

ORIGINAL PAPER

Open Access



# Investigating key explanatory factors for safer long-distance bus services

Shaghayegh Rahnama<sup>1\*</sup>, Adriana Cortez<sup>2</sup> and Andres Monzon<sup>3</sup>

## Abstract

Buses are among the most accessible and frequently used means of transport. Due to its importance, road safety analysis is frequently conducted to reduce accidents. This paper studied the relationship between weather conditions and the causes of accidents to improve road safety, focusing on long-distance services between Madrid and Bilbao (Spain). We employed Latent Class Clustering (LCC) and Hierarchical Ordered Logit models to identify these factors' relationships. Additionally, Kaplan-Meier survival analysis was adopted to provide temporal insights into accident occurrences.

The main results show a downward trend in accidents since 2019, with manoeuvres being the most frequent cause. LCC reveals that "manoeuvres and car invading lanes in the opposite direction" in "clear and cloudy weather" has the highest probability of occurrence (63%). The hierarchical-ordered logit model indicates that rainy weather significantly affects all accident causes. Kaplan-Meier survival analysis reveals a vertical initial decline in survival probability within the first ten days, emphasizing a high initial accident risk.

The integrated approach used in this work provides a thorough understanding of accident hazards, which is its main contribution. By integrating LCC, Hierarchical Ordered Logit models and Kaplan-Meier survival analysis; we could offer a comprehensive and nuanced interpretation of the connection between weather and bus accidents. The findings highlight the need for rapid and sustained safety interventions, enhancing robustness and providing actionable insights for improving bus safety.

**Keywords** Bus transport, Road Safety, Weather conditions, Latent class clustering (LCC), Hierarchical ordered logit method, Kaplan-Meier survival analysis

## 1 Introduction

Public transportation is safer than other motorized transportation options. Nevertheless, a significant number of bus-related incidents occur each year. The World Health Organization stated that 84,500 people died on Europe's roadways in 2015 [1]. Road transportation accounts for 85.5% of all transportation in Spain, including buses and private automobiles. Furthermore, buses accounted for 34% of all collective transportation in 2019 [2]. Since long-distance buses are among the most widely available and frequently utilized forms of transportation, ensuring these journeys are safe is crucial. Accident data from

\*Correspondence:

Shaghayegh Rahnama  
shaghayegh.rahnama91@gmail.com

<sup>1</sup>Escuela de Ingeniería y Arquitectura, Universidad de Zaragoza, Zaragoza 50018, Spain

<sup>2</sup>Higher Polytechnic School, Universidad Francisco de Vitoria, Pozuelo de Alarcón, Madrid 28223, Spain

<sup>3</sup>Transport Engineering, Urban and Regional Planning Department, Universidad Politécnica de Madrid, Madrid 28040, Spain



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

prior years can be analysed to identify the primary causes of accidents.

Previous research has identified weather, speed, and accident location as some of the most significant factors influencing bus accidents [3]. However, the assessment of bus safety has varied due to differences in data availability and the methodologies used [4]. To develop effective safety measures, it is essential to base analyses on real accident data, providing concrete foundation for understanding and mitigating risks.

Accident data provides empirical evidence essential for conducting robust and reliable analyses. By examining real-world data, researchers can uncover trends, correlations, and causative factors that might not be apparent through theoretical models or simulations alone. Empirical data, like the “Road Accidents in India-2022” report, helps identify critical factors contributing to accidents, enabling detailed examination of trends and demographic impact.

Analysing real accident data to identify the most relevant explanatory factors associated with traffic accidents, by examining factors such as weather conditions, driver behaviour, and road infrastructure, helps us gain valuable insights into the root causes of accidents. This knowledge can then be used to develop targeted interventions to improve road safety and save lives.

For instance, Hammad et al. [5], found that weather conditions like rainfall, temperature, fog, and windstorms directly affect road traffic accidents (RTAs). Age, vehicle size, drivers’ maturity, road conditions, and environmental impacts were significant factors in RTAs. Another study analysed the correlation between rainfall intensity and water depth in Seoul City, Korea, and traffic accident severity. It found that traffic, environmental, human, rain, water depth, and road factors correlate with accident severity [6].

This study aims to analyze the safety of long-distance bus services in Spain by focusing on the relationship between weather conditions and accident causes. This focus is particularly relevant because long-distance buses are one of the most accessible and used transport modes in Spain, and ensuring the safety of these trips is essential. By utilizing real accident data, from long distance bus services between Madrid and Bilbao, we can accurately identify the primary factors contributing to accidents under different weather conditions. These insights have the potential to shape the development of more effective strategies for preventing and responding to bus accidents. Moreover, the new methodology proposed can be applied to analyze bus safety not only in other regions of Spain but also in other countries worldwide, promising a better future for bus safety.

This paper is structured in six sections. After this brief introduction, the second section reviews previous

research on road safety. The third section describes the case study and data sources. Section four presents the methodology employed, and the discussion and results are presented in section five. Finally, the most relevant findings and conclusions are presented in Sect. 6.

## 2 Literature review

Traveling by bus is recognized as one of the safest modes of transport. According to a study published in the *Journal of Public Transportation*, public transit travel, including bus travel, has about one-tenth the traffic casualty (injury or death) rate compared to automobile travel [7]. However, bus safety remains a critical issue affecting passengers, operators, and governments and can have serious repercussions. Public Transport Companies’ (PTC) perspective says bus accidents increase the costs of an industry with high operational costs and low fare revenues daily [8, 9]. Furthermore, it has been demonstrated that passengers’ travel decisions are influenced by their sense of bus safety [10]. To capture the effects of the several factors that determine their frequency and severity, a great deal of studies have been done on finding patterns of bus accidents [11]. Data analysis can assist in determining the variables, such as traffic, weather, and road conditions, that contribute to accidents. The research by Kumar et al. [12], suggests a technique for examining Gujarati hourly traffic accident data. The cophenetic correlation coefficient chooses the best distance metric for clustering comparable accident patterns. The outcomes demonstrate how effectively the suggested strategy clusters districts with comparable accident patterns, which can be utilized for trend analysis and other purposes. To determine the primary causes and variables influencing bus accident data, Goh et al. [13] state that a literature analysis can be used. Previous research was based on aggregate data on several bus models and the current bus accident classifications used in various nations [14]. Studies have indicated that the number of accidents is impacted by the type of vehicle, road parameters, weather, and traffic [15]. Data on bus accidents in Denmark from 2002 to 2011 was examined in the study by Prato & Kaplan [16]. The severity of bus accidents was analysed using a generalized ordered logit model, and the likelihood of passenger injuries was investigated using logistic regression. According to the study, the severity of bus accidents is positively correlated with the elderly driver of the third party involved, high-speed restrictions, nighttime, vulnerable road users, bus drivers, and other drivers who cross at yellow or red lights. Heavy trucks, running red or yellow lights at junctions, wide spaces, fast speed restrictions, and slick roads all correlate positively with bus passenger injuries.

Weather conditions drastically impact the number of road accidents and fatalities [17]. Poor weather reduces

friction and hinders visibility, which contributes significantly to crashes. Numerous studies conducted in recent years have incorporated weather into road safety assessments [18]. The findings indicated that the weather could account for roughly 5% of the fluctuation in collisions and fatalities per month [19].

The number of accidents reported has also been shown to be influenced by road conditions, particularly when paired with other outside variables like speed and weather. It has been observed that haul portions, intermediate stations, and intersections are the primary locations for bus accidents [18]. Another element that may have an impact on accidents is road infrastructure. Gitelman et al. [20] assessed the safety efficacy of infrastructure upgrades on Israel's non-urban roadways. The study's data collection included about 200 sites with roughly 30 treatments. Accident changes were investigated using after-before comparisons that considered variations in comparison-group sites and a regression-to-the-mean. The study generated a weighted efficiency index value for sites receiving comparable treatments. The results were compared with prior local and worldwide experience to determine a final list of safety consequences. Accident reduction factors were obtained for 19 road infrastructure improvements. The study estimated that the black-spots treatment project was associated with an annual average saving of 224 injuries and 531 total accidents, which had a tangible economic value. Overall, the study provides valuable insights into the effectiveness of infrastructure improvements in reducing accidents on non-urban roads in Israel.

When evaluating road safety, a causal relationship between accidents and a set of explanatory variables – generally discrete ones – must be defined. Once the nature of the variables is known, different statistical methods can be used to analyse the data and to identify the relationship among different factors influencing the accidents [21]. Through these methods, cluster analysis (CA) Karlaftis et al. [22], Latent Class Cluster (LCC) and Multinomial Logit (MNL) Depaire et al. [23] are the most used. Once the variables are clustered, the main problem is to understand the relationship of the influencing factors. For this, different models such as the Generalized Ordered Logit model, the Ordered Multilevel Cross-Classification model, and the Hierarchical Ordered Logit (HOL) model, have been used by different authors [24, 25].

On the other hand, Kaplan-Meier Survival analysis estimates the probability of survival (i.e., no accidents) over time, highlighting periods with higher or lower risk of accidents, aimed to assess survival rates and predictors of mortality among road traffic accident victims in Kinshasa, DRC. Using a historical cohort design (2011–2016), it found a 19.6% mortality rate, with socioeconomic

status and specific body injuries predicting mortality. The Kaplan-Meier method analysed survival, with SPSS 21.0 used for statistical analyses. This underscores the significant toll of road accidents in the DRC and similar contexts, emphasizing the need to address preventable causes of death [26]. Another study using data from a digital tachograph device aimed to evaluate the effects of traffic safety education on abnormal driving behaviours among bus drivers. A 9-week survey was conducted on 61 bus drivers, and a survival model was used to analyse the effects. Results showed that the effects of education decreased more quickly on acceleration-related abnormal driving behaviours than steering- and deceleration-related ones. The effects for the dangerous group lasted longer than the normal group, suggesting that tailored education could reduce abnormal driving behaviours. The study recommends specialized education and periodic training to prevent recurrence of abnormal driving behaviours [27]. In urban transit area the study of Lipton, Cunradi & Chen [28] examines the role of smoking in all-cause mortality among urban transit operators, focusing on a minority group. Data from 1,785 workers, including 61% African American and 9% female, was analysed. The study found that 45% of the workers were current smokers, 30% were former smokers, and 25% had never smoked. The probability of survival was lower for former smokers but not for current smokers. Years of smoking significantly contributed to mortality, with African American and white operators having higher mortality risks than Asian-American operators. Gender and average weekly number of drinks were not significantly associated with mortality. Although smoking rates have declined among blue-collar workers, former smoking prevalence may contribute to excess mortality [28].

Previous studies have used aggregate data from different types of buses and classifications across countries. However, the degree of influence and relationship with other factors in bus accidents has not been thoroughly explored. Our study focuses on analyzing accident patterns in long-distance services between Madrid and Bilbao, using the LCC methodology and Hierarchical Ordered Logit method. It seeks to identify the relationship between weather conditions and bus accident causes, filling a gap in current understanding. This analysis creates on previous studies and offers an up-date view of bus safety research.

### 3 Case study

This study is centered on intercity bus transportation linking the Spanish cities of Madrid and Bilbao, separated by 395 km. The decision to use this road and its bus services as the case study was made with three key factors in mind. Firstly, the route links two major Spanish cities: Madrid, the country's capital and most populous

city in the centre, and Bilbao, the largest city in the country's north. Second, this corridor is one of Spain's busiest roads, carrying a variety of vehicles, including cars, lorries, and buses. And third, the diverse landscapes along this route, including towns, valleys, and two significant mountain ranges, present unique challenges for road safety, making it an ideal case for this study. See Madrid-Bilbao Road corridor in Fig. 1.

Additionally, the weather in the northern area of the corridor is marked by cloudy and rainy days, posing an additional safety hazard. The months that exhibit the best visibility range from February to December (10 km). January holds the record for the lowest visibility at 9 km. The yearly mean precipitation in Bilbao is 1200 mm. Throughout the year, it experiences rainfall, with November being the wettest month (103 mm) and September the driest with only 31 mm of rainfall. January is the month that has the highest level of cloudiness, at 51% while September and October are the least cloudy months, with a cloudiness of 34%. The State Meteorological Agency (AEMET) provides all of this meteorological information.

#### 4 Data and methods

In the initial stage, a general descriptive analysis was carried out to obtain a summary of the data. Weather conditions and accident causes were then grouped using

Latent Class Clustering (LCC). After all variables were found to be related, a Hierarchical Ordered Logit model was used to validate the relationships that did not exist between the variables. Additionally, Kaplan-Meier survival analysis was included to offer temporal insights into the risk of accidents over time. Using a survival function estimate, this method shows when accidents are more likely and less likely to occur.

#### 4.1 Data

The required data for this work were collected from two primary sources: the bus operator ALSA and the State Meteorological Agency (AEMET). The dataset includes the report of the 115 accidents registered between January 2019 and November 2021. This number of accidents referred to all services on this route. ALSA's bus accident records contain the following information: date, location, and causes. The factors considered in this study were accident characteristics and weather conditions. Accident characteristics included the year of the accident, the season in which the accident occurred, and the causes of the accident. ALSA categorizes those accidents under 9 main causes: A1-Manoeuvres, A2-Car invading lane in the opposite direction, A3-Others, A4-Spin, A5-Passengers (inside the vehicle), A6-Lane change, A7-Crossing hit, A8-Fixed objects, and A9-Not respecting preference

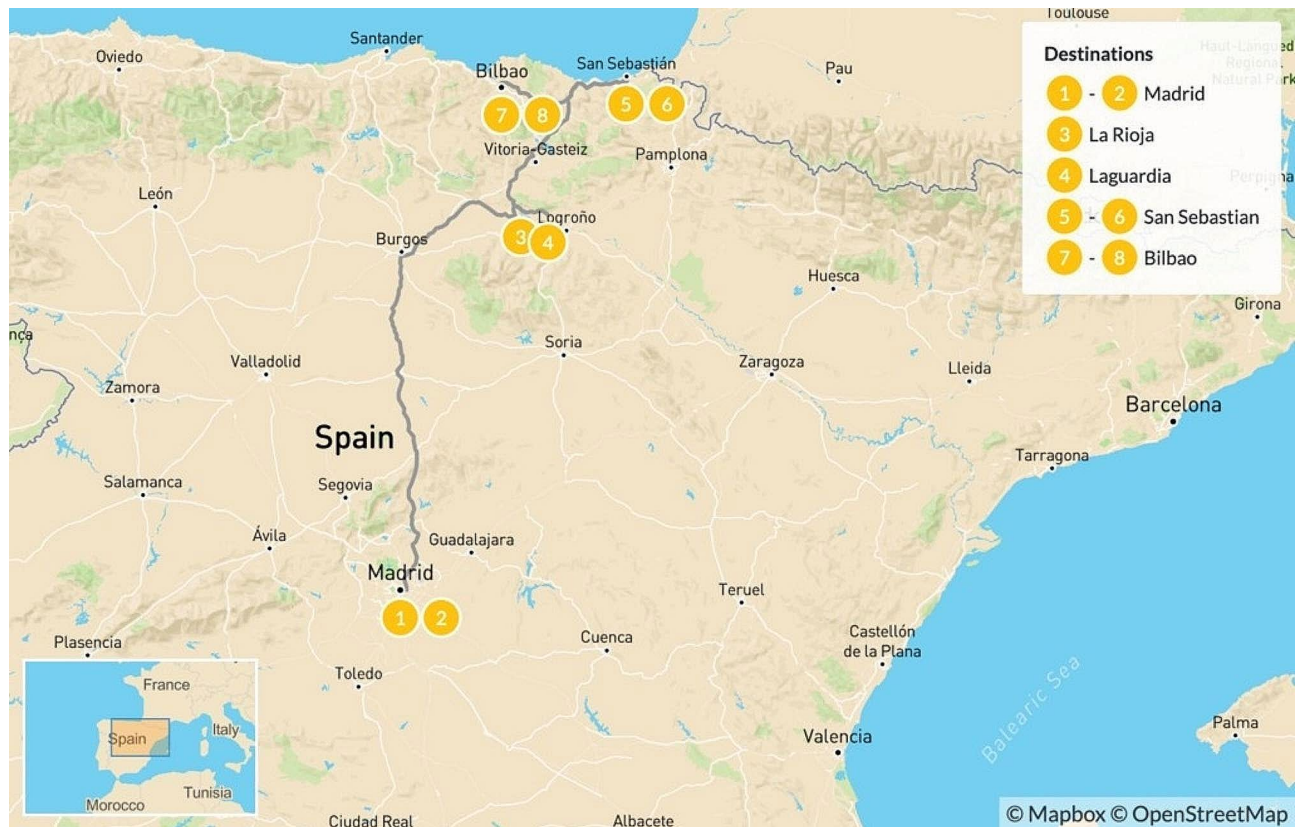


Fig. 1 Map of Madrid-Bilbao Road

sign. Weather conditions included: W1-Visible (clear), W2-Cloudy, W3-Rainy and W4-Foggy.

Table 1 shows the total km-driven and the number of expeditions recorded by the operator ALSA in the three years of this study. The reduction in the number of expeditions registered in 2020 and 2021 was due to mobility restrictions caused by the COVID-19 pandemic.

Initially, a general descriptive analysis was conducted to obtain a summary of the data. This involved categorizing accident data by year and by the frequency of various causes. This preliminary step provided an overview of the accident trends and the predominant causes over the study period.

#### 4.2 Latent class clustering (LCC) analysis

Cluster analysis was employed to identify patterns and groupings within the data. Specifically, LCC was selected for its ability to handle heterogeneous data and uncover latent structures. To determine which model best captures reality, we applied the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) [29]. A lower number for each of these criteria indicates a better fit, and they are crucial in figuring out which cluster selection method works best.

As the observations inside a cluster are as similar as possible and as different from each other [9], it served us to determine the primary causes of accident occurrences.

#### 4.3 Hierarchical ordered logit method

The data suggests that there is a relationship between weather and accident causes. Nonetheless, there could be a connection between two or more meteorological factors that affect the reasons behind accidents. There is no correlation to describe this relationship. Therefore, at several levels, the link between independent variables (weather) and dependent variables (accident causes) needs to be assessed.

To further investigate the relationship between weather conditions and accident causes, a Hierarchical Ordered Logit model was applied. This method was chosen for its ability to handle data with a hierarchical structure and assess the impact of multiple independent variables (weather conditions) on the dependent variables (accident causes). It is assumed that there is no association between the samples in the standard regression approach. The residuals are independent as a result. Since a hierarchical model can account for both within-group

and across-group effects simultaneously, the model is suitable for analysing data with a hierarchical structure.

#### 4.4 Kaplan-Meier survival analysis

To provide temporal insights into accident risks, Kaplan-Meier Survival Analysis was conducted. The Kaplan-Meier estimator is a non-parametric statistic used to estimate the survival function from time-to-event data. For this study, the time-to-event data consisted of the duration between the start of the observation period and the occurrence of an accident. The survival function  $S(t)$  represents the probability that a bus will operate without an accident up to time  $t$ . This method estimates the probability of no accidents occurring over time and highlights periods with higher or lower accident risk. The analysis helps identify high-risk periods and informs the need for targeted safety interventions.

Researchers obtained temporal insights into accident dangers by applying Kaplan-Meier survival analysis. The likelihood of not having an accident over time is estimated using this method. To account for censored data (individuals who did not have any accidents throughout the research), accident dates were converted into a time-to-event format. Although the Kaplan-Meier approach offers significant insights into trends that vary over time, complementing models such as LCC and Hierarchical Ordered Logit models can be employed to investigate the impact of additional variables on the likelihood of accidents.

### 5 Results

The results of the analysis conducted in this study are summarized in this section. Section 5.1 examines the general descriptive analysis of the accident data. Section 5.2 illustrates the results of the cluster analysis and the number of clusters gathered by the algorithm. Finally, the last section shows the relationship between weather conditions and the cause of the accident obtained with the Hierarchical Ordered Logit method.

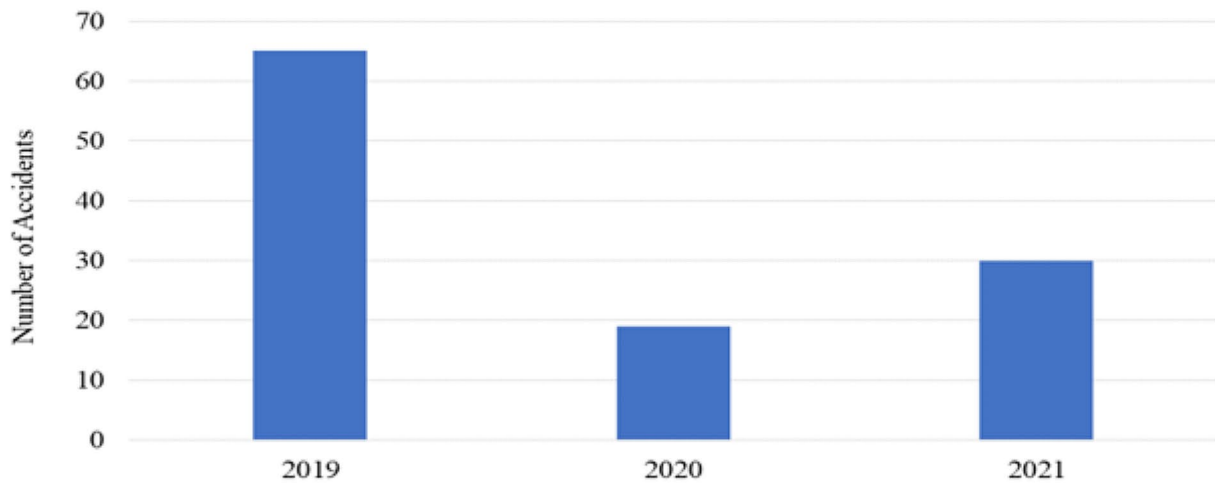
#### 5.1 Descriptive analysis

For the general descriptive analysis, accidents were grouped by year, and the occurrence of each cause of accident in these three years is shown to have a first overview. Figure 2 shows the number of accidents between 2019 and 2021. A deep exploration shows that the highest accident frequency in this timespan was in 2019. Due to the pandemic lockdown and reduction in the number of services, the lowest number of accidents occurred in 2020 (see Table 2).

Figure 3 shows the frequency of each of the 9 causes of accidents considered by the operator in the three years studied. The most frequent cause of accidents is “manoeuvres” with a 46% share. Followed by “car

**Table 1** ALSA buses activities 2019–2021

Year	Km driven	Services
2019 (normal services)	3,958,217	10,021
2020 (Pandemic-4 months block-down)	1,742,009	4,410
2021 (Pandemic-restrictions)	2,374,930	6,012



**Fig. 2** Number of accidents y ALSA buses operating Madrid-Bilbao services along years 2019–2021

**Table 2** Accidents per km

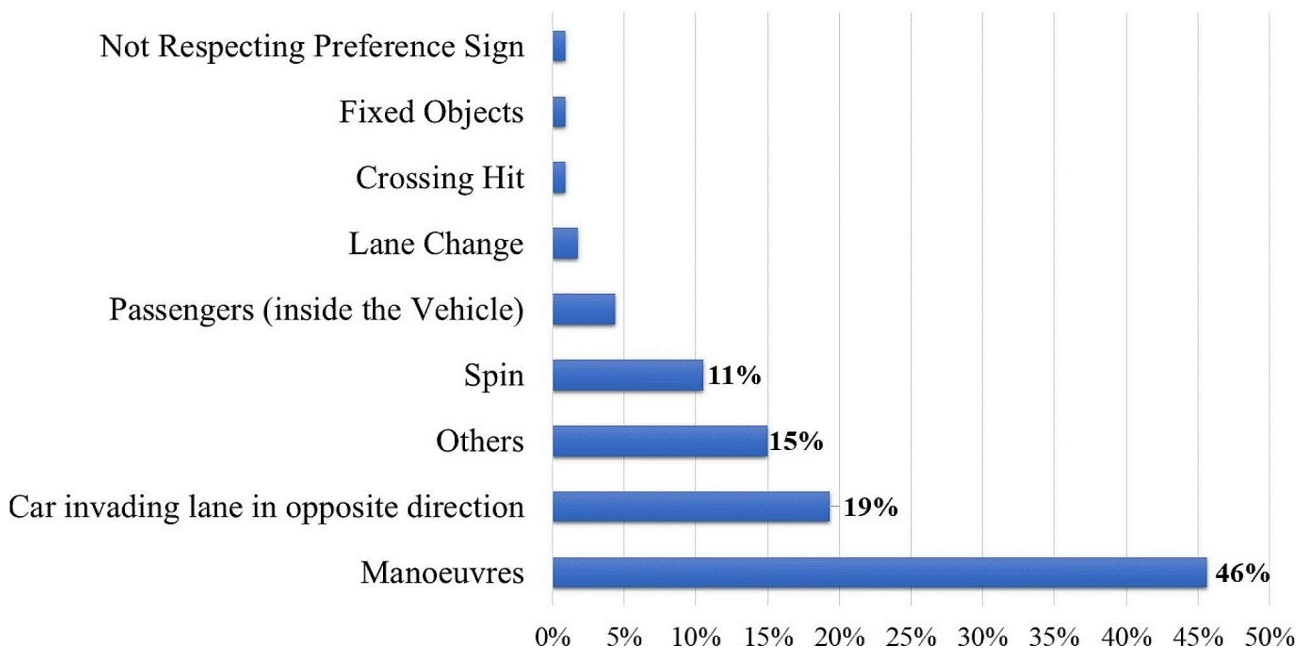
Year	N° Accident	Accidents per km
2019 (normal services)	65	0.0000164
2020 (Pandemic-4 months block-down)	19	0.0000109
2021 (Pandemic-restrictions)	30	0.0000126

causes” (15%).

Weather conditions were divided into four categories: visible, cloudy, rainy, and foggy. Table 3 shows the individual causes of accidents that occurred in different weather conditions. As the trends show, manoeuvring is the most frequent type of accident in all types of weather. For rainy conditions, spin and car invading lanes in the opposite direction of an oncoming vehicle have the same

invading lane in opposite direction” (19%) and “other

### Causes of Accidents



**Fig. 3** Percentage of accidents based on causes

**Table 3** Causes of accidents y relation to weather conditions

Causes of accident	Visible (clear)	Cloudy	Rainy	Foggy
A1-Manoeuvres	44%	58%	37%	36%
A2-Car invading lane in opposite direction	19%	19%	16%	18%
A3-Others	18%	4%	11%	27%
A4-Spin	12%	4%	16%	9%
A5-Passengers (inside the vehicle)	5%	8%	5%	0%
A6-Lane change	2%	8%	0%	9%
A7-Crossing hit	0%	0%	5%	0%
A8-Fixed objects	0%	0%	5%	0%
A9-Not respecting preference sign	0%	0%	5%	0%

**Table 4** BIC and AIC values for Cluster models

Number of Clusters	BIC	AIC
1	630.57	617.55
2	626.17	597.51
3	642.58	598.29
4	-84.02	-143.94
5	-563.29	-638.84

**Table 5** Contribution of weather condition and causes of accidents to each cluster (%)

Variables		Cluster 1 (C1)	Cluster 2 (C2)	Cluster 3 (C3)
Weather Conditions	W1-Visible (clear)	55%	0%	64%
	W2-Cloudy	45%	0%	0%
	W3-Rainy	0%	100%	0%
	W4-Foggy	0%	0%	31%
Causes of Accidents	A1-Manoeuvres	59%	41%	27%
	A2-Car invading lane in opposite direction	19%	18%	21%
	A3-Others	4%	12%	35%
	A4-Spin	6%	18%	14%
	A5-Passengers (inside the vehicle)	9%	0%	0%
	A6-Lane change	3%	0%	0%
	A7-Crossing hit	0%	0%	3%

percentage. In contrast, spin has a lower percentage in other types of weather.

**5.2 Latent class clustering analysis**

An LCC analysis was done using JAMOVI (v.2.2.5). The optimum number of clusters (3) was obtained by estimating BIC and AIC criteria evaluating from 1 to 10 clusters. Table 4 presents the results of the first 5. These clusters represent different combinations of weather conditions and accident causes.

Table 5 shows the probability of each variable belonging to each cluster. It is important to note that, due to negligible correlation, the variables “fixed objects” and “not respecting preference sign” were excluded from the

**Table 6** Probability of assignment of one accident to each cluster

Cluster	Probability
C1. Manoeuvres and Car invading lane in opposite direction in Clear and Cloudy weather	63%
C2. Manoeuvres, Car invading lane in opposite direction of an oncoming and Spin in Rainy weather	22%
C3. Spin and Other causes in Foggy and Visible weather	15%

analysis. Consequently, the remaining parameters were grouped into seven accident causes and four weather conditions, resulting in three distinct clusters.

Cluster 1(C1). In this cluster, accidents happen when the weather is visible or cloudy. The main causes are manoeuvres and Car invading lane in the opposite direction, with 78% of its accidents. Based on these facts, Cluster1 is called: “Manoeuvres and Car invading lane in opposite direction in clear and cloudy weather”.

Cluster 2(C2). In this cluster, 100% of the accidents correspond to rainy weather, and they are due to “manoeuvres”, “car invading lane in opposite direction” and “spin”, with 77% of the total number of accidents registered. Because of this reason, Cluster2 is called: “Manoeuvres, Car invading lane in opposite direction of an oncoming vehicle and spin in rainy weather”.

Cluster 3(C3). In this cluster, two-thirds of accidents happen in clear weather, while the other third in foggy conditions. The causes are mainly “spin” and “other reasons”, with 50% of the total accidents registered. Based on these frequencies, Cluster3 is called: “Spin and other causes in foggy and visible weather”.

Table 6 shows the probability of assignment to each cluster. The table shows that the C1 has the highest likelihood of occurrence of (63%). The two other clusters, with 22% and 15% respectively, are less likely to occur.

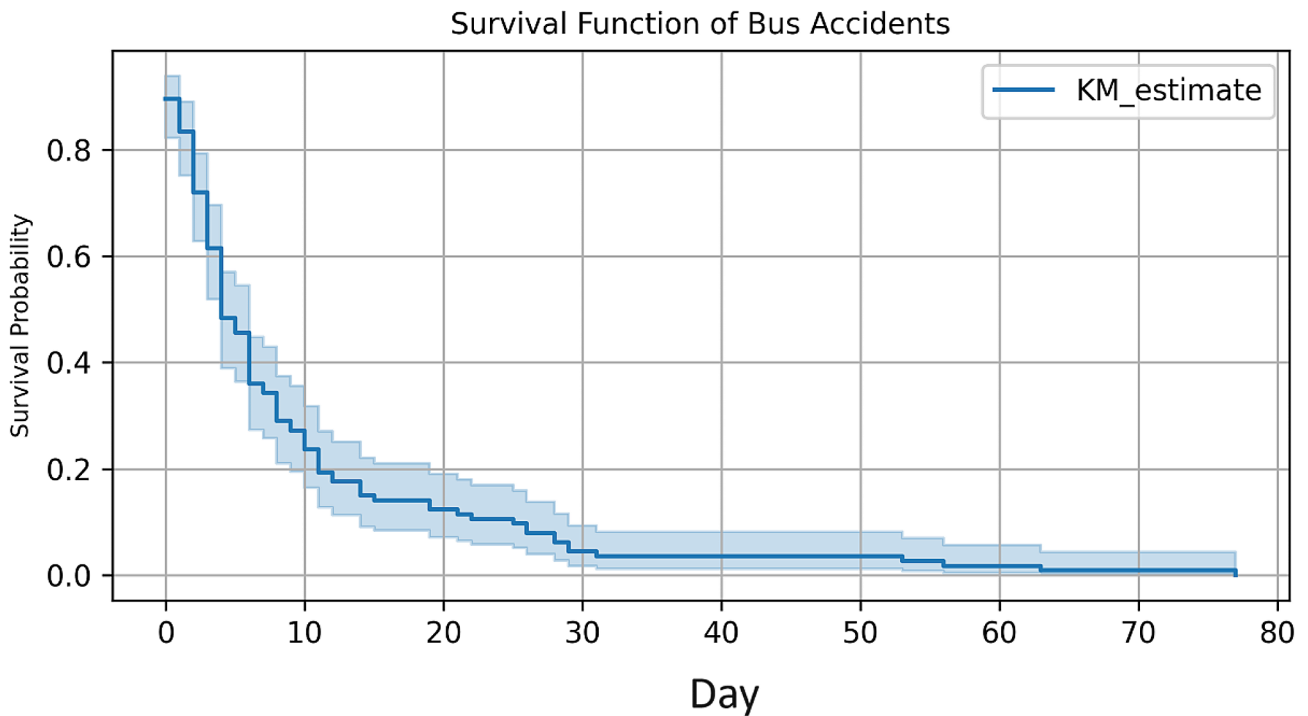
**5.3 Correlation analysis hierarchical ordered logit**

A Hierarchical method is implemented using STATA. Table 7 shows the summary of *p*-values for the seven causes of accident in each weather condition. It is important to note that the method in this analysis was based on four different weather conditions. In each evaluation, there were seven iterations corresponding to the number of accident causes. It is important to mention that two causes of accidents which are namely: fixed objects and not respecting preference signs were ignored. This was because they were not significantly related to the weather. This estimation was done for each weather condition and cause of accident one by one. Each complete evaluation includes seven *p*-values plus the last iteration by total *p*-value for that group. Also, all-weather conditions were examined together for realizing the disappearing correlation.

**Table 7** Hierarchical results summary

Causes of accident Weather Conditions	Manoeuvres	Car invading lane in opposite direction	Others	Spin	Passengers (inside the vehicle)	Lane change	Crossing hit
Visible	0.24	0.23	0.16	0.15	0.21	-	-
Cloudy	0.04*	0.35	-	-	-	-	-
Rainy	0.001***	0.001***	0.001***	0.005***	0.001***	0.005***	0.02**
Foggy	0.72	0.68	0.32	0.71	1	1	0.008***

\*  $p$ -Value $\leq$ 0.1 \*\*\* $p$ -Value $\leq$ 0.05 \*\*\*\* $p$ -Value $\leq$ 0.001



**Fig. 4** Kaplan-Meier Survival Function Plot

Table 7 presents the estimation results of regional effects by weather types. In the analysis, the seven causes of accident were evaluated to the 95% confidence level. Rainy conditions have a significant impact on all accident causes, with  $p$ -values less than 0.05 across the board. Additionally, cloudy conditions significantly affect manoeuvres ( $p=0.04$ ) while the analysis did not find statistically significant correlations for most accident causes under foggy and visible conditions.

**5.4 Kaplan-Meier survival analysis**

As previously mentioned, the Kaplan-Meier survival analysis was conducted to provide temporal insights into the likelihood of bus accidents occurring over time. This method is valuable for estimating the survival function, which in this context represents the probability of no accidents occurring within a specific time frame. Figure 4 illustrates the probability of no bus accidents occurring over time in the study period.

*Survival probability* The y-axis represents the survival probability or the likelihood of no bus accidents up until a specific point. The x-axis represents time in days.

*Survival curve* The blue line represents the estimated survival function, which illustrates the gradual decrease in the likelihood of surviving. The dark region surrounding the blue line shows the confidence interval for the estimated survival. The estimate’s degree of uncertainty increases with the width of the darkened area.

*Initial period (0–10 days)* The survival probability drops dramatically during the first ten days. This suggests that there is a reasonable chance of a bus accident during this early stage. During the first ten days, the survival probability decreases from around 1 (100%) to about 0.4 (40%).

*Mid Period (10–30 days)* Compared to the first Period, survival probability still drops, but it does so more slowly.

The survival probability drops to about 0.2 (20%) by day 20.

*Later period (30–60 days)* As time passes, there are fewer accidents since the chance of survival declines dramatically. The probability of surviving after 50 days is almost 0.1 (10%).

*End period (60–80 days)* Almost all intervals have had at least one accident, as the survival probability stabilizes and becomes closer to zero.

## 6 Discussion

The results provide critical perception into the patterns and causes of bus accidents on the Madrid-Bilbao route. By integrating Latent Class Clustering (LCC), Hierarchical Ordered Logit models, and Kaplan-Meier survival analysis, we achieve a comprehensive understanding of the influence of weather conditions on accident risks. These findings highlight the importance of weather-specific safety protocols and prompt interventions to mitigate accident risks.

The analysis of ALSA accidents on the Madrid-Bilbao bus line between 2019 and 2021 shows significant patterns, causes, and potential modifying factors. These findings have considerable implications for transport operators, policymakers, and society.

The initial analysis revealed that accidents occurred most frequently in 2019, with a significant reduction in 2020, mainly due to pandemic-related mobility restrictions. This highlights the influence of external factors on accident rates and underscores the need for transport operators to adapt safety measures dynamically in response to changing conditions. The breakdown of accidents into nine types, with “manoeuvring” as the leading cause, followed by “Car invading lane in opposite direction” and “other causes,” provides a clear basis for targeted interventions.

The relationship between weather conditions and accident causes is particularly noteworthy. The Latent Class Clustering (LCC) analysis identified three distinct clusters based on weather and accident causes, revealing that specific weather conditions significantly influence the likelihood and types of accidents. For instance, Cluster 1, associated with clear and cloudy weather, primarily involved “manoeuvring” and “Car invading lanes in opposite direction.” In contrast, Cluster 2, linked to rainy weather, included “manoeuvring,” “cars entering the roadway in the opposite direction,” and “spin.” Cluster 3, involving foggy and clear weather, was characterized by “spin” and “other causes.”

The Hierarchical Ordered Logit model highlights the significant influence of rainy weather on bus accidents, affecting all major causes of accidents. This finding aligns

with previous studies like Malin et al. [18] that have demonstrated the adverse effects of poor weather conditions on road safety. The significant correlation between cloudy weather and manoeuvres indicates that even less severe weather conditions can impact driving safety, necessitating appropriate measures. The lack of significant correlations in foggy and visible conditions for most accident causes suggests that these weather conditions may not be as critical as rainy weather. However, the significant relationship between foggy weather and crossing hits indicates specific risks that need targeted interventions.

The Kaplan-Meier survival analysis revealed a high initial risk of accidents within the first ten days, reinforcing the need for immediate safety interventions. Combined with the LCC and Hierarchical Ordered Logit findings, this temporal insight provides a comprehensive understanding of accident risks and underscores the importance of sustained safety measures over time.

Long-term policymaking should focus on weather-specific safety protocols, driver training, infrastructure improvements, and data-driven decision-making. These include developing and implementing stringent protocols to reduce accident rates, investing in better drainage systems and clearer road markings, and using advanced data analytics to monitor and predict accident trends.

The study’s limitations include its focus on accidents involving the ALSA operator on the Madrid-Bilbao route, its short data period (2019–2021), its inability to capture longer-term trends or rare weather events, and its inability to analyse factors like driver fatigue and road conditions.

## 7 Conclusions

This study takes a special approach to analysing traffic accidents on long-distance bus services between Madrid and Bilbao. We utilize data from 115 traffic accidents registered by the ALSA operator between 2019 and 2021, focusing on the interplay between weather conditions and accident causes. Our novel approach incorporates the LCC and Hierarchical methods, providing a fresh view of this critical issue.

While previous studies have examined the impact of weather on traffic accidents, our study uniquely integrates temporal insights from Kaplan-Meier analysis with clustering methods, providing a more nuanced understanding of accident risks. This integration allows us to determine weather conditions significantly influencing accident causes, offering actionable recommendations for improving bus safety on long-distance routes.

One of the novel results of this study is the identification of rainy weather as the most critical factor associated with bus accidents, which is consistent across both LCC and Hierarchical analyses. Road safety measures should mainly focus on rainy days to mitigate accident

risks. The clustering analysis further revealed that Cluster 1, characterized by manoeuvres and cars invading lanes in clear and cloudy weather, has the highest probability of occurrence (63%), emphasizing the need for targeted interventions.

Our study also provides significant temporal insights. We found that 2019 had the highest accident rate, while 2020 saw the lowest, likely due to COVID-19 mobility restrictions. This temporal variation underscores the importance of considering external factors when analysing accident trends and the potential impact of such factors on bus safety.

By integrating multiple analytical techniques, our study offers a comprehensive approach to understanding the relationship between weather conditions and bus accidents, thereby contributing significantly to the current body of literature. Further research could expand on these findings by incorporating data from multiple operators across different countries and extending the study period to identify long-term trends. Additionally, exploring the interaction of weather conditions with other factors such as driver fatigue, road conditions, and vehicle maintenance could provide a more holistic understanding of the causes of bus accidents.

#### Acknowledgements

The authors acknowledge the Spanish Ministry of Science and Innovation MCIN/AEI/<https://doi.org/10.13039/501100011033> financial support as this research was possible thanks to data collected in the project TrackBest-3 S RTC2019-007041-4. They also want to acknowledge the bus operator ALSA for the data provided and for supporting to do the surveys in its buses.

#### Author contributions

Shaghayegh Rahnama: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. Adriana Cortez: Methodology, Formal analysis, Writing - original draft. Andres Monzon: Conceptualization, Writing - review & editing, Supervision.

#### Funding

This manuscript is the result of the research work that was part of the TrackBest-3 S project and that received funds for the preparation from the Spanish Ministry of Science and Innovation.

#### Declarations

#### Competing interests

None.

Received: 24 March 2023 / Accepted: 6 August 2024

Published online: 14 September 2024

#### References

- (2015). Road traffic injuries. In: World Health Organization (WHO).
- The Spanish Transport and Logistics Observatory (OTLE) (2022). <https://observatoriortransporte.mitma.gob.es/en/mobility>. In: Ministry of transport and sustainable mobility.
- Koetse, M. J., & Rietveld, P. (2009). The impact of climate change and weather on transport: An overview of empirical findings. *Transp Res D Transp Environ*, 14, 205–221. <https://doi.org/10.1016/j.trd.2008.12.004>
- Ye, Z., Wang, C., Yu, Y., et al. (2016). Modeling level-of-safety for bus stops in China. *Traffic Injury Prevention*, 17, 656–661. <https://doi.org/10.1080/15389588.2015.1133905>
- Hammad, H. M., Ashraf, M., Abbas, F., et al. (2024). Retraction note: Environmental factors affecting the frequency of road traffic accidents: A case study of sub-urban area of Pakistan. *Environmental Science and Pollution Research*, 31, 27492–27492. <https://doi.org/10.1007/s11356-024-33079-2>
- Lee, J., Chae, J., Yoon, T., & Yang, H. (2018). Traffic accident severity analysis with rain-related factors using structural equation modeling – a case study of Seoul City. *Accident Analysis and Prevention*, 112, 1–10. <https://doi.org/10.1016/j.aap.2017.12.013>
- Litman, T. (2014). A New Transit Safety Narrative. *J Public Trans*, 17, 114–135. <https://doi.org/10.5038/2375-0901.17.4.7>
- Barabino, B., Cabras, N. A., Conversano, C., & Olivo, A. (2020). An Integrated Approach to Select Key Quality indicators in Transit services. *Social Indicators Research*, 149, 1045–1080. <https://doi.org/10.1007/s11205-020-02284-0>
- Bonera, M., Maternini, G., Parkhurst, G., et al. (2020). Travel experience on board urban buses: A comparison between Bristol and Brescia. *Eur Transp Trasp Eur*, 76, 1–12.
- Gärling, T. (2004). Changes of Private Car Use in Response to Travel Demand Management. In: ICTTP 2004. Elsevier Oxford. pp 1–22.
- Chang, L-Y., & Wang, H-W. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis and Prevention*, 38, 1019–1027. <https://doi.org/10.1016/j.aap.2006.04.009>
- Kumar, S., & Toshniwal, D. (2016). Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPC). *J Big Data*, 3, 13. <https://doi.org/10.1186/s40537-016-0046-3>
- Goh, K., Currie, G., Sarvi, M., & Logan, D. (2014). Factors affecting the probability of bus drivers being at-fault in bus-involved accidents. *Accident Analysis and Prevention*, 66, 20–26. <https://doi.org/10.1016/j.aap.2013.12.022>
- Chimba, D., Sando, T., & Kwigizile, V. (2010). Effect of bus size and operation to crash occurrences. *Accident Analysis and Prevention*, 42, 2063–2067. <https://doi.org/10.1016/j.aap.2010.06.018>
- Mohammed, A. A., Ambak, K., Mosa, A. M., & Syamsunur, D. (2019). A review of the traffic accidents and related practices Worldwide. *The Open Transportation Journal*, 13, 65–83. <https://doi.org/10.2174/1874447801913010065>
- Prato, C. G., & Kaplan, S. (2014). Bus accident severity and passenger injury: Evidence from Denmark. *European Transport Research Review*, 6, 17–30. <https://doi.org/10.1007/s12544-013-0107-z>
- Bergel-Hayat, R., Debbarh, M., Antoniou, C., & Yannis, G. (2013). Explaining the road accident risk: Weather effects. *Accident Analysis and Prevention*, 60, 456–465. <https://doi.org/10.1016/j.aap.2013.03.006>
- Malin, F., Norros, I., & Innamaa, S. (2019). Accident risk of road and weather conditions on different road types. *Accident Analysis and Prevention*, 122, 181–188. <https://doi.org/10.1016/j.aap.2018.10.014>
- Abdullaev, B., Yuldoshev, D., Muminov, T., & Axmedov, D. (2021). Improving the method of assessing road safety at intersections of single-level highways. *E3S Web of Conferences*, 264(05027). <https://doi.org/10.1051/e3sconf/202126405027>
- Gitelman, V., Carmel, R., & Pesahov, F. (2014). The evaluation of safety efficiency of non-urban infrastructure improvements; a case-study. *European Transport Research Review*, 6, 477–491. <https://doi.org/10.1007/s12544-014-0145-1>
- Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis and Prevention*, 43, 1666–1676. <https://doi.org/10.1016/j.aap.2011.03.025>
- Karlaftis, M. G., & Tarko, A. P. (1998). Heterogeneity considerations in accident modeling. *Accident Analysis and Prevention*, 30, 425–433. [https://doi.org/10.1016/S0001-4575\(97\)00122-X](https://doi.org/10.1016/S0001-4575(97)00122-X)
- Depaire, B., Wets, G., & Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. *Accident Analysis and Prevention*, 40, 1257–1266. <https://doi.org/10.1016/j.aap.2008.01.007>
- Yoon, S., Kho, S-Y., & Kim, D-K. (2017). Effect of Regional characteristics on Injury Severity in Local Bus crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2647, 1–8. <https://doi.org/10.3141/2647-01>
- Fountas, G., & Anastasopoulos, P. C. (2017). A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. *Anal Methods Accid Res*, 15, 1–16. <https://doi.org/10.1016/j.amar.2017.03.002>
- Ndongila, J. M., Natuhoyila, A. N., Nkumu, M. L., et al. (2021). Survival and predictors of patient mortality during Road Traffic accidents in the Democratic

- Republic of Congo: Historical cohort study. *OALib*, 08, 1–11. <https://doi.org/10.4236/oalib.1108059>
27. Kim, D-G., Lee, C., & Park, B-J. (2016). Use of Digital Tachograph Data to provide Traffic Safety Education and Evaluate effects on Bus driver behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2585, 77–84. <https://doi.org/10.3141/2585-09>
28. Lipton, R., Cunradi, C., & Chen, M-J. (2008). Smoking and all-cause mortality among a cohort of Urban Transit operators. *Journal of Urban Health*, 85, 759–765. <https://doi.org/10.1007/s11524-008-9295-6>
29. Akaike, H., Parzen, E., Tanabe, K., & Kitagawa, G. (1998). *Selected papers of hirotugu akaike*. Springer Science & Business Media.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.