

Research Paper

A Comparative Study of Road Traffic Mortality Rates Using Time Series Forecasting Models



Manoochehr Babanezhad^{1*}, Hassan Khorsha², Ali Mohajervatan³

1. Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Iran.

2. Health Management and Social Development Research Center, Golestan University of Medical Sciences, Gorgan, Iran.

3. Department of Anesthesia and Pre-Hospital Emergency Care, School of Paramedical Sciences, Golestan University of Medical Sciences, Gorgan, Iran.



Citation Babanezhad M, Khorsha H, Mohajervatan A. A Comparative Study of Road Traffic Mortality Rates Using Time Series Forecasting Models. *Iranian Journal of Health Sciences*. 2025; 13(3):221-232. <http://dx.doi.org/10.32598/ijhs.13.3.567.4>

doi <http://dx.doi.org/10.32598/ijhs.13.3.567.4>

ABSTRACT

Background and Purpose: The increase in road traffic accidents (RTAs) due to urban and suburban expansion and rising vehicle numbers complicates mortality trend analysis. This study aims to compare RTA mortality rates in urban and suburban areas of Golestan Province, north of Iran, using time-series forecasting models.

Materials and Methods: This retrospective study examined all RTA data (n=37, 409) recorded by the emergency medical service system of Golestan Province from March 2021 to March 2023 for urban and suburban areas. We employed three forecasting models, including multiple logistic regression (MLR), autoregressive integrated moving average (ARIMA), and propensity score matching (PSM), within a time-series data framework, to evaluate their performance in predicting RTA mortality rates in urban and suburban areas and find the demographic factors predicting the mortality rates. We calculated the root mean square error (RMSE) and the mean absolute percentage error (MAPE) to evaluate the prediction accuracy of each model.

Results: The survival rate was 98.7% (n=36931) and only 1.3% (n=478) led to death. Over the two-year period, the RTA mortality rate was significantly higher in suburban areas (1.6%) than in urban areas (0.8%) (P=0.001). The PSM model outperformed other models with lower RMSE and MAPE for both urban and suburban areas. Age and oxygen saturation (SPO₂) were the significant predictors of RTA mortality rate.

Conclusion: The RTA mortality rate is higher in suburban areas of Golestan Province than in its urban areas. The PSM model provides higher prediction accuracy than the RTA and MLR models for both urban and suburban areas.

Keywords: Time series, Road traffic accidents (RTAs), Forecasting, Propensity score matching (PSM), Autoregressive integrated moving average (ARIMA)

Article info:

Received: 21 Feb 2025

Accepted: 12 Mar 2025

Available Online: 01 Jul 2025

* Corresponding Author:

Manoochehr Babanezhad, Associate Professor.

Address: Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Iran.

Tel: +98 (17) 32245882

E-mail: m.babanezhad@gu.ac.ir



Copyright © 2025 The Author(s);

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY-NC: <https://creativecommons.org/licenses/by-nc/4.0/legalcode.en>), which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Introduction

Road traffic accidents (RTAs) are a major global health challenge, accounting for approximately 1.19-1.3 million deaths each year. Alarming, 93% of these deaths occur in low- and middle-income countries, even though these countries own only 60% of the world's vehicles [1-6]. Iran is among the countries with a high mortality rate from RTAs, with nearly 16,000 lives lost annually. The accidents are especially prevalent in high-traffic areas of the country, like the roads of northern provinces, where poor road conditions, frequent traffic violations, and insufficient law enforcement exacerbate the incidence of RTAs [6-8]. The consequences of these accidents, in addition to death, include physical disabilities, emotional trauma, and financial burdens on families and the healthcare system [7-12]. Contributing factors include human error (e.g. exceeding speed), vehicle defects, inadequate infrastructure, and environmental hazards [13-17].

While prior studies have examined risk factors and policies for RTAs, critical gaps still remain, including a lack of comparative urban-suburban analyses [5-15] and insufficient attention to temporal trends [18-22]. A comparative analysis of RTAs and related mortality rates between urban and suburban areas provides essential insights into disparities in risk factors, infrastructural challenges, and policy requirements. However, most studies rely on cross-sectional data, missing seasonal variations and long-term patterns that can enhance efforts to prevent accidents [7-10, 21-25]. Time-series data, though underutilized, offers valuable predictive insights into accident dynamics [21-29]. For instance, urban areas may experience peak RTAs during rush hours due to congestion, whereas suburban fatalities might surge on weekends or holidays as a result of high-speed driving. Forecasting models can predict these fluctuations, enabling tailored traffic control measures for each setting. Similarly, suburban roads, which are often less maintained than urban roads, may see a higher RTA-related mortality rate during adverse weather conditions such as rainy or snowy days. Additionally, urban and suburban expansion may alter accident rates over time. Forecasting models can help planners to anticipate future risks and optimize infrastructure development before the occurrence of accidents. However, few studies have used advanced forecasting methods such as multiple logistic regression (MLR), the autoregressive integrated moving average (ARIMA), and propensity score matching (PSM)

models [29-31] to analyze urban and suburban RTA trends closely.

This study addresses these gaps by analyzing temporal trends of RTAs in Golestan Province, north of Iran, providing evidence for better traffic management and emergency response. This study aims to explore various aspects of RTA mortality rates resulting in urban and suburban areas of the province using MLR, ARIMA, and PSM forecasting models. The MLR model is useful for studying how confounding factors relate to a binary response variable, such as mortality across urban and suburban areas. The ARIMA model provides insights into temporal trends and patterns of mortality across urban and suburban regions. We also employed the semi-parametric PSM method to identify comparable features between urban and suburban areas. This involves distinguishing groups based on their propensity score (PS), which are defined as random effects within a mixed-effects model. The novelty of this study is the integration of PSM with a time-series data framework, offering a fresh perspective on forecasting RTA mortality rates across urban and suburban areas. The key questions are: What are the differences in RTA mortality rates between urban and suburban areas? How can the proposed forecasting models help in decision-making for staffing in emergency response centers in urban and suburban areas? Which model is more effective for forecasting RTA mortality rates in these areas?

Materials and Methods

Study design and data

This retrospective cohort study was carried out using the RTA data (n=37,409) reported by [Emergency Medical Services \(EMS 115\)](#) from March 2021 to March 2023 in Golestan Province, located in northeast Iran. The inclusion criteria were cases such as RTAs injuries that need medical care, fatalities at the scene, dangerous situations, and mass casualty events. On the other hand, the exclusion criteria were the cases that are usually not recorded by the [EMS 115](#), such as minor accidents without injuries, medical emergencies not related to traffic, non-urgent transportation, and incidents not occurring on the road.

This study measured multiple factors related to RTA scenarios, including age and gender of the individual involved in the accident, the duration between the ambulance being dispatch and its arrival at the scene (D1), the duration between receiving the message and

the ambulance's arrival at the treatment center (D2), peripheral oxygen saturation level (SPO₂) at the time of accident, location of accident (urban or suburban area), year of accident (from March 2021 to March 2023), month of accident, and time of accident (daytime or nighttime). Descriptive statistical methods were utilized to summarize the data, including calculations of Mean±SD.

Forecasting methods

We used forecasting models MLR, ARIMA, and PSM within a time-series data framework. Although MLR and PSM are not conventionally standard models for time-series forecasting, they can be effectively applied to analyze the time-series data. The ARIMA model excels in forecasting continuous time series, while the MLR model is suited for classification tasks. The choice depends on whether the focus is prediction of values (ARIMA) or probabilities (MLR) and how well temporal dependencies are handled. For rigorous time-series classification, extending MLR with autoregressive terms or PSM model may be necessary, because, the MLR method may involve the inclusion of lagged variables to tackle autocorrelation by accounting for the temporal structure [29-31]. Regarding the PSM model, it is essential to consider time-related confounding factors such as age, SPO₂, and gender; and also, potentially match within time strata or employ time-dependent covariates [31]. We apply these three different models to our dataset and assess their performance using the root mean square error (RMSE) and the mean absolute percentage error (MAPE) criteria [22]. These two errors criteria are calculated as Equations 1 and 2:

$$1. RMSE = \sqrt{\frac{1}{n-k} \sum_{t=k+1}^n (Y_t - \hat{Y}_t)^2}$$

and

$$2. MAPE = \left(\frac{1}{k} \sum_{t=1}^k \frac{|Y_t - \hat{Y}_t|}{Y_t} \right) \times 100,$$

where Y_t and \hat{Y}_t are the observed and predicted values, respectively, for all time t , n is the total number of observations and k is the samples size ($k < n$). Furthermore, the auto-correlation functions of the residuals for diagnostic testing were conducted.

ARIMA model

ARIMA is a popular time-series model designed for forecasting using historical data. It extends the ARMA model by including differencing, making it suitable for non-stationary data. If no differencing is needed ($d=0$),

ARIMA reduces to ARMA. While ARMA works specifically with stationary time series, ARIMA is tailored for handling non-stationary ones. This model helps to identify and forecast trends over time, providing a temporal context that is crucial for understanding how urban/suburban area development and increased vehicle numbers affect RTA rates. By applying the ARIMA model, researchers analyze temporal patterns, which are crucial for understanding the influence of the two above-mentioned factors on RTA rates over time. Conducting residual diagnostics is crucial for evaluating the validity of the ARIMA model, ensuring that the residuals exhibit independence. For this purpose, the Ljung-box portmanteau test was utilized to assess the adequacy of the fitted ARIMA model. This statistical test is based on the autocorrelation of the squared residuals. The Ljung-box portmanteau test statistic is calculated as Equation 3:

$$3. Q = n(n+2) \sum_{k=1}^h \frac{r_k^2}{n-k},$$

where n is the total number of observations, h is the number of lags to include in the test, and r_k indicates the sample autocorrelation coefficients at lag k . The test statistic Q follows a chi-square distribution with $h-w$ degrees of freedom where h is the number of lags the test is computed for and w is the number of parameters used in fitting the model. Our data consists of information spanning two years. This data is regarded as a sample, and the calculated r_k at different lags ($k=1, 2, \dots, h$) represents the sample autocorrelation of the residuals. The range of their variation is $(-1, 1)$. To perform the Ljung-box test, we first determine the sample autocorrelations r_k for the specified lags up to h . We assess the P ; if the $P > 0.05$, it confirms the null hypothesis, indicating that the residuals do not have significant autocorrelation and that the ARIMA model is appropriate. In contrast, if the $P < 0.05$, we reject the null hypothesis, suggesting that significant autocorrelation exists in the residuals, and the ARIMA model is inappropriate.

The Ljung-box portmanteau test results of our study revealed that the residuals from our fitted ARIMA model do not exhibit significant autocorrelation, supporting the model's adequacy for the time-series data. They follow an ARIMA (p, d, q) model, where p represents the Autoregressive order, d denotes the number of differencing steps applied to the data to achieve stationarity, and q signifies the Moving Average order. Our dataset was non-stationary. By applying a single time difference ($d=1$), we transformed it into stationary data. Subsequently, we generated plots for the

autocorrelation function and the partial autocorrelation function to determine the suitable orders p and q . Consequently, we concluded that our data was well-suited for the ARIMA (1, 1, 1) model. Finally, differentiating the autoregressive moving average model (ARIMA) (1, 1, 1) using the Ljung-box test, it was significant ($P > 0.05$). The key advantage of the Ljung-box test is that it helps assess whether the model has adequately captured the structure of the data or whether significant patterns remain unexplained.

MLR model

The MLR model analyzes the impact of multiple independent variables on a binary outcome (e.g. accident outcome such as survival/death), allowing researchers to control for confounding factors. This approach can be adjusted for time-series analysis, particularly when the outcome is binary and time-dependent, and the predictors involve lagged variables [29]. The model is described as Equation 4:

$$4. \text{Logit } P(Y_t=1) = \beta_0 + \beta_1 X_t + \epsilon_t$$

Where Y_t is the actual observed response variable (death or survival), X_t is the set of explanatory variables, and ϵ_t is the error term at time t .

PSM method

PSM is a semi-parametric approach that matches groups with similar attributes, enabling a more accurate prediction of RTAs. This method distinguishes between urban and suburban areas based on their PS, which are treated as random effects in a mixed-effects model. By identifying and matching similar characteristics among different groups, this method aims to reduce selection bias and enhance the validity of comparisons. This estimation technique allows for the examination of data over time (longitudinal data setting), which is important for capturing changes in RTA rates relative to urban growth and policy changes. This innovative combination allows for a more nuanced understanding of the factors influencing RTAs, facilitating comparisons across different urban and suburban contexts. In such instances, it is essential to ensure that the distribution of covariates such as SPO_2 is balanced at each time point.

We used the PS derived from the Cox proportional hazards model, which is employed in risk set matching. We employed the PS to predict RTA mortality rates. The outcome was binary; 1 (death) or 0 (survival), determined by factors including D1, D2, age, gender, SPO_2 , location (urban, suburban), year, month, and time factors. Logistic regression was employed, and the nearest neighbor matching technique was applied in this context [31]. Then, individuals were matched to assess mortality risks among comparable risk profiles

Table 1. Number of RTA patients (survived or dead) from March 2021 to March 2023 based on confounding factors

Confounders		No. (%) / Mean \pm SD		P
		Survived (n=36931)	Death (n=478)	
Gender	Male	28827(98.7)	392(1.3)	0.038
	Female	8104(98.9)	86(1.1)	
Arrival time	Day (6 AM–6 PM)	20618(98.9)	238(1.1)	0.008
	Night (6 PM–6 AM)	16313(98.6)	240(1.4)	
Accident area	Urban	15151(99.2)	127(0.8)	2×10^{-5}
	Suburban	21780(98.4)	351(1.6)	
Age		30.01 \pm 15.98	40.67 \pm 20.72	4×10^{-7}
SPO_2 level		97.74 \pm 3.29	72.1 \pm 36.87	5×10^{-5}
D1 (hour: minute)		10:47 \pm 19:02	09:55 \pm 07:21	0.327
D2 (hour: minute)		41:01 \pm 23:36	35:36 \pm 21:45	0.842

Note: When time, measured in minutes, is treated as a quantitative variable within our dataset, analyzing the central tendency becomes important. This process becomes even more meaningful when examined in conjunction with the Mean \pm SD.

Table 2. The MLR coefficients and odds ratios before matching

Confounders	Coefficient	Adjusted OR	95% CI	P
Constant	-53.56	5×10^{-24}	-	0.139
Age	0.05	1.05	(1.04, 1.05)	0.000
Gender	-0.46	0.63	(0.46, 0.87)	0.005
D1	-3×10^{-6}	1	(0.99, 1)	0.159
D2	-2×10^{-7}	1	(1, 1)	0.072
SPO ₂	0.3	1.35	(0.92, 2.01)	0.000
Location	0.73	2.08	(1.57, 2.77)	0.000
Year	0.38	1.46	(0.88, 2.42)	0.139
Month	-0.04	0.96	(0.92, 0.99)	0.022
Time	0.23	1.34	(1.05, 1.72)	0.018

(e.g. high-SPO₂ vs low-SPO₂ groups). In the regression model, the variable was considered 0 or 1 (survival or death), while other factors were confounding factors. Our analysis demonstrated that matching based on the PS successfully achieved a balanced distribution of covariates between individuals who died and those who survived after RTAs.

Data analysis

The statistical analysis was carried out in R software, version 4.5.0. To evaluate differences in demographic characteristics and RTA injuries between urban and suburban areas, independent t-test was employed, while the chi-square test was used to examine difference in categorical variables. $P < 0.05$ was considered statistically significant.

Results

This study analyzed 37,409 records of RTAs collected from March 2021 to March 2023. As shown in Table 1, the survival rate was 98.7% ($n=36931$) and only 1.3% ($n=478$) led to death. The mortality rate of RTAs occurred in suburban areas was twice the rate in urban areas (1.6% vs 0.8%). The accidents happened more at daytime, while the mortality rate was higher for RTAs occurred at nighttime ($P=0.008$). Also, most of accidents occurred in suburban areas than in urban areas. The age of deceased individuals was significantly higher than that of survivors, while their mean SPO₂ level was lower than that of survivors. Moreover, there was a statistically significant difference between men and

women ($P=0.038$), where men experienced more RTAs compared to women.

The results of MLR before PSM presented in Table 2 indicated the statistically significant association of RTA mortality rate with age, gender, D2, SPO₂ level, location, month, and the time of the accident (daytime vs. nighttime) ($P < 0.05$). Additionally, the analysis reveals that a decrease in SPO₂ is associated with a 35% increase in the risk of RTA mortality. Furthermore, males have 37% lower odds of RTA mortality rate compared to females (OR=0.63, 95% CI, 0.46%, 0.87%). Location of accident predicted a change by 2.08 units in RTA mortality rate (OR=2.08, 95% CI, 1.57%, 2.77%). The year of accident predicted the risk of RTA mortality rate by 46%. The time of accidents predicted the risk of RTA mortality rate by 1.34 units.

The standard error (SE) in the estimated effect size (log-odds coefficients) for all covariates can be derived from the ORs at 95% CI as Equation 5:

$$5. SE = \frac{\ln(\text{upper bound CI}) - \ln(\text{lower bound CI})}{2 \times 1.96}$$

We used the PSM method to analyze RTA data while adjusting for four non-modifiable confounders, which were selected since the emergency center was unable to modify them and also due to their significant baseline differences ($P < 0.05$). Table 3 shows the PSM results for the distance (convergence metric) and confounders between the death and survival groups before and after successful matching. The mean values

Table 3. The results of PSM before and after matching

Variables	Before Matching for All Data					
	Mean			Var. Ratio	eCDF Mean	eCDF Max
	Survived	Dead	Std. Diff.			
Distance	0.0227	0.0125	0.5161	3.4733	0.1569	0.2871
Age	40.8122	30.0094	0.5208	1.6851	0.1162	0.2463
Gender	1.1667	1.2194	-0.1414	0.8126	0.0264	0.0527
Location	1.7278	1.5899	0.3097	0.8205	0.069	0.138
Time	1.4831	1.4252	0.1159	1.0239	0.029	0.058

Variables	For Matched Data					
	Mean			Var. Ratio	eCDF Mean	eCDF Max
	Survived	Dead	Std. Diff.			
Distance	0.0227	0.0226	0.0006	1.0057	0	0.0021
Age	40.8122	40.7068	0.0051	1.0196	0.0011	0.0063
Gender	1.1667	1.1646	0.0057	1.0103	0.0011	0.0021
Location	1.7278	1.73	-0.0047	1.0049	0.0011	0.0021
Time	1.4831	1.4873	-0.0084	0.9995	0.0021	0.0042

Note: The eCDF outlines the degree of compliance, which reflects the overall tendency.

Table 4. The MLR coefficients and odds ratios after matching

Confounders	Coefficient	OR	95% CI	P
Constant	76.54	2×10^{-13}	-	0.87
Age	0.01	1.01	(1, 1.02)	0.05
Gender	0.05	0.95	(0.61, 1.48)	0.83
D1	-5×10^{-8}	1	(1, 1)	0.68
D2	-1.4×10^{-9}	1	(1, 1)	0.09
SPO ₂	0.19	1.21	(0.96, 1.53)	0.00
Location	-0.22	0.8	(0.55, 1.16)	0.24
Year	-0.04	0.96	(0.48, 1.89)	0.89
Month	-0.16	0.85	(0.61, 1.19)	0.34
Time	0.003	1	(0.95, 1.05)	0.92

Table 5. RMSE and MAPE values of forecasting models for urban and suburban areas

Models	Suburban		Urban	
	RMSE	MAPE (%)	RMSE	MAPE (%)
MLR	32.76	18.58	25.13	15.58
ARIMA	15.26	10.79	14.75	9.29
PSM	5.12	4.42	6.79	3.42

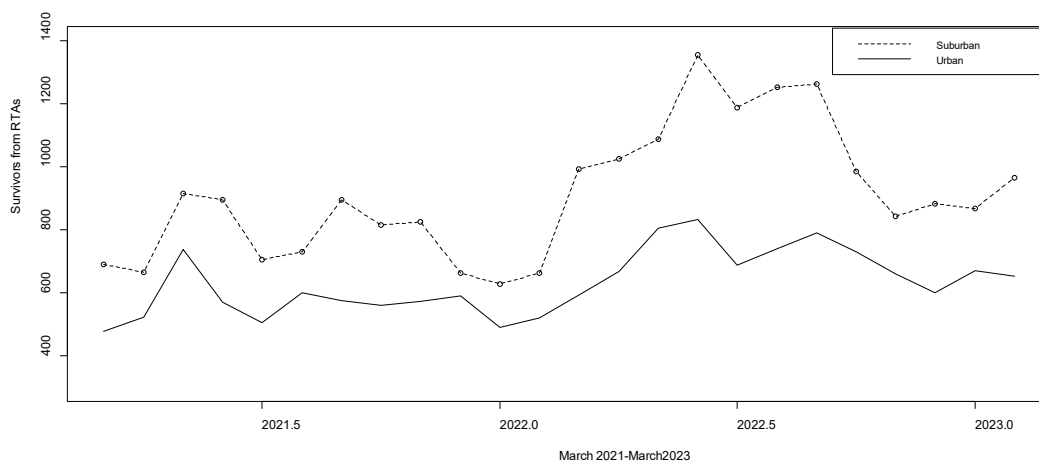
for distance and confounders showed improved balance and a substantial reduction in the standardized mean difference (from 0.52 for the total dataset to 6×10^{-4} in the matched dataset), confirming the algorithm's effectiveness in matching of the characteristics of the two groups. This approach in fact allows for a more equitable comparison by pairing patients with similar characteristics, thereby minimizing biases that could affect the mortality [30, 31]. It should be noted that, the nearest neighbor matching technique was utilized to create a balanced comparison between two groups.

Table 4 shows the MLR coefficients and adjusted ORs after PSM. This analysis entails modeling the probability that an individual involved in accidents may belong to one group as opposed to another, based on the confounding factors. The results revealed that SPO_2 and age were the only significant factors, with SPO_2 being the most significant predictor of RTA mortality.

Table 5 shows the RMSE and MAPE values of the forecasting models MLR, ARIMA, and PSM for urban and suburban areas. The ARIMA and PSM models demonstrate higher predictive accuracy compared to the MLR model, since they had lower RMSE and MAPE values.

The selection of confounding factors has a direct impact on these metrics, as omitting essential predictors results in underfitting, which subsequently increases RMSE and MAPE due to systematic bias. On the other hand, the inclusion of irrelevant confounders may lead to overfitting, which can reduce in-sample results but deteriorate out-sample results. Covariate balance assessments indicate whether the matching process has enhanced the predictive fairness of the PS. This, in turn, provides indirect insights into RMSE and MAPE if balance is successfully achieved.

The lowest RMSE recorded in both urban (5.12) and suburban (6.79) areas for the PSM indicates that the PSM exhibits the smallest prediction errors. On average, its mortality rate predictions deviate by 5–7 units from the actual values, which is considerably more accurate than the MLR/ARIMA models. The ARIMA model had an RMSE of 15.26 for urban areas and 14.75 for suburban areas, suggesting a moderate level of accuracy. In contrast, the MLR model had a higher RMSE of 32.76 in urban settings and 25.13 in suburban settings, indicating the worst performance among the models evaluated. These results suggest that PSM more effectively captures the predictors of RTA mor-

**Figure 1.** Total number of RTA survivors in suburban vs urban areas from March 2021 to March 2023

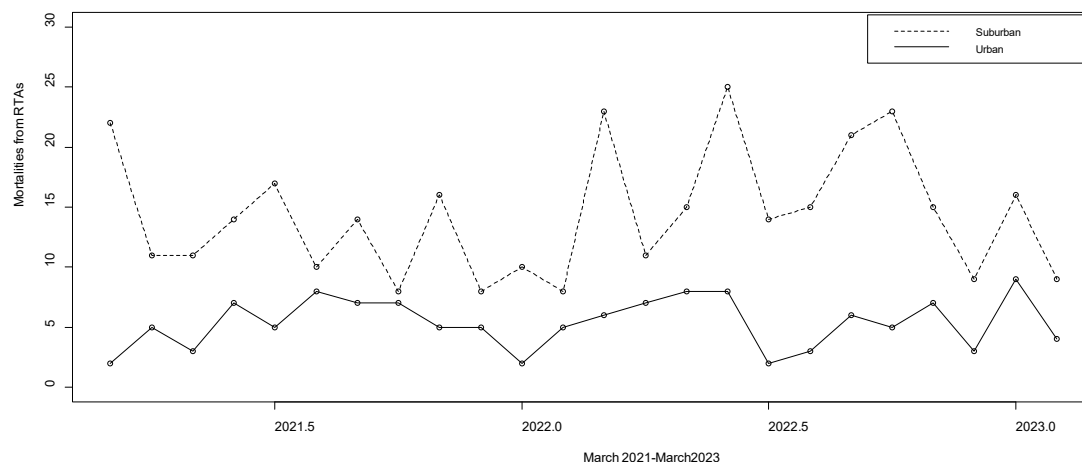


Figure 2. Total number of RTA mortalities in suburban vs. urban areas from March 2021 to March 2023

tality rates, while the high errors of the MLR indicate an inadequate fit. Additionally, the data presented in Table 5 indicated that all models exhibited higher RMSEs in urban areas, implying that RTA mortality rate is more difficult to predict in urban areas, potentially due to the complexities of traffic dynamics.

Figure 1 illustrates the total number of RTA survivors in suburban and rural areas from March 2021 to March 2023. As can be seen, the number of survivors was lower in urban areas compared to suburban areas, due to a lower incidence rate. Due to the inherent characteristics of suburban driving conditions, which can include higher driving speeds, less traffic control, and greater distances from emergency services, suburban

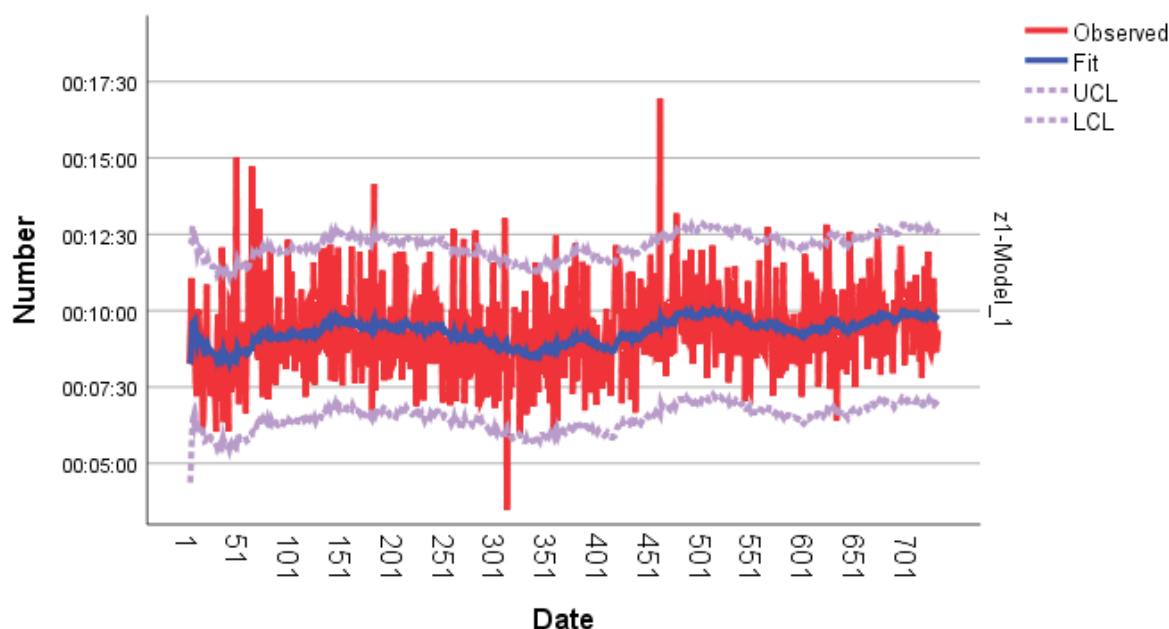


Figure 3. The average time for the EMS ambulance to arrive at the scene on a daily basis from March 2021 to March 2023

UCL: Upper-bound confidence interval; LCL: Lower-bound confidence interval.

Note: The red lines represent the actual observed data, while the blue lines indicate the predicted data.

areas are at higher risk of RTAs. Factors associated with crashes can also be effective, including road conditions, visibility, and the availability of immediate medical assistance following an accident. Overall, while both suburban and urban areas experience RTAs, the degree and context of these accidents differ, leading to variations in survival rate. Figure 2 plots the total number of RTA mortalities in suburban and rural areas from March 2021 to March 2023. Over the two-year period, suburban areas experienced a higher number of mortalities from RTAs compared to urban areas.

Figure 3 illustrates the daily average time for the EMS ambulance to arrive at the scene from March 2021 to March 2023 (700 days). The higher times observed in Figure 3 show critical cases when the EMS ambulance arrives late at the scene. By exploring the various factors that contribute to these delays and conducting a thorough analysis, authorities should implement effective measures to improve emergency response time and ensure better outcomes for RTAs.

Discussion

This study aimed to compare mortality rates of RTAs between urban and suburban areas in Golestan Province, north of Iran, using time-series data and forecasting models MLR, ARIMA, and PSM, which have not been previously addressed. Our findings revealed that the mortality rate was significantly higher in suburban areas (1.6%) compared to urban areas (0.8%), indicating a twofold increased risk of death in suburban areas. Key factors contributing to this result may be longer emergency response times, reduced access to healthcare facilities, and differences in road infrastructure and traffic conditions in suburban areas. Previous studies have indicated that delayed medical intervention worsens trauma outcomes [4, 5]. Studies in other parts of Iran have also reported increased RTA mortality in rural or suburban areas due to longer travel times to hospitals [8, 9]. The EMS ambulances take longer time to arrive at the accident scene in suburban areas, increasing the risk of death for critically injured individuals, particularly those with low SpO_2 .

The urban areas exhibited lower RTA mortality rates, maybe due to higher traffic density leading to slower speeds and less severe collisions, as well as better access to trauma care. This result is consistent with results of studies demonstrated that urban RTAs often involve low-speed impacts, resulting in non-fatal injuries [5, 21] (e.g. whiplash or minor fractures) rather than life-threatening trauma. This aligns with the speed-fatality

gradient well-documented in road safety literature, where higher speeds exponentially increase the risk of fatal outcomes. Additionally, stricter traffic laws in cities (e.g. speed limits, pedestrian crossings, and traffic calming measures), may contribute to lower accident severity. Differences in road infrastructure and vehicle types may also have a role. Conversely, suburban roads often have higher speed limits and fewer safety measures, increasing the likelihood of severe crashes. Previous research has highlighted that high-speed collisions in less regulated areas significantly elevate fatality risks [7, 12]. Our use of PSM helped control for confounding variables, strengthened the validity of these comparisons [31], confirming that suburban accidents involve higher-risk conditions (e.g. uncontrolled intersections, poor lighting, and delayed EMS ambulance access). A study reported that the incidence of RTAs during the summer months was nearly twice that in other seasons. Their results revealed a greater number of male victims with lower average age compared to females. Additionally, the study indicated a mortality rate of approximately 1.4% in Golestan Province [7]. Consistent with these results, we found an RTA mortality rate of 1.3%. Furthermore, our findings showed that the average age of male patients was less than that of female patients, and the number of male patients was higher. Notably, in our two study areas, a gender difference was evident; in suburban regions, the number of male victims (59.2%) was more than that of female victims (58.9%), while in urban areas, females (41.1%) were more than males (40.8%).

Deretić et al. [22] utilized the seasonal ARIMA model (SARIMA) for forecasting of traffic accidents and showed a notable seasonal trend using the time-series data. The model had an MAPE of 5.22%, which implies that the forecasts generated by the model were sufficiently accurate. The authors explained that an $MAPE < 10\%$ is indicative of highly accurate predictions, 10-20% suggests good forecasting, 20-50% represents reasonable forecasting, and $> 50\%$ shows inaccurate forecasting. In our study, the findings demonstrated that the prediction of RTA mortality rates by the PSM model was highly accurate in both urban and suburban areas, based on the MAPE value.

Mohammadi [19] revealed that 66% of the injuries involved car occupants, including both drivers and passengers, with a male-to-female ratio of 5:1. The most pronounced male-to-female ratio was observed among drivers (12.2:1), whereas pedestrians exhibited the lowest ratio (1.8:1). The majority of injuries occurred 16:00 and 20:00, followed by the period from

08:00 AM to 12:00 PM. The highest incidence of injuries was recorded among males aged 18-24 years. In our study, in suburban areas, a higher number of RTAs occurred at daytime more than at night, while urban areas experienced a higher number of accidents at night. This trend can be due to several key factors. Suburban roads often serve as connectors for commuters, commercial transport, and school-related travel, leading to higher traffic volumes during daylight hours. Also, there are usually fewer late-night or shift workers and entertainment centers in suburban areas, reducing nighttime traffic. In urban areas, there are vibrant nightlife, extended business hours, and shift work (e.g. healthcare work), increasing nighttime traffic density.

Conclusion

Our findings showed the RTA forecasting performance of ARIMA and PSM models, with PSM proving higher accuracy for both urban and suburban areas. The results highlight the promise of data-driven modeling. The RTA mortality rate is higher in suburban areas of Golestan Province than in its urban areas. This difference indicates the need for deeper exploration of region-specific factors, including emergency service coverage, road infrastructure, and socioeconomic elements, which may not be captured by the forecasting models. Future research should benefit from incorporating spatial analytics or hybrid modeling approaches to better understand these disparities.

Study limitations

This study had some limitations and disadvantages. First, varying service quality across health centers may have led to inconsistent medical reporting. To mitigate this, future studies should employ standardized trauma registries or accredited emergency departments, with sensitivity analyses to assess and exclude lower-quality health centers. Second, integrating diverse data sources (e.g. traffic patterns, accident reports) was challenging due to inconsistent formats and referencing systems. Adopting interoperable data standards, automated cleaning tools, and their testing in pilot studies can improve integration. Third, the under-reporting of minor traffic injuries (often due to the absence of medical treatment) may have led to selection bias. Supplementing hospital data with community surveys or law enforcement records, along with statistical methods such as capture-recapture analysis, can address this issue. A prospective study with standardized protocols can further enhance data consistency.

Ethical Considerations

Compliance with ethical guidelines

This study was approved by the Ethics Committee of [Golestan University of Medical Sciences](#), Gorgan, Iran (Code: IR.GOUMS.REC.1402.127). The study followed rigorous ethical standards in examining RTA data through time-series forecasting models. To safeguard victim privacy, all datasets were anonymized, and sensitive personal details were excluded.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors contributions

Conceptualization, study design and initial draft preparation: Manoochehr Babanezhad; Data collection, data analysis and interpretation: Hassan Khorsha and Ali Mohajervatan; Review, editing and final approval: All authors.

Conflict of interest

The authors declared no conflict of interest.

Acknowledgements

The authors would like to thank [Golestan University](#) and the [Golestan University of Medical Sciences](#).

References

- [1] World Health Organization (WHO). Road traffic injuries. 2022 [Updated 2025 August 30]. Available from: [\[Link\]](#)
- [2] Agyemang EF, Mensah JA, Ocran E, Opoku E, Nortey ENN. Time series based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints. *Heliyon*. 2023; 9(12):e22544. [DOI:10.1016/j.heliyon.2023.e22544] [PMID] [PMCID]
- [3] Pardeh MM, Yarmohammadian MH, Atighechian G, Ebrahimi A, Maracy MR. 11-year trend of mortality from fatal road traffic injuries in the center of Iran; A cross-sectional study. *Archives of Academic Emergency Medicine*. 2024; 12(1):e27. [DOI:10.22037/aaem.v12i1.2205] [PMID]

- [4] Hasani J, Ali Khorshidi A, Erfanpoor S, Nazparvar B, Hashemi Nazari SS. Comparison of risk factors for pedestrian fatality in urban and suburban traffic accidents. *Archives Trauma Research*. 2018; 7:39-44. [DOI:10.4103/atr.atr_7_18]
- [5] Khosravi Shadmani F, Soori H, Ainy E, Zayeri F, Mehmandar MR. [Comparison of road traffic death occurrence within urban and metropolitan roads focusing on environmental factors (Persian)]. *Hakim*. 2013; 15(4):339-45. [Link]
- [6] Babanezhad M, Khorsha H, Mohajervatan A, Choori A. Estimating the demand for ambulances in traffic accidents. *Health in Emergencies and Disasters Quarterly*. 2025; 10(4):247-58. [DOI:10.32598/hdq.10.4.149.8]
- [7] Khorshah H, Eri M, Honarvar MR, Mirkarimi SK, Abbasi M, Badiei F, et al. Epidemiological study of road traffic accidents and detection of accident hot spot in Golestan province, Northern Iran. *Journal of Clinical and Basic Research*. 2018; 2(4):15-22. [DOI:10.29252/jcbr.2.4.15]
- [8] Afkhaminia F, Yazdani Charati J, Rahimi E, Mousavi Nasab N. [Epidemiological study of the suburban accident mortalities recorded in Golestan, Iran in 2015 (Persian)]. *Jorjani Biomedicine Journal*. 2018; 6(1):67-73. [DOI:10.29252/jorjanibiomedj.6.1.67]
- [9] Yazdani Cherati J, Ahmadi Baseri E, Ghadami M. Mapping of mortality rate in suburban accidents, Mazandaran province, 2007-2010. *Journal of Mazandaran University of Medical Sciences*. 2013; 22(97):50-8. [Link]
- [10] Mohtasham-Amiri Z, Dastgiri S, Davoudi-Kiakalyeh A, Imani A, Mollarahimi K. An epidemiological study of road traffic accidents in Guilan Province, Northern Iran in 2012. *Bulletin of Emergency and Trauma*. 2016; 4(4):230-5. [PMID]
- [11] Abbasi HR, Mousavi SM, Taheri Akerdi A, Niakan MH, Bolandparvaz S, Paydar S. Pattern of traumatic injuries and injury severity score in a major trauma center in Shiraz, Southern Iran. *Bulletin of Emergency and Trauma*. 2013; 1(2):81-5. [PMID]
- [12] Yousefifard M, Toloui A, Ahmadzadeh K, Gubari MI, Neishaboori AM, Amraei F, et al. Risk factors for road traffic injury-related mortality in Iran; A systematic review and meta-analysis. *Archives of Academic Emergency Medicine*. 2021; 9(1):e61. [DOI:10.22037/aaem.v9i1.1329] [PMID]
- [13] Bakhtiyari M, Mehmandar MR, Riahi SM, Mansournia MA, Sarpti M, Bahadorimonfared A. Epidemiologic Pattern of Fatal Traffic Injuries among Iranian Drivers; 2004-2010. *Iranian Journal of Public Health*. 2016; 45(4):503-14. [PMID]
- [14] Mehmandar M, Soori H, Amiri M, Norouzrad R, Khabzkhooob M. Risk factors for fatal and nonfatal road crashes in Iran. *Iranian Red Crescent Medical Journal*. 2014; 16(8):e10016. [DOI:10.5812/ircmj.10016] [PMID] [PMCID]
- [15] Najimi-Varzaneh A, Gholami Fesharaki M. Prevalence of road traffic accidents in Iran: A systematic review, GIS and meta-analysis. *Iranian Red Crescent Medical Journal*. 2018; 20(10):e83852. [DOI:10.5812/ircmj.83852]
- [16] Nassiri H, Mohammadpour SI, Dahaghin M. Forecasting time trend of road traffic crashes in Iran using the macro-scale traffic flow characteristics. *Heliyon*. 2023; 9(3):e14481. [DOI:10.1016/j.heliyon.2023.e14481] [PMID] [PMCID]
- [17] Ghaneian MT, Tarfiei A, Ehrampoush MH, Lotfi MH, Namayandeh SM, Adamizadeh AR, et al. An epidemiological survey on factors related to traffic accidents in Yazd city, center of Iran (2016-2018). *Journal of Occupational Health and Epidemiology*. 2021; 10(4):231-8. [DOI: 10.52547/johe.10.4.231]
- [18] Sadeghi-Bazargani H, Ayubi E, Azami-Aghdash S, Abedi L, Zemestani A, Amanati L, et al. Epidemiological patterns of road traffic crashes during the last two decades in Iran: A review of the literature from 1996 to 2014. *Archives of Trauma Research*. 2016; 5(3):e32985. [DOI:10.5812/atr.32985] [PMID] [PMCID]
- [19] Mohammadi G. Road traffic crash injuries and fatalities in the city of Kerman, Iran. *International Journal of Injury Control and Safety Promotion*. 2013; 20(2):184-91. [DOI:10.1080/17457300.2012.686039] [PMID]
- [20] Manouchehrifar M, Hatamabadi HR, Derakhshandeh N. Treatment costs of traffic accident casualties in a third-level hospital in Iran; A preliminary study. *Emergency*. 2014; 2(1):40-2. [PMID]
- [21] Meng G, Liu J, Feng R. Prediction of construction and production safety accidents in China based on time series analysis combination model. *Applied Sciences*. 2022; 12(21):11124. [DOI:10.3390/app122111124]
- [22] Deretić N, Stanimirović D, Awadh MA, Vujanović N, Djukić A. SARIMA modelling approach for forecasting of traffic accidents. *Sustainability*. 2022; 14(8):4403. [DOI:10.3390/su14084403]
- [23] Getahun KA. Time series modeling of road traffic accidents in Amhara Region. *Journal of Big Data*. 2021; 8(1):102. [DOI:10.1186/s40537-021-00493-z]
- [24] Corcuera Hotz I, Hajat S. The effects of temperature on accident and emergency department attendances in London: A time-series regression analysis. *International Journal of Environmental Research and Public Health*. 2020; 17(6):1957. [DOI:10.3390/ijerph17061957] [PMID] [PMCID]
- [25] Kumar S, Toshniwal D. A novel framework to analyze road accident time series data. *Kumar and Toshniwal Journal of Big Data*. 2016; 3(8):1-11. [DOI:10.1186/s40537-016-0044-5]
- [26] Huang H, Chen J, Sun R, Wang S. Short-term traffic prediction based on time series decomposition. *Physica A*. 2022; 585:126441. [DOI:10.1016/j.physa.2021.126441]
- [27] Diaz-Hierro JO, Martín JM, Arenas ÁV, Gonzalez MP, Arevalo JM, González CV. Evaluation of time-series models for forecasting demand for emergency health care services. *Emergencias*. 2012; 24(3):181-8. [Link]
- [28] Pitarque A, Guillen M. Interpolation of quantile regression to estimate driver's risk of traffic accident based on excess speed. *Risks*. 2022; 10(1):19. [DOI:10.3390/risks10010019]
- [29] Al-Azzani MAK, Davari S, England TJ. An empirical investigation of forecasting methods for ambulance calls - A case study. *Health Systems*. 2020; 10(4):268-85. [DOI:10.1080/20476965.2020.1783190] [PMID] [PMCID]
- [30] Bonifaz JL, Fasanand A. The use of propensity scores matching methodology for analysis of concession: The case of road networks in Peru. *Case Studies on Transport Policy*. 2023; 10:2350-7. [DOI:10.1016/j.cstp.2022.10.017]

- [31] Khorsha H, Babanezhad M, Behnampour N. Evaluation of the causal association of risk factors for death from COVID-19 patients admitted to Golestan hospitals by propensity score estimation method. *Iranian Journal Health Sciences*. 2023; 11(4):279-88. [\[DOI:10.32598/ijhs.11.4.567.3\]](https://doi.org/10.32598/ijhs.11.4.567.3)