Technical Memorandum

Hard-Braking Events Sample Data Evaluation

Jidong J. Yang, Ph.D., P.E.
Associate Professor

Jiachen Bian
Graduate Research Assistant

College of Engineering
University of Georgia

March 29, 2026

Table of Contents

1. Introduction

This technical memorandum presents an assessment of a vendor-provided sample of hard-braking events to determine whether the spatial distribution of braking-related events corresponds to the spatial distribution of crash frequency when both datasets are aggregated using a common roadway segmentation reference. The objective of this assessment is to determine whether vendor-derived braking events, when aggregated to roadway segments, produce rank-consistent spatial patterns relative to crash frequency. Evidence of such consistency would provide preliminary justification for (a) procuring the vendor dataset and/or (b) conducting subsequent roadway safety analyses that use these event data as a substitute for, or complement to, traditional crash data.

The analysis is explicitly framed as a distributional consistency assessment rather than a causal inference. Specifically, we examine whether roadway segments with higher intensities of braking events tend to correspond to segments with higher crash frequencies under a consistent spatial indexing scheme. Additionally, we evaluate the hard-braking events as an explanatory safety exposure through a negative binomial (NB) regression, conditional upon other relevant explanatory variables.

1.1 Test Dataset Description

Vendor hard-braking events data were provided as a one-week sample spanning November 10-16, 2025, comprising over 1.8 million individual records. The vendor dataset includes six event classes, namely *Hard Acceleration*, *Hard Braking*, *Severe Acceleration*, *Severe Braking*, *Very Hard Acceleration*, and *Very Hard Braking*. In this technical memorandum, we restrict the analysis to the three braking-related classes to maintain behavioral relevance to crash involvement mechanisms. For the selected braking events, the reported acceleration magnitude ranges from 0.30 to 2.14 g. Each event record contains geospatial coordinates, enabling point-based events to be spatially assigned to roadway segments on a GIS map.

Statewide crash data are provided by GDOT. To ensure definitional consistency with braking-related conflict and impact scenarios, crash records are filtered to five crash types: *Angle*, *Head On*, *Rear End*, *Sideswipe-Opposite Direction*, and *Sideswipe-Same Direction*. Crash counts are constructed under two temporal aggregations for subsequent comparison with the segment-level braking-event intensity: a weekly crash dataset covering November 11-17, 2024, and a quarterly crash dataset aggregated across October-December 2024. Both the vendor events and crash data are aggregated to a common analysis unit defined by HERE map roadway segmentation, where the maximum segment length is 200 m. This shared segmentation serves as the consistent spatial index for braking event counting and crash counting, ensuring that all distributional comparisons are performed on identical roadway segments.

1.2 Aggregation Protocol

Spatial aggregation was performed in ArcGIS by matching both braking events and crash records, each represented as point observations, onto the HERE map roadway segments using a within-distance assignment procedure. For this purpose, a 30-meter matching radius was applied, and a one-to-one rule was enforced such that each point observation was assigned to the single nearest HERE map segment within the search radius; when multiple candidate segments were present, the closest segment was selected, and observations without an eligible segment within 30 meters were treated as unmatched and excluded from segment-level aggregation. This spatial matching protocol enforces a shared roadway index by consistently assigning braking events and crashes to specific HERE map segments based on its geographic location, thereby converting the point layers into segment-referenced counts that are directly comparable across datasets and time windows.

1.3 Evaluation Criteria and Decision Rules

The primary evaluation goal of this validation is to test rank-order consistency between segment-level braking-event intensity and segment-level crash frequency. Because both measures are non-negative, strongly right-skewed, and characterized by substantial ties and zero values at the segment level, consistency is assessed using nonparametric rank correlations rather than relying on linear assumptions. Specifically, we compute Spearman's rank correlation (ρ) as the primary indicator of monotonic association between total braking-event counts and crash counts across HERE map segments, and we additionally report Kendall's (τ) coefficient (τ) coefficient as a complementary rank-based measure that is typically more robust in the presence of heavy ties and discrete count data. For each estimated ρ and τ, we report the corresponding p-value under the null hypothesis of no association, thereby quantifying whether the observed alignment in segment rankings could plausibly arise from random ordering. To transparently address zero inflation and distinguish broad network-level co-location patterns from association among crash-occurred locations, all metrics are reported under two parallel structures: one computed over the full network including segments with "*Crash Count = 0*", and one computed over "*Crash Count* > *0*." It should be noted that this assessment is intended as a screening-level consistency check. The vendor-provided event sample covers the one-week period of November 10–16, 2025, while the crash data used for comparison correspond to approximately the same calendar week in the previous year, specifically November 11–17, 2024. It was assumed that the spatial crash distribution patten would remain same or similar across consecutive years (i.e., 2024 and 2025). To support a more robust evaluation using a larger crash dataset, the event sample was also compared against a quarterly crash dataset aggregated for October–December 2024.

In practical terms, the association is considered supportive of distributional consistency when both ρ and τ are positive under both inclusion rules, when the accompanying p-values indicate that the association is unlikely to be explained by random segment ordering, and when the sign and relative magnitude of the correlations are not solely an artifact of the prevalence of zero-crash segments.

2. Comparison of Crash Frequency with Hard Braking Events by Temporal Window (Weekly vs. Quarterly)

Crash counts at the segment level depend on the length of the aggregation window. A weekly window yields sparse crash observations across the network, while a quarterly window captures more crashes per segment and increases the share of crash positive segments for a more robust analysis. This section evaluates whether the rank association between total braking events and crash frequency strengthens when crash frequency is measured over a larger temporal window, using identical spatial units and the same rank-based evaluation criteria.

2.1 Methods

Total braking events of the sample week were aggregated to the roadway segments. Crash frequency was aggregated to the same segments under two temporal windows, weekly crashes for November 11 to 17 in 2024 and quarterly crashes for October through December in 2024. For each temporal window, rank based association between segment-level total braking events and crash frequency was quantified using *Spearman rank correlation* and *Kendall* (τ) *coefficient*. Statistical significance was assessed using the corresponding p value under the null hypothesis of no association. Metrics were reported for all segments and for the subset of segments with crash count greater than zero.

2.2 Results

For the weekly crash window, the analysis covered 158,400 roadway segments. As shown in Table 1, the Spearman rank correlation between total braking events and crash frequency was 0.159 ($p < 0.001$), while the Kendall's (τ) coefficient (τ) coefficient was

0.145 ($p < 0.001$). Restricting the analysis to segments with crash count > 0 resulted in 3,156 segments. Within this crash-positive subset, the Spearman correlation was 0.169 ($p < 0.001$) and Kendall's (τ) coefficient was 0.140 ($p < 0.001$).

Results for the quarterly crash window are summarized in Table 2. As expected, the correlation strengthened when a larger crash dataset was used. The Spearman correlation increased to 0.360 ($p < 0.001$), and Kendall' coefficient increased to 0.325 ($p < 0.001$). Restricting the analysis to segments with crash count > 0 yielded 26,536 segments. Within this crash-positive subset, the Spearman correlation is 0.347 ($p < 0.001$), and Kendall is 0.282 ($p < 0.001$).

Table 1 Rank-Based Association Between Weekly Braking Events and Weekly Crash Frequency

Sample	**N (Segments)**	**Spearman**	**p-value**	**Kendall**	**p-value**
All road segments	158,400	0.159	< 0.001	0.145	< 0.001
Segments with crash count > 0	3,156	0.169	< 0.001	0.14	< 0.001

Table 2 Rank-Based Association Between Weekly Braking Events and Quarterly Crash Frequency

Sample	**N (Segments)**	**Spearman**	**p-value**	**Kendall**	**p-value**
All road segments	158,400	0.36	< 0.001	0.325	< 0.001
Segments with crash count > 0	26,536	0.347	< 0.001	0.282	< 0.001

2.3 Discussion

Both temporal windows exhibit a positive rank association between total braking events and crash frequency at the segment level, and this relationship is statistically supported in both samples. The quarterly window produces substantially higher rank correlations than the weekly window for both the Spearman and Kendall measures. In addition, quarterly aggregation increases the number of crash-positive segments from 3,156 to 26,536, thereby reducing sparsity in the crash outcome. Overall, these findings indicate that the alignment between braking-event intensity and crash-frequency rankings becomes stronger when crash frequency is measured over a longer temporal window, while the direction of the association remains consistent across both temporal specifications.

3. Comparison of Crash Frequency and Hard Braking Events by Functional Classification

Functional class (FC) stratification was conducted using the HERE roadway segmentation functional class attribute. Segments classified as FC 1, 2, and 3, corresponding to interstate, freeway, and arterial roadways, were extracted and analyzed separately.

3.1 Methods

Following the same analytical framework described in Section 2, total braking events and crashes were aggregated to the corresponding roadway segments within each functional class. The same rank-based association statistics, namely Spearman rank correlation and Kendall's (τ) coefficient, were computed and evaluated against the null hypothesis of no association. Results are reported for all segments as well as for the subset of segments with crash count greater than zero.

3.2 Results

For Interstate segments, the sample includes 20,060 segments. As shown in Table 3, the Spearman correlation between total braking events and crash frequency is 0.406 ($p < 0.001$), and Kendall's (τ) coefficient is 0.367 ($p < 0.001$). Restricting the analysis to the 4,888 Interstate segments with crash count greater than zero, the Spearman correlation increases to 0.455 ($p < 0.001$), while Kendall's (τ) coefficient is 0.375 ($p < 0.001$).

For freeway segments, the sample includes 45,776 segments. As shown in Table 4, the Spearman correlation is 0.352 ($p < 0.001$), and Kendall's (τ) coefficient is 0.321 ($p < 0.001$). Restricting the analysis to the 6,655 Freeway segments with crash count greater than zero, the Spearman correlation is 0.363 ($p < 0.001$), and Kendall's (τ) coefficient is 0.298 ($p < 0.001$).

For arterial segments, the sample includes 92,564 segments. As shown in Table 5, the Spearman correlation is 0.363 ($p < 0.001$), and Kendall's (τ) coefficient is 0.326 ($p < 0.001$). Restricting the analysis to the 14,993 Arterial segments with crash count greater than zero, the Spearman correlation is 0.318 ($p < 0.001$), and Kendall's (τ) coefficient is 0.256 ($p < 0.001$).

Table 3 Rank-Based Association Between Braking Events and Crash Frequency on Interstates

Sample	N (Segments)	Spearman	p-value	Kendall	p-value
All road segments	20,060	0.406	< 0.001	0.367	< 0.001
Segments with crash count > 0	4,888	0.455	< 0.001	0.375	< 0.001

Table 4 Rank-Based Association Between Braking Events and Crash Frequency on Freeways

Sample	N (Segments)	Spearman	p-value	Kendall	p-value
All road segments	45,776	0.352	< 0.001	0.321	< 0.001
Segments with crash count > 0	6,655	0.363	< 0.001	0.298	< 0.001

Table 5 Rank-Based Association Between Braking Events and Crash Frequency on Arterials

Sample	N (Segments)	Spearman	p-value	Kendall	p-value
All road segments	92,564	0.363	< 0.001	0.326	< 0.001
Segments with crash count > 0	14,993	0.318	< 0.001	0.256	< 0.001

3.3 Discussion

All three functional classes exhibit a positive rank association between total braking events and crash frequency, and all associations are statistically significant. Interstate segments show the strongest association, with both Spearman and Kendall coefficients exceeding the corresponding values observed for freeway and arterial segments.

For the all-segment sample, the results for freeway and arterial segments are similar in magnitude. However, within the crash-positive subset, the association is weaker for Arterial segments than for Freeway segments. Overall, these results indicate that the alignment between braking-event intensity and crash-frequency rankings persists across functional classes, with the strongest correspondence observed on Interstate segments, which is intuitive as a braking event is more likely to cause a crash due to the high speed.

4. Comparison of Crash Frequency and Hard Braking Events by Area Type

Segments were classified as urban or rural using the U.S. Census Bureau's urban area delineations derived from the Census Urban Area shapefile. This segment-level urban–rural classification was then applied consistently when aggregating total braking events and crash frequency, and when reporting rank-based associations separately for urban and rural areas.

4.1 Methods

Total hard-braking events were aggregated to HERE map segments for the vendor sample week using the GIS spatial matching and aggregation procedure described earlier. Crash frequency was aggregated to the same segments for the crash window used in this context stratification analysis, restricted to the selected crash types. The rank-based association between segment-level total braking events and segment-level crash frequency was quantified separately for urban and rural segments using Spearman rank correlation and Kendall's (τ) coefficient, with the corresponding p-values reported for each metric under the null hypothesis of no association. Results are presented for all segments within each context, as well as for the subset of segments with crash count greater than zero.

4.2 Results

In urban areas, the analysis includes 59,409 segments. As shown in Table 6, the Spearman correlation between total braking events and crash frequency is 0.343 ($p < 0.001$), and Kendall's (τ) coefficient is 0.286 ($p < 0.001$). Restricting the analysis to the 19,105 urban segments with crash count greater than zero, the Spearman correlation is 0.299 ($p < 0.001$), and Kendall's (τ) coefficient is 0.236 ($p < 0.001$). In rural areas, the analysis includes 98,991 segments. As shown in Table 7, the Spearman correlation is 0.178 ($p < 0.001$), and Kendall's (τ) coefficient is 0.169 ($p < 0.001$). Restricting the analysis to the 7,431 rural segments with crash count greater than zero, the Spearman correlation is 0.230 ($p < 0.001$), and Kendall's (τ) coefficient is 0.204 ($p < 0.001$).

Table 6 Rank-Based Association Between Braking Events and Crash Frequency in Urban Areas

Sample	N (Segments)	Spearman	p-value	Kendall	p-value
All road segments	59,409	0.343	< 0.001	0.286	< 0.001
Segments with crash count > 0	19,105	0.299	< 0.001	0.236	< 0.001

Table 7 Rank-Based Association Between Braking Events and Crash Frequency in Rural Areas

Sample	N (Segments)	Spearman	p-value	Kendall	p-value
All road segments	98,991	0.178	< 0.001	0.169	< 0.001
Segments with crash count > 0	7,431	0.230	< 0.001	0.204	< 0.001

4.3 Discussion

Both urban and rural contexts exhibit a positive rank association between total braking events and crash frequency, and all reported associations are statistically significant. When evaluated across all segments, the urban subset shows stronger associations than the rural subset for both rank-based metrics. When the analysis is restricted to segments with crash count greater than zero, the association decreases in urban areas but increases in rural areas relative to their respective all-segment results. This is largely due to the excessive number of zero-crash rural segments.

5. Comparison of Crash Frequency and Hard Braking Events by Crash Type

This section examines whether the segment level association between hard-braking events and crash frequency is sensitive to crash types. Using the same HERE segmentation and the same ArcGIS spatial matching protocol, crash counts are partitioned by three crash type groups and compared against the total hard-braking frequencies. Results are reported using rank-based correlations for the entire network and for crash positive segments to assess whether the observed distributional alignment persists across crash type groups.

5.1 Methods

This section evaluates sensitivity of the rank association between total hard-braking events and crash frequency by crash type group. Total hard braking events were aggregated to HERE map segments for the vendor sample week using the same ArcGIS within distance matching procedure and one to one nearest segment assignment described earlier. Crash frequency was constructed on the same HERE segments under four groups of crash types. Group 1 includes all selected crash types used in the primary analysis. Group 2 includes only rear-end crashes. Group 3 contains angle and head-on crashes. Group 4 includes only angle crashes. For each outcome definition, segment level rank association between total braking events and crash frequency was quantified using Spearman rank correlation and Kendall (τ) coefficient, and the corresponding p value was reported under the null hypothesis of no association. Metrics were reported for the full network and for the subset of segments with crash count greater than zero under each outcome definition.

5.2 Results

As shown in Table 8, under Group 1 (all crash types), the sample includes 142,819 segments. The Spearman correlation between total braking events and crash frequency is 0.360 with p-value less than 0.001, and the Kendall (τ) coefficient is 0.325. Restricting to crash-positive segments yields 23,997 segments. For this subset, the Spearman correlation is 0.347, and the Kendall (τ) coefficient is 0.282. As shown in Table 9, under Group 2 (only rear-end crashes), the Spearman correlation is 0.323, and the Kendall (τ) coefficient is 0.294. Restricting to crash positive segments yields a sample of 16,765 segments. For this subset, the Spearman correlation is 0.280, and the Kendall (τ) coefficient is 0.224. As presented in Table 10, under Group 3 (angle and head-on crashes), the Spearman correlation is 0.242, and the Kendall (τ) coefficient is 0.221. Restricting to crash-positive segments yields a sample of 8,071 segments. For this subset, the Spearman correlation is 0.203, and the Kendall (τ) coefficient is 0.166. As presented in Table 11, under Group 4 (only angle crashes), the Spearman correlation is 0.240, and the Kendall (τ) coefficient is 0.219. Restricting to crash-positive segments yields a sample of 7,385 segments. For this subset, the Spearman correlation is 0.193, and the Kendall (τ) coefficient is 0.158.

Table 8 Rank-Based Association Between Braking Events and Crash Frequency (All Crash Types)

Sample	**N (Segments)**	**Spearman**	**p-value**	**Kendall**	**p-value**
All road segments	142,819	0.360	< 0.001	0.325	< 0.001
Segments with crash count > 0	23,997	0.347	< 0.001	0.282	< 0.001

Table 9 Rank-Based Association Between Braking Events and Crash Frequency (Rear-end Crashes)

Sample	**N (Segments)**	**Spearman**	**p-value**	**Kendall**	**p-value**
All road segments	142,819	0.323	< 0.001	0.294	< 0.001
Segments with crash count > 0	11,708	0.280	< 0.001	0.224	< 0.001

Table 10 Rank-Based Association Between Braking Events and Crash Frequency (Angle and Head-on Crashes)

Sample	N (Segments)	Spearman	p-value	Kendall	p-value
All road segments	142,819	0.242	< 0.001	0.221	< 0.001
Segments with Crash Count > 0	8,071	0.203	< 0.001	0.166	< 0.001

Table 11 Rank-Based Association Between Braking Events and Crash Frequency (Angle Crashes)

Sample	N (Segments)	Spearman	p-value	Kendall	p-value
All road segments	142,819	0.240	< 0.001	0.219	< 0.001
Segments with Crash Count > 0	7,385	0.193	< 0.001	0.158	< 0.001

6. Comparison of Crash Frequency and Hard Braking Events by Social Vulnerability Index

This section investigates the rank-based correlation between hard braking events and crash frequencies by socioeconomic context. The correlation analysis was conducted at both county (macroscopic) and road segment (microscopic) levels. The following subsections detail the methodology and present findings.

6.1 Methods

This section evaluates the relationship between hard-braking events and crash frequency across socioeconomic contexts. The Social Vulnerability Index (SVI) serves as the stratification variable. The dataset was divided into 10 equal intervals based on the SVI from 0 to 1. Spearman rank correlation and Kendall rank correlation tests assess the monotonic relationship between crash counts and braking events within each stratum. The analysis is conducted at two spatial aggregation scales (i.e., the county level and the road segment level) to examine the effect of different spatial scales.

6.2 Results

Table 12 and Table 13 present the correlation coefficients at the county and segment levels respectively. All statistical tests indicate a significant positive correlation with p-values below 0.001. At the county aggregation scale, the Spearman coefficients range from 0.8014 to 0.9728 and Kendall coefficients range from 0.6250 to 0.9121 indicating a strong monotonic relationship. When evaluated at the microscopic segment scale, the correlation strength attenuates. The Spearman coefficients drop to a range between 0.2456 and 0.4195 while Kendall coefficients fall between 0.2209 and 0.3666. Despite the reduction in magnitude the positive association remains statistically significant across all SVI intervals.

Table 12 Correlation Analyses of Crash Counts and Braking Events across SVI Bins (County Level)

SVI Bin	County Count	Spearman ρ	p-value	Kendall τ	p-value
(0, 0.1]	16	0.8971	< 0.001	0.7500	< 0.001
(0.1, 0.2]	16	0.9706	< 0.001	0.9000	< 0.001
(0.2, 0.3]	16	0.9728	< 0.001	0.9121	< 0.001
(0.3, 0.4]	16	0.8565	< 0.001	0.6611	< 0.001
(0.4, 0.5]	16	0.9581	< 0.001	0.8619	< 0.001
(0.5, 0.6]	15	0.9526	< 0.001	0.8613	< 0.001
(0.6, 0.7]	17	0.9626	< 0.001	0.8635	< 0.001
(0.7, 0.8]	15	0.8014	< 0.001	0.6250	< 0.001
(0.8, 0.9]	16	0.9183	< 0.001	0.7782	< 0.001
(0.9, 1.0]	16	0.9088	< 0.001	0.8167	< 0.001

Table 13 Correlation Analyses of Crash Counts and Braking Events across SVI Bins (Segment Level)

SVI Bin	Segment Count	Spearman ρ	p-value	Kendall τ	p-value
(0, 0.1]	13352	0.3028	< 0.001	0.2651	< 0.001
(0.1, 0.2]	12676	0.2456	< 0.001	0.2209	< 0.001
(0.2, 0.3]	12689	0.3856	< 0.001	0.3415	< 0.001
(0.3, 0.4]	17211	0.4116	< 0.001	0.3606	< 0.001
(0.4, 0.5]	16540	0.3630	< 0.001	0.3219	< 0.001
(0.5, 0.6]	16831	0.4195	< 0.001	0.3666	< 0.001
(0.6, 0.7]	13863	0.2909	< 0.001	0.2727	< 0.001
(0.7, 0.8]	12592	0.2516	< 0.001	0.2384	< 0.001
(0.8, 0.9]	14157	0.2654	< 0.001	0.2499	< 0.001
(0.9, 1.0]	12886	0.3423	< 0.001	0.3169	< 0.001

7. Crash Frequency Modeling with Hard-Braking Exposure

To extend the analysis beyond nonparametric association measures, we conducted crash frequency modeling that treats hard-braking events as a proxy for risk exposure, conditional on other explanatory variables. In this framework, the dependent variable is the quarterly crash count aggregated over October through December 2024.

The number of hard-braking events from the vendor sample is included as a covariate representing relative exposure to risky driving conditions at the segment level. In addition, weather-related covariates were constructed by aggregating hourly meteorological observations over October–December 2024, ensuring temporal consistency with the crash aggregation window.

The model incorporates a range of predictors describing segment exposure and operational characteristics, roadway geometry and condition, and meteorological conditions. Continuous predictors include AADT, segment length, free-flow speed, speed difference, longitudinal grade, cross slope, cracking percent, sharp curve ratio, total rainfall, 95th percentile rain intensity, heavy-rain hours, near-freeze hours, high relative humidity (RH) hours, and 95th percentile wind speed.

7.1 Negative Binomial Regression Results

The crash frequency modeling results are summarized in Table 14. The negative binomial regression identifies a clear set of variables associated with segment-level crash counts.

Table 14 Negative Binomial Regression Results

Variable	Coefficient	z statistic	p-value
Intercept	-11.6321	-31.35	<0.001
Log total braking	***0.5007***	***32.786***	***<0.001***
Length	0.0015	4.143	<0.001
Log AADT	0.5031	24.436	<0.001
Free flow speed	-0.0097	-5.325	<0.001
Speed difference	0.0543	10.048	<0.001
Grade	-0.0078	-0.928	0.353
Cross slope	-0.0330	-2.555	0.011
Cracking percent	-0.0083	-5.001	<0.001
Sharp curve ratio	0.6322	3.327	0.001
Rain total	0.0008	2.151	0.032
Rain intensity (95th percentile)	-0.1478	-8.582	<0.001
Heavy rain hours	0.0500	5.226	<0.001

Near freeze hours	0.0052	7.772	<0.001
High relative humidity hours	0.0020	5.958	<0.001
Wind speed (95th percentile)	0.1506	6.741	<0.001

As shown in Table 14, the coefficient estimates indicate that hard-braking events have a positive and statistically significant association with crash frequency. The estimated coefficient is 0.5007 ($p < 0.001$), indicating that braking exposure provides substantial explanatory information after accounting for traffic exposure, operating speed, roadway condition, geometric characteristics, and weather-related factors. Among all explanatory variables, log total hard-braking events emerges as one of the strongest positive predictors in the model.

AADT is also positive and highly significant, with a coefficient of 0.5031 ($p < 0.001$), confirming that higher traffic exposure is associated with higher expected crash counts. Segment length is likewise positive and significant, with a coefficient of 0.0015 ($p < 0.001$), indicating that longer segments tend to accumulate more crashes when other variables are held constant.

The speed-related variables show distinct effects. Free-flow speed has a negative coefficient (− 0.0097) and is significant at the 0.001 level, suggesting that segments with higher free-flow speeds tend to have lower expected crash counts after controlling for the remaining variables in the model. In contrast, speed difference has a positive coefficient of 0.0543 ($p < 0.001$), indicating that greater deviation from free-flow conditions is associated with higher crash frequency. These results suggest that crash occurrence is more closely related to operational disturbance than to nominal speed level alone.

Among the roadway geometry and surface condition variables, cross slope is negative and statistically significant (−0.0330, $p = 0.011$), which helps with drainage during the wet weather conditions. Cracking percent is also negative and significant (−0.0083, $p < 0.001$). In contrast, sharp curve ratio is positive and significant, with a coefficient of 0.6322 ($p = 0.001$), indicating a substantial increase in crash counts for segments with a higher proportion of sharp curvature.

The rainfall-related variables also exhibit meaningful associations. Total rainfall is positive and significant (0.0008, $p = 0.032$), while heavy-rain hours is positive and highly significant (0.0500, $p < 0.001$). These results indicate that both accumulated rainfall and the duration of heavy precipitation are associated with increased crash frequency. Rain intensity (95th percentile) is negative and significant (−0.1478, $p < 0.001$), which may be due to additional cautions taken by the drivers (e.g., reduce speed, stop driving) during the extreme rainfall events.

The cold- and moisture-related variables are all positive and statistically significant. Near-freeze hours has a coefficient of 0.0052 ($p < 0.001$), high relative humidity (RH) hours has a coefficient of 0.0020 ($p < 0.001$), and wind speed (95^{th} percentile) has a coefficient of 0.1506 ($p < 0.001$). These findings indicate that near-freezing conditions, sustained high humidity, and elevated wind speeds are associated with higher crash counts after controlling for the other explanatory variables.

Overall, the regression results support three main conclusions. First, braking exposure, measured by log total hard-braking events, is a strong predictor of crash counts. Second, conventional exposure and operational variables, particularly AADT and speed difference, remain important contributors to crash occurrence. Third, roadway curvature and several adverse weather indicators retain significant effects.

7.2 Model Comparison with and without Hard-Braking Exposure

Table 15 compares the baseline model with the full model that includes hard-braking exposure using the same sample of 142,819 observations. The results show a clear improvement in model fit after including the hard-braking exposure. Log likelihood increases from negative 14,256.82 to negative 13,742.17. AIC decreases from 28,543.63 to 27,516.34. BIC decreases from negative 1,675,201.55 to negative 1,676,218.97. The McFadden pseudo R^2 increases from 0.1283 to 0.1598. The likelihood ratio test confirms the significance of the hard braking exposure in explaining variability in crash counts at the segment level.

Table 15 Incremental Contribution of Hard-braking Exposure

	Baseline Model	**Full Model**	**Change (Δ)**
Number of observations	142,819	142,819	0
Degrees of freedom	14	15	1
Log-likelihood	-14,256.82	-13,742.17	514.649
AIC	28,543.63	27,516.34	-1,027.29
BIC	-1,675,201.55	-1,676,218.97	-1,017.42
McFadden pseudo R^2	0.1283	0.1598	0.0315
Likelihood Ratio (LR) statistic:	1,029.30		
p-value:	<0.001		

Note: The total hard-braking events were excluded in the baseline model.

7.3 Findings and Discussion

The results clearly indicate that hard-braking exposure is significantly correlated with crash frequency at the segment level. In the full negative binomial specification, log total braking events is positive and highly significant even after controlling for traffic exposure, speed conditions, roadway geometry, pavement condition, and weather-related variables. This demonstrates that hard-braking events capture crash-relevant operational information that is not fully represented by conventional covariates alone.

Comparison of the baseline and full models further reinforces this finding. The likelihood-ratio test shows a statistically significant improvement in model performance when log total braking events is included as a measure of risk exposure. At the same time, the model retains the expected effects of conventional predictors, including positive associations for AADT, speed difference, sharp curve ratio, heavy-rain hours, near-freeze hours, high relative humidity hours, and wind speed. These results indicate that hard-braking events serve as an effective behavior-derived exposure indicator, enhancing both the explanatory power and the practical utility of segment-level crash frequency modeling.

8. Conclusions

This technical memorandum documents an assessment of the vendor-provided hard-braking event sample against crash data by aggregating both datasets to the same roadway segmentation using a unified GIS spatial matching procedure. Rank-based analyses indicate positive and statistically significant associations between total braking events and crash counts at the segment level, with stronger alignment observed when crash frequency is aggregated over the October-December 2024 quarterly window compared with a single week in November 2024.

Stratified analyses confirm that this alignment persists across functional classes FC 1 through FC 3. The strongest rank association is observed on Interstate segments, with lower associations on Freeway and Arterial segments. Context stratification shows positive associations in both urban and rural settings, with stronger network level alignment in urban areas and stronger alignment within crash positive segments in rural areas.

Crash type sensitivity analyses show that the rank association is significantly positive across crash type groups. When restricting to angle crashes only, the association decreases, but remains significantly positive. Additionally, correlation analyses by SVI reveal a consistently significant positive association between hard braking events and crash frequencies across all SVI bins. While spatial aggregation at the county level tends to inflate these correlations, the positive rank association remains statistically robust when evaluated at the road segment level, confirming the potential of hard-braking as a surrogate safety measure across diverse socioeconomic contexts.

Within the negative binomial modeling framework for quarterly crash counts, hard-braking exposure remains a statistically significant predictor after controlling for traffic exposure, operating speed measures, roadway geometry and condition, and quarter-aggregated weather variables. Inclusion of hard-braking exposure leads to substantial improvements in log likelihood and corresponding reductions in AIC and BIC relative to a baseline model without braking events.

Collectively, these findings support the conclusion that the vendor braking-event sample provides a coherent segment-level behavioral signal that aligns with observed crash frequency patterns. Moreover, its inclusion materially improves crash count model fit under the tested specification.