Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Analysis of Crash Severity Influencing Factors on Mountainous Expressways Based on Ordinal Logistic Regression

Junhui Zou
North China University of Technology
Changan Liu
North China University of Technology
Yuntao Shi
shiyuntao@ncut.edu.cn
North China University of Technology
Peng Ge
Guizhou Institute of Technology
Hailong Dong
Guizhou Institute of Technology
Ying Liao
Guizhou Communications Polytechnic University
Tao Lv
Guizhou Institute of Technology

Article

Keywords: crash severity, ordinal logistic regression, mountainous expressways, high-risk segments, horizontal-vertical alignment coupling, freezing pavement hazards

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11385137/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Analysis of Crash Severity Influencing Factors on Mountainous Expressways Based on Ordinal Logistic Regression

Junhui Zou[1,2], Changan Liu[1], Yuntao Shi[1,*], Peng Ge [2], Hailong Dong[2], Ying Liao[3], Tao Lv[2]

[1] School of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China

[2] College of Civil Engineering, Transportation and Water Conservancy, Guizhou Institute of Technology, Guiyang 550021, Guizhou, China

[3] School of Art and Tourism, Guizhou Communications Polytechnic University, Guiyang 551400, Guizhou, China

** Corresponding author: Yuntao Shi(e-mail: shiyuntao@ncut.edu.cn)*

Abstract

Road traffic injuries remain a leading cause of death worldwide, and mountainous expressways in subtropical karst regions pose distinctive safety challenges related to complex horizontal–vertical alignments and freezing pavement. Using 7,739 crash records (2019–2023) from high-risk segments of expressways in Guizhou Province, China, we developed an ordinal logistic regression model to identify factors associated with crash severity, classified as property damage only, injury, or fatal. Six categories of explanatory variables were considered: holiday status, time of day, segment type, weather, vehicle combination, and a newly proposed 14-grade horizontal–vertical alignment coupling classification. The parallel lines test (p = 0.078) supported the proportional odds assumption, and the likelihood ratio test (χ^2 = 272.978, $p < 0.001$) confirmed overall model significance. At the conventional α = 0.05 level, motorcycle involvement (OR = 24.31), truck–truck collisions (OR = 6.32), freezing pavement (OR

= 1.95), midnight driving (OR = 1.29), and other vehicle configurations were associated with elevated severity, whereas holiday periods (OR = 0.61) and the G6 alignment (OR = 0.53) were protective. Closed-form probability equations were derived for scenario-based severity prediction. These findings provide quantitative benchmarks for safety governance of mountainous expressways.

Keywords: crash severity; ordinal logistic regression; mountainous expressways; high-risk segments; horizontal-vertical alignment coupling; freezing pavement hazards

According to the World Health Organization, road traffic injuries claim approximately 1.19 million lives worldwide each year, with more than 90% of these fatalities occurring in low- and middle-income countries[1]. As a rapidly developing mountainous province in southwestern China, Guizhou faces acute highway safety challenges. Its karst topography—rugged terrain, frequent elevation changes, and complex horizontal–vertical alignments—creates driving conditions fundamentally different from those on temperate plain highways. On expressways with speed limits exceeding 100 km/h, locations with high crash frequencies do not necessarily correspond to the most severe outcomes.

Mountainous expressways in subtropical regions are also exposed to unique weather hazards. Seasonal black ice, formed through freezing rain and temperature-inversion effects, can reduce pavement friction from a typical range of 0.6–0.8 to as low as 0.1–0.2, causing sudden and visually undetectable loss of vehicle control. This hazard remains understudied relative to snow and ice in temperate regions, which are more predictable and visible to drivers.

Discrete response regression models are the dominant analytical paradigm in crash severity research. The proportional-odds ordinal logistic regression (OLR) model is the foundational framework for ordered outcomes (property-damage-only → injury → fatal): it preserves the ordinal structure of severity while remaining computationally

efficient and directly interpretable for engineering and policy decisions[2-4]. Although extended specifications—generalized ordered logit[2] and random-parameter ordered logit[3]—relax the proportional odds assumption or capture unobserved heterogeneity, the fixed-parameter OLR model remains the workhorse of applied crash severity research when the proportional odds assumption is satisfied[4].

Prior work on mountainous freeways has highlighted the synergistic risk of combined horizontal–vertical alignments: specific curve–slope combinations substantially elevate crash likelihood[5], and multi-truck crashes on mountain freeways are strongly influenced by geometric and traffic conditions[6]. On rural mountainous highways, a combined decision-tree and logistic modelling approach shows that horizontal curves on steep gradients and roadside features increase injury-producing crashes[7]. Partial proportional odds models applied to truck crashes on Chinese mountain freeways further demonstrate that curve and grade effects differ across severity thresholds[8]. Weather and geometric characteristics have been shown to influence crash severity on mountainous freeways[9-10], although most parametric studies treat weather and geometry in isolation.

Vehicle-related risks are consistently severe: crashes involving motorcycles and heavy commercial vehicles exhibit substantially higher injury–fatality risks than single-passenger-vehicle crashes[11-12]. Subdomain research identifies additional hazards: roadside barriers and emergency escape ramps mitigate run-off-road severity[13-14]; secondary collisions amplify consequences in tunnel environments[15], and crashes on mountainous freeway tunnel sections exhibit distinct severity patterns[16]; successive sharp curves with insufficient transition lengths increase driver workload[17]; and heavy-vehicle flows interacting with steep grades elevate rollover and jackknifing risks[18].

Despite these advances, three gaps remain in mountain expressway research. First, large-sample empirical evidence for subtropical karst expressways prone to seasonal freezing black ice is scarce. Second, most studies treat curve radius and slope gradient as separate variables, lacking a refined, multi-grade coupling classification that

quantifies combined geometric risk. Third, few studies provide deployable probability formulas for routine on-site risk screening. A recent systematic review confirms that mountain road safety research remains fragmented across single-factor studies[19]. To address these gaps, this study develops an OLR model using 7,739 crash records (2019–2023) from high-risk segments of Guizhou expressways, proposes a 14-grade horizontal–vertical alignment coupling classification, and derives closed-form probability equations for scenario-based severity prediction.

Results

Model diagnostics

Before parameter estimation, multicollinearity was assessed through variance inflation factors (VIFs) computed from auxiliary regressions. All 30 explanatory variables yielded VIF values between 1.02 and 3.97 (mean = 1.62), well below the conventional threshold of 5, indicating no problematic multicollinearity. The parallel lines test did not reject the proportional odds assumption ($\chi^2 = 41.586$, $df = 30$, $p = 0.078$), supporting the use of a standard proportional-odds specification. The likelihood ratio test confirmed that the full model significantly outperformed the intercept-only model ($\chi^2 = 272.978$, $df = 30$, $p < 0.001$). Pseudo-R^2 values (Cox–Snell = 0.035; Nagelkerke = 0.059; McFadden = 0.040) were low but within the range typically reported for crash severity models, reflecting the stochastic nature of crash outcomes; AIC = 3440.39 and BIC = 3649.01.

Parameter estimation

Table 1 reports the maximum likelihood estimates for all 30 dummy-coded explanatory variables and the two threshold intercepts. Odds ratios (ORs) were computed as $\exp(\beta)$, with 95% confidence intervals (CIs) derived from the standard errors. Significance was assessed with two-sided Wald tests at the conventional $\alpha = 0.05$ level. To preserve comparability with exploratory conventions in traffic-safety research, variables that

were significant only at a more lenient screening threshold ($\alpha = 0.15$) are also noted in the text and table notes; these effects are explicitly interpreted as exploratory and require confirmation in future work.

Table 1 Ordinal logistic regression parameter estimates

Category	Code	Variable	β	S.E.	OR (95% CI)	p-value
Holiday	A1	Holiday	-0.498	0.095	0.608 [0.504, 0.732]	<0.001
	A2	Non-holiday(ref.)	0	—	—	—
Time of Day	B1	00:00–06:00	0.252	0.112	1.287 [1.033, 1.602]	0.025
	B2	18:00–24:00	0.081	0.097	1.084 [0.897, 1.311]	0.403
	B3	12:00–18:00	-0.065	0.082	0.937 [0.798, 1.100]	0.428
	B4	06:00–12:00 (ref.)	0	—	—	—
Segment Type	C1	Bridge	0.091	0.210	1.095 [0.726, 1.653]	0.663
	C2	Service area	0.095	0.175	1.100 [0.780, 1.550]	0.587
	C3	Interchange	-0.124	0.081	0.883 [0.754, 1.035]	0.123
	C4	Tunnel	-0.048	0.099	0.953 [0.785, 1.157]	0.625
	C5	Subgrade (ref.)	0	—	—	—
Weather Condition	D1	Sunny	-0.132	0.081	0.876 [0.748, 1.027]	0.104
	D2	Rain	-0.029	0.085	0.971 [0.822, 1.148]	0.730
	D3	Freezing	0.667	0.247	1.948 [1.201, 3.162]	0.007
	D4	Fog	0.144	0.292	1.155 [0.652, 2.047]	0.623
	D5	Overcast(ref.)	0	—	—	—
Vehicle Combination	F1	Motorcycle	3.191	0.347	24.313 [12.316, 47.996]	<0.001
	F2	Truck–truck	1.844	0.185	6.322 [4.399, 9.085]	<0.001
	F3	Truck–passenger	1.037	0.142	2.821 [2.135, 3.726]	<0.001
	F4	Passenger–passenger	0.621	0.108	1.861 [1.506, 2.299]	<0.001
	F5	Single truck	0.309	0.150	1.362 [1.015, 1.828]	0.039
	F6	Single passenger(ref.)	0	—	—	—
	G1	1+1	-0.181	0.278	0.834 [0.484, 1.439]	0.514
	G2	1+2	-0.248	0.151	0.780 [0.580, 1.049]	0.102

	G3	1+3	-0.422	0.338	0.656 [0.338, 1.272]	0.212
	G4	2+1	-0.197	0.517	0.821 [0.298, 2.262]	0.703
	G5	2+2	0.235	0.309	1.265 [0.690, 2.318]	0.446
	G6	3+1	-0.641	0.194	0.527 [0.360, 0.770]	0.001
	G7	3+2	-0.109	0.180	0.897 [0.630, 1.276]	0.546
Horizontal-Vertical Alignment	G8	3+3	0.285	0.231	1.330 [0.846, 2.091]	0.218
	G9	4+1	-0.240	0.202	0.787 [0.529, 1.169]	0.236
	G10	4+2	-0.086	0.150	0.918 [0.684, 1.231]	0.567
	G11	4+3	-0.100	0.201	0.905 [0.610, 1.342]	0.619
	G12	5+1	-0.166	0.224	0.847 [0.546, 1.314]	0.460
	G13	5+2	-0.294	0.159	0.745 [0.546, 1.018]	0.065
	G14	5+3(ref.)	0	—	—	—
Thresholds	—	Threshold 1 (Y ≤ 1)	2.052	0.186	7.783 [5.406, 11.207]	<0.001
	—	Threshold 2 (Y ≤ 2)	5.421	0.234	226.105 [142.931, 357.680]	<0.001

Note. CI = confidence interval; SE = standard error; ref. = reference category; OR = odds ratio = exp(β). The reference-category parameter is fixed at zero. p-values were computed from two-sided Wald tests; $p < 0.05$ is considered statistically significant (95% CI), and $0.05 \leq p < 0.15$ is interpreted as exploratory (85% CI).

At $\alpha = 0.05$, nine variables were statistically significant. Risk-amplifying factors were motorcycle involvement (OR = 24.31), truck–truck collisions (OR = 6.32), truck–passenger collisions (OR = 2.82), passenger–passenger collisions (OR = 1.86), single-truck collisions (OR = 1.36), freezing pavement (OR = 1.95), and midnight driving (OR = 1.29). Protective factors were holiday periods (OR = 0.61) and the G6 alignment (OR = 0.53). Bridge, service area, tunnel, rain, and fog showed no significant association. Under the exploratory threshold ($\alpha = 0.15$), sunny weather, interchange segments, and alignments G2 and G13 also indicated protective directions; these effects did not reach significance at $\alpha = 0.05$ and are interpreted cautiously. Figure 1 illustrates the injury–fatality rates of the factors discussed above ($\alpha = 0.15$).

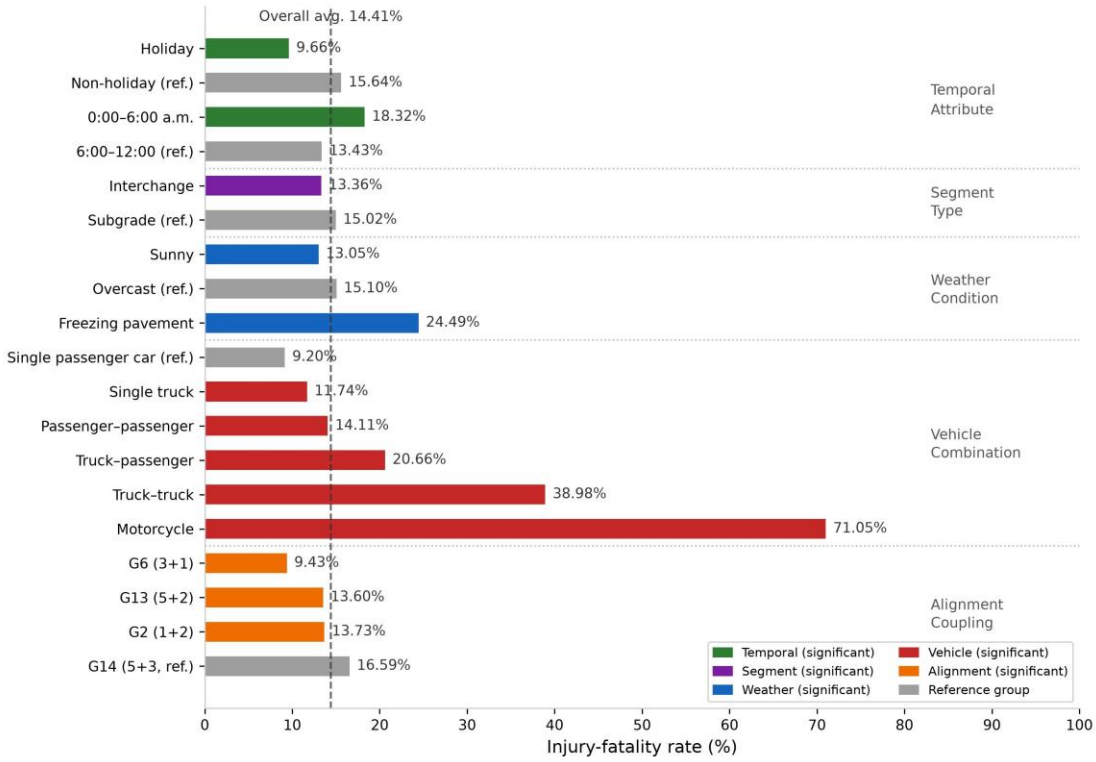


Figure 1 Injury-fatality rates of crash-related factors identified at the exploratory screening threshold (α = 0.15).

Temporal factors

Holiday periods were associated with significantly lower severity odds (OR = 0.61, 95% CI 0.50–0.73; $p < 0.001$), consistent with descriptive statistics (injury–fatality rate 9.66% vs 15.64% on non-holidays). This pattern may reflect intensified traffic enforcement, lower average speeds under congested flows, and a reduced share of commercial heavy vehicles during official holidays. The midnight-to-early-morning window (00:00–06:00) was the only time period significantly associated with elevated severity (OR = 1.29, 95% CI 1.03–1.60; injury–fatality rate 18.32%), consistent with driver fatigue and reduced visibility; evening and afternoon windows were not significant[20].

Segment infrastructure factors

After controlling for vehicle, weather, and temporal factors, segment type showed no statistically significant association with severity at α = 0.05. Interchange segments exhibited a protective direction (OR = 0.88, 95% CI 0.75–1.04, p = 0.126) that was

significant only under the exploratory threshold; consistent with evidence that interchange design features such as enhanced signage and lower operating speeds reduce severity[21], this effect should be confirmed in future work.

Weather factors

Freezing pavement was associated with markedly higher severity odds (OR = 1.95, 95% CI 1.20–3.16; injury–fatality rate 24.49%), confirming black ice as the most hazardous weather condition[22]. Sunny weather showed a protective direction (OR = 0.88, 95% CI 0.75–1.03, p = 0.103) that did not reach significance at $\alpha = 0.05$. Rain and fog were not significant.

Vehicle combination factors

Relative to single-passenger vehicles, all five other configurations significantly increased severity odds: motorcycle (OR = 24.31, 95% CI 12.32–48.00), truck–truck (OR = 6.32, 95% CI 4.40–9.09), truck–passenger (OR = 2.82, 95% CI 2.14–3.73), passenger–passenger (OR = 1.86, 95% CI 1.51–2.30), and single truck (OR = 1.36, 95% CI 1.02–1.83). The extreme odds ratio for motorcycles reflects the absence of occupant protection, while the high values for truck-involved configurations reflect the large impact energy of heavy vehicles[11].

Alignment coupling factors

Using the highest-risk grade G14 (400–1000 m radius, 4–5.99% slope; injury–fatality rate 16.59%) as the reference, the G6 grade (2001–3000 m radius, 0–1.99% slope) showed the strongest protective association (OR = 0.53, 95% CI 0.36–0.77, p = 0.001), corresponding to a 47% reduction in severity odds. Grades G2 (OR = 0.78, p = 0.100) and G13 (OR = 0.75, p = 0.064) also showed protective directions that were significant only at the exploratory threshold. These results support the hypothesis that well-matched radius–slope combinations reduce severity risk, although only G6 reached conventional significance[5,17].

Marginal effects

Marginal effects, computed at sample means, quantify probability changes in directly interpretable units. Motorcycle involvement raised injury probability by 22.5 percentage points (pp) and fatal probability by 2.1 pp; truck–truck collisions raised injury probability by 15.3 pp; freezing pavement and midnight driving raised injury probability by 4.2 pp and 2.8 pp, respectively. On the protective side, the G6 alignment reduced injury probability by 5.1 pp and holiday periods by 3.2 pp. The ranking of marginal effects differs from that of log-odds coefficients because of the non-linear link function, but the qualitative conclusions coincide.

Scenario-based probability prediction

To demonstrate practical utility, predicted severity probabilities were computed with the closed-form Equations (2)-(4). In a compound-risk scenario (holiday, 03:00, interchange segment, freezing pavement, passenger–passenger collision, G6 alignment), the predicted probabilities were P_{PDO} = 85.5%, P_{Injury} = 13.9%, and P_{Fatal} = 0.6%; the protective directions of the G6 alignment and, to a lesser extent, the interchange segment partly offset the adverse effects of freezing pavement and midnight driving. In the worst-case scenario (non-holiday, midnight, subgrade, freezing pavement, motorcycle, G14), predictions were P_{PDO} = 11.3%, P_{Injury} = 67.4%, and P_{Fatal} = 21.2%. In the best-case scenario (holiday, daytime, interchange, sunny, single-passenger vehicle, G6), predictions were P_{PDO} = 96.9%, P_{Injury} = 3.0%, and P_{Fatal} = 0.1%. The 21.1-pp spread in fatal probability between the extreme scenarios illustrates the strong combined effect of the modelled factors and supports the use of the derived formulas for on-site risk screening without specialised software.

Discussion

Our results corroborate international evidence identifying midnight driving, motorcycle involvement, and heavy-truck collisions as robust severity determinants across mountainous road environments[7,9,11]. The convergence of these patterns across distinct

geographical contexts suggests that vehicle- and time-based countermeasures are broadly transferable.

The study contributes two region-specific findings. First, freezing pavement was the most lethal weather condition (24.49% injury–fatality rate; OR = 1.95), exceeding rain and fog, and consistent with evidence that icy pavement sharply increases severe-crash odds on rural mountainous highways[9]. Black ice on subtropical karst expressways is visually undetectable, so conventional driver responses are ineffective; this hazard warrants dedicated engineering and operational countermeasures.

Second, the 14-grade horizontal–vertical alignment coupling classification provided fine-grained geometric risk stratification. G14 (small radius combined with steep slope) carried the highest empirical injury–fatality rate (16.59%), whereas G6 (medium radius combined with gentle slope) was the safest (9.43%) and the only grade reaching conventional significance as a protective factor (OR = 0.53). These results align with prior work showing that combined curve–slope configurations exert synergistic effects on severity[5,17] and with evidence from other mountainous contexts[7-8].

Notably, official holidays in China were associated with lower severity (9.66% vs. 15.64%), in contrast to some studies from other countries that report elevated holiday severity due to drunk driving and speeding. The difference likely reflects China's intensive holiday traffic enforcement, standardised traffic organisation, and reduced commercial heavy-vehicle shares, underscoring the need for context-specific safety strategies[20].

Four priority countermeasures emerge from the marginal-effect analysis. First, strengthen midnight mobile patrols and fatigue-warning facilities along long mountain segments (18.32% injury–fatality rate during 00:00–06:00). Second, install real-time pavement temperature and humidity sensors on shaded sections, bridge decks, and high-risk slopes, and establish tiered black-ice early-warning mechanisms (from advisory alerts to mandatory speed restrictions). Third, prioritise the G6 alignment standard (medium radius, gentle slope) in new construction, apply differentiated speed

limits on high-risk grades (G5, G8, G14), and reinforce roadside barriers. Fourth, enforce motorcycle eligibility requirements and consider dedicated truck lanes with physical separation of truck and passenger flows.

Several limitations should be acknowledged. First, the sample was drawn exclusively from high-risk segments, so generalisation to the entire mountain expressway network requires caution. Second, driver-specific variables (age, experience, fatigue, alcohol use) were unavailable, restricting causal interpretation. Third, the outcome is strongly imbalanced (0.6% fatal) and some geometric subgroups are sparse (e.g., G4, n = 32), which may bias estimates for low-frequency categories; sensitivity analyses using a binary injury-versus-non-injury outcome are recommended in future work. Fourth, the fixed-parameter specification assumes constant effects across observations and cannot capture unobserved heterogeneity; random-parameter ordered logit models should be estimated for robustness comparison. Fifth, the main confirmatory analysis used α = 0.05 with 95% CIs; variables significant only at the exploratory α = 0.15 threshold (interchange, sunny weather, G2, G13) were interpreted as exploratory and require confirmation.

In conclusion, using 7,739 crash records from high-risk segments of Guizhou's subtropical karst mountain expressways, we developed and validated an ordinal logistic regression model that identifies motorcycle involvement, multi-truck collisions, freezing pavement, and midnight driving as dominant severity risk factors, and holiday periods and the G6 alignment as protective factors. The proposed 14-grade alignment classification and the derived closed-form probability equations provide quantifiable, deployable benchmarks for alignment optimisation and prioritised safety investment in mountainous expressway networks.

Methods

Data source and study sample

The crash data were obtained from official operational accident records provided by Guizhou Expressway Group for the period January 2019–December 2023. High-risk segments were identified with an improved cumulative frequency curve method, which screens road sections whose crash frequencies exceed the statistical average for comparable mountain expressway segments[23]. After data quality control, 7,739 complete crash records were retained. Each record contains temporal (date, time of day, holiday status), spatial (mileage marker, segment type, curve radius, slope gradient), environmental (on-site weather), and vehicle (collision type, vehicle composition) attributes.

Outcome and explanatory variables

Following China's Provisions on Procedures for Handling Road Traffic Accidents (Order No. 146)[24], crash severity was classified into three ordered levels: Level 1, property damage only (PDO; 6,624 records, 85.6%); Level 2, injury (1,067 records, 13.8%); and Level 3, fatal (48 records, 0.6%). The overall injury–fatality rate was 14.41%. All explanatory variables were binary-coded (0 = absence, 1 = presence), with the last category of each variable group designated as the reference (coefficient fixed at zero). Six categories were considered: holiday status, time of day, segment type, weather condition, vehicle combination, and the 14-grade horizontal–vertical alignment coupling classification (Table 2).

Table 2 Explanatory variable dummy coding

Variable Category	Classification	Code	Reference Group
Holiday attribute	Holiday / Non-holiday	A1 / A2	A2 (Non-holiday)
Daily period	0–6 h / 18–24 h / 12–18 h / 6–12 h	B1–B4	B4 (6–12 h)
Segment type	Bridge / Service area / Interchange / Tunnel / Subgrade	C1–C5	C5 (Subgrade)
Weather	Sunny / Rain / Freezing / Fog / Overcast	D1–D5	D5 (Overcast)

Variable Category	Classification	Code	Reference Group
Vehicle combination	Motorcycle / Truck-truck / Truck-passenger / Passenger-passenger / Single truck / Single passenger	F1–F6	F6 (Single passenger)
Horizontal-vertical alignment	14 coupled grades G1–G14	G1–G14	G14 (Highest-risk grade)

Horizontal–vertical alignment coupling classification

Horizontal curve radius (five bands: >4000, 3001–4000, 2001–3000, 1001–2000, 400–1000 m) and longitudinal slope (three bands: 0–1.99%, 2–3.99%, 4–5.99%) were cross-classified into 14 grades (G1–G14; Table 3). The theoretical 3001–4000 m × 4–5.99% combination was excluded because no crashes occurred in this category. G14 (400–1000 m, 4–5.99%) was designated the reference grade as the highest-risk layout (injury–fatality rate 16.59%), while G6 (2001–3000 m, 0–1.99%) was the lowest-risk grade (9.43%).

Table 3 Fourteen-grade horizontal – vertical alignment coupling classification

Horizontal Radius (m)	Slope 0–1.99%	Slope 2–3.99%	Slope 4–5.99%
>4000	G1 (1+1)	G2 (1+2)	G3 (1+3)
3001–4000	G4 (2+1)	G5 (2+2)	No samples
2001–3000	G6 (3+1)	G7 (3+2)	G8 (3+3)
1001–2000	G9 (4+1)	G10 (4+2)	G11 (4+3)
400–1000	G12 (5+1)	G13 (5+2)	G14 (5+3)

Ordinal logistic regression model

Crash severity was modelled as an ordered categorical outcome with three levels. Let $P(Y \leq j)$ denote the cumulative probability that severity does not exceed level j (j = 1, 2); by definition $P(Y \leq 3) = 1$. The cumulative logit function for the proportional odds OLR model is defined as Equation (1):

$$\text{logit}\left[P\left(Y \le j\right)\right] = \ln\left(\frac{P\left(Y \le j\right)}{1-P\left(Y \le j\right)}\right) = \alpha_j - \sum_{k=1}^{K} \beta_K X_K \quad \textbf{(1)}$$

where α_j is the threshold intercept for level j ($\alpha_1 < \alpha_2$), β_k the coefficient of the k-th explanatory variable x_k, and K the number of explanatory variables. The proportional odds assumption—constant β_k across thresholds—was validated with the parallel lines test. The marginal occurrence probability of each severity level can be derived as Equations **Error! Reference source not found.**-(4):

$$P(Y=1) = P(Y \le 1) = \frac{1}{1+e^{-(\alpha_1 - \beta X)}} \quad \textbf{(2)}$$

$$P(Y=2) = P(Y \le 2) - P(Y \le 1) = \frac{1}{1+e^{-(\alpha_2 - \beta X)}} - \frac{1}{1+e^{-(\alpha_1 - \beta X)}} \quad \textbf{(3)}$$

$$P(Y=3) = 1 - P(Y \le 2) = 1 - \frac{1}{1+e^{-(\alpha_2 - \beta X)}} \quad \textbf{(4)}$$

These closed-form equations allow severity probabilities to be computed directly from estimated coefficients for any combination of explanatory variables, without specialised software.

Parameter estimation and diagnostics

Parameters were estimated by maximum likelihood in IBM SPSS Statistics[25]. Two-sided Wald tests evaluated individual significance; the main confirmatory analysis used $\alpha = 0.05$ (95% CIs), and an exploratory screening threshold of $\alpha = 0.15$ (85% CIs) was additionally applied to flag potentially meaningful effects for future confirmation. Multicollinearity was assessed via variance inflation factors computed from auxiliary regressions. For the k-th explanatory variable, the VIF is defined as Equation (5):

$$VIF_k = \frac{1}{1-R_k^2} \quad \textbf{(5)}$$

where R^2_k is the coefficient of determination obtained by regressing the k-th variable on the remaining predictors; all VIFs were below 5 (mean 1.62). The parallel lines test ($\chi^2 = 41.586$, $df = 30$, $p = 0.078$) supported the proportional odds assumption at $\alpha = 0.05$. The likelihood ratio test ($\chi^2 = 272.978$, $df = 30$, $p < 0.001$) confirmed that the full model significantly improved the fit relative to the intercept-only model. Because the outcome

was strongly imbalanced (0.6% fatal) and several geometric subgroups were sparse (e.g., G4, n = 32), we recommend supplementary binary-outcome sensitivity analyses and random-parameter specifications in future work.

Descriptive statistics

Table 4 presents sample sizes, proportions, and injury–fatality rates for all variable subgroups. Freezing pavement accounted for only 1.3% of crashes yet had the highest injury–fatality rate among weather categories (24.49%). Motorcycle-involved crashes (0.5%) showed an alarming injury–fatality rate of 71.05%. Among alignment grades, G14 recorded 16.59% and G6 9.43%, supporting the risk differentiation capacity of the coupling classification.

Table 4 Descriptive statistics and injury-fatality rates by variable subgroup

Variable Category	Code	Classification	n	%	Injury-Fatality Rate (%)
Crash Severity	1	Property-damage-only	6,624	85.6	—
	2	Injury (non-fatal)	1,067	13.8	—
	3	Fatal	48	0.6	—
Holiday	A1	Holiday	1,595	20.6	9.66
	A2	Non-holiday (Reference)	6,144	79.4	15.64
Time of Day	B1	0:00–6:00	808	10.4	18.32
	B2	18:00–24:00	1,368	17.7	16.23
	B3	12:00–18:00	3,084	39.9	13.36
	B4	6:00–12:00 (Reference)	2,479	32.0	13.43
Segment Type	C1	Bridge	281	3.6	13.88
	C2	Service area	274	3.5	17.88
	C3	Interchange	2,687	34.7	13.36
	C4	Tunnel	1,475	19.1	14.51
	C5	Subgrade (Reference)	3,022	39.0	15.02
Weather Condition	D1	Sunny	2,499	32.3	13.05
	D2	Rain	2,015	26.0	14.44
	D3	Freezing pavement	98	1.3	24.49
	D4	Fog	88	1.1	17.05
	D5	Overcast (Reference)	3,039	39.3	15.10
Vehicle Combination	F1	Motorcycle	38	0.5	71.05
	F2	Truck–truck	177	2.3	38.98
	F3	Truck–passenger	605	7.8	20.66

Variable Category	Code	Classification	n	%	Injury-Fatality Rate (%)
	F4	Passenger–passenger	4,840	62.5	14.11
	F5	Single truck	775	10.0	11.74
	F6	Single passenger (Reference)	1,304	16.8	9.20
	G1	1+1	154	2.0	15.58
	G2	1+2	1,529	19.8	13.73
	G3	1+3	104	1.3	12.50
	G4	2+1	32	0.4	15.63
	G5	2+2	85	1.1	18.82
	G6	3+1	615	7.9	9.43
Horizontal-Vertical Alignment	G7	3+2	526	6.8	15.78
	G8	3+3	203	2.6	18.72
	G9	4+1	396	5.1	13.13
	G10	4+2	1,771	22.9	15.30
	G11	4+3	344	4.4	16.86
	G12	5+1	243	3.1	15.23
	G13	5+2	1,279	16.5	13.60
	G14	5+3 (Reference)	458	5.9	16.59
Total	—	—	7,739	100.0	14.41

Note. Injury–fatality rate = (injury crashes + fatal crashes) / total crashes × 100%. The overall sample injury–fatality rate is 14.41%.

Ethics approval and consent to participate

All methods were carried out in accordance with the relevant guidelines and regulations. This retrospective study used de-identified official crash records (2019–2023) provided by Guizhou Expressway Group under a data-sharing agreement; the dataset contained no names, national identification numbers, vehicle licence plates, contact details, or any other personally identifiable information, and no contact with crash-involved individuals was involved. The study protocol was reviewed and granted an exemption from scientific ethics review by the Department of Science and Technology of Guizhou Institute of Technology, the institutional body responsible for research ethics review (exemption certificate issued on 23 October 2026). Because the data were analysed anonymously and retrospectively, the requirement for informed consent from the individuals involved was waived by the same institution.

References

[1] World Health Organization. Global status report on road safety 2023. (World Health Organization, Geneva, 2023).

[2] Zhao, L., Wang, C., Yang, H. et al. Exploring injury severity of non-motor vehicle riders involving in traffic accidents using the generalized ordered logit model. Ain Shams Engineering Journal 14, 101962 (2023).

[3] Okafor, S., Adanu, E. K., Lidbe, A. et al. Severity analysis of single-vehicle left and right run-off-road crashes using a random parameter ordered logit model. Traffic Injury Prevention 24, 251–255 (2023).

[4] Mannering, F. L. & Bhat, C. R. Analytic methods in accident research: methodological frontier and future directions. Analytic Methods in Accident Research 1, 1–22 (2014).

[5] Fu, H., Wang, X., Cafiso, S. D. et al. Safety threshold quantification for mountainous freeways: a CatBoost-SHAP framework for crash frequency modeling on combined horizontal and vertical curve alignments. Accident Analysis & Prevention 236, 108675 (2026).

[6] Chen, Z., Wen, H., Zhu, Q. et al. Severity analysis of multi-truck crashes on mountain freeways using a mixed logit model. Sustainability 15, 6499 (2023).

[7] Rusli, R., Haque, M. M., Saifuzzaman, M. et al. Crash severity along rural mountainous highways in Malaysia: an application of a combined decision tree and logistic regression model. Traffic Injury Prevention 19, 741–748 (2018).

[8] Wang, Y., Luo, Y. & Chen, F. Interpreting risk factors for truck crash severity on mountainous freeways in Jiangxi and Shaanxi, China. European Transport Research Review 11, 26 (2019).

[9] Ahmed, I. U. & Ahmed, M. M. Investigating the temporal instability in injury severity outcomes of clear and adverse weather crashes on rural mountainous highways. Transportation Research Record 2676, 107–132 (2022).

[10] Qi, Z., Yao, J., Zou, X. et al. Investigating factors influencing crash severity on mountainous two-lane roads: machine learning versus statistical models. Sustainability 16, 7903 (2024).

[11] Kar, P., Venthuruthiyil, S. P. & Chunchu, M. Crash risk estimation of heavy commercial vehicles on horizontal curves in mountainous terrain using proactive safety method. Accident Analysis & Prevention 199, 107521 (2024).

[12] Li, S. & Han, H. Analysis on causes and random parameters of truck accident severity in mountainous expressway. Journal of Highway and Transportation Research and Development 40, 228–236 (2023).

[13] Zhang, C., Wang, Y., Wu, K. et al. An assessment of the safety effects of roadside barriers on reducing mountainous crash injury severity. Journal of Advanced Transportation 2024, 5568950 (2024).

[14] Li, L., Li, G. Z., Zhang, D. et al. Effects of emergency escape ramps on crash injury severity reduction on mountain freeways: a case study in China. Journal of Transportation Safety & Security 14, 1269–1288 (2022).

[15] Chung, Y. & Kim, J. J. Exploring factors affecting crash injury severity with consideration of secondary collisions in freeway tunnels. International Journal of Environmental Research and Public Health 20, 3723 (2023).

[16] Wang, X., Azati, Y., Quddus, M. et al. Statistical analysis of traffic crashes on mountainous freeway tunnel sections. Transportation Research Record 2678, 1–10 (2024).

[17] Zhang, W., Gao, H., Tang, X. et al. Research on the severity of mountain highway accidents based on random parameter logit model. China Safety Science Journal 34, 101–110 (2024).

[18] Sun, K., Li, Y., Zeng, Q. et al. Exploring the factors affecting the injury severity of crashes on mountainous freeways with high proportions of heavy vehicles

considering their heterogeneous and interactive effects. Sustainability 17, 1624 (2025).

[19] Sharma, S. & Choudhary, P. Which factors compromise safety on mountainous roads: a systematic review. Journal of Traffic and Transportation Engineering (English Edition) (2026).

[20] Fountas, G., Anastasopoulos, P. C. & Abdel-Aty, M. Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. Analytic Methods in Accident Research 18, 57–68 (2018).

[21] Wang, F., Wang, C., Easa, S. M. et al. Do expressway interchanges increase crash injury severities? Insights using temporal instability and unobserved heterogeneity. Journal of Advanced Transportation 2023, 1682850 (2023).

[22] Li, J., Guo, F., Zhou, Y. et al. Predicting the severity of traffic accidents on mountain freeways with dynamic traffic and weather data. Transportation Safety and Environment 5, tdad001 (2023).

[23] Zou, J., Liu, C., Shi, Y. et al. Research on the identification of accident-prone locations on Guizhou expressways based on an improved cumulative frequency curve method. Proc. SPIE 13160, 306–313 (2024).

[24] Ministry of Public Security of the People's Republic of China. Provisions on Procedures for Handling Road Traffic Accidents (Order No. 146) (2018).

[25] Agresti, A. Categorical Data Analysis 3rd edn (John Wiley & Sons, Hoboken, 2013).

Acknowledgements

This work was supported by the Guizhou Provincial Theoretical Innovation Project (Grant No. GZLCLH-2026-78), the Science and Technology Project of the Guizhou Provincial Department of Education (Grant No. QJHKY-2021-259), the Guizhou Provincial Science and Technology Department Project (Grant No. QKHJC-2019-1151), the Guizhou Communications Polytechnic University Project (Grant No.

2024ZD01SK), and the Guizhou Provincial Department of Transportation Project (Grant No. 2022121042).

The authors sincerely thank Guizhou Expressway Group for providing the crash dataset used in this research. The authors acknowledge the use of Doubao and DeepSeek for language polishing and error checking. All scientific content and conclusions were generated and verified by the authors, who take full responsibility for the work.

Author contributions

Conceptualization: Z.J.H. and S.Y.T.

Data curation: D.H.L. and L.T.

Formal analysis: Z.J.H. and G.P.

Funding acquisition: Z.J.H. and G.P.

Investigation: D.H.L. and L.Y.

Methodology: S.Y.T. and L.C.A.

Project administration: Z.J.H. and G.P.

Resources: L.Y. and L.T.

Supervision: S.Y.T. and L.C.A.

Writing - original draft: Z.J.H. and G.P.

Writing - review & editing: S.Y.T. and L.C.A.

Competing interests

The authors declare no competing interests.

Data availability

The crash data were obtained from the operating agency of Guizhou Expressway Group under a data-sharing agreement and cannot be made publicly available. Aggregated statistics supporting the findings of this study are presented in the tables and figures of

this article. Reasonable requests for further information should be directed to the corresponding author.