Primljen / Received: 11.2.2026.
Ispravljen / Corrected: 15.4.2026.
Prihvaćen / Accepted: 29.4.2026.
Dostupno online / Available online: 10.5.2026.

Predicting the probability of occupational accident outcomes in the construction industry

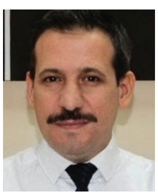
Authors:

[1]Şükrü Bulut, M.Sc.
sukrubulut@kku.edu.tr
Corresponding author

[2]Muhammed F. Kahraman, M.Sc.
kahramanfurkan@gmail.com

[3]IAssoc.Prof. Serap Yörübulut
serapyorubulut@kku.edu.tr

[4]Prof. Ahmet Kürşad Türker
kturker@kku.edu.tr



[4]Prof. Adnan Aktepe
aaktepe@gmail.com

[1]Kırıkkale University, Turkey
Department of Occupational Health and Safety Programme
[2]Ministry of Labor and Social Security, Turkey
[3]Kırıkkale University, Turkey
Faculty of Engineering and Natural Science
Department of Statistics
[4]Kırıkkale University, Turkey
Faculty of Engineering and Natural Science
Department of Industrial Engineering

Original research paper

Şükrü Bulut, Muhammed Furkan Kahraman, Serap Yörübulut, Ahmet Kürşad Türker, Adnan Aktepe

Predicting the probability of occupational accident outcomes in the construction industry

The construction industry experiences numerous accidents of varying severity. This study analyses 434,134 occupational accident records from the construction sector in Turkey (2012–2021) using ordinal logistic regression for four-category severity modelling. Unlike previous studies, the dataset was split into 70 % training and 30 % independent test sets. The model achieved a classification accuracy of 56.0 %, demonstrating strong sensitivity in identifying rare, high-severity outcomes, such as death and permanent disability. These results provide a validated predictive equation and identify critical risk factors, thereby offering a data-driven framework for prevention strategies. This comprehensive modelling approach bridges the gap between theoretical regression and practical safety decision-making.

Key words:

construction, occupational health and safety, occupational accident, ordinal logistic regression, risk factors, accident severity, predictive modelling

Izvorni znanstveni rad

Şükrü Bulut, Muhammed Furkan Kahraman, Serap Yörübulut, Ahmet Kürşad Türker, Adnan Aktepe

Predviđanje vjerojatnosti ishoda ozljeda na radu u građevinskoj industriji

Građevinska industrija bilježi brojne nesreće različitih težina. Ovo istraživanje analizira 434.134 zapisa o ozljedama na radu iz građevinskog sektora u Turskoj (2012. – 2021.) primjenom ordinalne logističke regresije za modeliranje težine u četiri kategorije. Za razliku od prethodnih radova, skup podataka podijeljen je na 70 % onih za treniranje i 30 % onih za neovisni skup za testiranje. Model je postigao točnost klasifikacije od 56,0 %, pokazujući visoku osjetljivost u prepoznavanju rijetkih, jako teških ishoda poput smrti i trajnog invaliditeta. Ti rezultati pružaju validiranu prediktivnu jednadžbu i identificiraju ključne čimbenike rizika, čime se nudi okvir za strategije prevencije temeljen na podacima. Takav sveobuhvatan pristup modeliranju premošćuje jaz između teorijske regresije i praktičnog donošenja odluka o sigurnosti.

Ključne riječi:

građevinska industrija, zaštita na radu, ozljeda na radu, ordinalna logistička regresija, čimbenici rizika, težina ozljede, prediktivno modeliranje

1. Introduction

Occupational accidents are defined as incidents occurring during or resulting from work that cause injury or death [1]. They incur costs that negatively affect employees, businesses, and national economies. These costs include loss of life, physical and mental health impairment, disability, decline in skilled labour and workforce, increased unemployment, temporary or permanent disability payments, disability or incapacity payments, increased healthcare costs, compensatory payments, monthly benefits paid to the employee's family, and time consumed in legal proceedings. In addition to direct costs, some costs cannot be determined immediately in the workplace and may arise after the accident [2, 3].

Every year, various industries worldwide suffer from numerous occupational accidents, resulting in millions of employees being injured or losing their lives. The construction industry accounts for 35 % of workplace fatalities [4]. Compared with other industries, the likelihood of fatal accidents in the construction industry may be four times higher than the average in developed countries and up to six times higher than the average in less-developed countries [5]. In 2022, 20 % of all work-related deaths occurred in the construction industry in the United States despite employing only 6 % of the total workforce [6]. Furthermore, it is among the most dangerous industries worldwide in terms of non-fatal occupational accidents [7]. In Türkiye, 12.1 % of the 681,401 occupational accidents and 28.1 % of the 1,966 fatal occupational accidents reported in 2023 occurred in this industry. According to public records, 225 occupational accidents occur daily at construction sites in Türkiye. These statistics highlight occupational health and safety (OHS) as a major challenge for the construction industry [8, 9].

Dynamic and complex interactions between people, machinery, and the environment during construction activities often result in workplace incidents such as falls, collapses, collisions, electric shocks, and mechanical injuries [10, 11].

Each related activity carries its own specific risks. For example, in high-rise building construction activities where various new digital techniques are used to significantly improve safety, there are eight main risk factors consisting of approximately 60 sub-factors that can cause occupational accidents, such as temporary electrical installations on-site, safety inspection and safety warning signs, personal protective equipment, work areas, personal factors, knowledge and skill levels of employees, formwork or crane dismantling, scaffolding, overload, safety management, and working at a height [12]. Each risk factor must be meticulously examined based on its merits. For example, the use of personal protective equipment, especially safety harnesses and independent lifelines, before starting work plays an important role in reducing worker-related risks and preventing falls associated with working at height; however, accident analyses show that a large proportion of accidents are related to the lack of personal protective equipment or improper use of this equipment [13, 14].

In order to prevent occupational accidents in construction sites, which are largely caused by the lack of safety awareness among workers, inadequate safety management in construction sites, insufficient or defective personal protective equipment, and the failure of operators to follow work procedures, it is necessary to plan activities based on lessons learned from past occupational accidents and to create safety policies to improve occupational health and safety conditions and prevent possible deaths and injuries [15].

Past accidents and OHS risks are strongly correlated, and reliable risk predictions can be made using data on past accidents. Consequently, authorities and OHS stakeholders encourage the reporting of occupational accidents and near-misses [16, 17].

Risk calculations are traditionally defined as a product of the probability of a defined hazard and its resulting severity. The accurate measurement and analysis of severity are more complex than probability assessment, and severity is a significant concern across various industries. Logistic regression, which is common in the medical and social sciences but rare in OHS studies, has been increasingly used to accurately predict the severity of occupational accidents. As an inferential statistical technique, logistic regression helps reveal the relationships between dependent and independent variables. These relationships may also support the prediction of accident severity [18, 19].

Previous studies have conducted standardised logistic regression analyses to investigate occupational accidents. Identifying the factors and relationships that cause occupational accidents can reduce the number and severity of accidents.

Standard regression analysis has been used to determine the factors affecting occupational accidents in the construction industry in Hong Kong [20], investigate the relationship between occupational accidents and health conditions among employees in Lebanon [21], examine the relationship between employee characteristics and the likelihood of injury in occupational accidents in the construction industry in the United States [22], and identify the factors affecting fatal occupational accidents in the mining industry [23]. Standard binary logistic regression was employed to predict the injury severity scores for occupational accidents in two categories (more than nine and nine or fewer) using variables from occupational accident data in the construction industry [18]. In the mining industry, work days lost due to nonfatal accidents in an open-pit mine were predicted using variables such as occupation, work area, employee age, accident cause, and injury location through standard binary logistic regression in two categories (longer than three days and three days or shorter) [24]. It has also been used in the construction industry to predict the outcomes of occupational accidents due to falls from a height in two categories (fatal and non-fatal) based on variables related to employees working on roofs and their working conditions [25]. It uses decision tree, random forest, and AdaBoost methods to determine the degree of impact of variables including age, gender, length of service of employees working on construction sites, type of construction, size of the employer, and date of the accident on the outcomes of accidents in two categories (fatal and nonfatal injury accidents) [26]. The outcomes of accidents were predicted in two categories (resulting or not resulting in lost workdays) by standard binary logistic regression analysis using variables related to the individual and occupational characteristics of road construction employees, such as age, gender, marital status, OHS training, experience, education, occupation, season, and location of the accident, and the material causing the accident [19]. It has also been employed to predict the outcomes of occupational accidents in two categories (fatal and nonfatal) using variables from a construction industry occupational accident dataset [27]. One study combined multiple logistic regression with support vector machines, the C5.0 decision tree, stochastic gradient boosting, and neural networks to identify how individual and occupational variables affected accident outcomes in two categories (fatal and nonfatal injuries) [28].

To the best of our knowledge, the literature contains few studies that employ regularised regression models designed to mitigate instability and multicollinearity in standard binary logistic regression, thereby improving model performance by yielding more robust regression coefficients. These studies used the Firth logistic regression model [29] and the Firth, Lasso, and Elastic Net logistic regression models [30] to predict accident outcomes in two categories (fatal/nonfatal) using data on all occupational accidents, regardless of the industry. Another study attempted to predict the accident frequency, lost workdays, and severity across two categories (fatal/nonfatal) using Lasso, Elastic Net, and Adaptive Lasso linear regression models with data on accidents, labour markets, economic factors, and production structures [31]. In the construction industry, we identified only one study that used regularised regression models. In this study, Firth, Ridge, Lasso, and Elastic Net regularised logistic regression models were applied alongside standard binary logistic regression to predict workday loss resulting from occupational accidents in two categories (more than three days, three days, or less) [32].

These studies show that logistic regression has been used effectively to identify the factors influencing occupational accidents and quantify their impact. Despite the measures taken and the studies conducted, the high incidence and severity of occupational accidents in the construction industry necessitate further research. In contrast to previous studies that used limited data from a single region or a few areas of the construction industry, this study examined data related to 434,314 officially reported occupational accidents in the construction industry between 2012 and 2021. This is the first study to employ ordinal logistic regression to predict the likelihood of occupational accidents based on individual and professional characteristics of employees, yielding four categories. These data encompassed the ten years following the enactment of Law No. 6331 on Occupational Health and Safety, a separate law on health and safety in Türkiye. Therefore, this study contributes to the literature by revealing the effectiveness of practices implemented in the industry and in accordance with legislation during the aforementioned period.

To the best of our knowledge, this study is the first to analyse occupational accidents in the construction industry using ordinal logistic regression with four categories of accident outcomes, thereby departing from the previously employed standard and regulatory binary logistic regression models.

2. Data and analysis method

This study examined officially reported occupational accidents in the construction industry in Türkiye between 2012 and 2021. The "Occupational Accident and Occupational Disease Report Form", which contains personal information about the victim as well as various details about the workplace and the accident, and must be submitted to the Social Security Institution (SSI) by employers following an occupational accident in Türkiye, serves as the only official source of data and therefore constitutes a valuable database for statistical analysis.

In this study, forms submitted to the Social Security Institution regarding accidents in the construction industry under permit number 112604753, dated 24 February 2025, were analysed. Each form was reviewed and encoded for statistical analysis, and the variables were categorised into subcategories to generate a dataset for the study.

The dataset comprised victim variables, such as educational background, marital status, age, overall experience, occupation, OHS training, and vocational training, as well as accident-related variables, such as the outcome of the accident, environment, location, province, time, year, and month of occurrence.

Accident severity was classified into the following four ordinal categories: (1) continued work, (2) injury, (3) loss of limb, and (4) death. The classification required ordinal logistic regression, because the dependent variables were both categorical and ordinal [33]. This comprehensive set of variables captures multidimensional effects on accident severity.

Following examination of the frequency tables for the variables, an ordinal logistic regression model was developed to predict the probability of the severity of occupational accidents in the construction industry.

The frequency analysis of the distribution of subcategories for the accident victim variables in the categorically organised dataset is presented in Table 1, and the frequency analysis results for the accident variables are presented in Table 2. This analysis served as the basis for subsequent multivariate analyses [34, 35].

The discrepancy between the number of accidents in the "Location of the Accident" subcategory ("at the workplace") and the "Accident Environment" subcategory ("while working") is related to SSI regulations. According to Social Insurance Law No. 5510, the SSI only recognises accidents reported to it as work-related accidents if they occur within workplace boundaries,

Table 1. Frequency analysis results of variables related to employees involved in accidents

Variable	Potkategorije varijabli	Frequency	Frequency ratio [%]
Marital Status (MS)	Single, Divorced, Widowed (MS1)	168.919	38.9
	Married (MS2)	265.395	61.1
Educational background (EB)	Illiterate (EB1)	2.998	0.7
	Literate (EB2)	73.505	16.9
	Elementary School (EB3)	132.061	30.4
	Middle-Primary School (EB4)	124.889	28.8
	Vocational High School (EB5)	8.833	2.0
	High school (EB6)	75.971	17.5
	Vocational School (EB7)	9.234	2.1
	University (EB8)	6.669	1.5
	Postgraduate (EB9)	154	0.0
OHS training (OT)	No (OT1)	31.128	7.2
	Yes (OT2)	403.186	92.8
Vocational training (VT)	No (VT1)	90.953	20.9
	Yes (VT2)	343.361	79.1
Occupation (O)	Managers (O1)	36.459	8.4
	Professionals (O2)	9.937	2.3
	Technicians, technical staff, and support professionals (O3)	41.852	9.6
	Office staff (O4)	1.985	0.5
	Service and sales staff (O5)	5.102	1.2
	Qualified agricultural, forestry, and fisheries staff (O6)	1.021	0.2
	Artisans and employees in related fields (O7)	183.817	42.3
	Plant and machine operators and assemblers (O8)	37.281	8.6
	Occupations not requiring qualifications (O9)	116.860	26.9
Age (A)	Under 25 (A1)	101.937	23.5
	25-34 (A2)	135.491	31.2
	35-44 (A3)	103.852	23.9
	45-54 (A4)	71.414	16.4
	55-64 (A5)	20.285	4.7
	65 and older (A6)	1.335	0.3
Overall experience (OE)	Less than 6 years (OE1)	104.190	24.0
	6-12 years (OE2)	122.930	28.3
	13-18 years (OE3)	68.103	15.7
	19-24 years (OE4)	54.736	12.6
	More than 24 years (OE5)	84.355	19.4

while the employee is being sent outside the workplace by the employer, during breastfeeding leave, inside or outside the workplace, or during the transportation provided by the employer.

Of the 358,577 accidents reported to have occurred in the workplace, 357,131 occurred during work hours (336,567) or breaks (20,564). Except for the "outside the workplace" subcategory, none of the other subcategories specify in detail whether the accidents occurred inside or outside the workplace boundaries. Therefore, the accident numbers in the "Accident Environment" and "Location of the Accident" subcategories do not fully match.

Table 2. Frequency analysis results of variables related to the accident

Variable	Subcategories of variables	Frequency	Frequency ratio [%]	Variable	Subcategories of variables	Frequency	Frequency ratio [%]
Time of the Accident (TA)	00.00 – 2.00 (TA1)	5419	1.2	Month of the Accident (MA)	Siječanj (MA1)	38.895	6.7
	2.00 – 4.00 (TA2)	5635	1.3		Veljača (MA2)	37.938	6.5
	4.00 – 6.00 (TA3)	4277	1.0		Ožujak (MA3)	45.747	7.9
	6.00 – 8.00 (TA4)	7670	1.8		Travanj (MA4)	44.475	7.7
	8.00 – 10.00 (TA5)	70.672	16.3		Svibanj (MA5)	49.236	8.5
	10.00 – 12.00 (TA6)	107.891	24.8		Lipanj (MA6)	48.628	8.4
	12.00 – 14.00 (TA7)	49.655	11.4		Srpanj (MA7)	54.497	9.4
	14.00 – 16.00 (TA8)	86.714	20.0		Kolovoz (MA8)	53.713	9.2
	16.00 – 18.00 (TA9)	62.010	14.3		Rujan (MA9)	54.633	9.4
	18.00 – 20.00 (TA10)	16.698	3.8		Listopad (MA10)	53.573	9.2
	20.00 – 22.00 (TA11)	10.243	2.4		Studeni (MA11)	51.797	8.9
	22.00 – 24.00 (TA12)	7430	1.7		Prosinac (MA12)	47.793	8.2
Accident Outcome (AO)	No incapacity (AO1)	197.003	45.4	Province of the Accident (PA)	Ankara (PA1)	37.526	8.6
	Injury (AO2)	232.027	53.4		Istanbul (PA2)	138.011	31.8
	Loss of limb (AO3)	1039	0.2		Izmir (PA3)	33.994	7.8
	Death (A4)	4245	1.0		Ostalo (PA4)	224.783	51.8
Accident Environment (AE)	During break (AE1)	20.564	4.7	Year of the Accident (YA)	2012 (YA1)	18.111	3.1
	While working (AE2)	336.567	77.5		2013 (YA2)	26.922	4.6
	On maternity leave (AE3)	214	0.0		2014 (YA3)	29.919	5.2
	Working outside the workplace (AE4)	20.890	4.8		2015 (YA4)	33.894	5.8
	Driving home from work (AE5)	3430	0.8		2016 (YA5)	44.362	7.6
	Driving to work from home (AE6)	49.687	11.4		2017 (YA6)	61.985	10.7
	On the way home from work by shuttle (AE7)	1261	0.3		2018 (YA7)	75.880	13.1
	On the way to work from home by shuttle (AE8)	1701	0.4		2019 (YA8)	46.038	7.9
Location of the Accident (LA)	At the workplace (LA1)	358.777	82.6		2020 (YA9)	42.004	7.2
	Outside the workplace (LA2)	75.537	17.4		2021 (YA10)	55.350	9.5

2.1. Ordinal logistic regression analysis

Ordinal logistic regression analysis is a common statistical method used to examine the relationship between a dependent variable measured on an ordinal scale with at least three categories and independent variables measured on a categorical or continuous scale [33]. The proportional odds assumption on which this model is based is a key assumption in the analysis [36]. According to this assumption, the relationship between the dependent and independent variables does not vary across the categories of the dependent variable [37, 38].

Ordinal logistic regression analysis assumes that an observed continuous random latent variable Y* underlies a categorically measured dependent Y variable. Consecutive intervals on a continuous plane, referred to as cutoff points, correspond to variable categories [33]. The unobserved variable Y* is expressed as shown in Equation (1), where
$\theta_{s-1} < Y^* < \theta_s$, $s = 1,\ldots,j$ i $\theta_0 = -\infty$ te $\theta_j = +\infty$ [39].

$$Y^* = \sum_{k=1}^{K} \beta_k \chi_k + \varepsilon \qquad (1)$$

In Equation (1), θ denotes the threshold value, χ_k the vector of independent variables, β_k the vector of parameters, and ε the error term. The relationship between the observed variable Y and unobserved variable Y* is given by Equation (2) [40].

$$y = \begin{cases} 1 & ako\, y_i^* \le \theta_1 \\ 2 & ako\, \theta_1 \le y_i^* \le \theta_2 \\ 3 & ako\, \theta_2 \le y_i^* \le \theta_3 \\ \vdots & \vdots \\ j & ako\, \theta_{j-1} \le y_i^* \end{cases} \tag{2}$$

The θ values in Equation (2) represent the threshold values distinguishing the categories of the dependent variable. The overall probability of the dependent variable falling into category k for the independent variables is given by Equation (3):

$$Prob(y = j \setminus x) = F\left[\theta_j - \sum_{k=1}^{K} \beta_k \chi_k\right] - F\left[\theta_{j-1} - \sum_{k=1}^{K} \beta_k \chi_k\right] \tag{3}$$

In Equation (3), F denotes the distribution function of the error term and is assumed to follow a logistic distribution. The link functions in ordinal logistic regression are transformations of the cumulative probabilities, as listed in Table 3 [41].

Equation (4) demonstrates the probabilities of the independent variables falling into dependent variable categories in ordinal regression analysis, where L is the logit distribution function [42].

$$\begin{aligned} Prob(y = 1) &= L\left(-\sum_{k=1}^{K} x_k \beta_k\right) \\ Prob(y = 2) &= L\left(\theta_2 - \sum_{k=1}^{K} x_k \beta_k\right) - L\left(-\sum_{k=1}^{K} x_k \beta_k\right) \\ Prob(y = 3) &= L\left(\theta_3 - \sum_{k=1}^{K} x_k \beta_k\right) - L\left(\theta_2 - \sum_{k=1}^{K} x_k \beta_k\right) \\ Prob(y = j) &= 1 - L\left(\theta_{j-1} - \sum_{k=1}^{K} x_k \beta_k\right) \end{aligned} \tag{4}$$

Table 3. The link functions of ordinal logistic regression

Function	Form	Details
Logit	$\log\left(\frac{x}{1-x}\right)$	The categories are distributed evenly
Complementary log-log	log (-log (1-x))	The probability of higher categories is higher
Negative log-log	-log (-log (x))	The probability of lower categories is higher
Probit	$F^{-1}(x)$	The variable is normally distributed
Couchit	tan (π(x – 0.5))	The variable has extreme values

The key point in interpreting parameter predictions is that positive β coefficients indicate that an increase in the independent variable increases the probability of observing the dependent variable in higher categories [43]. According to the literature, exp(β) values are used to calculate odds ratios, and other variables should also be considered when interpreting the results [44].

It is recommended that the model assumptions be tested using methods such as the Brant test [45] or score tests. When the proportional odds assumption is violated, alternative approaches such as generalised ordinal logit or partial proportional odds models should be considered.

Methodological considerations include ensuring an adequate sample size, assessing multicollinearity issues, and evaluating the model fit. Information criteria, such as the Hosmer–Lemeshow test, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), are also critical in selecting a model.

In conclusion, adhering to a rigorous methodology and conducting appropriate statistical checks in ordinal logistic regression analyses are essential to obtain valid results.

3. Analysis results

3.1. Trends in accident frequency and incidence rates

A descriptive analysis of occupational accidents in the Turkish construction industry between 2012 and 2021 revealed a significant upward trend in both the absolute number of accidents and incidence rates. As shown in Table 4, the number of accidents will increase from 18,110 in 2012 to 55,207 in 2021. More importantly, the incidence rate, which represents the number of accidents per 1,000 employees, tripled during this period, rising from 10.12 in 2012 to 33.85 in 2021. The peak incidence rate was recorded in 2018 (IR = 47.39), indicating that the risk of accidents increased at a much faster rate than the industry workforce expansion.

To provide a comprehensive overview of the regional risk distribution, Table 5 presents the annual incidence rates for the three major provinces (Istanbul, Ankara, and Izmir), along with the national average for other provinces. These longitudinal data demonstrate that while the absolute number of accidents is highest in Istanbul, the relative risk per 1,000 employees has followed a volatile but generally increasing trend across all regions.

As shown in Table 5, the incidence rates in Ankara and Izmir frequently surpassed those in Istanbul, particularly in 2018 and

Table 4. Annual accident counts and incidence rates in the construction industry (2012–2021)

Year	Accidents (n)	Total employees (N)	Incidence rate (IR)
2012.	18.110	1.789.487	10.12
2013.	26.921	1.851.942	14.54
2014.	29.918	1.876.129	15.95
2015.	33.893	1.980.130	17.12
2016.	44.362	1.887.099	23.51
2017.	61.982	2.083.431	29.75
2018.	75.880	1.601.184	47.39
2019.	46.038	1.294.688	35.56
2020.	42.003	1.587.666	26.46
2021.	55.207	1.630.740	33.85

Table 5. Regional accident incidence rates per 1,000 employees (2012–2021)

Year	Istanbul (IR)	Ankara (IR)	Izmir (IR)	Other provinces (IR)
2012.	7.42	10.47	9.90	11.46
2013.	10.45	17.30	14.97	16.45
2014.	12.76	18.31	15.69	17.56
2015.	17.25	19.94	16.02	16.74
2016	25.16	25.92	20.01	22.56
2017	35.22	27.03	34.08	26.20
2018	46.03	45.78	56.52	46.91
2019	29.99	37.09	26.24	41.79
2020	24.54	22.02	16.40	30.10
2021	26.05	32.17	23.19	42.01

Note: Incidence rates were calculated as (n/N) × 1.000. The fluctuations between 2018 and 2021 were correlated with shifting economic activities and pandemic-related workplace restrictions.

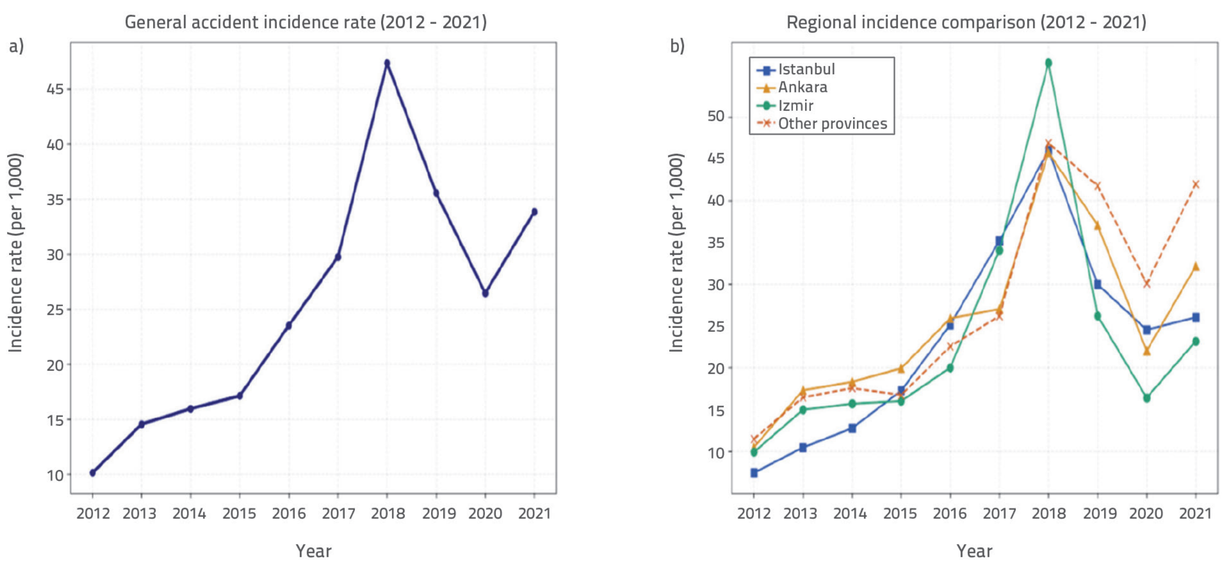


Figure 1. Temporal a) and regional b) distribution of accident incidence rates in the Turkish construction industry (2012–2021)

2021. This indicates that the high volume of accidents in Istanbul is proportional to its large workforce, whereas other regions face higher relative risks regardless of the total workforce size. This escalating risk landscape, which transcends simple demographic growth, effectively mitigates "size bias" concerns and justifies the transition from descriptive statistics to advanced predictive modelling through ordinal logistic regression to explore the complex determinants of accident severity.

As illustrated in Figure 1(a), the national incidence rate exhibited a sharp upward trajectory over the decade, peaking in 2018 before reaching a state of relative stabilisation. The sharp fluctuations observed between 2018 and 2021 were associated with shifting economic activities and the transformative impacts of the COVID-19 pandemic on construction operations. However, the consistent overall rise from 2012 levels confirms that occupational health and safety (OHS) challenges in the sector have intensified rather than diminished.

The regional comparison presented in the right panel of Figure 1(b) further dismantles simplistic "size bias" assumptions; although Istanbul records the highest absolute number of accidents, provinces such as Izmir and Ankara experienced significantly higher relative risks (incidence rates) in specific years. This volatility suggests that regional industrial densities and localised safety cultures are more potent drivers of accident frequency than total employee count alone.

3.2. Ordinal logistic regression analysis

This study employs ordinal logistic regression to examine the factors influencing the severity of occupational accidents in the Turkish construction industry between 2012 and 2021. In accordance with best methodological practices and to ensure the predictive validity of the findings, the dataset (N = 434,134) was partitioned into a 70 % training set (N = 304,019) for model estimation and a 30 % independent test set (N = 130,295) for validation. To capture the nuanced effects of temporal, demographic, and environmental factors, independent variables were included in the model as dummy-coded categorical predictors.

3.2.1. Model assumptions and goodness-of-fit

The fundamental "proportional odds assumption" (also known as the parallel lines assumption) was re-evaluated on the 70 % training set using the Brant test. As presented in Table 6, the results (χ^2 = 89.52, df = 73, p = 0.091) indicate that the parallel lines assumption was not violated. Although the test statistics changed owing to the revised sample size and the inclusion of detailed dummy variables, the p-value remained above the threshold of 0.05, justifying the use of the ordinal logistic regression framework.

Table 6. Proportional odds assumption test

Variable	Chi-square	P > χ^2	Stupnjevi slobode
All	89.52	0.091	73

The overall significance and suitability of the established model were evaluated using a Likelihood Ratio (LR) chi-square test on a training dataset. Table 7 summarises the goodness-of-fit statistics.

Table 7. Model goodness-of-fit test

Variable test	Value
Log-Likelihood	-225.331,70
LR Chi-Square (χ^2)	7.301,98
Degree of Freedom (df)	73
Probability > χ^2 (p-value)	< 0,001

Considering the 73 degrees of freedom, a high LR Chi-Square value (7,301.98) indicates that the model is statistically significant at the 0.001 level. These findings confirm the rejection of the null hypothesis, which states that the model with predictors provides no better fit than the intercept-only model. This demonstrates that the model is robust and that the included temporal, demographic, and environmental factors are strong predictors of accident severity within this large-scale dataset.

Table 8. Ordinal logistic regression analysis results (full model: N = 304,019)

Variable group	Variable	Coeff. (B)	Std. Error (SE)	z-value	p-value	Odds ratio	Significent
Year of the accident (Ref: YA1)	YA2	0.180	0.024	7.548	0.000	1.198	***
	YA3	0.148	0.024	6.253	0.000	1.159	***
	YA4	0.026	0.023	1.116	0.264	1.026	
	YA5	-0.353	0.022	-15.993	0.000	0.702	***
	YA6	-0.548	0.021	-25.785	0.000	0.578	***
	YA7	-0.506	0.021	-24.170	0.000	0.603	***
	YA8	-0.396	0.022	-18.006	0.000	0.673	***
	YA9	-0.316	0.022	-14.197	0.000	0.729	***
	YA10	-0.078	0.053	-1.484	0.138	0.925	

Table 8. Ordinal logistic regression analysis results (full model: N = 304,019) - continuation

Variable group	Variable	Coeff. (B)	Std. Error (SE)	z-value	p-value	Odds ratio	Significent
Month of the accident (Ref: MA1)	MA2	0.180	0.020	8.982	0.000	1.197	***
	MA3	0.194	0.019	10.111	0.000	1.214	***
	MA4	0.172	0.019	8.954	0.000	1.188	***
	MA5	0.123	0.019	6.518	0.000	1.131	***
	MA6	0.123	0.019	6.457	0.000	1.131	***
	MA7	0.140	0.018	7.593	0.000	1.150	***
	MA8	0.070	0.019	3.723	0.000	1.072	***
	MA9	0.129	0.019	6.829	0.000	1.137	***
	MA10	0.130	0.019	6.936	0.000	1.138	***
	MA11	0.148	0.019	7.894	0.000	1.160	***
	MA12	0.143	0.019	7.478	0.000	1.154	***
Time of the accident (Ref: TA1)	TA2	0.133	0.046	2.884	0.004	1.142	***
	TA3	0.187	0.049	3.796	0.000	1.206	***
	TA4	0.236	0.043	5.437	0.000	1.266	***
	TA5	0.212	0.034	6.243	0.000	1.236	***
	TA6	0.179	0.034	5.334	0.000	1.196	***
	TA7	0.125	0.034	3.627	0.000	1.133	***
	TA8	0.103	0.034	3.045	0.002	1.108	***
	TA9	0.086	0.034	2.517	0.012	1.090	**
	TA10	0.148	0.038	3.928	0.000	1.160	***
	TA11	-0.033	0.040	-0.815	0.415	0.968	
	TA12	0.050	0.043	1.171	0.242	1.052	
Province of the accident (Ref: PA1)	PA2	-0.007	0.014	-0.476	0.634	0.993	
	PA3	0.065	0.018	3.605	0.000	1.068	***
	PA4	0.098	0.014	7.258	0.000	1.103	***
Age (Ref: A1)	A2	-0.055	0.014	-3.900	0.000	0.946	***
	A3	0.004	0.018	0.254	0.800	1.004	
	A4	0.081	0.021	3.888	0.000	1.084	***
	A5	0.152	0.026	5.840	0.000	1.165	***
	A6	0.251	0.072	3.512	0.000	1.286	***
Marital status (Ref: MS1)	MS2	0.045	0.010	4.482	0.000	1.046	***
Educational background (Ref: EB1)	EB2	-0.114	0.045	-2.510	0.012	0.892	**
	EB3	-0.098	0.045	-2.182	0.029	0.906	**
	EB4	-0.026	0.045	-0.587	0.557	0.974	
	EB5	0.036	0.052	0.697	0.486	1.037	
	EB6	-0.044	0.045	-0.960	0.337	0.957	
	EB7	-0.087	0.051	-1.697	0.090	0.916	*
	EB8	-0.197	0.054	-3.624	0.000	0.821	***
	EB9	-0.207	0.200	-1.034	0.301	0.813	
OHS Training (Ref: OT1)	OT2	-0.126	0.017	-7.622	0.000	0.882	***
Vocational training (Ref: VT1)	VT2	-0.095	0.010	-9.465	0.000	0.910	***

Table 8. Ordinal logistic regression analysis results (full model: N = 304,019) - continuation

Variable group	Variable	Coeff. (B)	Std. Error (SE)	z-value	p-value	Odds ratio	Significent
Occupation (Ref: O1)	O2	-0.063	0.029	-2.195	0.028	0.939	**
	O3	0.052	0.024	2.212	0.027	1.054	**
	O4	0.181	0.059	3.095	0.002	1.199	***
	O5	0.186	0.041	4.475	0.000	1.204	***
	O6	0.053	0.077	0.688	0.491	1.054	
	O7	0.115	0.025	4.607	0.000	1.122	***
	O8	0.060	0.027	2.193	0.028	1.062	**
	O9	0.212	0.025	8.369	0.000	1.236	***
Overall experience (Ref: OE1)	OE2	0.095	0.013	7.125	0.000	1.100	***
	OE3	0.146	0.017	8.849	0.000	1.157	***
	OE4	0.160	0.019	8.427	0.000	1.173	***
	OE5	0.161	0.020	8.135	0.000	1.175	***
Location of the accident Ref: LA1)	LA2	-0.039	0.046	-0.840	0.401	0.962	
Accident environment (Ref: AE1)	AE2	0.084	0.018	4.635	0.000	1.087	***
	AE3	0.060	0.164	0.369	0.712	1.062	
	AE4	0.242	0.050	4.820	0.000	1.274	***
	AE5	0.074	0.059	1.257	0.209	1.077	
	AE6	-0.016	0.052	-0.302	0.763	0.985	
	AE7	-0.108	0.084	-1.284	0.199	0.898	
	AE8	-0.087	0.078	-1.115	0.265	0.916	
Thresholds	**Cut 1**	**-0.072**	0.005	-	-	-	
	Cut 2	**1.534**	0.005	-	-	-	
	Cut 3	**-1.506**	0.008	-	-	-	
Mean dependent variable		1.568		LR Chi-Square			7.301.98
SD dependent variable		0.556		Prob > \chi2			0.000
Number of observations (N)		304.019		Akaike crit. (AIC)			450.809.41
Pseudo R2 (McFadden)		0.016		Bayesian crit. (BIC)			451.585.02
*** p < 0.01. ** p < 0.05. * p < 0.1.							

3.2.2. Interpretation of regression coefficients

The results of ordinal logistic regression analysis (Table 8) identified several critical predictors of accident severity. The model's findings were interpreted using the estimated coefficients (B) and odds ratios, providing a multidimensional view of the risk factors.

The temporal analysis in Table 8 shows that accident severity initially increased in 2013 (B = 0.180, $p < 0.001$) and 2014 (B = 0.148, $p < 0.001$) compared with 2012, but a significant downward trend emerged after 2016. Specifically, 2019 (B = -0.396, $p < 0.001$) and 2020 (B = -0.316, $p < 0.001$) were associated with milder accident consequences. Monthly data indicate that February (B = 0.180) and March (B = 0.194) represent the highest-risk periods compared to January. Furthermore, accident severity peaked during the early morning hours (04:00–08:00) and remained significantly higher during the evening (B = 0.065) and night (B = 0.098) periods compared to the daytime, likely owing to reduced visibility and fatigue.

Regional findings directly address "size bias" concerns. Despite its high accident frequency, Istanbul showed no statistically significant difference in severity compared with Ankara ($p = 0.634$). In contrast, Izmir (B = 0.065, $p < 0.001$) and other provinces (B = 0.098, $p < 0.001$) had a higher probability of severe outcomes. This confirms that regional industrial density and safety enforcement levels are more critical determinants of severity than total workforce volume.

The demographic analysis showed that accident severity increased progressively with age, with the highest risk observed in the 65+ age group (B = 0.251, $p < 0.001$). Married employees (B = 0.045, $p < 0.001$) were also slightly more likely to experience severe consequences.

This model strongly validates the protective role of education and safety interventions. Compared to the illiterate group, workers with elementary ($B = -0.114$) or middle school ($B = -0.098$) education experienced lower severity. Most importantly, the administration of OHS training ($B = -0.126$, $p < 0.001$) and vocational training ($B = -0.095$, $p < 0.001$) significantly reduced the likelihood of severe consequences, emphasising that targeted training effectively mitigates high-risk incidents.

Occupational roles significantly influenced severity, with plant and machine operators ($B = 0.060$, $p < 0.05$) and unqualified workers ($B = 0.212$, $p < 0.001$) facing higher risks than managers. Professional experience exhibited a seniority trend; workers with over 24 years of experience faced higher severity risks ($B = 0.161$, $p < 0.001$), which may be attributed to senior personnel undertaking more hazardous operational tasks.

Environmental conditions at the time of an accident are crucial indicators of severity. Working outside the workplace ($B = 0.242$, $p < 0.001$) and performing activities while working ($B = 0.084$, $p < 0.001$) had the most severe consequences compared to breaks. Accidents occurring while driving from home ($B = 0.262$) or driving home from work ($B = 0.360$) were also associated with significantly increased severity, highlighting substantial risks outside traditional construction boundaries.

3.2.3. Cutoff points in the ordinal logistic regression model

In an ordinal logistic regression model, the cut-off points are the threshold values that determine the transition between successive categories. They represent the boundaries between the observed ordinal categories and the underlying continuous (latent) trend variables. The three cutoff values obtained in the model are presented and interpreted as follows:

Cut1 = -0,072, SE = 0,005, 95 % interval pouzdanosti = = [-0,082, -0,062]
Cut2 = 1,534, SE = 0,005, 95 % interval pouzdanosti = = [1,524, 1,544]
Cut3 = -1,506, SE = 0,008, 95 % interval pouzdanosti = = [-1,522, -1,490]

The cut-off values obtained using the ordinal logistic regression model were statistically significant. The predicted cutoff values for the model were: cut1 = -0.07 (SE = 0.005), cut2 = 1.53 (SE = 0.005), and cut3 = -1.51 (SE = 0.008), respectively. They demonstrated how an unobservable latent trend variable could be divided into four ordinal levels of the dependent variable [46]. All cutoff values were statistically significant and had narrow confidence intervals (Cut1: CI [-0.08, -0.06]; Cut2: CI [1.52, 1.54]; Cut3: CI [-1.52, -1.49]), suggesting that the ordinal structure of the model was appropriate and that there were significant transitions between the observed categories [47]. The Cut1 value represented the transition point between "continued work" (Category 1) and "minor injury" (Category 2) after the accident, Cut2 defined the distinction between "minor injury" and "limb loss" (Category 3), and Cut3 defined the distinction between "loss of limb" and "death" (Category 4).

The findings confirmed that the ordinal structure and cut-off points were statistically valid, thereby supporting the validity of the model. The significance of the cutoff points also indicated that the proportional odds assumption was valid in the model [48]. In conclusion, the cut-off values demonstrated the ability of the model to distinguish among ordinal categories, and the ordinal regression model was statistically significant and reliable. Therefore, the model was considered theoretically and practically valid.

3.2.4. Summary of the ordinal logistic regression model and results of the predictive margins analysis

Predictive margins were calculated to estimate the average probability of the four distinct accident outcomes based on the validated training dataset (N = 304,019). These values reflect the capacity of the model to identify the risk distribution at the mean values of the independent variables. Due to the large sample size, high-precision reporting was used for standard errors and confidence intervals to ensure statistical transparency.

As summarised in Table 9, nonfatal injuries (Category 2) were the most frequent outcome, with a predicted probability of 0.53400 (SE = 0.00075, $z = 714.75$, $p < 0.001$). This was followed by the probability of returning to work (Category 1) at 0.45400 (SE = 0.00075, $z = 608.79$, $p < 0.001$). Although severe outcomes were numerically low, their statistical significance was robust. The probability of death (Category 4) was estimated to be 0.01000 (SE = 0.00015, $z = 65.52$, $p < 0.001$) and the probability of limb loss (Category 3) was 0.00200 (SE = 0.00006, $z = 32.28$, $p < 0.001$).

Table 9. The results of the predictive margin analysis (N = 304,019)

Outcome category	Margin (probability)	Standard error - SE	Z-value	P > z	[95% Conf. Interval]
Category 1 (Return to Work)	0.45400	0.00075	608.79	0.000	0.45253 – 0.45547
Category 2 (Injury)	0.53400	0.00075	714.75	0.000	0.53253 – 0.53547
Category 3 (Loss of Limb)	0.00200	0.00006	32.28	0.000	0.00188 – 0.00212
Category 4 (Death)	0.01000	0.00015	65.52	0.000	0.00970 – 0.01030

High Z-values and extremely narrow confidence intervals across all categories demonstrate the robustness of the ordinal logistic regression model. These findings underscore the fact that even the most severe outcomes are systematically influenced by identified predictors, such as age, vocational training, and work environment, rather than by random events.

The distribution of these probabilities is shown in Figure 2, which highlights the severe class imbalance inherent in occupational accident data while demonstrating the model's ability to quantify rare but critical events.

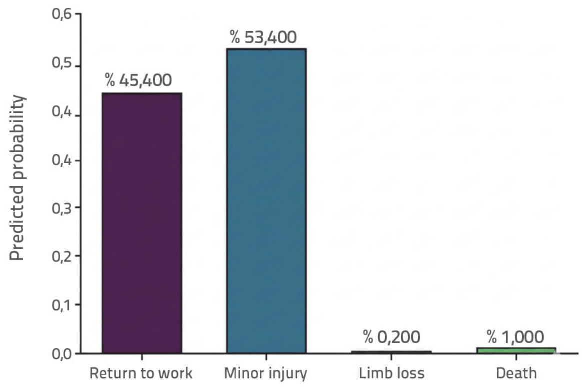


Figure 2. Predicted marginal probabilities of occupational accident outcomes in the Turkish construction industry (N = 304,019)

3.2.5. Validation performance of the ordinal logistic regression model

The model's predictive reliability and generalisability were rigorously evaluated using an independent holdout test dataset (N = 130,295). The classification performance, detailed through the confusion matrix in Table 10, demonstrates the model's success in distinguishing between the four levels of accident severity, with an overall accuracy rate of 56.0 %. In the context of large-scale administrative datasets involving highly imbalanced categories, an overall accuracy exceeding 50 % is considered significant, indicating that the model effectively manages the substantial inherent variance of sociodemographic and environmental predictors [49].

The model exhibited its highest classification efficiency in the "Return to Work" category (57.7 %) and maintained a consistent performance in the "Injury" category (54.6 %). Notably, the model successfully addressed the "rare event" bias by correctly classifying 54.1 % of both "Loss of Limb" and "Death" cases. This performance was particularly significant when compared with the theoretical random chance baseline of 25 % for the four-category model. Achieving an accuracy rate above 54 % for fatal accidents signifies a substantial improvement over random assignment, confirming the analytical depth of the model and its capacity to identify specific risk thresholds.

As illustrated in Figure 3, the consistent accuracy across all categories, specifically for high-severity outcomes, confirmed that the independent variables effectively captured the risk thresholds required to distinguish fatalities from minor incidents. The strong consistency between the validation accuracy and category-specific success rates indicates that the model is well-calibrated, providing a scientifically sound foundation for data-driven OHS interventions in the construction sector.

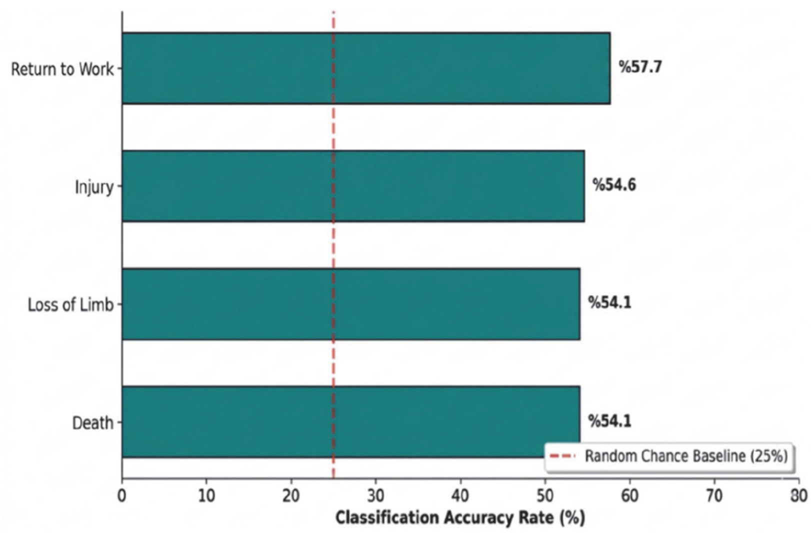


Figure 3. Classification accuracy by accident severity category on the independent validation set (N = 130,295) compared to the random chance baseline

Table 10. Classification performance (confusion matrix) on the independent test set (N = 130,295)

Observed severity	Pred: Category 1	Pred: Category 2	Pred: Category 3	Pred: Category 4	Percent correct
Category 1 (Return to work)	34.103	24.196	54	748	57.7 %
Category 2 (Injury)	30.753	37.926	82	661	54.6 %
Category 3 (Loss of limb)	59	74	166	8	54.1 %
Category 4 (Death)	353	316	3	793	54.1 %
Overall percentage	**50.1 %**	**47.9 %**	**0.2 %**	**1.8 %**	**56.0 %**

4. Discussion

The empirical findings of this decade-long analysis provide a robust framework for understanding the multidimensional predictors of occupational accident severity in the Turkish construction industry (N = 434,134). The primary methodological strength lies in the rigorous validation of the ordinal logistic regression model using an independent test set (N = 130,295), yielding an accuracy rate of 56.0 %. Although this may appear lower than that of the experimental models, it represents a significant achievement in large-scale administrative datasets with substantial variance and class imbalance [49, 50].

According to traditional fit measures like Pseudo R^2 values, they should be interpreted with caution in logistic regression, as low values do not necessarily imply poor utility when predictors show high significance and strong out-of-sample validation [51, 52]. Low R^2 values are statistically expected in very large datasets, and do not mean that the model is useless [53, 54]. To overcome this limitation, the power of the model was confirmed with a prediction success rate of 56.0 % in an independent test set of 130,295 accidents. This performance, which is more than double that of random chance (25 %), validates the practical validity of the model much more robustly than Pseudo R^2 [55].

As emphasised by He and Garcia [56], the ability to correctly classify rare and extreme outcomes, such as the high success rate in identifying fatalities and limb losses in this study, is a more critical performance indicator than the overall accuracy. However, as logistic regression can underestimate rare event probabilities in massive datasets [57], our specialised focus on the predictive accuracy of the model for these critical categories was justified. This capacity to overcome 'rare event bias' ensures that the identified risk factors are not only statistically significant, but also practically relevant for high-severity risk assessment. The validity of the model was further reinforced by the confirmation of the proportional odds assumption [33] and the high significance of the threshold cutoff points [46], proving its effectiveness in distinguishing among the four levels of accident severity.

The temporal and regional patterns identified in the regression analysis offer a strategic perspective on risk evolution over time. The transition from an increasing trend in severity between 2012 and 2014 to a notable decline after 2016 suggests the maturation of safety protocols and regulatory oversight. More importantly, the geographical findings effectively dismantle "size bias" assumptions by demonstrating that Istanbul, despite having the highest frequency of accidents, does not exhibit a significantly higher risk of severe outcomes compared to Ankara ($p = 0.634$). Instead, Izmir and other provinces emerged as higher-risk zones for severe consequences, implying that regional safety cultures and industrial densities were more potent drivers of severity than total employee count. This regional volatility, particularly the peak risks noted in certain periods and locations, highlights the need for dynamic and localised safety interventions, rather than a one-size-fits-all national policy.

Beyond geographical and temporal trends, the socio-demographic and vocational findings underscore the critical role of the "human factor" in accident outcomes. Scientific confirmation of the protective role of OHS and vocational training ($p < 0.001$) provided empirical evidence for the efficacy of targeted safety education as a mitigator of extreme severity [58]. However, the observed increase in severity with both age and professional experience (over 24 years) introduces a critical "seniority risk paradox." This suggests that seasoned workers may either be assigned to more hazardous operational tasks or develop a decreased perception of risk due to habituation, where overconfidence and a diminished sensitivity to routine hazards - often referred to as 'optimism bias' in safety literature - increase the likelihood of severe outcomes despite high technical competence, a finding that warrants a reassessment of OHS training for senior personnel [59]. Furthermore, the environmental influences identified, specifically, the increased risk during early morning hours and transitions outside traditional workplace boundaries, indicate that current safety measures must be expanded to address operational risks beyond the physical construction site.

The synthesis of the predictive marginal probabilities and cutoff thresholds provides a data-driven foundation for proactive safety management. By quantifying the average probabilities of injury (0.534) and death (0.010) with high precision, this study allowed stakeholders to move beyond descriptive statistics toward predictive risk prioritisation. The narrow confidence intervals and high Z-scores reported across the analysis [44] confirm that these outcomes were systematically tied to the identified predictors. Consequently, the findings suggest that reducing the severity of accidents in the construction sector requires a multi-pronged approach that integrates regional oversight, vocational training enhancements for high-risk occupations, and specialised safety protocols. Integrating these predictive insights into the national OHS monitoring systems could transform reactive oversight into a proactive risk management framework, aligned with international safety standards for older and more experienced workers. This comprehensive modelling approach, as supported by the existing literature, successfully bridges the gap between theoretical regression assumptions and practical OHS decision-making.

Despite its contributions, this study had several limitations. While this study utilises a massive administrative dataset, it is subject to the inherent limitations of such records, including the potential underreporting of minor incidents. Future studies should integrate real-time sensor data or qualitative psychological assessments to further refine the predictive power of the seniority paradox.

5. Conclusion

This study constitutes one of the most extensive longitudinal analyses of occupational accident severity in the Turkish construction industry and provides a scientifically validated predictive framework. By shifting the perspective from traditional descriptive statistics to a robust ordinal logistic regression model, this study successfully quantified the influence of temporal, regional, and socio-vocational factors on injury outcomes. The rigorous validation process, utilising a large independent test set (N = 130,295), confirmed that the identified risk determinants are not merely retrospective observations but reliable predictors capable of informing evidence-based safety policies. The model's consistent performance across all severity levels, particularly its ability to overcome rare-event bias for fatalities and permanent disabilities, bridges the critical gap between theoretical modelling and practical OHS decision-making.

The findings revealed that accident severity is systematically driven by the complex interplay between human and environmental factors. The identification of the "seniority risk paradox" is a primary conclusion of this research, suggesting that seasoned workers face higher risks of severe outcomes due to habituation and optimism bias. Furthermore, the identified regional volatility, where provinces such as Izmir and Ankara exhibit higher relative risks than Istanbul, demonstrates that safety oversight must move beyond simple employee counts to address localised industrial densities and safety cultures. These conclusions indicate that achieving a safer construction environment requires transitioning from reactive reporting to a proactive data-driven management approach.

Based on these results, industry stakeholders and policymakers should consider several strategic recommendations. First, the predictive marginal probabilities identified in this study could be integrated into national OHS monitoring systems to allow proactive risk prioritisation. Second, safety training programs must be redesigned to address the seniority paradox; senior personnel should undergo specialised "Refresher Safety Courses" focused on risk perception recalibration and behavioural safety. Additionally, empirical evidence of the protective role of vocational training supports the implementation of mandatory, high-standard safety certification across all construction subsectors. Finally, regional OHS directorates should be empowered to develop decentralised intervention strategies that target specific provincial risk profiles, particularly focusing on high-risk operational windows, such as early morning hours and off-site transitions.

In conclusion, this study provides a validated roadmap for reducing the severity of construction accidents in emerging markets. By implementing these data-driven recommendations, the industry can significantly mitigate extreme risks and progress toward a sustainable "Vision Zero" goal. The framework established here serves as a baseline for future studies that could further refine these predictive insights by integrating real-time sensor data or qualitative behavioural assessments into existing administrative datasets.

REFERENCES

[1] Istraživanje ozljeda na radu i profesionalnih bolesti: Praktični vodič za inspektore rada, Međunarodna organizacija rada, Ženeva, 2014.

[2] Çakar, A.E., Akbaba, T., Yazıcı, M., Tekin, B., Çakar, Y.: Izvješće Odbora za zdravlje i sigurnost radnika, 718, TMMOB Komora strojarskih inženjera, Ankara, 2020.

[3] Öztürk, T.: Procjena čimbenika koji utječu na ozljede na radu multivarijatnim statističkim metodama, Doktorska disertacija, Istanbul Aydın University, Istanbul, 2019.

[4] Chiang, Y.H., Wong, F.K.W., Liang, S.: Smrtne nesreće na gradilištima u Hong Kongu, Journal of Construction Engineering and Management, 144 (2018) 3, str. 04017121.

[5] Choi, J., Gu, B., Chin, S., Lee, J.S.: Prediktivni model strojnog učenja temeljen na nacionalnim podacima za smrtonosne nesreće građevinskih radnika, Automation in Construction, 110 (2020), str. 102974.

[6] Američki Zavod za statistiku rada, https://www.bls.gov/news.release/cfoi.nr0.htm, 19.12.2024

[7] Koc, K., Ekmekcioğlu, Ö., Gurgun, A.P.: Predviđanje nesreća u građevinskoj industriji primjenom hibridnog modela valićne analize i strojnog učenja, Automation in Construction, 133 (2022), str. 103987.

[8] Statistika Zavoda za socijalno osiguranje, https://guvenliinsaat.csgb.gov.tr/media/3w4lph4f/in%C5%9Faat-sekt%C3%B6r%C3%BC-2023-istatistikleri.pdf, 10.05.2025

[9] Podaci o ozljedama na radu u građevinskom sektoru, https://guvenliinsaat.csgb.gov.tr/media/2dwpjq5a/2023-in%C5%9Faat-genel-veriler-info-1.pdf, 10.05.2025

[10] Cabello, A.T., Martínez-Rojas, M., Carrillo-Castrillo, J.A., Rubio-Romero, J.C.: Analiza ozljeda na radu prema profesionalcima različitih faza gradnje pomoću pravila asocijacije, Safety Science, 144 (2021), str. 105457.

[11] Qi, H., Zhou, Z., Irizarry, J., Lin, D., Zhang, H., Li, N., Cui, J.: Automatsko prepoznavanje uzročnih čimbenika iz izvještaja o istragama nesreća povezanih s padovima korištenjem metoda strojnog učenja i skupnog učenja, Journal of Management in Engineering, 40 (2024) 1, str. 04023050.

[12] Rui Lei, W., Khoiry, M.A., Mutalib, A.A.: Sustavni pregled identifikacije, procjene i ublažavanja sigurnosnih rizika u visokogradnji, GRAĐEVINAR, 76 (2024) 12, str. 1099-1114, https://doi.org/10.14256/JCE.4073.2024.

[13] Yalçın, E.: Analiza nesreća zbog padova sa skele u građevinskoj industriji, GRAĐEVINAR, 76 (2024) 10, str. 907-917, https://doi.org/10.14256/JCE.4106.2024.

[14] Leonavičiūtė, G., Dėjus, T., Antucheviciene, J.: Analiza i prevencija nesreća na gradilištima, GRAĐEVINAR, 68 (2016) 5, str. 399-410, https://doi.org/10.14256/JCE.1428.2015.

[15] Andrić, J., Xiang, L.: Istraživanje uzroka nesreća, sigurnosnog ponašanja i ozljeda u tri regije Kine, GRAĐEVINAR, 78 (2026) 2, str. 107-117, https://doi.org/10.14256/JCE.4402.2025.

[16] Chua, D.K.H., Goh, Y.M., Model uzroka incidenata za poboljšanje povratnih informacija o sigurnosnom znanju, Journal of Construction Engineering and Management, 130 (2004) 4, str. 542-551.

[17] Goh, Y.M., Ubeynarayana, C.U.: Klasifikacija narativa o nesrećama u građevinarstvu: Evaluacija tehnika rudarenja teksta, Accident Analysis & Prevention, 108 (2017), str. 122-130.

[18] Kale, Ö.A., Baradan, S.: Identifikacija čimbenika koji doprinose težini ozljeda u građevinskoj industriji primjenom logističke regresije, Teknik Dergi, 31 (2020) 2, str. 9919-9940.

[19] Bilim, A., Çelik, O.N.: Procjena vjerojatnosti ozljeda na radu s gubitkom radnog dana u cestogradnji, GRAĐEVINAR, 75 (2023) 12, str. 1183-1192, https://doi.org/10.14256/JCE.3660.2022

[20] Chau, N., Mur, J.M., Benamghar, L., Siegfried, C., Dangelzer, J.L., Français, M., Jacquin, R., Sourdot, A.: Odnos između određenih individualnih karakteristika i ozljeda na radu za različita zanimanja u građevinskoj industriji: studija slučaja-kontrole, American Journal of Industrial Medicine, 45 (2004) 1, str. 84-92

[21] Nuwayhid, I.A., Usta, J., Makarem, M., Khudr, A., El-Zein, A.: Zdravlje djece koja rade u malim gradskim industrijskim radionicama, Occupational and Environmental Medicine, 62 (2005) 2, str. 86-94.

[22] Dong, X.: Dugi radni sati, raspored rada i ozljede povezane s radom među građevinskim radnicima u Sjedinjenim Državama, Scandinavian Journal of Work, Environment & Health, 2005, str. 329-335.

[23] Sanmiquel, L., Freijo, M., Edo, J., Rossell, J.M.: Analiza nesreća na radu u španjolskom rudarskom sektoru od 1982. do 2006. godine, Journal of Safety Research, 41 (2010) 1, str. 1-7.

[24] Onder, S.: Procjena ozljeda na radu s gubitkom radnih dana među radnicima u površinskim rudnicima ugljena primjenom logističkih regresijskih modela, Safety Science, 59 (2013), str. 86-92.

[25] Cakan, H., Kazan, E., Usmen, M.: Istraživanje čimbenika koji doprinose smrtonosnim i nesmrtonosnim nesrećama padova krovopokrivača, International Journal of Construction Education and Research, 10 (2014) 4, str. 300-317.

[26] Choi, J., Gu, B., Chin, S., Lee, J.S.: Prediktivni model strojnog učenja temeljen na nacionalnim podacima za smrtne nesreće građevinskih radnika. Automation in Construction, 110 (2020), str. 102974.

[27] Yahya, M.Y.A., Rohani, J. M.: Profesionalne nesreće i prevencija kroz dizajn u malezijskoj građevinskoj industriji, Zbornik radova Međunarodne konferencije o industrijskom inženjerstvu i operacijskom menadžmentu, 2021, str. 2289-2299.

[28] Recal, F., Demirel, T.: Usporedba metoda strojnog učenja u predviđanju binarne i višeklasne ozbiljnosti ozljeda na radu, Journal of Intelligent & Fuzzy Systems, 40 (2021) 6, str. 10981-10998.

[29] Gonzalez-Delgado, M., Gómez-Dantés, H., Fernández-Niño, J.A., Robles, E., Borja, V.H., Aguilar, M.: Čimbenici povezani sa smrtnim ozljedama na radu među meksičkim radnicima: nacionalna analiza, Plo S one, 10 (2015) 3, str. e0121490.

[30] Gavanji, R.: Metode penalizirane regresije za modeliranje podataka o rijetkim događajima s primjenom na istraživanje ozljeda na radu, doktorska disertacija, Sveučilište Saskatchewan, Kanada, 2019.

[31] Gallego, V., Sánchez, A., Martón, I., Martorell, S.: Analiza ozljeda na radu u Španjolskoj primjenom metoda shrinkage regresije, Safety Science, 133 (2021), str. 105000.

[32] Toptancı, Ş., Erginel, N., Acar, I.P.: Predviđanje težine ozljede na radu u građevinskoj industriji primjenom standardnih i regulariziranih logističkih regresijskih modela, Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 12 (2023) 3, str. 778-798.

[33] McCullagh, P.: Regresijski modeli za ordinalne podatke, Journal of the Royal Statistical Society: Series B (Methodological), 42 (1980) 2, str. 109–127.

[34] Babbie, E.R.: Osnove društvenih istraživanja, 5. izdanje, Wadsworth Publishing, Kalifornija, 2010.

[35] Akboğa, Ö.: Modeliranje težine nesreće u građevinskim ozljedama na radu pomoću logističke regresije, doktorska disertacija, Ege Üniversitesi Fen Bilimleri Enstitüsü, Izmir, 2014.

[36] Long, J.S., Freese, J.: Regresijski modeli za kategorijalne zavisne varijable pomoću Stata, 3. izdanje, Stata Press College Station, Texas, 2014.

[37] Hayat, E., Özden, A.S.: Utvrđivanje čimbenika koji utječu na razinu sreće pojedinaca pomoću generalizirane ordinalne logističke regresijske analize, OPUS International Journal of Society Researches, 18 (2021) 40, str. 2288-2316.

[38] Akın, H.B., Şentürk, E.: Ispitivanje razine sreće pojedinaca pomoću ordinalne logističke regresije, Öneri, 10 (2012) 37, str. 183-193.

[39] Anderson, J. A.: Regresija i uređene kategoričke varijable. Journal of the Royal Statistical Society: Series B (Methodological), 46 (1984) 1, str. 1-22

[40] Liao, T.F.: Tumačenje modela vjerojatnosti, Logit, Probit i drugi generalizirani linearni modeli, Sage Publications, u seriji Sage University Papers, Thousand Oaks, Kalifornija, 1994.

[41] Norusis, M.J.: IBM SPSS Statistics 19 Advanced Statistical Procedures Companion, Addison Wesley, Boston, 2012.

[42] Akkus, Ö., Türkan, S., Tatlidıl, H.: Primjena modela za sekvencijalno zavisne varijable i diskriminantne analize u klasifikaciji komercijalnih banaka prema njihovoj financijskoj uspješnosti, Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences, 15 (2010) 2, str. 319-332.

[43] Liu, X.: Primijenjena ordinalna logistička regresija pomoću Stata: Od jednorazinskog do višerazinskog modeliranja, Sage publications, 2015.

[44] Hosmer, D.W., Lemeshow, S.: Primijenjena logistička regresija, Wiley, New York, 2000.

[45] Brant, R.: Procjena proporcionalnosti u modelu proporcionalnih omjera za ordinalnu logističku regresiju, Biometrics, 1990, str. 1171-1178.

[46] Calvin, J.A.: Regresijski modeli za kategorijalne i ograničene zavisne varijable, 1998.

[47] Williams, R.: Generalizirani uređeni logit/djelomični modeli proporcionalnih omjera za ordinalne zavisne varijable, The Stata Journal, 6 (2006) 1, str. 58–82.

[48] Agresti, A.: Analiza ordinalnih kategorijalnih podataka, 2. izd., John Wiley & Sons, Inc., Hoboken, New Jersey, 2010.

[49] McFadden, D.: Kvantitativne metode za analizu putničkog ponašanja pojedinaca: Neki noviji razvoj, u Behavioural travel modelling, Routledge, (2021) str. 279–318.

[50] Menard, S.: Primijenjena analiza logističke regresije (2. izd.), SAGE Publication, 2001.

[51] Allison, P.D.: Mjere prilagodbe za logističku regresiju, u Zborniku radova sa SAS Global Forum 2014 konferencije, str. 1–13, 2014.

[52] McFadden, D.: Kondicionalna logit analiza kvalitativnog ponašanja pri odabiru, 1972.

[53] Hox, J., Roberts, J.K.: Priručnik za naprednu višerazinsku analizu, Psychology Press, 2011.

[54] Weisberg, S.: Primijenjena linearna regresija (Vol. 528), John Wiley & Sons, 2005.

[55] Harrell Jr, F.E.: Ordinalna logistička regresija. U: Strategije modeliranja regresije: s primjenama na linearne modele, logističku i ordinalnu regresiju te analizu preživljavanja (str. 311-325), Cham: Springer International Publishing, 2015.

[56] He, H., Garcia, E.A.: Učenje iz neuravnoteženih podataka, IEEE Transactions on Knowledge and Data Engineering, 21 (2009) 9, str. 1263–1284

[57] King, G., Zeng, L.: Logistička regresija kod podataka o rijetkim događajima, Political Analysis, 9 (2001) 2, 137–163

[58] Cohen, J.: Statistička analiza snage za bihevioralne znanosti, Routledge, 2013.

[59] Tutz, G.: Regresija za kategorijalne podatke (Vol. 34), Cambridge University Press, 2011.