'\beta$.

Zero-inflated count regression models

In the real life data the major source of over-dispersion is a relatively large number of zero counts and the resulting over-dispersion cannot be modeled accurately with the negative binomial model. In such cases, one may use zero-inflated Poisson or zero-inflated negative binomial models to fit the data [27]. Such models assume that the data are a mixture of two separate data generation processes: onegenerates only zeros, and the other is either a Poisson or negative binomial data-generating process.

Zero- inflated Poisson Regression Model

The excess zeros are a form of over dispersion andfitting a zero inflated Poisson model can account for the excess zeros, but there are also other sources of over dispersion that must be considered. If there are sources of over dispersion that cannot be attributed to the excess zeros, failure to account for them constitutes a model misspecification, which results in biased standard errors.In ZIP models, the underlying Poissondistribution for the first subpopulation is assumed to have a variance that is equal to the distribution’s mean. If this is an invalid assumption, the dataexhibit over dispersion (or under dispersion).The probability distribution of a zero inflated Poisson random variable is given by:

See formula 3 in the supplementary files.

The response variable y_i is a non-negative integer, $\mu_i$is the expected Poisson count for the i^{th} individual; $\omega_i$is the probability of extra zeros.The mean and variance of ZIP distribution are$E(Y_i) = (1 - \omega_i)$ and$var(Y_i) = E(Y_i)(1 + \omega_i\mu_i)$. The parameters and ω_i depend oncovariates xi and z_i respectively, wherelog( ) = $x_i^{'}\beta$ and log ( )= $z_i^{'}\gamma$

Zero-Inflated Negative Binomial Regression Model

Zero-inflated negative binomial regression isoften used for modeling over-dispersed count outcome variables with excessivezeros. Furthermore, theory suggests that the excess zeros are generatedby a separate process from the count valuesand that the excess zeros can be modeledindependently. The probability distribution of a zero inflated negative binomial response variable is given by:

See formula 4 in the supplementary files.

where $\delta > 0$ is an over-dispersion parameter. The meanand variance of the ZINB model are $E(Y_i) = (1 - \omega_i)$ and$Var(Y_i) = (1 - \omega_i)(1 + \omega_i\mu_i + \delta\mu_i)\mu_i$. The parameters μ_i and $\omega_i$depend on vectors of covariates $x' = [x_1, x_2,\dots,x_p]$ and z_i, respectively.The method of Fisher scoring is more appropriate to obtain the parameter estimates of ZINB regression models.

Hurdle Models

The concept underlying the hurdle model is that a binomial probability model governs thebinary outcome of whether a count variable has a zero or a positive value. If the value is positive, the "hurdle is crossed," and the conditional distribution of the positive values is governed bya zero-truncated count model [28].Hurdle count models are two-component models with a truncated count component for positive counts and a hurdle component that models the zero counts. The count model is typically a truncated Poisson or negative binomial regression (with log link). The probability mass function of the response variable y in the hurdle model is given by:

See formula 5 in the supplementary files.

where$y_i$is the value of the dependent variable for the i^{th} person $i = 1, \dots, n$), $z_i$is a vector denoting the number of predictor variables in the zero counts, $x_i$represents a vector denoting the number of predictor variables in the hurdle part, γ is a vector of coefficients belonging to z, and β denotes a vector of coefficients related to *x, f zero* is a probability density function at least binary outcome (0, 1) or counts (0, 1, 2, 3…), and *f count* is a probability density function of counts(0, 1, 2, 3 …).

The regression coefficients of the model are estimated bymaximum likelihood. The f_{zero} part, where $y_i = 0$ is typically modeled with a binary logit (logistic regression) model, where all counts greater than 0 are given a value of one [32].

Using a binary logistic regression model for thispart, the probability of $y_i = 0$ is denoted as;

See formula 6 in the supplementary files.

where$z_i$represents the observed data and γ the vector of coefficientsbelonging to z_i. Obviously, the probability of a nonzerocount is given by $1 - \psi_i$.The non-zero count part *(fcount)*is modeled with a truncated (y_i>0) count model. This is typically a truncated Poissonmodel or a negative binomial model in case of over-dispersion.

Results

Descriptive Statistics

The total number of road traffic accidents that happened in the study period was about 3, 900, out of which 1,188 accidents involved the death of 1,541 people. Asshown in Table 1, the variance of the response variable (number of human deaths per accident) is greater than its mean, suggesting a possibility of over-dispersion. Table 1 showed more than 69% of the accidents were nonfatal suggesting excess zeros in the dataset(**Table 1).**

Further, the histogram in Figure1 is highly skewed to the right showing massive counts of zero outcomes. However, a large number of human deaths per accident were less frequently observed, implying that dual regime event count models such as zero-inflated models and hurdle models will often tend to indicate over-dispersion in the data as a result of a large number of zero counts (**Figure 1).**

Table 2 shows that the mean number of human deaths per accident for male drivers is almost the same as that for female drivers.The average number of human deaths per accident was inversely related with driver's age, where the highest average number of human deaths per accidents (mean = 0.4874) was seen among the drivers in the youngest age group of 18-30 years. Considering the experience of the drivers, the highest mean number of human deathsoccurredamong drivers with less than 5 years of driving experience (mean = 0.4362), while the smallest mean number of human deaths (mean = 0.2361) was reported among drivers with more than 10 years of driving experience.Employed drivers had the highestimpactof the mean number of human deaths (mean = 0.4399); whilethe smallest contribution was made by owner drivers (mean = 0.1161). Considering the vehicle type,the highest mean number of human deaths of (mean = 1.2517) was attributable to buses with 13-45 seats, whilst the smallest mean number of human deaths per accident (mean = 0.1983) wasdue to minibuses and automobiles.Although itshowed that the highest vehicle service years (10 year or above), was related with the highest mean number of human deathsper road traffic accidents (mean = 0.5997).

Table 2illustrated that drinking during cloudy, rain, cold, or hotweather conditionswas associated with experienced a higher mean number of human deaths per accident (mean = 0.4146). Concerning the accident time, the highest mean number of human deaths per road traffic accident occurs in the evening time. Regarding the geographical distribution,the highest mean number of human deaths per accident occurred in central zones (special zones surrounding Finfinne and, Shoa zones mean =0.4496), whereas the lowest mean number of human deaths per accidentoccurs in the western zones of the region (mean = 0.2113). Furthermore, the highest mean number of human death was happeningnear residential areas. As type of accident was concerned, accidentcollision between vehicle and pedestrianhad the highest mean number of human deaths (mean = 0.5366), followed by collision between vehicles and animals (mean = 0.4826), and collision between vehicles(mean = 0.4361). Comparatively, accidents that turned vehicles upside down had the smallest mean number of human deaths per accident (mean = 0.2145).

Regarding the educational level of driver's the highest mean number of human deaths (mean = 0.4449) was reported among driverswith at most primary education, whilethe smallest mean number of human death was (mean = 0.0564) seen among drivers having more than high school education level. Moreover,the highest mean number of human deaths per accident occurred during the weekend as compared to the other days.The highest mean number of human deaths per accident happened on site clearing roads and gravel roads as compared to the other road pavements such as asphalt roads. Similarly, a higher mean number of human deathswere observed in roads that were not straight (sloppy, curved, sharp, uphill) than straight roads. Concerning the causes of accidents, the highest mean number of human deaths occurredduetosteeringproblem (mean = 0.8034) followed by drivers denyingpriority to pedestrian (mean = 0.7085). On the contrary, over-speed and overload accounted for the lowest mean number of human deaths. Whereas, turning to illegal direction, brake problem, release of tyreand not keeping the minimum required distance betweenhad an intermediate mean human death per accident(Table 2).

Test of over dispersion, Goodness-of-fit test, and comparison

Initially we fitted Poison model. The fitted Poison model was tested for over-dispersion. It was found that the over-dispersion parameter is significant (deviance statistics = 4060.98, p-value = 0.035) and (Pearson Chi-square = 5679.44, p-value < 0.0001). Further, the six different models namely; Poisson, NB, ZIP and ZINB, HP and HNB models were compared using Log pseudo likelihood, AIC and BIC in order to select an appropriate model which fits the data well. Moreover, the Vuong statistic (Voung = 27.74, p-values <

0.0001) for comparing ZIP versus Poisson model wassignificant, implying that the ZIP model is preferred to the Poisson model for predicting the human death per accident. Likewise, the calculated value of the Vuong test statistic (Voung = 28.45, p-values < 0.0001) for comparing ZINB versus NB models was significant, indicating that the ZINB model is preferred to NB regression model. Finally, the Log pseudo likelihood, AIC and BICfound that the HP model was the most appropriate model to fit the number of human deaths per road traffic accident than other models sinceit has the smallest AIC, smallest BIC and maximum log pseudo likelihood(**Table 3).**

Factors associated with the number of human deaths per road traffic accidents

The estimated Hurdle Poison regression model was presented in Table 4. The model has two parts. The first part is from the equation predicting counts for numbers of human death per road traffic accidents (truncated Poisson with log link) interpreted similarly to the standard Poison model. The second part of the model predicts the zero hurdle model (binomial with logit link) zero deaths versus not zero deaths.

The covariates age of driver, driver experience, type of vehicle, vehicle service years, road condition, time of accident, location, environment of accident, type of accident and accident cause were statistically significant predictors of human deaths in the zero truncated Poisson part.The multivariable HP analysis presented in Table 4, illustrated that drivers aged 30-50 years old were associated with a 71% (adjusted odds ratio (AOR) = 0.289; 95% confidence interval (CI):0.175, 0.479) decreased risk of experiencing number of human death per accident due to road traffic accident, whereas those drivers aged at least 50 years old were associated with a 69% decreased risk of experiencing human deaths (AOR = 0.311; 95%CI: 0.129, 0.751)as compared to drivers 18-30 years oldholding all other variables in the model constant. Driver’s year of experience wasfound inversely statistically significant. Driverswith experience of 5-10 years had a98% (AOR = 0.014; 95%CI: 0.007, 0.027)lowerrisk experiencing the number of human deaths per road traffic accident, likewise drivers with at least 10years of experience had an 89.9% (AOR = 0.101; 95%CI: 0.057, 0.176)lowerrisk of expected number of human deaths due to road traffic accidents compared to drivers with less than 5 years driving experience.

As vehicle type was concerned, compared to taxi minibus with up to 12 seatautomobiles were 8.642(AOR =8.642; 95%CI: 2.764, 27.023) times more likely to experience number of human deaths per road traffic accident. Moreover, vehicle length of service 5-10 years (AOR = 2.484; 95%CI: 1.194, 5.169) and at least 10 years (AOR = 2.639; 95%CI: 1.268, 5.497) was associated with a higher risk of experiencing number of human deaths per road traffic accident as compared to vehicles served not more than five years. Furthermore, theexpected number of human deaths per road traffic accident was 4.739(AOR = 4.739; 95%CI: 2.400, 9.356) times higher for dry roads as compared towet roads.The average number of human deaths per road traffic accident occurred during the evening and night time was statistically significantly lower than the number of human deathsoccurred in the afternoons(AOR = 0.472; 95%CI: 0.369, 0.603) and (AOR = 0.480; 95%CI: 0.370, 0.623), respectively.

The result in Table 4, also illustrated that geographical region, western zone, which include the four Wollega zones was associated with 72% (AOR = 0.275; 95%CI: 0.167, 0.455) decreased risk expected number of human deaths while the south Eastern zone, including Arsi and Guji was associated with 36% (AOR = 1.363; 95%CI: 1.084, 1.887) and 1.380 (AOR =0.624; 95%CI: 0.410, 0.950) reduced risk of expected number of human deaths compared to the special zones surrounding Finfinne. Furthermore, the results showed that the odds of experiencing number of human deaths per road traffic accident was higher among the accidents occurred near residential (AOR = 108.506; 95%CI: 13.725, 857.798), government office areas(AOR = 129.606; 95%CI: 16.448, 1021.263)and around hospitals (AOR = 23.789; 95%CI: 3.038, 186.298) compared to those accidents occurred near factories. But it was lower in the rural village areas (AOR = 0.152; 95%CI: 0.028,0.810)

The result also illustrated that the accident type vehicle upside down(AOR = 5.560; 95%CI: 2.506, 12.336) and vehicle crash to inert (AOR = 1.453; 95%CI: 1.009,2.092)was associated with experienced a higher expected number of human deaths per road traffic accidentcompared collision between vehicles. In contrast, the number of human deaths per road traffic accident due to illegal turning was 54.5% (AOR = 0.454; 95%CI: 0.226, 0.913) lower as compared to the number of human deaths by accidents due to drivers not giving priority to pedestrians.

The second parts of Table 4 providesestimated odds for the factor change in the odds of being in the zero counts group (no human death per road traffic accident faced) compared to the non-zero countgroup (at least one human death per road traffic accident). In the zero counts group, there is no human death per road traffic accident but there are human injuries and property damages. Age of driver, vehicle-driverrelationship, road condition, time of accident, location of accident, environment of accident, road pavement and accident cause had significant associated with the probability of being in the zero counts group. The odds of being in the zero countsgroup was 19.89 (AOR = 19.897; 95%CI: 3.175, 124.685) times more likely among drivers in the age range at least 50 years as compared to those aged18-30 controlling other variables in the model.With regard to vehicle-driver relationship, the odds of being in the zero counts group was 0.096 (AOR = 0.096; 95%CI: 0.028, 0.336) times less likelyamong owner drivers as compared to employee drivers. In contrast, the odds of being in the zero countgroup were increased by a factor of 91.464(AOR = 91.464; 95%CI: 46.138, 181.317) for road traffic accidentson dry roads as compared to accidentson wet roads.

Time of the accident had significant associationwith the odds of being in the zero counts group. For instance, as compared to accidents occurred in the afternoon, the odds of being in the zero counts group reduced by 86.6% (AOR = 0.134; 95%CI: 0.020, 0.609), and 99% (AOR = 0.009; 95%CI: 0.0002, 0.411) for accidents occurred in the evening and night time,respectively.Similarly, as the location of accidents was concerned the odds of being in the zero counts, was reduced by 98.7% (AOR = 0.013; 95%CI: 0.0007, 0.224), 95.6% (AOR = 0.044; 95%CI: 0.010, 0.192),92.8% (AOR = 0.072; 95%CI: 0.021, 0.253), and 92.6% (AOR = 0.074; 95%CI: 0.014, 0.400), for western zones, south western zones, eastern zones and south eastern zones respectively so as compared to the central zones. On the other hand, the odds of being in the zero countsgroup was 174.96 (AOR = 174.96; 95%CI: 25.519, 1199.53) and 23.079(AOR = 23.079;

95%CI: 1.397, 381.369), times more likely for road traffic accidents in commercial areas and worship areas respectively, as compared to road traffic accidents around factory areas.The odds of being in the zero countsgroup was reduced by 99 % of road traffic accidents at road pavements on defected asphalt roads as compared to those on road pavements on good asphalt roads holding all other variables in the model constant.Whereas, the odds of being in the zero countsgroup was reduced by 99.9% of road traffic accident due to over-speeding as compared to those due to denying priority to pedestrian (Table 4).

Discussion

This study was carried out to identify the major factors associated with the number of human deaths per road traffic accident based on Oromia region road traffic accident data. The total number ofaccidentsfromJuly 2016 to July 2017 was included in the present study. Of the total of 3,900 accidents 1188 (30.5%) had experienced at least one human death per accident. The hurdle Poisson regression model was found the most appropriate model from other possible count models.

Driversaged18-30 had a higher impact to the number of human deaths per road traffic accident as compared to those driversabove the age of 30 years. This result was similar to previous studies in Addis Ababa [13, 19]and Bahir Dar city [12].The driving experience of less than 5 years wasassociatedwith a higher human death perroad traffic accidents as compared to those who have more than 5 yearsdriving experience. Similarly, prior studies reported that the smaller driving experience the higher probability of experiencing human deaths per accident [16-18]. Among vehicle types, numbers of deaths due to road traffic accidents by automobiles weresignificantly greater than other vehicle types. This finding is consistent with the results of Towelde [13] andAhmed [34].Consistent to the findings by Fikadu [20], vehicles served for a long time were associated with accidents resulting in a higher number of human deaths per road traffic accident as compared to accidents tovehicleshaving served less than 5 years. The possible reason for thismight be attributed to the poor maintenance of vehicles.

Accident time was another factor significantlyassociated with the number of human deaths per road traffic accident. The less number of human deaths per road traffic accident took place in the evenings and night times as compared to afternoons. This finding is similar to the findings of previous studies [20, 21, 35] and this variation in road traffic accidents by times of the day reflects variations in traffic volumes.Furthermore, the higher number of human deaths per accident was occurring in the residential areas, hospital areas and government office areas, whereas, fewer number of human deaths per accident happened in rural villages than near factory areas**.**This finding is congruentwith the results of previous studies[21, 34]**.**Consistent with the findings of Getahun [19], vehicle upside down and others accident types such as vehicle down, vehicle crash to inertwere associated with the increased number of human deaths per roadtraffic accident as compared to the number of human deaths due to collision between Vehicles.

The number of human deaths per road traffic accident that took place due to turning in anillegal direction was lower as compared to the number of deaths due to drivers denying priority to pedestrians. It was observed that the number of human deaths per road traffic accident due to denying priority to pedestrians was more than the number of deaths by accidents due to other causes. This result is similar to the finding of studies conducted in central Ethiopia [20] and Bahir Dar[12].Similar to a study in Dire Dawa[22], the cause of about 80% of car accidents was attributed todriver faults among which denying priority for pedestrian was the leading.

Conclusion

The study attempted to identify the major factors associated with the number of human deaths per road traffic accident using an appropriate count regression model. The findings from this study show a high prevalence of human deaths per road traffic accidents, in the study area which is 30.5% of accidents were give rise to at least one human deaths and different associated number of human deaths per road traffic accident factors have been identified.

Thus, interventions by the bodies concerned withintroduction educational programs that will create awareness about road traffic accidents and the associated human deaths, especially targeting road users, young drivers, passengers and pedestrians. All concerned bodies should pay special attention to vulnerable road users. There is a need to focus on road infrastructure investments that allow the separation of vulnerable from other road users forimproved road rule enforcement. Besides, as longvehicle years of service and accidents occurred near residential areas, workplaces and hospitals are significantly associated witha higher expected number of human deaths per road traffic accident, attention should be given to control used vehicles and old vehicles on the road, and controlling traffic system near residences and workplacesto reduce the number of human deaths pear road traffic accident. Finally, furtherstudy should be conducted to identify mechanisms for addressing the causes of road traffic accidents and consequences in the study area that will show the seasonal variation.

Declarations

Limitation of the study

This study is based on secondary data obtained from the Oromia Police Commission Bureau. However, the reliability of the data related to driver's age, vehicle service years and the driving experience was questionable as these were collected by interviewing the drivers themselves. Moreover, some very useful data for variables such as the utilization of seat belts and helmets, use of alcohol and talking ona mobile phone while driving were not available.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no financial support for the research, authorship, and/or publication of this article.

Authors' contributions

MAA conceived the original idea of the study, design the study, analyzed the data and drafted the manuscript. MT contributed to the design of the study,interpretation of findings and revision of the manuscript. BTW involved with the conception of the study, statistical analysis, interpretation and revision of the manuscript. All authors read and approved the final manuscript

Acknowledgements

The Authors acknowledge the Oromia regional state Police Commission Bureautraffic control and crime investigation departmentfor kind permission to use the all registered road traffic accidents data for this study.

Abbreviations

AIC:Akaike Information Correction

AOR: Adjusted odds ratio

BIC: Bayesian Information Correction

CI: Confidence Interval

GDP: Gross Domestic Product

GLMs: Generalized Linear Models

HNB: Hurdle Negative Binomial

HP: Hurdle Poisson

NB: Negative Binomial

RTA: Road Traffic Accidents

SD: Standard Deviation

ZIP: Zero-InflatedPoisson

ZINB: Zero-Inflated Negative Binomial.

References

1. Margie Peden, World Report on Road Traffic Injury Prevention,2004.
2. Global plan: decade of action for road safety 2011-2020. Geneva: WHO, 2010.
3. Road Traffic Injuries Home Page: WWW.Who.Int/Violence-Injury-Prevention. 2015
4. UNECA United Nation Economic Commission for Africa. Case Study: Road Safety in Ethiopia. 2009
5. Central Statistics Agency, Number of Traffic Accidents Recorded and Victims by Type of Accident in Ethiopia. Addis Ababa, 2011.
6. Persson A.**Road traffic accidents in Ethiopia: Magnitude, causes and possible interventions.***Advances in Transportation Studies an International Journal*,15 `pp. 5-18, 2008
7. Getu S.Causes of Road Traffic Accidents and possible counter measures on Addis Ababa-Shashemene Roads,Ethiopia. MSc Thesis Addis Ababa University department of Civil 2007
8. AmanuelKassa. Trends, Causes, and Costs of Road Traffic Accidents in Ethiopia, College of Civil Service,Ethiopia. 2017
9. Anteneh K.Road Traffic Accident related Fatalities in Addis Ababa City, Ethiopia. Addis Ababa University School of medicine Department of Emergency Medicine. 2015
10. Dawit O. Road Traffic Accident and Safety Evaluation Case of Addis Ababa Bole Sub City, Ethiopia. Addis Ababa University School of Civil Engineering. 2016
11. Fikadu M.Road Traffic Accident Causes and Control Mechanisms in Addis Ababa City. Addis Ababa University Department of Public Administration and Development Management. 2015
12. Haile M. and Demeke L. Analysis of Factors that affect Road Traffic Accidents in BahirDar City, North Western Ethiopia. *Science Journal of Applied Mathematics and Statistics*, 2:5: 91-96. 2014
13. Tewolde M. Empirical Analysis on Traffic Accidents Involving Human Injuries. MSc. Thesis Presented Addis Ababa University, Department of Statistics. 2017
14. Central Statistics Agency.Population Projection for Ethiopia 2007-2037 trends of population growth rate. Addis Ababa, 2013
15. FDRE States: Basic Information, Oromia. The Federal Democratic Republic of Ethiopia. 2008.

16. Niezgoda M, Kamiński T, Kruszewski M, Tarnowski A. Self-reported drivers' behaviour: an application of DBQ in Poland. J KONES. 2013;20(1):233–8.
17. Machin MA, Sankey KS. Relationships between young drivers' personality characteristics, risk perceptions, and driving behaviour. Accid Anal Prev. 2008;40(2):541–7.
18. Bener A, Yildirim E, Özkan T, Lajunen T. Driver sleepiness, fatigue, careless behavior and risk of motor vehicle crash and injury: population based case and control study. J Traffic Transport Eng (English Edition). 2017;4(5):496–502.
19. Getahun W.(2015). Count Regression Models of Human Death by Road Traffic Accidents in Addis Ababa, Ethiopia .Msc. Thesis presented Addis Ababa University.
20. Fekede A., Demeke A. and Gezahegn T. (2014).Magnitude of, trends in, and associated factors of road traffic collision in central Ethiopia. Department of Public Health, College of Health and Medical Sciences, Haramaya University, Ethiopia
21. Bisrat, M.(2010). Determinants of Traffic Fatalities and Injuries in Addis Ababa.UnpublishedMsc thesis, Department of Statistics, Addis Ababa University.
22. BelachewMelese H. and ZelekeDutamo A.(2015), Statistical Analysis of Road Traffic Car Accident in Dire Dawa Administrative City, Eastern Ethiopia. Science Journal of Applied Mathematics and Statistics. Vol. 3, 250-256
23. Stata Corp. Statastatistica software: release 13. College: Station: StataCorp LP; 2013.
24. Hur K, Hedeker D, Henderson W, Khuri S, Daley J. Modeling clustered count data with excess zeros in health care outcomes research. *Health Service Outcomes Research Methodology*. 2002;3(1):5–20.
25. Mccullagh P. and Nelder J.A. Generalized Linear Models. Chapman and Hall, London. 1989
26. Nelder, J .A. and Wedderburn R.W.M., Generalized linear models, *Journal of the statistical society*. 1972
27. Lambert, D. Zero-inflated Poisson regression, with an application to defects in Manufacturing, *Technometrics*; 34:1–14.1992
28. Mullahy, J. Specification and Testing of Some Modified Count Data Models.*Journal of Econometrics,* 33, 341-365.1986
29. Cameron A.C. and Trivedi P.K. Regression Analysis of Count Data. Second edition Cambridge University Press. 2013
30. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics. 2001;57(1):120–5.
31. Vuong Q.H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:2:307-333. 1989
32. Min and A. Agresti. Random effect models for repeated measures of zero-inflated count data. *Statistical Modeling* 5: 1–19.2005
33. Alan Agresti: Categorical data analysis. 3rd edition, Wiley Inter science, New York. 2002

34. Ahmed Abdella.Statistical analysis of correlates of number of fatalities per traffic accident in Addis Ababa using count data models. Unpublished MSc thesis, Addis Ababa University Department of Statistics, Ethiopia. 2013

35. Getu S. Tulu, Simon W. and Mark J. King. Characteristics of Police-reported Road Traffic Crashes in Ethiopia over a Six Year Period. Queensland University of Technology. 2013

Tables

Table 1: Distribution of number of human deaths per road traffic accident (n = 3900).

Count (No. of deaths)	N	%
0	2712	69.5
1	961	24.6
2	169	4.3
3	29	.7
4	14	.4
5	8	.2
6 or higher	7	.18
Mean	0.3951	
Variance	0.7586	
Skewness	4.211	
Kurtosis	37.398	

Table 2: Descriptive statistics RTAby predictor variables

Predictor variables	Categories	n	Mean	Variance
Gender of the Driver	Male	3605	0.3953	0.5066
	Female	295	0.3932	1.4231
Age of Driver	18-30	1986	0.4874	0.9109
	31-50	1121	0.3568	0.2297
	51 and above	793	0.2181	0.1707
Driving Experience	Less than 5 years	2696	0.4362	0.7325
	5-10 years	698	0.3510	0.2281
	Above 10 years	504	0.2361	0.1807
Vehicle driver relationship	Employee	3355	0.4399	0.6364
	Owner	379	0.1160	0.1028
	Others	166	0.3072	0.2141
Vehicle Type	Taxi-minibuses (up to 12 Seats)	1457	0.1983	0.1591
	Automobile	90	0.5555	0.2497
	Pick up (up to 10 quintals)	375	0.2960	0.2089
	Cargo (11-40 quintals)	416	0.2764	0.2005
	Cargo (41-100 quintals)	284	0.3838	0.2373
	Cargo with trailer (up to 400 quintals)	503	0.3837	0.2369
	Buses (13-45 seats)	290	1.2517	1.1094
	Buses (above 46 seats)	114	0.9824	3.9288
	Others	371	0.5364	1.4872
Vehicle service	Up to 5 years	2296	0.3649	0.2318
	6-10 years	927	0.3204	0.4058
	Above 10 years	677	0.5997	1.928
Road Condition	Wet	1817	0.4259	0.6863
	Dry	414	0.4082	0.2421
	Muddy	1669	0.3583	0.5358
Weather Condition	Not normal(cloudy, rainy, cold, hot)	3432	0.4146	0.6252
	Normal	468	0.2521	0.1889
Accident Time	Afternoon	1309	0.4553	0.9928
	Morning	705	0.3049	0.2122
	Evening	1303	0.4029	0.3728
	Night	583	0.3516	0.5136
Location of the Accident	Central zones	2371	0.4496	0.7960
	Western zones	407	0.2113	0.1670
	South west zones	311	0.3633	0.2320

	Easter Zones	359	0.4011	0.2409
	South east zones	452	0.2920	0.2338
Environment of the Accident	Factory place	303	0.1320	0.1149
	Residence place	447	0.8009	2.6037
	Commercial place	491	0.4562	0.2486
	Religious place	340	0.3471	0.2273
	Entertainment place	355	0.3662	0.2327
	School place	361	0.2770	0.2008
	Hospital place	325	0.3323	0.2225
	In village rural area	454	0.2378	0.1817
	Out of village in rural area	442	0.2262	0.1754
	Government office place	382	0.6675	0.8052
Accident Type	Collision between Vehicles	2004	0.4361	0.9011
	Collision between vehicle and pedestrian	341	0.5366	0.2494
	Collision between vehicle and animal	201	0.4826	0.2509
	vehicle upside down	769	0.2145	0.1687
	Others	585	0.3795	0.2358
Educational level of the driver	Elementary school and below	2778	0.4449	0.6885
	High School	679	0.4123	0.3695
	Above high school	443	0.0564	0.0533
Day of the weeks	Monday	659	0.3050	0.2123
	Tuesday	557	0.3052	0.2124
	Wednesday	529	0.3043	0.2121
	Thursday	574	0.3048	0.2123
	Friday	495	0.3050	0.2124
	Saturday	640	0.6547	1.7788
	Sunday	446	0.5919	0.9612
Road pavement	Good Asphalt	1405	0.3388	0.2242
	Defected Asphalt	1220	0.3877	0.7691
	Gravel	1152	0.4392	0.7174
	Site clearing	123	0.6992	1.2120
Road inclination	Not straight(sloped, curved, scarp, uphill)	3416	0.4124	0.6265
	Straight	484	0.2727	0.1987
Accidentt cause	not given priority to pedestrian	525	0.7085	0.2068

Over load	222	0.0990	0.0896
Over speed	246	0.0203	0.0199
Turning illegal position	480	0.2229	0.1736
Steering problem	580	0.8034	2.2272
Brake problem	546	0.3095	0.2141
Release of tyre	614	0.3306	0.4076
following not keeping distance between	687	0.2867	0.2048

Table 3: Model selection and fit statistics for model comparison

Selection criteria	Models					
	Poisson	NB	ZIP	ZINB	HP	HNB
Logpseudo likelihood	-1616.283	-1571.679	-1307.637	-1395.554	**-914.64**	-1219.76
AIC	3400.567	3331.359	2759.279	2989.108	**2037.28**	2649.52
BIC	3927.14	3920.62	3210.623	3609.713	**2627.41**	3144.73
Vuong test (p-value)			27.74 (<0.0001)	28.45(<0.0001)		

Table 4: Factors associated with number of human deaths per road traffic accident, Hurdle Poison model

riables	Count Model coefficients (truncated Poisson with log link)			Zero hurdle model (binomial with logit link)		
	Odds ratio	95% CI for Odds ratio	p-values	Odds ratio	95% CI for Odds ratio	p-values
nder of driver ale	1					
male	1.113	0.969, 1.28	0.1290	0.132	0.0008, 20.918	0.4340
e of driver						
-30	1					
-50	0.289**	0.175, 0.479	< 0.0001	11658.88	4.891E-07, 2.78E+14	0.4420
and above	0.311*	0.129, 0.751	0.0090	19.897**	3.175, 124.685	0.0010
perience of ver						
5 years	1					
.0 years	0.014**	0.007, 0.027	< 0.0001	0.0005	2.698E-17, 1.03E+10	0.6290
ove 10	0.101**	0.057, 0.176	< 0.0001	0.0348	3.236E-11, 37426516	0.7520
hicle-driver nership						
nployee	1					
vner	0.822	0.289, 0.429	0.7130	0.096**	0.028, 0.336	0.0000
hers	1.168	0.607, 2.246	0.6420	54.833	0.007, 404707.6	0.3780
ucation level of ver						
imary/below	1					
gh school	1.029	0.975, 1.087	0.2940	0.980	0.224, 4.287	0.9790
ove high school	1.003	0.573, 1.753	0.9920	0.096	3.8007E-07, 24074.54	0.7110
pe of vehicle						
xi-minibuses (up 12 seat)	1					
tomobile	8.642**	2.764, 27.023	< 0.0001	131.837	1.763E-07, 9.85E+10	0.6400
:k up (up to 10 intals)	0.379	0.085, 1.679	0.2010	747.915	1.606E-06, 3.48E+11	0.5160
rgo (11-40	0.453	0.094, 2.182	0.3240		4.686E-	0.5160

intals)				$1.28e^{+7}$	15, 3.5E+28	
rgo (41-100 intals)	1.059	0.197, 5.692	0.9460	$2.29e^{+11}$	2.759E-08, 1.9E+30	0.2390
rgo with trailer ɔ to 400 intals)	3.372	0.506, 22.477	0.2090	$5.20e^{+16}$	2.767E-06, 9.79E+38	0.1410
ses (13-45 seat)	4.614	0.626, 33.989	0.1330	$2.44e^{+20}$	1.656E-41, 3.6E+81	0.5140
ses (above 46 at)	6.499	0.863, 48.938	0.0690	$2.65e^{+20}$	1.615E-41, 4.35E+81	0.5130
hers	4.328	0.584, 32.073	0.1520	$5.37e^{+19}$	2.503E-42, 1.15E+81	0.5280
hicle length of rvice						
5 years	1					
.0 years	2.484*	1.194, 5.169	0.0150	66.691	4.623E-36, 9.62E+38	0.9230
ove 10 years	2.639*	1.268, 5.497	0.0090	0.978	0.241, 3.975	0.9750
ad condition						
et	1					
y	4.739**	2.400, 9.356	< 0.0001	91.464**	46.138, 181.317	0.0000
uddy	0.937	0.829, 1.058	0.2930	0.158	0.015, 1.652	0.1230
ne of accident						
ternoon	1					
orning	0.709	0.179, 2.806	0.6240	0.736	0.020, 27.153	0.8680
ening	0.472**	0.369, 0.603	< 0.0001	0.134*	0.020, 0.609	0.0090
ght	0.480**	0.370, 0.623	< 0.0001	0.009*	0.0002, 0.411	0.0160
cation of cident						
ntral zones	1					
estern zones	0.275**	0.167, 0.455	<0.0001	0.013*	0.0007, 0.224	0.0030
uth west zones	0.991	0.779, 1.259	0.9400	0.044**	0.010, 0.192	0.0000
stern zones	0.810	0.608, 1.081	0.1520	0.072**	0.021, 0.253	0.0000
uth east zones	0.624*	0.410, 0.950	0.0280	0.074*	0.014, 0.400	0.0030
vironment of cident						

ctory place	1					
sidence place	108.506**	13.725, 857.798	< 0.0001	58.843	5.795E-22, 5.97E+24	0.8800
mmercial place	0.475	0.109, 2.061	0.3200	174.960**	25.519, 1199.53	0.0000
ligious place	0.378	0.116, 1.235	0.1070	23.079*	1.397, 381.369	0.0280
tertainment ice	0.439	0.067, 2.859	0.3890	63.152	5.150E-19, 7.74E+21	0.8610
hool place	0.707	0.381, 1.313	0.2720	.0005	1.677E-17, 1.78E+10	0.6360
spital place	23.789*	3.038, 186.298	0.0030	48.039	3.59E-23, 6.43E+25	0.8910
ral village area	0.152*	0.028, 0.810	0.0270	$1.28e^{-09}$	1.55E-21, 1064.976	0.1440
t of village in ral area	0.268	0.064, 1.128	0.0730	$1.41e^{-07}$	1.6229E-19, 121649.6	0.2610
vernment office ea	129.606**	16.448, 1021.263	<0.0001	64.255	2.2128E-22, 1.87E+25	0.8800
ad Pavement						
od asphalt	1					
fected asphalt	0.947	0.441, 2.035	0.8900	0.00001*	3.098E-09, 0.039	0.0060
avel	0.497	0.220, 1.125	0.0930	3.15e-06	8.161E-19, 12174328	0.3920
e clearing	0.567	0.253, 1.267	0.1670	6.12e-06	7.081E-19, 52909317	0.4300
pe of accident						
llision between nicles	1					
llision between nicle and destrian	1.408	0.861, 2.303	0.1730	6.872	1.9758E-08, 2.39E+09	0.8480
llision between nicle and animal	1.298	0.784, 2.147	0.3110	0.813	0.109, 6.095	0.8410
hicle upside wn	5.560**	2.506,12.336	0.0000	14.758	2.929E-08, 7.44E+09	0.7920
hers	1.453*	1.009, 2.092	0.0440	0.485	0.105, 2.239	0.3530
cident cause						
t given priority pedestrian	1					
er load	0.749	0.359, 1.559	0.4390	7.76e-09	2.3157E-39,	0.6030

					2.6E+22	
er speed	0.401	0.077, 2.096	0.2790	1.24e-13**	8.178E-17, 1.89E-10	0.0000
rning illegal sition	0.454*	0.226, 0.913	0.0270	7.24e-14*	4.574E-23, 0.000115	0.0050
eering problem	0.693	0.249, 1.925	0.4820	1.87e-21	2.160E-84, 1.61E+42	0.5190
ake problem	0.726	0.219, 2.411	0.6010	6.13e-18	1.428E-40, 262663.8	0.1360
lease of tyre	0.708	0.256, 1.963	0.5070	6.56e-22	3.312E-85, 1.3E+42	0.5120
llowing not eping distance tween	0.699	0.415, 1.179	0.1800	1.31e-07	5.887E-19, 29339.79	0.2350

** Significant p-value < 0.001, * significant p-value < 0.05

Figures

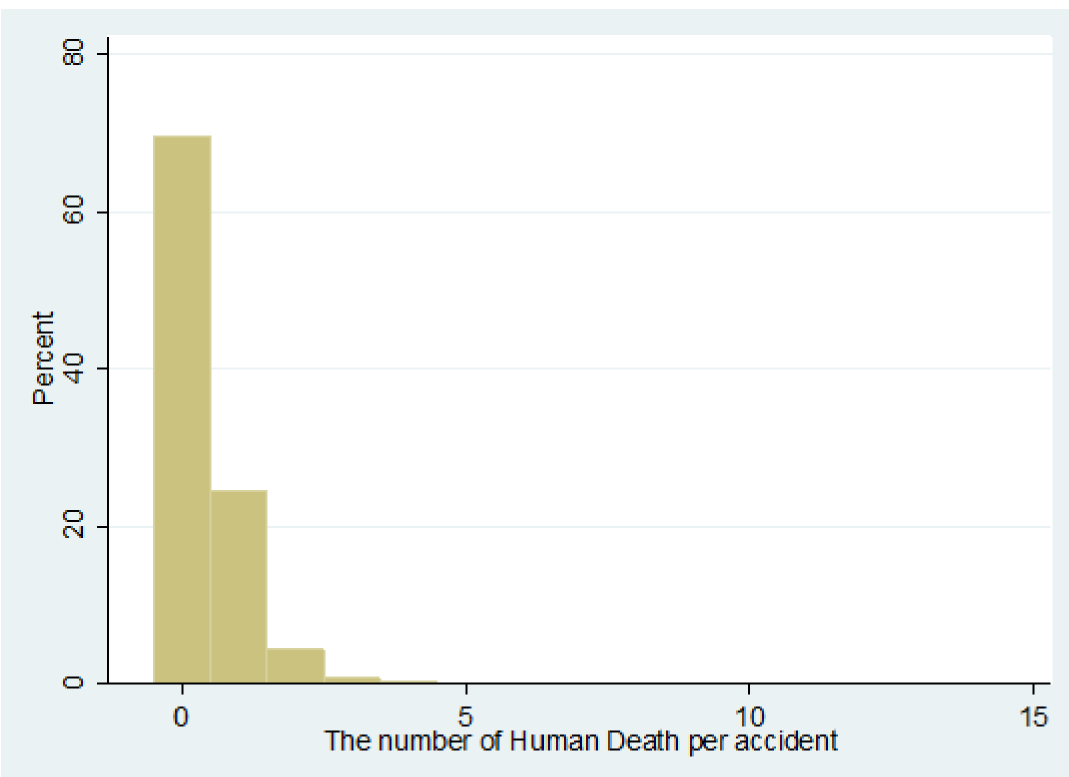

Figure 1

Distribution of response variable, count of human death per accident