

Article

Crash Risk Analysis in Highway Work Zones: A Predictive Model Based on Technical, Infrastructural, and Environmental Factors

Sofia Palese ¹, Margherita Pazzini ¹, Davide Chiola ², Claudio Lantieri ¹, Andrea Simone ¹
and Valeria Vignali ^{1,*}

¹ Department of Civil, Chemical, Environmental and Materials (DICAM) Engineering, University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy; sofia.palese2@unibo.it (S.P.); margherita.pazzini2@unibo.it (M.P.); claudio.lantieri2@unibo.it (C.L.); andrea.simone@unibo.it (A.S.)

² R&D and Innovation, Movyon S.p.a., 50013 Firenze, Italy; davide.chiola@movyon.com

* Correspondence: valeria.vignali@unibo.it

Abstract

Road infrastructure is the foundation of the predominant modes of transport, and its effective management is crucial to meet mobility needs. Although necessary for reconstruction, maintenance, and expansion projects, roadworks produce negative impacts, resulting in further risk for workers and drivers and failing to ensure sustainable development. The objective of this paper is twofold: Firstly, investigate the contributing factors to the occurrence of crashes in roadworks. Secondly, develop a model to estimate crash numbers in these areas. The results, which could support municipalities at the planning stage and implement policies for safe and sustainable development, are achieved by examining 121 sites, where 549 crashes occurred, and 25 contributing factors. The variables are divided into three categories: technical characteristics of the site, infrastructural, and environmental. Besides the conventional variables, a risk-increasing factor is calibrated. It assesses the impact of roadworks according to the manoeuvres imposed and the number of lanes. Consistent with previous findings, several variables related to the work zone layout, traffic conditions, infrastructure, and surrounding environment are correlated with the crash number. After performing a further statistical analysis, a multiple linear regression model, statistically significant (0.000) and suitable for accurately estimating the possible number of crashes ($R^2_{\text{adj}} = 0.41$), is determined.

Keywords: roadworks; highways; crash analysis; risk evaluation; work zone



Academic Editor: Victoria Gitelman

Received: 20 May 2025

Revised: 18 June 2025

Accepted: 29 June 2025

Published: 3 July 2025

Citation: Palese, S.; Pazzini, M.; Chiola, D.; Lantieri, C.; Simone, A.; Vignali, V. Crash Risk Analysis in Highway Work Zones: A Predictive Model Based on Technical, Infrastructural, and Environmental Factors. *Sustainability* **2025**, *17*, 6112. <https://doi.org/10.3390/su17136112>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Relevance of the Topic and Introduction

Road infrastructures are of essential importance for the guarantee of mobility, and the effective management and development of such infrastructures enable the transportation of both people and goods [1]. The necessity of moving has been further underlined by the United Nations (UN) in the context of the Sustainable Development Goals (SDGs) [2]: Specifically, within Goal 3—Ensure healthy lives and promote well-being for all at all ages, and Goal 11—Make cities and human settlements inclusive, safe, resilient, and sustainable. Nevertheless, a variety of factors contribute to the deterioration of road infrastructure, thereby hindering the maintenance of a high performance over time and affecting the comfort, safety, and execution of movements: Firstly, the transport activity, which is projected to

more than double from 2015 to 2050. In detail, passenger transport is expected to increase by a factor of 2.3, and freight transport by 2.6 [3], aggravating loads on pavements. Secondly, climate change has been shown to adversely affect pavements' lifetime and performance, worsening the rutting phenomena due to extreme temperatures [4]. The confluence of these factors is resulting in a substantial increase in the number of work zones on roads. This increase is driven by several factors, including but not limited to increasing traffic volumes, the implementation of enhanced safety standards, and extreme temperatures, which require reconstruction, maintenance, and expansion projects.

Although necessary, the presence of work zones negatively affects several aspects, leading to congestion, increasing environmental pollution and emissions [5,6], and elevating the risk, considering both workers and drivers [7]. Regarding the former aspect, transport is one of the most impactful activities on the environment in terms of land use, greenhouse gas (GHG) emissions, pollution, and toxic effects on the ecosystem [8]. Especially in areas adjacent to work zones, several stop-and-go phenomena occur for different reasons, such as a reduced number of lanes or distraction by drivers. These continuous accelerations and decelerations cause a significant increase in carbon emissions [9]. Road crashes are considered a public health priority worldwide since, according to the World Health Organization, road injuries represent one of the main causes of death, especially among young people [10]. The areas adjacent to work zones represent critical locations of the road network, where the crash risk and frequency are higher, raising the question of safety for all road users [5]. The subject of crashes is of global relevance and represents a major social disaster of particular interest. In detail, in 2022, among the 27 Member States of the European Union, road crashes resulted in about 20,634 deceases and more than 1.13 million injuries. Regarding the different road types, in the same year, 62% of deaths occurred on rural roads and motorways, characterised by higher travel speeds and the presence of heavy vehicles [11]. This is a subject of interest in Italy, characterised by one of the highest motor vehicle ownership rates in Europe, with 693.2 cars per 1000 inhabitants [12].

The negative impact of work zones is a subject of increasing relevance to researchers, institutions, and public administrations, and the attention to this topic is growing. Consequently, research is investigating the issue across a range of road types, as they involve different user categories, including pedestrians, cyclists, motorcyclists, car and truck drivers, and others. On highways, the number of journeys on the network in 2023 for all types of vehicles (heavy and light) increased compared to 2022, reaching the highest values ever recorded along the entire Italian highway network [12]. In addition, comparing 2021 and 2022, accidents and victims on highways increased by 9.7% and 19.9%, respectively [13]. Although the negative influence of work zones' presence is evident in all road types, it is more impactful on highways, where the capacity is drastically reduced and drivers are forced to perform unusual manoeuvres while driving at high speed [14]. Researchers have noted a correlation between the presence of roadworks and crashes, especially on highways, where the likelihood of injury increases for both drivers and workers [15,16]. In addition, highway maintenance works involve several unsafe aspects, such as high flows and speeds, stressed pavements, and weather-related risks to personnel. These features require a high level of attention from management departments—in particular, on safety evaluation during maintenance works [17]. In light of the substantial demand for highways, it is a common practice for the infrastructure to remain operational during the construction phase. This, however, often results in disruptions to the driving environment due to the implementation of various countermeasures, such as lane closures or road diversions [18]. Indeed, it is not always feasible to stop traffic to perform roadwork, and the common practice consists of closing only the lane in which the work is being carried out and guiding vehicles to the

adjacent lane during maintenance periods. This period varies in duration—it may last hours, weeks, or months, depending on the type of work and the boundary conditions [19].

For the aforementioned reasons, the objective of this paper is twofold: firstly, to quantify the impact of work zones in terms of possible crashes, and secondly, to identify the factors that most affect this phenomenon in order to minimise it. Indeed, although many existing contributions in the literature have studied the impact of work zone areas on the crash rate, these studies have several shortcomings, including the limited availability of data and a lack of specific information on the roadwork plan adopted backed up by a large and varied database.

1.2. Literature Review

Given the paramount importance of the subject matter and the necessity of implementing work zones on infrastructural elements, numerous contributions have been made over the years. These contributions have explored the subject, with a particular emphasis on the safety of both drivers and workers. When analysing the latest knowledge regarding the risk of users driving through roadworks, various contributions are available both from institutions and the scientific community. Regarding the former, many European projects investigated road safety in a work zone environment, evaluating its impact and drivers' perceptions. Three examples of significant projects are ARROWS (Advanced Research on Road Work Zone Safety standards in Europe) [20], STARS (Scoring Traffic At Roadworks) [21], and ASAP (Appropriate Speed Saves All People) [22]. The ARROWS project aims to develop measures and principles related to roadwork safety to better manage the planning, project, implementation, and operation of work zones (European Commission—Advanced research on road work zone safety standards in Europe). The STARS project develops a methodology to score roadwork plans on three interdependent aspects, which are usually considered independently: road user safety, road worker safety, and network performance [21]. Last, the ASAP project deals with speed management in work zones to ensure that drivers can safely navigate the vehicle through the work zone routing [22]. As for the scientific community, many contributions have focused their attention on the impact of work zones. The researchers have shown that roadworks affect traffic conditions, causing congestion and raising levels of environmental pollution and emissions [5,6]. The environmental impact is not the only consequence of the presence of the work zones on drivers. It has been determined that they also influence the large-scale performance of the road network since the average travel time of vehicles increases by 20–50% and the capacity is reduced by 10–20% [16].

A further noteworthy consequence of the presence of work zones is associated with the phenomenon of road traffic crashes. When conducting a study of crashes, it is imperative to consider the crash rate as a fundamental element. A substantial body of research has indicated that users traversing work zones are more likely to be involved in crashes and that the severity of these collisions is increased. When considering the crash rate, a significant number of studies examining crashes before and during construction periods have pointed out that they are more likely to occur during roadworks' execution [23,24] and that the risk of crashes is higher in work zone areas [25]. Nevertheless, some findings appear inconsistent with the outcome above since they show no statistically significant variations [26], determine a lower crash rate during the roadwork period [27], or find a higher probability of crashes along sections without a work zone when comparing sections of the same infrastructure with and without a work zone [28]. A crucial point to consider for a comprehensive safety analysis is that the crash rates before and during the installation of a work zone appear to be higher than that recorded post-intervention [27], confirming the importance of implementing roadworks. To mitigate both the frequency and severity of

crashes near roadworks and to enhance their safety, the scientific community is addressing this issue through various approaches. Among these, three main topics are roadworks' features, driving behaviour, and determining the factors that most affect safety, linked to the implementation of mathematical models.

Firstly, the research on work zone configurations is being expanded, with a focus on roadwork plans as well as the characteristics, functionality, and placement of individual traffic signals. Considering the overall configuration, several contributions investigated the impact of an innovative layout [29] or compared similar work zone plans, including different merging strategies [30] and their impact on the workload [14]. In addition, researchers determined that layout configurations involving a crossover are extremely critical and have the worst effects in terms of safety [24]. The configuration and positioning of signs can also influence the incidence of crashes. The fundamental aspects are providing distance information and several warning signs. Researchers have demonstrated that increasing distance information to support a driver's visual perception can enhance their level of vigilance. However, the number of signs is a critical issue since an excessive number of signs can increase the workload of drivers and decrease driver alertness. Indeed, studies found that the number of warning signs significantly influences the warning effect and that the wrong amount of road signs can distract drivers and lead them to miss important information [15]. According to the research, installing the optimal number of warning signs will lead to about 6% lower accident costs and operating costs [31]. Specific areas of interest for research involve the impact analysis of images and texts in signalling systems, and also assessing the effect of new technologies to inform drivers, such as Variable Message Signs (VMSs) [32–34].

The second topic pertains to driving behaviour in areas adjacent to work zones, incorporating aspects such as driving style, user perception, vigilance, and workload. The human factor constitutes a crucial element in the analysis of crashes, a consideration that assumes even greater significance when the environment of the work zone is taken into account. Previous studies have demonstrated that the geometry of the work zone impacts driving behaviour [29]. Research efforts are focusing on this topic by conducting several tests both in the real environment and through driving simulators. Workload data can be collected using different methodologies, such as eye tracking, heartbeat, and neural activity analysis, and measuring the mental workload while driving indicates the cognitive demands placed on the driver [19]. Different road, traffic, and environmental conditions in work zones provide different amounts of stimulation to the driver, which determines the driving workload [5]. According to the literature, in work zones, the frequency of road sign glances is higher than in the normal road section, confirming the importance of a clear and visible site layout [35]. In addition, previous research also demonstrated an interaction between longitudinal control and the standard deviation of horizontal gaze [36].

Lastly, several contributions to the research focused on identifying factors affecting safety to predict the likelihood of crashes on highway infrastructure with statistical models. Despite the great efforts at crash prediction, especially for freeway conditions, a limited number of studies have conducted crash prediction for traffic in work zones [16]. The performance of statistical models strongly relies on the quality and quantity of collected crash data. Road accidents are rare and random events [37–42], and the presence of roadwork is a determinant condition for this kind of study that is not included as frequently as other traffic analysis conditions. Furthermore, given the complexity and diversity of possible work zone plans, the acquisition of a sufficiently large, varied, complete, and detailed database represents an obstacle to deepening research on this theme. The characteristics of the required data have led to specific and sectoral studies, and this issue is often overcome by collecting data through traffic simulations with dedicated software [14,43,44] or driv-

ing simulators [29,32,36]. Many contributions investigated crash-related factors through different approaches. When excluding causes linked to drivers' behaviour and vehicle characteristics, and objects of other specific analyses, the remaining causes can be divided into three macro categories: work zone-related factors, infrastructure-related factors, and environmental factors. Within the work zone-related factors, the literature includes the duration and timing (day/night) [45], the plan, its features, and complexity [34,46–48], the speed limit [43,49], the work zone length [48], the warning zone length [44], and others. The road type [50], the road geometry [51], the traffic volume [45], and the percentage of heavy vehicles [14,47,48] are some of the factors related to infrastructure. Finally, considering the variables included in the environmental category, weather conditions and temporal and lighting characteristics [14,34,46,51] have been demonstrated to influence crash occurrence in work zones.

Although the correlation between the occurrence of crashes in areas adjacent to work zones and the above factors has been extensively assessed, the methods of detection adopted should be further developed. Indeed, only a few studies focus on quantifying the number of crashes resulting from these factors. In addition, considering that the available contributions consider a limited number of causes, one of the main objectives of this study is to simultaneously take into account a number of widely recognised variables, which, in most existing research, tend to be investigated singularly.

A relevant model estimating the number of crashes depends on the road type, traffic volume (Annual Average Daily Traffic—AADT), and length and duration of the roadwork [50]. The scientific community considers the above variables to be the most significant. Nevertheless, this model does not include many of the technical characteristics of the work zone and infrastructure and does not involve any features of the surrounding environment. On the other hand, other models take account of the technical characteristics of the work zone and do not include a wide variety of roadwork plans or other variables mentioned above. Based on the methodology established by Meng et al., the present research aims to expand the developed model by incorporating additional variables related to the work zone, the infrastructure, and the external environment. Moreover, this study includes detailed aspects related to the complexity of the work zone's layout.

The primary objective of this paper is to ascertain the correlation between the number of road traffic crashes that occur at roadworks and several of the most common variables that have been documented in the extant literature, incorporating all three existing macro areas. The second objective is to quantify this relationship by developing a statistically significant indicator capable of estimating the number of possible crashes in a work zone, given its characteristics.

To achieve this goal, the paper is structured as follows: Section 2—Materials and Methods—describes the early data processing and the statistical analysis conducted, Section 3—Results—illustrates the results obtained, Section 4—Discussion—evaluates and discusses the outcomes, and Section 5—Conclusions—sums up the main findings achieved by the current research.

2. Materials and Methods

2.1. Regression Process

Multiple linear regression (MLR) was selected as the statistical technique to reveal the relations between the number of crashes in work zone areas and their characteristics and external conditions. Previous studies have employed multiple linear (or linearised) regression models to estimate the association between the number of crashes that occurred in areas adjacent to the work zone and the variables identified as significant. In particular, previous studies have applied Equation (1) [50], which expresses the relation between the

logarithm of the number of crashes at a given roadwork and its characteristics. A model is developed, characterised by the functional relationships shown below. In the formula, Y is the dependent variable and specifically in the current study corresponds to the expected number of crashes (C_n), X_i represents the independent variables identified as significant for the occurrence of the phenomenon, the coefficient α_0 is the constant, α_i indicates the values to be estimated that weight each regressor X_i , and ε is the random error term.

$$\ln(Y) = \alpha_0 + \sum_i \alpha_i \cdot X_i + \varepsilon \quad (1)$$

The variables considered for the building of the regression model are selected among all those either found to correlate with the number of crashes or identified as significant by previous studies in the literature. Before performing the regression, it was determined whether each of the independent variables involved in the model was normally distributed. For large samples, this check is conducted using the Kolmogorov–Smirnov test. Furthermore, to evaluate the trend toward normality, the skewness and kurtosis of the distributions of the independent variables are assessed. In cases where certain variables deemed significant for the model fail to meet the normality requirement, logarithmic and inverse transformations are employed. After the application of the aforementioned transformations, if the variables are found to be normally distributed, their transformed version is entered into the regression model. This condition is not verified for binomial variables, which are never normally distributed, due to their inherent structure.

Since the frequency of work zone crashes depend on the selected variables, the function form can be identified by using the least squares method based on the available database. A trial-and-error process was used to construct the model to find the best combination of variables and their functional forms (logarithmic, inverse, or linear) to represent the crash phenomenon at roadworks.

2.2. Database Creation

The database describes various aspects of the work zones and is composed of data on crashes that occurred from 2010 to 2023 on the Italian highway network. This study provides a comprehensive overview of the events under investigation, incorporating crashes of all severity levels in order to ensure a realistic and representative sample. The work zones, characterised by specific features, were selected from the total available data. In detail, the current analysis is based on the work zones where two or more crashes occurred. Indeed, considering the long duration of some roadworks and the randomness of the crash phenomenon, a certain number of crashes could have occurred even under standard conditions, in the absence of the work zone. For that reason, this approach aims to ensure that only work zones with the greatest impact in terms of crashes, and with a stronger cause–effect relationship between the number of crashes and the presence of the site, are considered. The upper limit of the dependent variable was 46, which is the highest number of crashes that occurred at a work zone in the database. In addition, for similar reasons, the selection of the sample was performed based on duration, length, and flow per lane.

As regards duration, roadworks present for more than three days were included in the analysis, to ensure that the alternating day–night cycle occurred repeatedly and both light and dark conditions were experienced by drivers. The upper limit of the duration interval is 174 days, corresponding to the longest duration. As for the length, to ensure that the work zones were long enough to be correctly perceived by drivers, the roadworks selected are characterised by a minimum length of 1.5 km. The maximum length is 9.0 km, corresponding to the value of the longest length. Finally, regarding the flow per lane, only work zones where the average daily flow per lane was between 5000 and 15,000 [veh/(day*lane)] were

included in the analysis. This approach ensured that roadworks characterised by absent flows or overcongested flows were not considered, thus avoiding false crash numbers that did not reflect the real risk generated by the work zone. After performing the selection process, the resulting database was composed of 121 sites, where a total of 549 crashes occurred. In light of the filters adopted and the resulting dataset, it should be highlighted that the model developed in this paper is valid and applicable only to work zones whose characteristics in terms of length and duration fall within the explicit ranges outlined in the current section.

For each of the construction sites forming this database, several features were determined, which are illustrated in the following sections.

2.3. Work Zone Plan Analysis

In light of the wide range of potential work zone configurations, the analysis incorporates diverse layouts, ranging from simple to complex, with the objective of evaluating a heterogeneous sample. In this study, six distinct plans for lane closure are investigated, with the specific type of work performed in the closed areas excluded from the current analysis. These six plans correspond to the roadworks identified in the database creation phase. Generally, the plans may involve closing, flexing, and deviating lanes; a combination of these configurations is allowed, too. Depending on the type of work and specific needs, the plans could be used on highways with different numbers of lanes, affecting both the traffic quality and the risk to users. Considering that the same lengths of work zones may have different impacts depending on their layouts, a parameter was identified to take this into account. Indeed, the closure or the deflection of a single lane affects the above-mentioned aspects differently if it is performed on a two-lane or a four-lane highway. Therefore, a parameter was calibrated to evaluate the impact of roadworks based both on the types of manoeuvres imposed by the plan and the number of lanes on the motorway where the plan is applied. The procedure for identifying a parameter, as the product of several multiplicative factors, is based on a previous study [21]. As suggested in that project, the ideal method for determining the multiplying factors is to perform a regression analysis. However, due to a lack of data, this was not possible. Consequently, this was achieved based on the literature and our judgement.

This parameter, identified by the letter M , represents a risk-increasing factor. The value of M is equal to 1 when there are neither closures nor unusual manoeuvres due to the work zone, and thus the risk level remains unvaried. Conversely, the value of M increases as the number of closures, deflections, and deviations increases about the total number of lanes, according to Equation (2), displayed below. In this equation, six coefficients can increase the risk resulting from the presence of the work zone on the infrastructure. All the coefficients involved are greater than or equal to one. The site layout can include three different configurations and their combinations (closing lanes, flexing lanes, and deviating lanes). These three options are expressed, respectively, by the coefficients L_1 , L_2 , and L_3 , which determine the additional risk due to the configuration of the site layout. Specifically, a higher multiplication factor is assigned when more lanes are closed (1.05 for an emergency lane closure, 1.1 for a one-lane closure, 1.2 for a two-lane closure, and so on), determining L_1 . A similar procedure was also applied to flexing and deviating lane manoeuvres, determining L_2 and L_3 , respectively. Considering that the number of lanes on Italian motorways currently varies between 2 and 4, L_1 and L_3 are between 1 and 1.3, and L_2 is between 1 and 1.4. This risk is evaluated by the lane modification mechanisms and the number of lanes affected by such changes. Moreover, a coefficient is employed to consider the number of closed lanes not as an absolute value, but about the total number of lanes initially available (P_{CL}), which is between 1 and 1.3. Finally, two risk factors were considered: one related to moving sites (1 for stationary sites, 1.1 for slow-moving sites, and

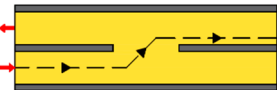
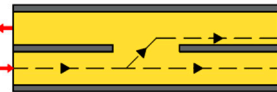
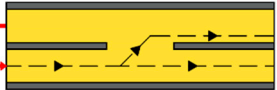

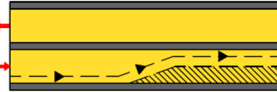

1.2 for moving sites), and one related to moving sites with link roads (1 for sites without a link road and 1.2 for sites with a link road).

$$M = P_{CL} \cdot A \cdot D \cdot \prod_{i=1}^3 L_i \quad (2)$$

The evaluation of the parameter M applied to the six plans reveals values ranging from 1.32 to 2.07. These values denote additional risk factors in comparison to the scenario where M is 1, for a section without work zone areas under analogous conditions. Considering a section of length L of the same infrastructure, it is assumed that the additional risk created by the roadwork is expressed as a linear combination of L and M . The risk of the site depends on L and M and consequently it can be reduced by either shortening the site or simplifying the manoeuvres (i.e., lowering M). As it is often not possible to change the number of lanes affected by the work zone for technical reasons, expressed through L_1 , L_2 , and L_3 , other factors are considered instead. In fact, for roadworks that are fixed and do not extend onto the opposite carriageway, the values of A and D are lower. Specifically, A is 1.2 for mobile sites, 1.1 for slow-moving sites, and 1 for fixed sites, while D is 1.2 with a link road and 1 without. This results in lower values of M and, consequently, G for situations closer to standard traffic conditions.

The six plans involved in the current analysis are presented in Table 1 below. The table gives a detailed graphical representation of the layout for each working area, a brief description, the number of work zones included in the database under examination, and the M -values for two-, three- and four-lane highways. The empty M fields in the table correspond to cases where it is not possible to determine M , such as a three-lane bend on a two-lane highway. The order in which the plans appear in the table is random; they are not listed in order of increasing complexity.

Table 1. Illustration of the six site layouts involved in the analysis and the corresponding M -values. The labels 1, 2, and 3 lanes, adjacent to the single hatches, represent the number of lanes affected by the manoeuvre and which remain open to traffic.

Plan ID	Image	Description	Roadwork Number	M		
				2 Lanes	3 Lanes	4 Lanes
1	 1 Lane	One-lane deviation	57	1.58	1.87	2.03
2	 1 Lane	Link road with one lane deviated	33	1.44	1.74	2.07
3	 2 Lanes	Link road with two lanes deviated	2	/	1.56	1.89
4	 3 Lanes	Three lanes flexed	4	/	1.37	1.57
5	 1 Lane	Reduction up to a single lane	17	1.32	1.56	1.69
6	 2 Lanes	Reduction up to two lanes	8	/	1.21	1.44

2.4. Identification of Variables and Correlation

Following a comprehensive review of the extant literature on factors contributing to crashes in work zone areas, and an analysis of the data available in the database, the variables illustrated in Table 2 were selected for further investigation. Among the various causes contributing to the occurrence of crashes at work zones (WZ in table), the selected ones can be divided into three macro categories, related to the technical characteristics of the work zone, the infrastructure, or the external environment.

The first category includes variables such as Occupancy of the carriageway (O), Duration (D), Length (L), and Risk-increasing factor (M). When considering the complexity of the site, the Carriageway occupancy variable specifies whether the site only involves one direction of travel or whether it also occupies the adjacent carriageway, impacting the flow in the opposite direction. Duration and Length reflect the occupation of the work zone in space and time, respectively. The Risk-increasing factor, described extensively in the previous section, identifies additional risk factors due to the complexity of the roadwork plan and the characteristics of the infrastructure on which it is applied. In addition, a third variable is introduced to represent the additional risk due to the complexity of the work zone along all its length. This variable is determined as the linear combination of M and L and is named Global complexity (G).

The variables related to the *Infrastructure's* features are Annual average traffic volume (AADT) per lane (F), Heavy vehicle AADT per lane (HV), Number of lanes (N), Two-lane highway (N_{L2}), Three-lane highway (N_{L2}), Four-lane highway (N_{L3}), Geometry-related variables (straight (S_t), curved (C), or mixed (X) work zone), and Junction-related variables (work zone involving entry junctions (I), exit junctions (Q), both (E), or any junctions (Z)). In detail, AADT per lane and Heavy vehicle AADT per lane, respectively, represent the total and heavy vehicle flow on an average day within a single lane. The heavy vehicle category includes vehicles over 1.3 m in height and vehicles with 3 or more axles. The total AADT per lane also comprehends all vehicles less than 1.3 m in height and two-wheeled vehicles. The variable Number of lanes expresses the number of lanes in the infrastructure affected by the site. The Geometry variable identifies the geometric elements along which the site is located, that is, in curves, straight lines, or both. The variable Junctions determines whether there are any junctions along the extension of the site, specifying their type (entry or exit junctions).

Last, the *External Environment* category involves Yearly seasons (Y), which denote the seasons over which the work zone runs, and Unusual locations (U), which are used to express whether the development of the site experiences specific conditions such as bridges, tunnels, or if it is along the shore. Finally, the table below also presents the characteristics of the main variable of the current study, namely, the number of crashes.

Table 2. Definitions of the variables.

Category	Variable	Type	Acronym	Unit or Description of Values
<i>Crash</i>	Crash number	Continuous	C_n	/
<i>Work zone</i>	Carriageway occupancy	Binary	O	0 for WZs involving one carriageway; 1 for WZs involving two carriageways
	Duration	Continuous	D	days
	Length	Continuous	L	km

Table 2. Cont.

Category	Variable	Type	Acronym	Unit or Description of Values	
<i>Work zone</i>	Risk-increasing factor	Continuous	M	/	
	Global complexity	Continuous	G	km	
<i>Infrastructure</i>	AADT per lane	Continuous	F	Vehicles per lane per day	
	Heavy vehicle AADT per lane	Continuous	HV	Heavy vehicles per lane per day	
	Number of lanes	Categorical	N	2, 3, or 4 lanes	
	Two-lane highway	Binary	N _{L2}	0 if false, 1 if true	
	Three-lane highway	Binary	N _{L3}	0 if false, 1 if true	
	Four-lane highway	Binary	N _{L4}	0 if false, 1 if true	
	Geometry	Expressed as the composition of two binary variables:			
		The WZ is along straight lanes		St	0 if false, 1 if true
		The WZ is along curves		C	0 if false, 1 if true
	Junctions	Expressed as the composition of four binary variables:			
The WZ meets only entry junctions			I	0 if false, 1 if true	
The WZ meets only exit junctions			Q	0 if false, 1 if true	
The WZ meets entry and exit junctions			E	0 if false, 1 if true	
	The WZ meets no junctions		Z	0 if false, 1 if true	
<i>External environment</i>	Yearly seasons	Expressed as the composition of four binary variables:		More than one can be true if the WZ is performed for more than one season	
		Summer	SU	0 if false, 1 if true	
		Autumn	AU	0 if false, 1 if true	
		Winter	WI	0 if false, 1 if true	
	Spring	SP	0 if false, 1 if true		
	Unusual locations	Expressed as the composition of three binary variables:			Both can be true if the WZ meets both the bridge and tunnel during its path
		The WZ is along a bridge		B	0 if false, 1 if true
The WZ is along a tunnel			T	0 if false, 1 if true	
	The WZ is along a shore		Sh	0 if false, 1 if true	

After identifying the variables, the correlation between them and the number of crashes per work zone is investigated. In detail, the values of these variables and the correspondent number of actual crashes are determined for each site in the database. This preliminary step of the analysis was performed through Pearson correlation and allowed us to outline the most influential variables to be involved in the next phase, characterised by building the regression model.

3. Results

3.1. Correlation

In order to ascertain which of the variables demonstrated a significant impact on the occurrence of crashes in proximity to work zones, and to elucidate the interrelationships among these variables, the Pearson correlation coefficient was analysed. The first step was to determine whether the risk-increasing factor was related to the number of crashes at a site, to assess the effectiveness of this parameter in representing the phenomenon. To extensively test the reliability of the M parameter, the correlation between it and the number of crashes (the logarithm of this value) was first assessed on a larger dataset than that obtained after the database creation process. The dataset considered for this first correlation consisted of 229 work zones where 827 crashes occurred. It should be noted that the 121 sites selected for the in-depth analysis presented below are a subset of the original 229 sites, obtained after the filtering process described in Section 2.2 Database Creation. The results show that the Pearson coefficient is 0.187 for the correlation with the logarithm of the number of crashes. In addition, this correlation has a strong significance (sig. < 0.01). Furthermore, when applying this procedure only to the sites included in the database under analysis, a Pearson correlation coefficient of 0.220 is obtained. This value appears to be significant (sig. < 0.05). In line with expectations, the correlation is positive for both datasets. In fact, due to the way it was constructed, M represents a risk-increasing factor, and it is therefore coherent that the correlation with the number of crashes is positive. Therefore, based on our first analysis, the parameter M seems to effectively reflect the additional level of risk posed by different plans applied to various infrastructures.

The results of the correlation analysis for the three macro categories are, respectively, provided in Table 3, Table 4, and Table 5. Starting from the variables belonging to the category *Work zone's* characteristics (Table 3), they appear to be moderately correlated with the logarithm of the number of crashes. In particular, the Length, Duration, and Carriageways' occupancy were positively correlated with the occurrence of crashes at roadworks. As outlined in previous sections, a variable resulting from the linear combination of the length and the risk-increasing factor, defined as Global complexity, was also considered, since the additional risk due to the complexity of the plan was assumed to be distributed throughout the entire length of the site. The correlation between the variable G and the occurrence of crashes in work zones, although still moderate, is more pronounced than the correlation between L and crashes.

Table 3. Results of the correlation between the logarithm of the number of crashes and variables belonging to the *Work Zone* category.

	$\ln(C_n)$	O	D	L	M	G
$\ln(C_n)$	1	0.209 *	0.360 **	0.182 *	0.220 *	0.213 *
O	-	1	0.147	0.266 **	0.762 **	0.407 **
D	-	-	1	-0.155	-0.211 *	-0.170
L	-	-	-	1	0.301 **	0.973 **
M	-	-	-	-	1	0.494 **
G	-	-	-	-	-	1

** Sig. < 0.01. * Sig. < 0.05.

Considering the variables belonging to the category *Infrastructure's* characteristics, Table 4 shows the Person correlation coefficients between all the variables included in the group and the number of crashes. This analysis establishes a slight positive correlation between the number of crashes and both the AADT per lane (F) and the Heavy vehicles AADT per lane (HV). These variables are interrelated, given that HV constitutes a subset of F . Although the correlation is moderate in both cases, it is higher for the HV variable. When

examining the variables related to the existing lanes in the infrastructure under standard conditions, the variable generally representing the number of lanes (N) does not appear to be correlated. On the other hand, there is a significant correlation between the number of crashes and the presence of four lanes under standard conditions. Contrary to expectations, the number of crashes does not appear to be correlated with either the variables linked to geometry or the variables linked to the presence of junctions within the construction worksite, which in general generate a disturbing condition. Finally, the variables belonging to the same field of study, i.e., flow (F and HV), lanes (N, NL₂, NL₃, and NL₄), geometry (St, C, and X) and junctions (I, Q, E, and Z), were strongly correlated. Other correlations shown between different variables were purely accidental, such as the correlation between the number of lanes and the fact that the work zone extension involves both curved and straight segments of road.

Table 4. Results of the correlation between the logarithm of the number of crashes and variables belonging to the *Infrastructure* category.

	ln(C _n)	F	HV	N	NL2	NL3	NL4	S _t	C	X	I	Q	E	Z
ln(C _n)	1	0.191 *	0.224 *	0.050	0.062	−0.194 *	0.284 **	0.123	−0.127	−0.018	−0.047	0.025	−0.030	0.039
F	−	1	0.935 **	0.246 **	−0.078	−0.148	0.508 **	0.443 **	−0.089	−0.304 **	−0.142	0.075	0.032	0.002
HV	−	−	1	0.275 **	−0.077	−0.187 *	0.591 **	0.544 **	−0.074	−0.396	−0.045	0.041	−0.010	0.011
N	−	−	−	1	−0.940 **	0.704 **	0.617 **	0.551 **	0.013	−0.460 **	−0.155	−0.091	0.049	0.071
NL2	−	−	−	−	1	−0.903 **	−0.313 **	−0.407 **	−0.040	0.359 **	0.164	0.088	−0.117	−0.011
NL3	−	−	−	−	−	1	−0.124	0.155	0.067	−0.171	−0.149	−0.068	0.184 *	−0.068
NL4	−	−	−	−	−	−	1	0.598 **	−0.055	−0.453 **	−0.051	−0.051	−0.135	0.174
S _t	−	−	−	−	−	−	−	1	−0.092	−0.757 **	−0.086	−0.086	−0.051	0.129
C	−	−	−	−	−	−	−	−	1	−0.581 **	−0.066	−0.066	−0.100	0.156
X	−	−	−	−	−	−	−	−	−	1	0.114	0.114	0.107	−0.207
I	−	−	−	−	−	−	−	−	−	−	1	−0.061	−0.161	−0.295 **
Q	−	−	−	−	−	−	−	−	−	−	−	1	−0.161	−0.295 **
E	−	−	−	−	−	−	−	−	−	−	−	−	1	−0.776 **
Z	−	−	−	−	−	−	−	−	−	−	−	−	−	1

** Sig. < 0.01. * Sig. < 0.05.

Table 5. Results of the correlation between the logarithm of the number of crashes and variables belonging to the *External Environment* category.

	ln(C _n)	SU	AU	WI	SP	B	T	S _h
ln(C _n)	1	0.277 **	0.079	0.145	0.022	0.091	0.180 *	0.027
SU	−	1	−0.125	−0.229 *	−0.079	−0.019	0.054	0.004
AU	−	−	1	−0.088	−0.660 *	0.014	−0.180 *	0.071
WI	−	−	−	1	0.043	0.041	0.187 *	0.074
SP	−	−	−	−	1	−0.052	0.051	−0.161
B	−	−	−	−	−	1	0.217 *	0.222 *
T	−	−	−	−	−	−	1	0.309 **
S _h	−	−	−	−	−	−	−	1

** Sig. < 0.01. * Sig. < 0.05.

Table 5 presents the results of the correlation performed within the variables related to the *External environment*. In terms of seasonality, crashes are more likely to occur during the summer, as a moderate correlation appears between the number of crashes and this season. Considering external environmental factors that have the potential to distract drivers due to their unusual nature, the analysis exhibits a marginal correlation with crashes. The only variable that manifests a slight correlation with the number of crashes is the presence of work zones within tunnels. As in the previous case, variables belonging to the same study area (annual season) tend to be correlated. Different results were obtained for variables corresponding to unusual locations, as the presence of one does not imply the presence of the other.

3.2. Regression Model

In the previous subsection, a comprehensive investigation was presented on all the variables that were deemed to be significant predictors of the number of crashes in a work zone, given its characteristics. In detail, Carriageways' occupancy, Duration, Global complexity, AADT per lane, Heavy vehicle AADT per lane, Three-lane highway, Four-lane highway, Summer, and Tunnels are found to be correlated in a statistically significant way with the logarithm of the number of crashes.

Among the selected variables, Carriageways' occupancy, Three-lane highway, Four-lane highway, Summer, and Tunnels are binomial variables. On the other hand, Duration, Global complexity, AADT per lane, and Heavy vehicle AADT per lane are continuous variables that need to meet the normal distribution requirements presented in Section 2.1. For this reason, inverse and logarithmic transformations are applied to the variables Duration, AADT per lane, and Heavy vehicle AADT per lane, which do not meet this condition. For each variable, the best transformation is determined, based on the fulfilment of the normality requirements and the values of Pearson's correlation coefficients. This results in the following variables: logarithm of duration ($\ln(D)$), the inverse of AADT ($1000/F$), and Heavy vehicle AADT ($1000/HV$) per lane. The results of the normality tests for the variables included in the model are presented in Table 6. In consideration of the scale of the latter two variables, a factor of 1000 is applied and employed in the following analysis. This results in an order of magnitude comparable to that of the other variables under examination. Finally, MLR is implemented to determine an initial model, which involves all the aforementioned variables, and the results are presented in Table 7.

Table 6. Results of the normality tests of the variables included in the MLR.

Variable	Kolmogorov—Smirnov		
	Statistics	DoL	Sig.
G	0.080	121	0.057
$\ln(D)$	0.080	121	0.056
1000/F	0.065	121	0.200
1000/HV	0.061	121	0.200

When observing the data from the model, the *Three-lane highway* variable does not appear significant in the MLR, since the significance far exceeds the reference values (>0.05). In addition, for the Heavy vehicle AADT per lane, the sign of the regression coefficient does not reflect what would be expected. Previous studies have found an increase in crashes as the number of heavy vehicles increases, which should be reflected by a negative coefficient of the variable $1000/HV$ in the model. Furthermore, the variable's significance, though only slightly, exceeds the ranges. In addition, the total and heavy vehicle flow variables demonstrate pronounced multicollinearity. The high VIF (Variance Inflation Factor) values for $10,000/F$ and $10,000/HV$ suggest strong collinearity, consistent with the expectation since heavy vehicle flow is a component of total flow.

In light of these considerations, a new MLR model was evaluated without considering these two variables, and the reference values of the final model are shown in Table 8. Consequently, the model presented in Equation (3) was derived, and the model that represents the frequency of crashes on construction sites can be expressed in accordance with Equation (4).

$$\ln(A_n) = -0.272 + 0.034 \cdot G + 0.249 \cdot \ln(D) - 0.304 \cdot \frac{1000}{F} + 0.349 \cdot O + 0.231 \cdot SU + 0.962 \cdot N_{L4} + 0.263 \cdot T \quad (3)$$

$$A_n = D^{0.249} \cdot e^{-0.272+0.034 \cdot G - 0.304 \cdot \frac{1000}{F} + 0.349 \cdot O + 0.231 \cdot SU + 0.962 \cdot N_{L4} + 0.263 \cdot T} \quad (4)$$

Table 7. Results of the initial regression model.

Model Recap							
R	R ²		R ² _{adj}		Std. error	Durbin Watson	
0.676	0.457		0.413		0.506	1.838	
ANOVA							
	Sum of squares		Mean square		F	Sig.	
Regression	23.964		2.663		10.399	0.000	
Residual	28.422		0.256		-	-	
Total	52.386		-		-	-	
Table of Coefficients							
	Non-standardised coefficients		Standardised coefficients		Collinearity		
	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Constant	−0.569	0.387	-	−1.470	0.144		
G	0.029	0.019	0.124	1.486	0.140	0.704	1.421
ln(D)	0.258	0.045	0.425	5.712	0.000	0.881	1.135
1000/F	−1.002	0.480	−0.424	−2.089	0.039	0.119	8.415
1000/HV	2.192	1.383	0.336	1.584	0.116	0.109	9.189
O	0.406	0.128	0.266	3.176	0.002	0.695	1.438
SU	0.243	0.097	0.182	2.503	0.014	0.923	1.084
N _{L4}	−0.017	0.118	0.338	3.915	0.000	0.780	1.282
N _{L3}	1.117	0.285	−0.011	−0.144	0.886	0.656	1.525
T	0.324	0.115	0.246	2.827	0.006	0.648	1.544

Table 8. Results of the final regression model.

Model Recap							
R	R ²		R ² _{adj}		Std. error	Durbin Watson	
0.667	0.445		0.411		0.507	1.904	
ANOVA							
	Sum of squares		Mean square		F	Sig.	
Regression	23.318		3.331		12.949	0.000	
Residual	29.068		0.257		-	-	
Total	52.386		-		-	-	
Table of Coefficients							
	Non-standardised coefficients		Standardised coefficients		Collinearity		
	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Constant	−0.272	0.333		−8.818	0.415		
G	0.034	0.019	0.148	1.845	0.068	0.760	1.316
ln(D)	0.249	0.045	0.411	5.560	0.000	0.898	1.114
1000/F	−0.304	0.183	−0.128	−1.656	0.100	0.817	1.224
O	0.349	0.120	0.229	2.910	0.004	0.794	1.259
SU	0.231	0.097	0.174	2.387	0.019	0.930	1.076
N _{L4}	0.962	0.262	0.291	3.666	0.000	0.779	1.283
T	0.263	0.098	0.199	2.682	0.008	0.888	1.126

4. Discussion

Variables in Table 3, especially those related to Length, Duration, and Carriageways' Occupancy, were positively correlated with the occurrence of crashes at roadworks. This finding is consistent with the results of earlier studies in the relevant literature, which showed a positive correlation between length and the occurrence of crashes [23,52–54] and between duration and the occurrence of crashes [23]. When considering the Global Complexity parameter in further detail, we can see that it is positively correlated with the

occurrence of crashes. This finding indicates that the parameter M , which was calibrated in the present study, is a reliable indicator of the overall impact of the work zone, extending beyond mere length. This result is consistent with previous findings in the literature. Indeed, several research contributions have determined that the number of closed lanes and the general complexity of the work zone, two determining factors in the calculation of the value of M , are positively correlated with the number of crashes [34,48,52,53]. This consideration also explains the elevated correlation values observed between O and M (as well as G). Moreover, high values of correlation coefficients of M and L with G can be attributed to the methodology employed in constructing the G variable.

The results from Table 4, in accordance with the findings of previous studies [14,45,47], also establish a slight positive correlation between the number of crashes and both the AADT per lane and the Heavy vehicles AADT per lane. As mentioned in the previous section, the number of crashes does not appear to be correlated with the outcomes related to geometry and the presence of junctions. These results diverge from previous findings that observed dependencies between the number of crashes and geometry [51]. Indeed, a demonstrable correlation has been identified between the incidence of crashes at roadworks and the geometry of the road. This correlation exhibits a negative relationship in the context of curved roads, where drivers are known to exercise greater caution when approaching [49], and a positive dependence for roadworks along straight roads [55].

A review of the results presented in Table 5 reveals an alignment with previously documented findings. Indeed, other studies found similar results, such as a higher probability of crashes at roadworks in summer. This result can be outlined by the greater frequency of roadworks and the higher volume of traffic during the summer season [49]. Additionally, considering the factors related to the external environment, particularly the presence of tunnels, this variable exhibits a significant correlation. Although this variable has not been extensively studied in the literature about the crash risk at roadworks, scientific research suggests that the tunnel environment has a determinant effect on drivers, provoking unpleasant feelings and affecting their psychological condition [56,57].

When considering the values of the coefficients, the correlation appeared slight or modest in most cases, according to the reference range [58]. Given the variability and the randomness of the accident phenomenon [40–42], the correlation coefficient values close to or greater than 0.2 were deemed to be representative of the existence of a relationship between the variables under examination, in light of both the statistical significance of these values and the considerations emerging from the literature.

Finally, according to the values in the ANOVA section, the final model seemed to be significant. Indeed, the results revealed $F = 12.95$ and $\text{sig.} < 0.01$. This finding, with a Durbin Watson value very close to two, confirmed that the set of regressors collectively contributed to the prediction of the number of crashes and that the overall model is significant and effectively fitted to the data. In general, the individual variables considered were significant, apart from variables G and F , which were outside the standard significance ranges. However, their significance was less than or equal to 0.10, which is not excessive. These parameters were retained in the model and their importance decided according to studies in the literature. Furthermore, the multicollinearity of the variables was good, as the VIF (Variance Inflation Factor) was around 1 for all the independent variables. The coefficients of the independent variables are consistent with the above discussion of correlations. In detail, all predictors exhibit a positive relationship with the expected number of crashes, indicating that an increase in the variables generally results in an increased risk. This is immediate for variables with a positive coefficient. As for the flow variable, although the coefficient is negative, its effect aligns with the increase in crashes. Since an inverse transformation has been applied, an increase in actual flow per lane

corresponds to a decrease in the variable in the model and, consequently, an increase in crashes. In terms of influence, the variable with the highest standardised coefficient is $\ln(D)$, indicating that changes in duration significantly impact the expected number of crashes. The variables NL4 and O follow in order of influence, based on the standardised coefficients. Finally, T and SU are also positively associated with the number of crashes, although their impact is weaker than that of the other predictors in the model.

After the construction of the model, its goodness is also assessed by performing a residual analysis, which is generally satisfactory. Finally, the MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) are assessed for the model in Equation (4), which explicitly considers the number of crashes. The values obtained are 1.94 and 3.82, respectively. Both values suggest that the model has a high level of accuracy and is predicted with minimal error. Indeed, a range of 44 is identified for the dependent variable, and the MAE is determined to be less than 5% of this range, precisely 4.4%, indicating that the realised model exhibits minimal absolute errors. In addition, the RMSE was less than 10% of the range, precisely 8.7%, meaning that even the largest errors committed by the model are marginal.

5. Conclusions

Recognising the importance of road infrastructural development and maintenance, this paper has examined the critical issue of roadworks, which, for various reasons, may remain present on our roadways over multiple years. To add to the field of on this topic, the present study concentrated on evaluating the impact of roadworks on road traffic crashes, with a particular focus on highway settings. The objective of this research was to reduce the vulnerability of all road users and the risk to them. Indeed, highways are characterised by particularly high travel speeds and traffic flows, and their infrastructure needs to be maintained to ensure an adequate performance, resulting in work zones, which have a strong impact on various aspects, including crash rates. The objective of the present study was to ascertain the factors that contribute to the occurrence of crashes in the proximity of work zone areas. With this purpose, a detailed analysis was conducted, encompassing the work zone's layout, the characteristics of the infrastructure on which the site is located, and the features of the surrounding environment. In addition, the current study aimed to quantify the impact of highway sites by developing a model for calculating the possible number of crashes related to a work zone with given characteristics.

The methodology was initially based on Pearson's correlation analysis, taking into account several variables related to the work zone, infrastructure, and external environment. Some of these variables had already been explored in the literature, but often separately. A variable of particular importance and calibrated within this study is parameter M , defined as the risk-increasing factor, which represents the risk associated with specific site configurations both in absolute value and related to the characteristics of the infrastructure on which it is applied. The analysis subsequently focused on the development of an MLR based on the variables defined as significant by the previous Pearson correlation study and contributions from the academic research. This model aims to determine the possible number of crashes at a given site. Early findings determine that Carriageways' occupancy, Duration, Global complexity, AADT per lane, Heavy vehicle AADT per lane, Three-lane highway, Four-lane highway, Summer, and Tunnels are correlated in a statistically significant way with the logarithm of the number of crashes. This result appears consistent with many contributions in the literature. In addition, a final MLR model was determined, including the Carriageways' occupancy, Duration, Global complexity, AADT per lane, Four-lane highway, Summer, and Tunnels variables; that model is statistically significant and suitable for estimating the possible number of crashes with a low error rate. The results

obtained should inform private companies and municipalities of the characteristics within worksites that determine a higher risk, allowing them to identify the aspects that need to be carefully managed when planning works, thereby reducing the risk of crashes and fatalities on the roads.

This study could be further developed by applying the MLR model to a new and larger database to validate the results obtained, and by investigating the application of other mathematical models to the same set of variable categories. Future developments may involve further variables to be included in the correlation analysis and implementation of the regression model, as well as taking into account the severity of the crashes that occur. Moreover, in the context of behavioural studies, it is of significant value to analyse the impact that the variables investigated in the present study have on driving behaviour and the perception of drivers, independently of the occurrence of crashes.

Author Contributions: Conceptualisation, A.S. and D.C.; methodology, V.V. and C.L.; validation, A.S. and D.C.; formal analysis, M.P. and S.P.; investigation, C.L. and S.P.; data curation, M.P., C.L., and S.P.; writing—original draft preparation, S.P.; writing—review and editing, M.P. and V.V.; supervision, V.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are partially available on request.

Acknowledgments: This study received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA-PNRR), Missione 4, Componente 1, Investimento 4.1 (D.M. 118/2023, CUP J33C23001380002). In addition, this study was also funded by Spoke 7 of the MOST-Sustainable Mobility National Research Center and also received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA-PNRR), Missione 4, Componente 2, Investimento 1.4 (D.D. 1033 17/06/2022, CN00000023). This manuscript reflects only the authors' views and opinions. Neither the European Union nor the European Commission can be considered responsible for them.

Conflicts of Interest: Author Davide Chiola was employed by R&D and Innovation, Movyon S.p.A. The remaining authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UN	United Nations
SDGs	Sustainable Development Goals
GHG	Greenhouse Gas
ARROWS	Advanced Research on Road Work Zone Safety standards in Europe
STARS	Scoring Traffic at Roadworks
ASAP	Appropriate Speed Saves All People
VMSs	Variable Message Signs
AADT	Annual Average Daily Traffic
MLR	Multiple Linear Regression
WZ	Work Zone
VIF	Variance Inflation Factor
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error

References

1. Martínez-Sánchez, J.; Piñeiro-Montegudo, H.; Balado, J.; Soilán, M.; Arias, P. Improving Safety in the Maintenance of Infrastructures: Design of a UAV-Based System for Work Zone Monitoring. *Transp. Res. Procedia* **2023**, *72*, 2518–2525. [CrossRef]
2. United Nations, Sustainable Development, the 17 Goals. Available online: <https://sdgs.un.org/goals> (accessed on 1 May 2025).
3. ITF. *International Transport Forum ITF Transport Outlook 2021*; OECD: Paris, France, 2021; ISBN 978-92-821-7497-5.
4. Qiao, Y.; Dawson, A.R.; Parry, T.; Flintsch, G.W. Evaluating the Effects of Climate Change on Road Maintenance Intervention Strategies and Life-Cycle Costs. *Transp. Res. Part D Transp. Environ.* **2015**, *41*, 492–503. [CrossRef]
5. Ma, S.; Hu, J.; Wang, R.; Qu, S. Study on the Median Opening Length of a Freeway Work Zone Based on a Naturalistic Driving Experiment. *Appl. Sci.* **2023**, *13*, 851. [CrossRef]
6. Zhang, K.; Batterman, S.; Dion, F. Vehicle Emissions in Congestion: Comparison of Work Zone, Rush Hour and Free-Flow Conditions. *Atmos. Environ.* **2011**, *45*, 1929–1939. [CrossRef]
7. Rahman, M.M.; Strawderman, L.; Garrison, T.; Eakin, D.; Williams, C.C. Work Zone Sign Design for Increased Driver Compliance and Worker Safety. *Accid. Anal. Prev.* **2017**, *106*, 67–75. [CrossRef] [PubMed]
8. Barth, M.; Boriboonsomsin, K. Real-World Carbon Dioxide Impacts of Traffic Congestion. *Transp. Res. Rec. J. Transp. Res. Board* **2008**, *2058*, 163–171. [CrossRef]
9. Demir, E.; Bektaş, T.; Laporte, G. A Comparative Analysis of Several Vehicle Emission Models for Road Freight Transportation. *Transp. Res. Part D Transp. Environ.* **2011**, *16*, 347–357. [CrossRef]
10. Vieira, A.; Santos, B.; Picado-Santos, L. Modelling Road Work Zone Crashes' Nature and Type of Person Involved Using Multinomial Logistic Regression. *Sustainability* **2023**, *15*, 2674. [CrossRef]
11. European Commission. Annual Statistical Report On Road Safety in the EU 2024. Available online: https://road-safety.transport.ec.europa.eu/document/download/b30e9840-4c22-4056-9dab-0231a98e7356_en?filename=ERSOnext_AnnualReport_20240229.pdf (accessed on 1 May 2025).
12. ACI—Automobile Club d'Italia and Istat—Incidenti Stradali Anno. 2023. Available online: <https://www.istat.it/wp-content/uploads/2024/07/REPORT-INCIDENTI-STRADALI-2023.pdf> (accessed on 1 May 2025).
13. ISTAT—Istituto Nazionale di Statistica. Comunicato stampa, Incidenti Stradali in Italia, Anno. 2022. Available online: <https://www.istat.it/comunicato-stampa/incidenti-stradali-in-italia-anno-2022/> (accessed on 1 May 2025).
14. Hou, G.; Chen, S. Study of Work Zone Traffic Safety under Adverse Driving Conditions with a Microscopic Traffic Simulation Approach. *Accid. Anal. Prev.* **2020**, *145*, 105698. [CrossRef]
15. Yang, Y.; Liu, X.; Easa, S.M.; Feng, Y.; Zheng, X. Effect of Distance Information and Number of Warning Signs on Driving Safety of Young Adults near Road Work Zones in China. *Accid. Anal. Prev.* **2023**, *192*, 107230. [CrossRef]
16. Wang, J.; Song, H.; Fu, T.; Behan, M.; Jie, L.; He, Y.; Shangguan, Q. Crash Prediction for Freeway Work Zones in Real Time: A Comparison Between Convolutional Neural Network and Binary Logistic Regression Model. *Int. J. Transp. Sci. Technol.* **2022**, *11*, 484–495. [CrossRef]
17. Zhao, J.; Fu, X.; Zhang, Y. Research on Risk Assessment and Safety Management of Highway Maintenance Project. *Procedia Eng.* **2016**, *137*, 434–441. [CrossRef]
18. Yang, Y.; Ye, Z.; Easa, S.M.; Feng, Y.; Zheng, X. Effect of Driving Distractions on Driver Mental Workload in Work Zone's Warning Area. *Transp. Res. Part F Traffic Psychol. Behav.* **2023**, *95*, 112–128. [CrossRef]
19. Shakouri, M.; Ikuma, L.H.; Aghazadeh, F.; Punniaraj, K.; Ishak, S. Effects of Work Zone Configurations and Traffic Density on Performance Variables and Subjective Workload. *Accid. Anal. Prev.* **2014**, *71*, 166–176. [CrossRef]
20. European Commission. Advanced Research on Road Work Zone Safety Standards in Europe. Available online: <https://trimis.ec.europa.eu/project/advanced-research-road-work-zone-safety-standards-europe> (accessed on 1 May 2025).
21. Nuallain, N.N.; Sarrazin, R.; Wennström, J.; Weekley, J. The STARs Evaluation Tool: Optimising Network Performance, Road Worker Safety and Road User Safety during Roadworks and Maintenance. In Proceedings of the Transport Research Arena, Paris, France, 14–17 April 2014.
22. Vadeby, A.; Sörensen, G.; Bolling, A.; Cocu, X.; Saleh, P.; Aleksa, M.; La Torre, F.; Nocentini, A.; Tucka, P. Towards a European Guideline for Speed Management Measures in Work Zones. *Transp. Res. Procedia* **2016**, *14*, 3426–3435. [CrossRef]
23. Khattak, A.J.; Khattak, A.J.; Council, F.M. Effects of Work Zone Presence on Injury and Non-Injury Crashes. *Accid. Anal. Prev.* **2002**, *34*, 19–29. [CrossRef] [PubMed]
24. La Torre, F.; Domenichini, L.; Nocentini, A. Effects of Stationary Work Zones on Motorway Crashes. *Saf. Sci.* **2017**, *92*, 148–159. [CrossRef]
25. Gan, X.; Weng, J.; Zhang, J. Evaluation of Travel Delay and Accident Risk at Moving Work Zones. *J. Transp. Saf. Secur.* **2021**, *13*, 622–641. [CrossRef]
26. Jin, T.G.; Saito, M.; Eggett, D.L. Statistical Comparisons of the Crash Characteristics on Highways Between Construction Time and Non-Construction Time. *Accid. Anal. Prev.* **2008**, *40*, 2015–2023. [CrossRef]

27. McClure, D.; Siriwardene, S.; Truong, L.; Debnath, A.K. Examination of Crash Rates and Injury Severity Before, During, and After Roadworks at High-Speed Regional Roads. *Transp. Res. Rec. J. Transp. Res. Board* **2023**, 2677, 351–359. [[CrossRef](#)]
28. Bidkar, O.; Arkatkar, S.; Joshi, G.; Easa, S.M. Effect of Construction Work Zone on Rear-End Conflicts by Vehicle Type under Heterogeneous Traffic Conditions. *J. Transp. Eng. Part A Syst.* **2023**, 149, 05023001. [[CrossRef](#)]
29. Nahed, R.; Nassar, E.; Khoury, J.; Arnaout, J.-P. Assessing the Effects of Geometric Layout and Signing on Drivers' Behavior through Work Zones. *Transp. Res. Interdiscip. Perspect.* **2023**, 21, 100901. [[CrossRef](#)]
30. Siriwardene, S.; Ashraf, M.; Debnath, A.K. Driver Preference Regarding Merging Strategies at Work Zones. *Transp. Res. Part F Traffic Psychol. Behav.* **2024**, 104, 217–233. [[CrossRef](#)]
31. Jørgensen, F.; Wentzel-Larsen, T. Optimal Use of Warning Signs in Traffic. *Accid. Anal. Prev.* **1999**, 31, 729–738. [[CrossRef](#)] [[PubMed](#)]
32. Almallah, M.; Hussain, Q.; Alhajyaseen, W.K.M.; Pirdavani, A.; Brijs, K.; Dias, C.; Brijs, T. Improved Traffic Safety at Work Zones through Animation-Based Variable Message Signs. *Accid. Anal. Prev.* **2021**, 159, 106284. [[CrossRef](#)] [[PubMed](#)]
33. Xu, W.; Zhao, X.; Chen, Y.; Bian, Y.; Li, H. Research on the Relationship Between Dynamic Message Sign Control Strategies and Driving Safety in Freeway Work Zones. *J. Adv. Transp.* **2018**, 2018, 9593084. [[CrossRef](#)]
34. Rathnasiri, N.; De Silva, N.; Wijesundara, J. State of the Art in Work Zone Safety: A Systematic Review. *Int. J. Transp. Sci. Technol.* **2024**, 13, 14–28. [[CrossRef](#)]
35. Vignali, V.; Bichicchi, A.; Simone, A.; Lantieri, C.; Dondi, G.; Costa, M. Road Sign Vision and Driver Behaviour in Work Zones. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, 60, 474–484. [[CrossRef](#)]
36. Kummetha, V.C.; Kondyli, A.; Chryssikou, E.G.; Schrock, S.D. Safety Analysis of Work Zone Complexity with Respect to Driver Characteristics—A Simulator Study Employing Performance and Gaze Measures. *Accid. Anal. Prev.* **2020**, 142, 105566. [[CrossRef](#)]
37. Theofilatos, A.; Yannis, G.; Kopelias, P.; Papadimitriou, F. Predicting Road Accidents: A Rare-Events Modeling Approach. *Transp. Res. Procedia* **2016**, 14, 3399–3405. [[CrossRef](#)]
38. Theofilatos, A.; Yannis, G.; Kopelias, P.; Papadimitriou, F. Impact of Real-Time Traffic Characteristics on Crash Occurrence: Preliminary Results of the Case of Rare Events. *Accid. Anal. Prev.* **2019**, 130, 151–159. [[CrossRef](#)] [[PubMed](#)]
39. Prieto Curiel, R.; González Ramírez, H.; Bishop, S.R. A Novel Rare Event Approach to Measure the Randomness and Concentration of Road Accidents. *PLoS ONE* **2018**, 13, e0201890. [[CrossRef](#)] [[PubMed](#)]
40. Fridstrøm, L.; Ifver, J.; Ingebrigtsen, S.; Kulmala, R.; Thomsen, L.K. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accid. Anal. Prev.* **1995**, 27, 1–20. [[CrossRef](#)]
41. Okamoto, H.; Koshi, M. A Method to Cope with the Random Errors of Observed Accident Rates in Regression Analysis. *Accid. Anal. Prev.* **1989**, 21, 317–332. [[CrossRef](#)] [[PubMed](#)]
42. Mičić, S.; Vujadinović, R.; Amidžić, G.; Damjanović, M.; Matović, B. Accident Frequency Prediction Model for Flat Rural Roads in Serbia. *Sustainability* **2022**, 14, 7704. [[CrossRef](#)]
43. Ouyang, N. Comprehensive Operation Risk Assessment of a Highway Maintenance Area Based on Reliability. *Sustainability* **2021**, 13, 8744. [[CrossRef](#)]
44. Ge, H.; Yang, Y. Research on Calculation of Warning Zone Length of Freeway Based on Micro-Simulation Model. *IEEE Access* **2020**, 8, 76532–76540. [[CrossRef](#)]
45. Zhang, Z.; Akinci, B.; Qian, S. Inferring the Causal Effect of Work Zones on Crashes: Methodology and a Case Study. *Anal. Methods Accid. Res.* **2022**, 33, 100203. [[CrossRef](#)]
46. Rangaswamy, R.; Alnawmasi, N.; Wang, Z.; Lin, P.-S. Exploring Contributing Factors to Improper Driving Actions in Single-Vehicle Work Zone Crashes: A Mixed Logit Analysis Considering Heterogeneity in Means and Variances, and Temporal Instability. *J. Transp. Saf. Secur.* **2024**, 16, 768–797. [[CrossRef](#)]
47. Zhang, C.; Wang, B.; Yang, S.; Zhang, M.; Gong, Q.; Zhang, H. The Driving Risk Analysis and Evaluation in Rightward Zone of Expressway Reconstruction and Extension Engineering. *J. Adv. Transp.* **2020**, 2020, 8943463. [[CrossRef](#)]
48. Nasrollahzadeh, A.A.; Sofi, A.R.; Ravani, B. Identifying Factors Associated with Roadside Work Zone Collisions Using Machine Learning Techniques. *Accid. Anal. Prev.* **2021**, 158, 106203. [[CrossRef](#)]
49. Mohammed, H.J.; Chang, Y.I.; Schrock, S.D. Factors Associated with Work Zone Crashes. *Transp. Res. Rec. J. Transp. Res. Board* **2023**, 2677, 224–235. [[CrossRef](#)]
50. Meng, Q.; Weng, J.; Qu, X. A Probabilistic Quantitative Risk Assessment Model for the Long-Term Work Zone Crashes. *Accid. Anal. Prev.* **2010**, 42, 1866–1877. [[CrossRef](#)] [[PubMed](#)]
51. Harb, R.; Radwan, E.; Yan, X.; Pande, A.; Abdel-Aty, M. Freeway Work-Zone Crash Analysis and Risk Identification Using Multiple and Conditional Logistic Regression. *J. Transp. Eng.* **2008**, 134, 203–214. [[CrossRef](#)]
52. Yang, H.; Ozbay, K.; Ozturk, O.; Yildirimoglu, M. Modeling Work Zone Crash Frequency by Quantifying Measurement Errors in Work Zone Length. *Accid. Anal. Prev.* **2013**, 55, 192–201. [[CrossRef](#)] [[PubMed](#)]
53. Yang, H.; Ozbay, K.; Xie, K.; Bartin, B. Transportation Research Record. In Proceedings of the Transportation Research Board's 94th Annual Meeting, Washington, DC, USA, 11–15 January 2015.

54. Theofilatos, A.; Ziakopoulos, A.; Papadimitriou, E.; Yannis, G.; Diamandouros, K. Meta-Analysis of the Effect of Road Work Zones on Crash Occurrence. *Accid. Anal. Prev.* **2017**, *108*, 1–8. [[CrossRef](#)] [[PubMed](#)]
55. Silverstein, C.; Schorr, J.; Hamdar, S.H. Work Zones versus Nonwork Zones: Risk Factors Leading to Rear-End and Sideswipe Collisions. *J. Transp. Saf. Secur.* **2016**, *8*, 310–326. [[CrossRef](#)]
56. Lee, J.; Kirytopoulos, K.; Pervez, A.; Huang, H. Understanding Drivers' Awareness, Habits and Intentions inside Road Tunnels for Effective Safety Policies. *Accid. Anal. Prev.* **2022**, *172*, 106690. [[CrossRef](#)]
57. Arias, A.V.; Lopez, S.M.; Fernandez, I.; Rubio, J.L.M.; Magallares, A. Psychosocial Factors, Perceived Risk and Driving in a Hostile Environment: Driving through Tunnels. *Int. J. Glob. Environ. Issues* **2008**, *8*, 165. [[CrossRef](#)]
58. Schober, P.; Boer, C.; Schwarte, L.A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.