Safety21
INNOVATING FOR SAFETY

US DOT National
University Transportation Center for Safety

Carnegie Mellon University

SafeSpeed: Enhancing Work Zone Safety through Speed Enforcement / Impact of Work Zones on Crash Occurrence: A Texas Case Study

PI: Sean Qian (ORCID: 0000-0001-8716-8989)

Lead Research Assistants: Diego Bermudez (ORCID: 0009-0000-8028-8363), Tao Tao (ORCID: 0000-0001-9865-5985)

Final Report – July 30, 2025

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under [grant number 69A3552344811 / 69A3552348316] from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

Technical Report Documentation Page

<table>
<tr><td>1. Report No.
-</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Impact of Work Zones on Crash Occurrence: A Texas Case Study</td><td>5. Report Date
July 30, 2025</td></tr>
<tr><td>6. Performing Organization Code
-</td></tr>
<tr><td colspan="2">7. Author(s)
Diego Bermudez, MS, (ORCID: 0009-0000-8028-8363)
Tao Tao, PhD, (ORCID: 0000-0001-9865-5985)
Sean Qian, PhD, (ORCID: 0000-0001-8716-8989)</td><td>8. Performing Organization Report No.
Enter any/all unique alphanumeric report numbers assigned by the performing organization, if applicable.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213</td><td>10. Work Unit No.</td></tr>
<tr><td>11. Contract or Grant No.
Federal Grant #: 69A3552344811 / 69A3552348316</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
Safety21 University Transportation Center
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213</td><td>13. Type of Report and Period Covered
Final Report (July 1, 2024 – June 30, 2025)</td></tr>
<tr><td>14. Sponsoring Agency Code
USDOT</td></tr>
<tr><td colspan="3">15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.</td></tr>
</table>

16. Abstract
Work zones are essential for maintaining and improving road infrastructure. However, their presence can alter traffic conditions and increase the risk of crashes. This research project applies a causal inference method—Regression Discontinuity Design—to examine the impact of work zones on crash risk using multiple data sources from Texas, US. The findings indicate that work zones significantly increase crash risk. Multiple robustness checks confirmed the consistency of this effect. Further heterogeneous impact analysis revealed that the risk is more pronounced on road segments with lower traffic volumes, higher speeds, and shorter work zones. These results highlight the need for targeted safety policies, including enhanced safety measures in low-traffic areas, improved speed management within work zones, and better design strategies for shorter work zones.

17. Key Words	**18. Distribution Statement**
work zone safety, speed, causal analysis, regression discontinuity design	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161. Enter any other agency mandated distribution statements. Remove NTIS statement if it does not apply.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	19	-

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Impact of Work Zones on Crash Occurrence: A Texas Case Study

Diego Bermudez[*], Tao Tao[†], Sean Qian[‡§]

August 1, 2025

Abstract

Work zones are essential for maintaining and improving road infrastructure. However, their presence can alter traffic conditions and increase the risk of crashes. This research project applies a causal inference method—Regression Discontinuity Design—to examine the impact of work zones on crash risk using multiple data sources from Texas, US. The findings indicate that work zones significantly increase crash risk. Multiple robustness checks confirmed the consistency of this effect. Further heterogeneous impact analysis revealed that the risk is more pronounced on road segments with lower traffic volumes, higher speeds, and shorter work zones. These results highlight the need for targeted safety policies, including enhanced safety measures in low-traffic areas, improved speed management within work zones, and better design strategies for shorter work zones.

Key words: work zone safety, speed, causal analysis, regression discontinuity design

[*]Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, United States

[†]Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States

[‡]Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, United States

[§]Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States

1 Introduction

Work zones are critical interventions for maintaining and enhancing the performance of road networks, thereby contributing significantly to economic development. However, they also introduce temporary changes to the driving environment—such as lane reductions, narrowed roadways, and traffic delays—that may increase crash risk. These disruptions can heighten the likelihood of crashes resulting in severe injuries or fatalities (FHWA, 2024). Consequently, it is essential to rigorously examine the causal impact of work zones on crash risk along roadway segments.

Texas has the longest public road network in the United States (Carney, 2010) and also reports the highest number of fatal crashes occurring within work zones (ARTBA, 2025). Alarmingly, there is a growing trend in work zone-related fatalities in the state. Between 2021 and 2023, the annual average number of fatal crashes in work zones reached 190, a significant increase compared to the previous three-year average of 147 (ARTBA, 2025). As the need for road maintenance continues to grow, it is critical to identify the risk factors contributing to work zone-related crash occurrences in Texas.

In this research project, we adopt a data-driven approach to examine the impact of work zones on crash risk, using work zone records and several other open-source datasets from Texas covering the period 2022 to 2024. We employ a causal inference method—Regression Discontinuity Design (RDD)—to rigorously identify the causal effect of work zones on crash occurrences. In addition, we investigate how this impact varies under different conditions, including average annual daily traffic (AADT), time of day, speed, and work zone length. To ensure the robustness of our findings, we conduct a series of sensitivity analyses and robustness checks.

The rest of the report is organized as follows. We review the related literature on work zone impact on crash risk in the next section. We introduce the datasets and method in Section 3. We present and discuss the model results in Section 4. In the last section, we conclude this study and offer several implications based on the results.

2 Literature review

Over the past three decades, numerous studies have examined the impact of work zones on crash risk (Yang et al., 2015), underscoring the significance and continued relevance of this topic in the field.

A majority of the work zone literature has examined the risk factors that may affect the work zone-related crash risk, including work zone configurations, road features, weather, and other variables. Work zone configurations indicate characteristics of the work zones. These factors mainly include work zone location, length, duration, and number of closed lanes. For example, Garber and Zhao (2002) compared the crash rates occurred in different sections of the work zones and found that activity area had the highest crash rate with the data from Virginia, US. Khattak et al. (2002) applied the data from California, US, and found that longer work zones were positively correlated with crash risk. Road features indicate the characteristics of the road segments where the work zones are located, including road classification, number of lanes, speed limit, and road width. For instance, Li and

Bai (2009) examined the impact of several road characteristics, such as number of lanes, road classification, and speed limit, on work zone crash risk with the data from Kansas, US. They found that number of lanes was negatively correlated with crash severity (Li and Bai, 2009). Weather related variables include temperature, wind speed, precipitation, and weather condition. One example is that Ghasemzadeh and Ahmed (2019) studied the impact of weather condition on crash risk in six states of the US and found that worse weather condition could increase crash risk in work zones during rush hours. Other variables include travel speed, time of day, *etc.* For example, several studies have found that speed is correlated with crash risk in work zones (Yang et al., 2015; Weng et al., 2016). Akepati and Dissanayake (2012) found the most work zone crashes occurred during the daylight.

The applied methods to examine the work zone impact on crash risk mainly include three types: descriptive analysis, correlation analysis, and causal analysis. Descriptive analysis main focuses on the simple comparison of crash rates before and during the work zones. For example, Khattak et al. (2002) compared the crash frequencies in the before- and during-work zone periods to examine the presence of work zones on crash occurrences. Descriptive analysis is simple and easy to apply. However, it does not control for the effects of other risk factors. Correlation analysis is to explore the correlation between the crash rate and multiple risk factors with regression models. For example, Li and Bai (2009) applied a logistic regression to study the effects of multiple factors on crash severity. Khattak et al. (2002) used a Negative Binomial regression to examine risk factors on crash frequencies. Correlation analysis controls for the effects of multiple variables, which help improve the accuracy of estimation. However, correlation analysis, as indicates by its name, can only uncover the correlational relationships instead of causal impacts. Causal analysis utilizes specific research design to explore the causal impact of the work zones and related risk factors on crash risk. For example, Zhang et al. (2022b) applied a RDD approach to examine the causal work zone effect on crash risk with the data from Pennsylvania, US. Causal analysis provides a robust estimate of the work zone impact but requires a careful research design and a large amount of datasets. Among the literature, most of the studies have used either descriptive and correlation analysis approaches, which sometimes generate mixed results (Zhang et al., 2022b). However, the literature using causal analysis is very limited, especially in the context of Texas.

3 Data and methodology

This section aims to describe the temporal and spatial data collection process and the empirical framework on which our results are based. This project applied a RDD setting to evaluate the impact of the presence of a work zone on the crash risk. To achieve this goal in the context of the Texas roadway system, we followed the multi-source data structure proposed by Zhang et al. (2022b). Table 1 shows the five categories of variables we applied in this project.

Table 1: Variable description.

Variable name	Type	Description	Source	Period
Dependent			C.R.I.S	2022.01 - 2024.12
Crash (Y)	Binary	1 if crash occurred and 0 otherwise		
Work zone Features			H.C.R.S	2022.01 - 2024.12
Roadwork (D)	Binary	1 if work zone is present and 0 otherwise		
Duration	Discrete	Duration of the work zone		
Length	Continuous	Length of the work zone		
Weekday	Binary	1 if work zone happened during weekdays and 0 otherwise		
Daytime	Binary	1 if work zone happened during daytime and 0 otherwise		
Road Features			TxDOT	2022.01 - 2024.12
AADT	Continuous	Annual Average Daily Traffic		
Lane counts	Discrete	Number of road lanes		
Speed limit	Continuous	Road segment speed limitation		
Number of intersections	Discrete	Number of roads crossing the segment		
NHS major roads	Binary	1 if it is marked as a major road in the National Highway system and 0 otherwise		
Weather Features			N.C.E.I	2022.01 - 2024.12
Average wind speed	Continuous	Hourly average wind speed		
Average temperature	Continuous	Hourly dry bulb temperature		
Average precipitation	Continuous	Hourly average precipitation index		
Speed			INRIX	2022.01 - 2024.12
Actual speed processed	Continuous	30-min transformed average speed		
Time Features				
Week (m)	Discrete	The number of weeks before (< 0), during (0), and after (> 0) the roadwork is present		

3.1 Data

As we applied the data sources that are not directly related to each other, we needed to correlate the datasets based on their spatial and temporal characteristics to obtain a structure that allows us to apply the methodological specification described in section 3.2.

First, we requested information on work zones from the Texas Department of Transportation (TxDOT, https://drivetexas.org/). This information is to keep users updated on road conditions and potential situations that may impact their travel plans in real time. In our project, we collected historical information on all work zones from 2022 to 2024. We then applied spatial analysis to correlate the work zones with the Texas road network. In addition, we merged duplicated work zones based on their locations and starting and ending times. An example of a map-matched and merged work zone is shown in Figure 1 below.

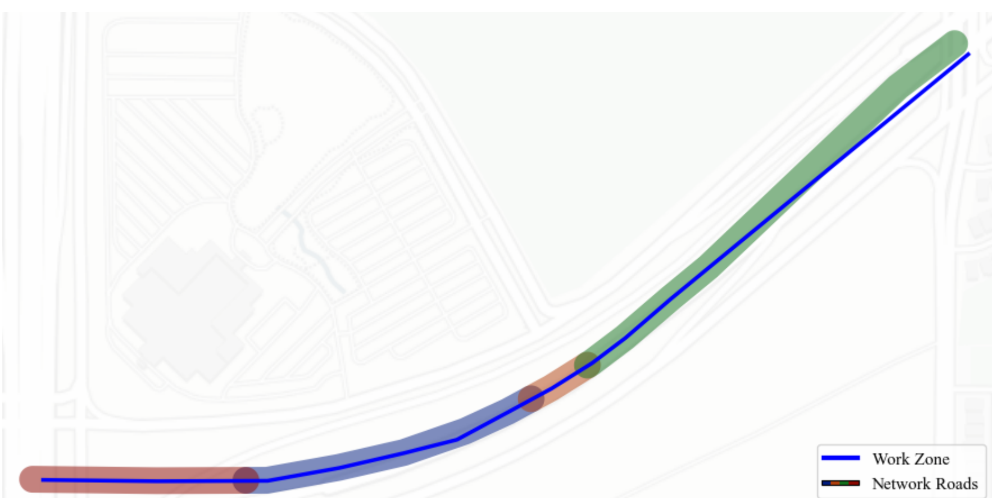

Figure 1: An example of map matched and merged work zone.
Note: Work zone ID = {992A1A11-609D-40C2-834E-5F359B0D0A50}

We requested the crash data from the Crash Record Information System (CRIS) through the API provided by Cox et al. (2019). From this source, we collected information on all traffic accidents in the state during the study period (*i.e.*, 2022-2024). We then correlated the crashes with the work zones based on their spatial and temporal information. In Figure 2, we presented the processed work zones and crashes on the road network.

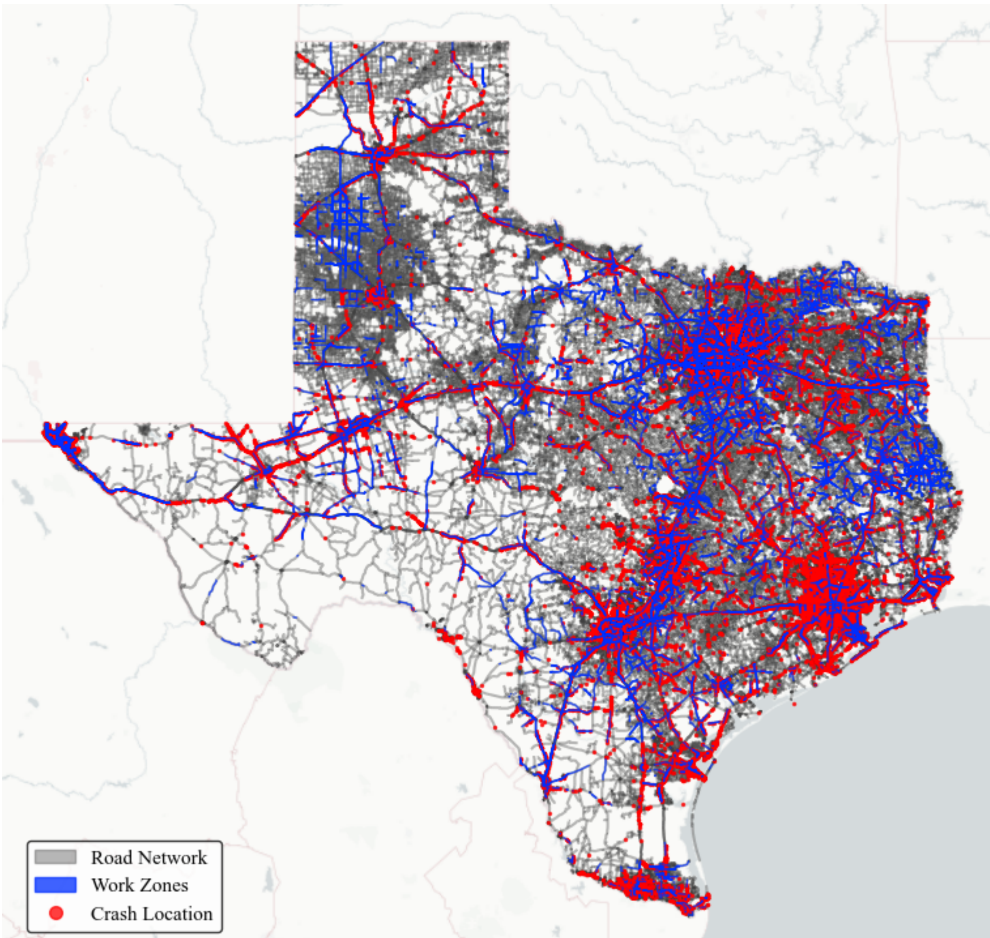

Figure 2: Distribution of work zones and crash events across TxDOT road network.

We used two additional data sources that provide national coverage, enabling us to complement the observable characteristics of the work zones. First, we collected the climatological characteristics from the National Centers for Environmental Information (NCEI). This agency collects information from approximately 1,000 stations throughout the country, where information is available with a granularity of up to one hour.

Second, we requested the speed data from INRIX (https://inrix.com/), which covers more than 70,000 roads in Texas. Given that speed is highly variable due to road conditions, this study seeks to capture as much detail as possible while balancing processing costs and consistency with other sources of information. We transformed the speed data by aggregating and averaging it over a 30-minute range. In Table 1, we presented all variable descriptions, sources, and periods. We tested the correlations among the variables (Figure 5 in the appendix) and removed the variables with their correlation greater than 0.6.

3.2 Research Design

To evaluate the effect that work zones have on the risk of crash, we used the RDD research design (Thistlethwaite and Campbell, 1960; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). We assumed that the trend of crash risk is consistent or follows a certain relationship during a short period when controlling for multiple types of variables, including road characteristics, weather, speed, *etc.* With that said, during several weeks before and after the presence of the work zone, the crash risk is consistent and only affected by the work zone. The advantage of RDD is that it does not need to search for comparable control groups, which is generally impossible in the real world (Zhang et al., 2022a; Tao and Qian, 2024).

With RDD, we will have a model capable of calculating the difference between the road segments that were treated - when had a working area - and those that were not. This difference is known as the average treatment effect (ATE) and is defined as Rubin (1974):

$$ATE = E[Y(1) - Y(0)] = E[Y \mid D = 1] - E[Y \mid D = 0]$$

By deriving this equation we get:

$$E[Y \mid m] = E[Y^0 \mid m] + E[Y^1 - Y^0 \mid m]$$

And, the RDD structure allows us to use a logistic model to capture ρ which acts as the ATE and the underlying function:

$$Y = [f_0(m) + \rho * D] + \beta * X$$

Where the probability of crash is represented by Y, the treatment D and X represent the matrix of control variables that are part of the project, and $f_0(\cdot)$ the functional form of the running variable m. We define $i \in I^n$ as the segments of the road and $t \in T$ as the time interval being studied.

In our specification, we include a number m of weeks before and after the work zone, allowing us to emulate normal road conditions when there is no construction zone. This variable m is, in turn, our running variable, so $m = 0$ for the week in which the work zone occurs, $m > 0$ for the weeks following and $m < 0$ for the weeks before the work on the route. It should be noted that for each week, the number of days equivalent to the duration of the work zone was taken (*that is,*, if the work zone lasted 1 day, for $m = 0$ there will be one observation, for $m = 1$ there will only be the same day of the previous week, *etc.*).

Using the granularity of our information, we divide the duration of the work zones into 30-minute intervals. These intervals enable us to ensure that minimal changes in factors, such as speed and weather variables, are not overestimated in the model. Thus, each work zone i within our study has information for m weeks, and for each day of the week, the information is separated into homogeneous intervals t (Figure 3).

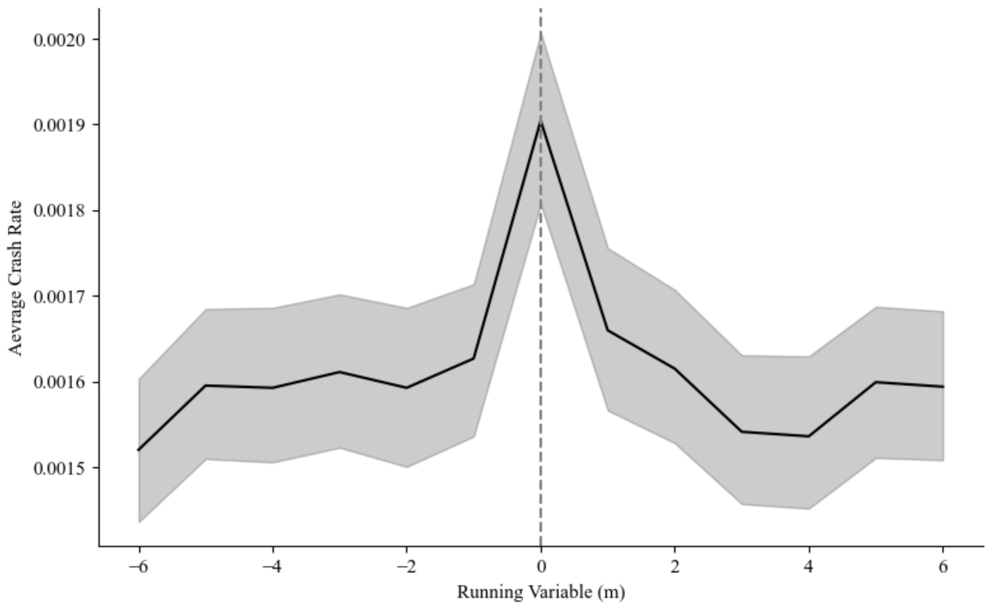

Figure 3: Average crash rate trends. *Note:* The gray shaded are represent the 95% confidence interval for the weekly average crash rate of the control and treatment weeks.

Regarding local randomization, as defined by Zhang et al. (2022b), this could pose a problem, since this assumption cannot guarantee that the speed is independent of the treatment. In particular, it is expected that the average speed could decrease in the period following road repair. This is why, instead of directly using the speed (S_{it}), we use a sequential estimate that incorporates the fraction of the speed not explained by the treatment D_{it}. This is seen as:

$$\hat{S}_{it} = \hat{\alpha}_0 + \hat{\alpha}_1 D_{it}$$
$$S'_{it} = S_{it} - \hat{S}_{it}$$

The last term (S'_{it}) is the actual speed processed, which represents the error between the speed estimated by the treatment and the actual values. This feature is intended to serve as a proxy for the fraction of variance in speed that treatment cannot explain, but that can still affect the risk of crashes.

3.3 Robustness Checks

To verify the robustness of the results, we performed several checks. First, we checked the impact of various bandwidths on the estimates. Bandwidth indicates the number of weeks before and after the work zones included in the analysis. Second, we applied different forms of the underlying functions, including linear and quadratic functions. Finally, we tested different temporal units when estimating the models.

4 Results

In this section, we presented the model estimate when the bandwidth is six weeks, *i.e.*, six weeks before the work zone and six weeks after the work zone were included in the model estimate. In addition, the underlying function is linear and the temporal analysis unit is 30 minutes. Table 2 showed the descriptive statistics of the variables in this setting. After illustrating the related results, we presented the robustness checks based on different bandwidths, underlying functions, and temporal analysis units.

Table 2: Descriptive Statistics

Variable name	N	Mean	St. Dev	Min	Pctl(25)	Pctl(75)	Max	Units
Crash	9,877,296	0.002	0.040	0	0	0	1.000	-
Treatment	9,877,296	0.077	0.266	0	0	0	1.000	-
Week	9,877,296	0	3.742	-6.000	-3.000	3.000	6.000	number of weeks
$log(duration)$	9,877,296	12.481	0.631	11.367	12.060	12.976	13.313	seconds; log transformed
$log(length)$	9,877,296	8.591	0.751	7.602	8.008	9.039	12.066	meters; log transformed
Weekday of week	9,877,296	0.897	0.304	0	1.000	1.000	1.000	-
Daytime of day	9,877,296	0.214	0.410	0	0	0	1.000	-
$log(AADT)$	9,877,296	10.216	1.721	2.398	9.221	11.557	12.409	number of cars, log transformed
$log(intersections)$	9,877,296	2.571	0.991	0	1.946	3.258	6.172	number of intersections, log transformed
Number of lanes	9,877,296	2.901	1.099	1.000	2.000	4.000	9.000	number of lanes
Speed limit	9,877,296	66.566	8.006	20	60	75.000	85.000	miles per hour
Hourly wind speed	9,877,296	3.446	2.375	0	2.050	4.770	362.400	meters per second
Hourly temperature	9,877,296	22.309	9.333	-95.000	16.500	28.900	396.000	degrees celsius
Hourly precipitation	9,877,296	0.541	2.176	0	0	0.200	79.800	millimeters
Average speed processed	9,877,296	0	14.949	-40.342	-12.102	11.028	36.908	miles per hour

Note: “N” number of observations, “St. Dev” standard deviation, “Pctl(25)” 25th percentile, “Pctl(75)” 75th percentile; ”—” means that the variable is a binary indicator, and it has no unit.

4.1 Overall and heterogeneous effects

Table 3 showed the full model which estimated the overall impact of work zone on crash risk. The table presented the estimated coefficients, standard errors (in parentheses), and their significance levels. The estimate of the work zone impact is positive with a coefficient of 0.195 and statistically significant. The result suggested that the presence of work zone increases the crash risk on the road networks in Texas.

In addition, Table 3 also presented the heterogeneous effects of work zones on crash risk conditional on AADT. Each model is estimated with the samples selected on different conditions of AADT. According to the results, when AADT is smaller than 11,580, the estimate of work zone impact is 1.064, which is much larger than the other two conditions. When AADT is between 11,580 and 93,060, the estimate is only 0.156, and when AADT is larger than 93,060, the estimate is 0.207. The reason for this might be that, as there are fewer vehicles nearby, drivers pay less attention to their surrounding traffic environments and increase the risk of collisions.

Table 3: The causal effect of the presence of a work zone on crash occurrence on roadways classified by AADT.

	Work zones classified by AADT			
	Crash occurrence			
	AADT = all	$0 < \text{AADT} \leq 11{,}580$	$11{,}580 < \text{AADT} \leq 93{,}060$	AADT > 93,060
	(1)	(2)	(3)	(4)
Work zone presence	0.195***	1.064***	0.156***	0.207***
	(0.028)	(0.179)	(0.041)	(0.038)
Week	0.002	−0.004	−0.004	0.006**
	(0.002)	(0.020)	(0.003)	(0.003)
Duration (second; log)	0.148***	0.687***	0.059***	0.319***
	(0.016)	(0.122)	(0.021)	(0.025)
Length (meter; log)	0.262***	0.676***	0.275***	0.330***
	(0.014)	(0.114)	(0.021)	(0.022)
Weekday of week	−0.215***	1.079***	−0.017	−0.330***
	(0.022)	(0.388)	(0.036)	(0.028)
Daytime of day	1.092***	0.878***	1.068***	1.126***
	(0.016)	(0.144)	(0.023)	(0.022)
AADT (vehicles per day; log)	0.935***	0.740***	1.410***	−0.244***
	(0.016)	(0.108)	(0.031)	(0.060)
Number of intersections (log)	0.195***	0.067	0.197***	0.174***
	(0.013)	(0.114)	(0.019)	(0.019)
Lane counts = 1	−0.343***	0.397***	−0.461***	−0.065***
	(0.010)	(0.129)	(0.017)	(0.015)
Speed limit	0.020***	−0.002	−0.015***	0.078***
	(0.002)	(0.009)	(0.002)	(0.003)
Average wind speed (mph)	−0.052***	0.020	−0.050***	−0.059***
	(0.004)	(0.030)	(0.005)	(0.005)
Average temperature (F)	−0.028***	−0.048***	−0.034***	−0.023***
	(0.001)	(0.011)	(0.002)	(0.002)
Average precipitation (inch)	0.033***	−0.176**	0.044***	0.023***
	(0.003)	(0.081)	(0.004)	(0.007)
Actual speed processed (mph)	−0.012***	−0.003	−0.003***	−0.014***
	(0.001)	(0.008)	(0.001)	(0.001)
Constant	−21.438***	−32.679***	−22.940***	−15.226***
	(0.276)	(2.108)	(0.447)	(0.743)
Monthly dummies	Y	Y	Y	Y
Yearly dummies	Y	Y	Y	Y
Observations	9,877,296	4,402,320	2,838,576	
R^2	0.083	0.076	0.061	0.046
χ^2	19,611.833*** (df = 27)	333.877*** (df = 27)	6,769.310*** (df = 27)	5,239.429*** (df = 27)

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01; "—" means the variable is deleted due to collinearity; "Y" means the corresponding dummies are controlled in the regression.

Figure 4 showed the heterogeneous work zone impacts conditional on the other three important variables, including time of day, speed, and work zone length. When the segment is crossed with the horizontal dash line of zero, it indicates the estimate is significant.

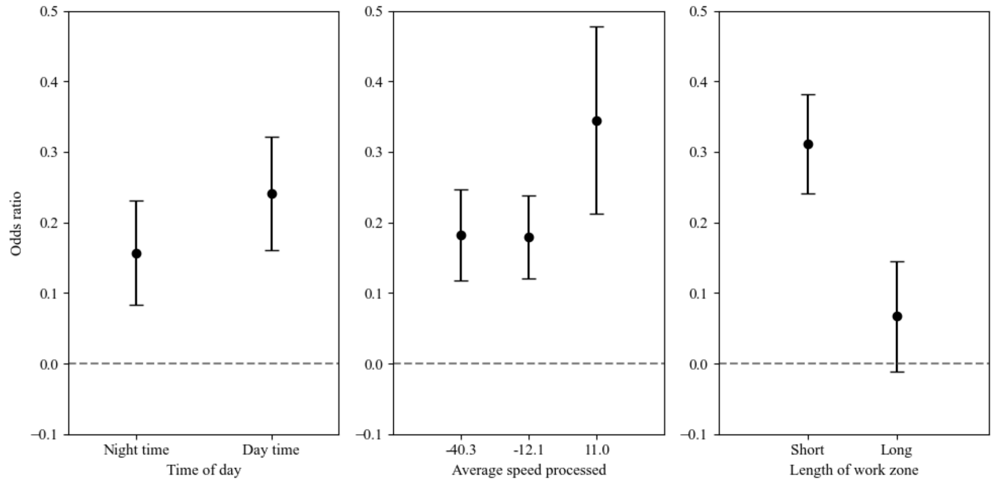

Figure 4: Estimated impact of the work zone presence by feature. *Note:* The dot represents the estimated value, and the bars indicate the 95% confidence intervals. The results correspond to the causal effect calculations presented in the Appendix.

As to time of day, we tested the work zone impacts during the day time and night time. The corresponding table is listed in Table 7 in the appendix. Both two estimates are both positive and significant. Although the estimate during the day time is slightly higher than that during the night time, the two estimates are similar to each other. In addition, we tested three categories of average speed processed. The detailed result is shown in Table 8 in the appendix. When the processed average speed is larger than 11 mph, there is a significant increase of work zone crash risk. Furthermore, we divided the work zones into two categories based on their length: those shorter and those long than than 5,400 meters. The detailed result is in Table 9 of the appendix. Only the estimate of the short work zones show significant result and the value is positive. Short work zones have limited space to put sufficient warnings in advance, which potentially lead to higher risk.

4.2 Robustness check results

We did several test of the robustness of the overall work zone impact on crash risk. First, we tested various bandwidths from one week to ten weeks before and after the work zone. The detailed result is listed in Table 4. All estimates are positive and significant across the models.

Table 4: The causal effect of the presence of a work zone on crash occurrence for different bandwidths.

	Work zones classified by different bandwidths									
	Crash Occurrence									
	M = 1	M = 2	M = 3	M = 4	M = 5	M = 6	M = 7	M = 8	M = 9	M = 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Work zone presence	0.108***	0.136***	0.146***	0.165***	0.170***	0.194***	0.195***	0.196***	0.197***	0.195***
	(0.024)	(0.024)	(0.025)	(0.025)	(0.026)	(0.027)	(0.028)	(0.028)	(0.029)	(0.030)
Week	−0.018	−0.00002	0.001	0.002	0.002	0.002	0.004**	0.006***	0.004***	0.004***
	(0.014)	(0.007)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Duration (second; log)	0.155***	0.159***	0.167***	0.179***	0.161***	0.146***	0.143***	0.143***	0.153***	0.174***
	(0.023)	(0.020)	(0.018)	(0.016)	(0.015)	(0.015)	(0.014)	(0.014)	(0.013)	(0.013)
Length (meter; log)	0.194***	0.193***	0.219***	0.224***	0.246***	0.241***	0.242***	0.260***	0.249***	0.268***
	(0.021)	(0.018)	(0.016)	(0.015)	(0.014)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
Weekday of week	−0.125***	−0.152***	−0.186***	−0.191***	−0.228***	−0.219***	−0.234***	−0.253***	−0.266***	−0.241***
	(0.032)	(0.028)	(0.025)	(0.023)	(0.022)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
Daytime of day	1.097***	1.095***	1.099***	1.094***	1.107***	1.088***	1.086***	1.080***	1.076***	1.076***
	(0.023)	(0.020)	(0.018)	(0.017)	(0.016)	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)
AADT (vehicles per day; log)	0.786***	0.801***	0.804***	0.808***	0.824***	0.820***	0.810***	0.811***	0.801***	0.792***
	(0.020)	(0.017)	(0.015)	(0.014)	(0.013)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
Number of intersections (log)	0.155***	0.174***	0.188***	0.209***	0.192***	0.209***	0.187***	0.192***	0.197***	0.186***
	(0.018)	(0.015)	(0.014)	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
Lane counts = 1	−0.295***	−0.290***	−0.296***	−0.304***	−0.317***	−0.309***	−0.303***	−0.295***	−0.303***	−0.284***
	(0.013)	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Speed limit	0.016***	0.018***	0.018***	0.020***	0.020***	0.019***	0.018***	0.017***	0.017***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Average wind speed (mph)	−0.042***	−0.051***	−0.056***	−0.056***	−0.055***	−0.051***	−0.053***	−0.057***	−0.055***	−0.054***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Average temperature (F)	−0.033***	−0.027***	−0.027***	−0.027***	−0.028***	−0.028***	−0.028***	−0.026***	−0.027***	−0.026***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Average precipitation (inch)	0.025***	0.020***	0.024***	0.028***	0.032***	0.033***	0.031***	0.035***	0.035***	0.034***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Actual speed processed (mph)	−0.010***	−0.011***	−0.012***	−0.012***	−0.012***	−0.011***	−0.010***	−0.011***	−0.012***	−0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	−19.005***	−19.382***	−19.804***	−20.113***	−20.165***	−19.988***	−19.615***	−19.760***	−19.635***	−19.920***
	(0.377)	(0.323)	(0.288)	(0.265)	(0.248)	(0.237)	(0.228)	(0.220)	(0.213)	(0.207)
Monthly dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Yearly dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3,691,728	5,175,360	6,512,016	7,733,232	8,928,480	9,877,296	10,784,880	11,591,280	12,262,752	12,959,856
R^2	0.070	0.074	0.078	0.081	0.082	0.083	0.081	0.083	0.084	0.085
χ^2	7,357.655*** (df = 27)	10,169.470*** (df = 27)	13,218.540*** (df = 27)	16,006.610*** (df = 27)	18,282.460*** (df = 27)	19,611.830*** (df = 27)	20,356.900*** (df = 27)	22,089.840*** (df = 27)	23,404.670*** (df = 27)	24,632.010*** (df = 27)

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01; "—" means the variable is deleted due to collinearity; "Y" means the corresponding dummies are controlled in the regression.

Second, we tested different forms of underlying functions, including linear and quadratic. The detailed result is shown in Table 5. For different underlying functions, the models show similar fitness to the data as their AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are almost the same. All estimates are significant and positive in these models.

Table 5: Functional form sensitivity test

Bandwith	AIC			BIC			ATE (ρ)	
	Linear	Quadratic	Delta value	Linear	Quadratic	Delta value	Linear	Quadratic
2	127,986.0	127,985.2	0.8	128,376.3	128,389.0	-12.7	0.136***(0.024)	0.104***(0.031)
3	156,578.0	156,578.7	-0.7	156,975.0	156,989.4	-14.4	0.146***(0.025)	0.130***(0.028)
4	182,338.5	182,339.5	-1.0	182,740.5	182,755.3	-14.8	0.165***(0.025)	0.153***(0.028)
5	204,256.5	204,257.8	-1.3	2046,62.7	204,677.9	-15.2	0.170***(0.026)	0.161***(0.028)
6	217,337.9	217,338.5	-0.6	217,747.0	217,761.6	-14.6	0.194***(0.027)	0.182***(0.028)
7	229,587.0	229,588.5	-1.5	229,998.6	230,014.3	-15.7	0.195***(0.028)	0.188***(0.029)
8	246.005.2	246,006.7	-1.5	245,342.6	245,342.1	0.5	0.196***(0.028)	0.189***(0.030)
9	256,758.5	256,759.2	-0.9	255,988.6	255,987.3	1.3	0.197***(0.029)	0.187***(0.030)
10	268,010.0	268,008.0	2.0	267,156.5	267,152.5	4.0	0.195***(0.030)	0.178***(0.031)

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01. ATE (ρ) is the effect of the presence of a work zone on the crash risk.

Finally, we tested different temporal analysis units in addition to 30 minutes, ranging from 60 to 240 minutes (Table 6). According to the results, the estimate is positive and significant across all analysis units.

Table 6: The causal effect of the presence of a work zone on crash occurrence for different time intervals.

	Work zones classified by different time intervals			
	Crash occurrence			
	60 minutes (1)	120 minutes (2)	180 minutes (3)	240 minutes (4)
Work zone presence	0.220***	0.205***	0.230***	0.295***
	(0.034)	(0.048)	(0.059)	(0.066)
Week	0.002	0.007*	0.009*	0.006
	(0.003)	(0.004)	(0.005)	(0.005)
Duration (second; log)	0.178***	0.162***	0.211***	0.144***
	(0.019)	(0.027)	(0.034)	(0.038)
Length (meter; log)	0.223***	0.233***	0.262***	0.238***
	(0.017)	(0.024)	(0.030)	(0.035)
Weekday of week	−0.190***	−0.094**	−0.146***	−0.074
	(0.027)	(0.039)	(0.047)	(0.056)
Daytime of day	1.118***	1.061***	1.143***	1.100***
	(0.020)	(0.028)	(0.035)	(0.040)
AADT (vehicles per day; log)	0.780***	0.847***	0.776***	0.815***
	(0.016)	(0.023)	(0.028)	(0.033)
Number of intersections (log)	0.196***	0.188***	0.141***	0.267***
	(0.015)	(0.021)	(0.026)	(0.031)
Lane counts = 1	−0.283***	−0.317***	−0.248***	−0.322***
	(0.011)	(0.016)	(0.020)	(0.023)
Speed limit	0.018***	0.019***	0.015***	0.017***
	(0.002)	(0.003)	(0.003)	(0.004)
Average wind speed (mph)	−0.063***	−0.063***	−0.072***	−0.055***
	(0.004)	(0.006)	(0.008)	(0.009)
Average temperature (F)	−0.010***	−0.009***	−0.010***	0.003
	(0.001)	(0.002)	(0.002)	(0.002)
Average precipitation (inch)	0.028***	0.023***	0.032***	0.034***
	(0.005)	(0.007)	(0.008)	(0.008)
Actual speed processed (mph)	−0.013***	−0.011***	−0.012***	−0.014***
	(0.001)	(0.001)	(0.001)	(0.002)
workzone_month	0.003	0.006	0.009	0.014**
	(0.003)	(0.005)	(0.006)	(0.007)
workzone_year	−0.085***	−0.115***	−0.099***	−0.148***
	(0.013)	(0.018)	(0.022)	(0.025)
Constant	152.727***	212.591***	180.262***	279.338***
	(26.027)	(36.158)	(44.939)	(51.539)
Monthly dummies	Y	Y	Y	Y
Yearly dummies	Y	Y	Y	Y
Observations	5,443,776	2,730,000	1,820,832	1,366,482
R^2	0.078	0.078	0.076	0.083
χ^2	10,951.160*** (df = 16)	5.675.625*** (df = 16)	3,592.933*** (df = 16)	2,947.224*** (df = 16)

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01; "—" means the variable is deleted due to collinearity; "Y" means the corresponding dummies are controlled in the regression; ASP refers to the Average Speed Processed variable.

5 Conclusions and implications

In this research project, we examined the causal work zone impacts on crash risk with multiple data sources from the state of Texas. We applied the RDD research design to comprehensively estimate the overall causal impact and heterogeneous impacts conditional on several variables, including AADT, time of day, speed, and work zone length. We carried out several additional checks of the robustness of the results.

The results indicated that the presence of work zones significantly increases crash risk, with the estimated coefficient being statistically significant. Robustness checks across various temporal analysis units, functional forms, and bandwidths confirmed the consistency of this finding, reinforcing the conclusion that work zones can lead to increased crash risk.

Further analysis of heterogeneous effects reveals that the influence of work zones on crash risk varies with AADT , vehicle speed, and work zone length. Specifically, AADT has a greater impact when it is below 11,580, likely because drivers are less vigilant in low-traffic conditions and therefore pay less attention to their surroundings. Additionally, higher speeds—particularly when the average processed speed exceeds 11 mph—are associated with increased crash risk, as elevated speeds raise both the likelihood and severity of collisions. Moreover, crash risk is higher when work zones are shorter than 5,400 meters, possibly due to limited space for adequate signage and guidance, which reduces drivers' preparedness.

The findings offer several important implications for improving work zone safety. First, safety measures should be strengthened on low-AADT roads. Installing additional signage, advance warning systems, dynamic message signs, or rumble strips can help attract drivers' attention and enhance awareness in these environments. Second, speed management within work zones should be improved. Strategies such as temporary speed limit reductions, automated speed enforcement, and increased police presence can effectively regulate travel speeds and reduce crash risk. Finally, the design of short work zones should be optimized. For instance, extending the advance warning area—even when the work zone itself is brief—can improve driver preparedness and safety.

6 References

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*(6), 309–317. https://doi.org/10.1037/h0044319

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Garber, N. J., & Zhao, M. (2002). Distribution and Characteristics of Crashes at Different Work Zone Locations in Virginia [Publisher: SAGE Publications Inc]. *Transportation Research Record*, *1794*(1), 19–25. https://doi.org/10.3141/1794-03

Khattak, A. J., Khattak, A. J., & Council, F. M. (2002). Effects of work zone presence on injury and non-injury crashes. *Accident Analysis & Prevention*, *34*(1), 19–29. https://doi.org/10.1016/S0001-4575(00)00099-3

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Li, Y., & Bai, Y. (2009). Highway Work Zone Risk Factors and Their Impact on Crash Severity [Publisher: American Society of Civil Engineers]. *Journal of Transportation Engineering*, *135*(10), 694–701. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000055

Carney, K. (2010). Road Miles by State–A chart of states and number of road miles — Cubit's Blog. Retrieved July 25, 2025, from https://blog.cubitplanning.com/2010/02/road-miles-by-state/

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

Akepati, S. R., & Dissanayake, S. (2012). Characteristics of the Work Zone Crashes [Publisher: American Society of Civil Engineers]. *Integrated Transportation and Development for a Better Tomorrow*, 1286–1295. https://doi.org/10.1061/41167(398)122

Yang, H., Ozbay, K., Ozturk, O., & Xie, K. (2015). Work Zone Safety Analysis and Modeling: A State-of-the-Art Review [Publisher: Informa UK Limited]. *Traffic Injury Prevention*, *16*(4), 387–396. https://doi.org/10.1080/15389588.2014.948615

Weng, J., Zhu, J.-Z., Yan, X., & Liu, Z. (2016). Investigation of work zone crash casualty patterns using association rules. *Accident Analysis & Prevention*, *92*, 43–52. https://doi.org/10.1016/j.aap.2016.03.017

Cox, L., Fields, K., & Castillo, J. (2019). *Frequently asked questions for cris query* (tech. rep.). Texas Department of Transportation. https://ftp.txdot.gov/pub/txdot-info/trf/crash_statistics/query-faq.pdf

Ghasemzadeh, A., & Ahmed, M. M. (2019). Exploring factors contributing to injury severity at work zones considering adverse weather conditions. *IATSS Research*, *43*(3), 131–138. https://doi.org/10.1016/j.iatssr.2018.11.002

Zhang, Z., Akinci, B., & Qian, S. (2022a). Inferring heterogeneous treatment effects of work zones on crashes. *Analytic Methods in Accident Research*, *34*, 100230. https://doi.org/10.1016/j.amar.2022.100230

Zhang, Z., Akinci, B., & Qian, S. (2022b). Inferring the causal effect of work zones on crashes: Methodology and a case study. *Accident Analysis & Prevention*, *190*, 107235. https://doi.org/10.1016/j.aap.2023.107235

FHWA. (2024). FHWA Work Zone Facts and Statistics - FHWA Office of Operations. Retrieved July 25, 2025, from https://ops.fhwa.dot.gov/wz/resources/facts_stats.htm

Tao, T., & Qian, S. (2024). Do Smart Loading Zones help reduce traffic congestion? A causal analysis in Pittsburgh. *Transportation Research Part E: Logistics and Transportation Review*, *192*, 103796. https://doi.org/10.1016/j.tre.2024.103796

ARTBA. (2025). National & State Traffic Data. Retrieved July 25, 2025, from https://workzonesafety.org/work-zone-data/work-zone-fatal-crashes-and-fatalities/

7 Appendix

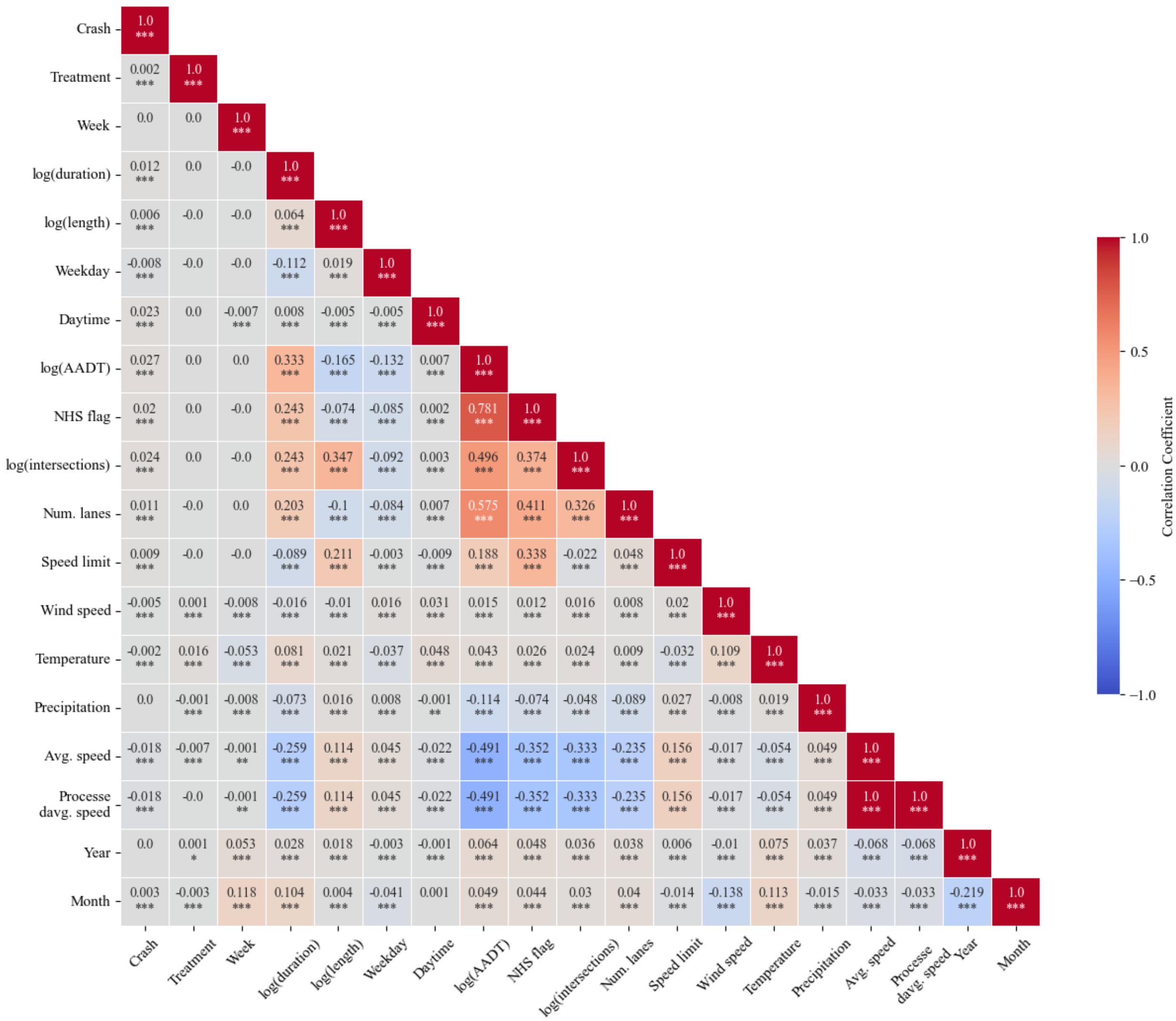

Figure 5: Correlation between the model variables. *Note:* Significance level t-test results are included bellow the correlation: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Red color represents a high correlation of 1 or -1 and gray shows a correlation of 0.

Table 7: The causal effect of the presence of a work zone on crash occurrence on roadways classified by work zone daytime of day.

	Work zones classified by daytime	
	Crash occurrence	
	Night time	Day time
	(1)	(2)
Work zone presence	0.157***	0.241***
	(0.038)	(0.041)
Week	0.0002	0.004
	(0.003)	(0.003)
Duration (second; log)	0.125***	0.182***
	(0.021)	(0.024)
Length (meter; log)	0.273***	0.243***
	(0.019)	(0.022)
Weekday of week	−0.327***	−0.074**
	(0.028)	(0.034)
AADT (vehicles per day; log)	0.962***	0.899***
	(0.021)	(0.025)
Number of intersections (log)	0.163***	0.241***
	(0.017)	(0.020)
Lane counts = 1	−0.371***	−0.309***
	(0.013)	(0.014)
Speed limit	0.017***	0.025***
	(0.002)	(0.002)
Average wind speed (mph)	−0.055***	−0.035***
	(0.005)	(0.006)
Average temperature (F)	−0.049***	0.007***
	(0.002)	(0.002)
Average precipitation (inch)	0.030***	0.038***
	(0.005)	(0.005)
Actual speed processed (mph)	−0.007***	−0.018***
	(0.001)	(0.001)
Constant	−20.911***	−21.299***
	(0.366)	(0.420)
Monthly dummies	Y	Y
Yearly dummies	Y	Y
Observations	7,768,234	2,109,062
R^2	0.065	0.073
χ^2	8.965.278*** (df = 27)	6.970.426*** (df = 27)

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01; "—" means the variable is deleted due to collinearity; "Y" means the corresponding dummies are controlled in the regression.

Table 8: The causal effect of the presence of a work zone on crash occurrence on roadways classified by speed range.

	Work zones classified by actual speed processed		
	Crash occurrence		
	-40.3 < ASP ≤ -12.1 (1)	-12.1 < ASP ≤ 11.0 (2)	11.0 < ASP ≤ 36.9 (3)
Work zone presence	0.182***	0.179***	0.344***
	(0.033)	(0.030)	(0.068)
Week	0.005*	0.002	0.002
	(0.003)	(0.002)	(0.006)
Duration (second; log)	0.214***	0.121***	−0.062*
	(0.020)	(0.017)	(0.038)
Length (meter; log)	0.240***	0.239***	0.465***
	(0.019)	(0.015)	(0.031)
Weekday of week	−0.206***	−0.216***	−0.470***
	(0.026)	(0.023)	(0.054)
Daytime of day	1.120***	1.095***	1.058***
	(0.019)	(0.017)	(0.042)
AADT (vehicles per day; log)	0.858***	0.875***	1.111***
	(0.022)	(0.017)	(0.032)
Number of intersections (log)	0.337***	0.184***	0.008
	(0.016)	(0.014)	(0.026)
Lane counts = 1	−0.310***	−0.288***	−0.453***
	(0.012)	(0.010)	(0.026)
Speed limit	0.033***	0.022***	0.008**
	(0.002)	(0.002)	(0.004)
Average wind speed (mph)	−0.050***	−0.053***	−0.010
	(0.004)	(0.004)	(0.010)
Average temperature (F)	−0.029***	−0.026***	−0.037***
	(0.002)	(0.002)	(0.004)
Average precipitation (inch)	0.025***	0.035***	0.038***
	(0.004)	(0.003)	(0.007)
Actual speed processed (mph)	0.004***	−0.013***	−0.038***
	(0.001)	(0.001)	(0.002)
Constant	−22.280***	−20.415***	−20.325***
	(0.370)	(0.291)	(0.639)
Monthly dummies	Y	Y	Y
Yearly dummies	Y	Y	Y
Observations	4,625,712	7,882,992	3,577,392
R^2	0.061	0.080	0.121
χ^2	9,739.891*** (df = 27)	16.246.916*** (df = 27)	4,709.204*** (df = 27)

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01; "—" means the variable is deleted due to collinearity; "Y" means the corresponding dummies are controlled in the regression; ASP refers to the Average Speed Processed variable.

Table 9: The causal effect of the presence of a work zone on crash occurrence on roadways classified by work zone length.

	Work zones classified by workzone length	
	Crash occurrence	
	Length $\leq$ 5,400	Length $>$ 5,400
	(1)	(2)
Work zone presence	0.313***	0.067
	(0.037)	(0.041)
Week	0.004	−0.0002
	(0.003)	(0.003)
Duration (second; log)	0.203***	0.073***
	(0.022)	(0.022)
Length (meter; log)	0.453***	−0.018
	(0.042)	(0.027)
Weekday of week	−0.140***	−0.263***
	(0.031)	(0.031)
Daytime of day	1.117***	1.067***
	(0.023)	(0.023)
AADT (vehicles per day; log)	0.752***	1.121***
	(0.022)	(0.025)
Number of intersections (log)	0.166***	0.282***
	(0.017)	(0.021)
Lane counts = 1	−0.179***	−0.543***
	(0.013)	(0.015)
Speed limit	0.019***	0.028***
	(0.002)	(0.002)
Average wind speed (mph)	−0.045***	−0.062***
	(0.005)	(0.005)
Average temperature (F)	−0.034***	−0.022***
	(0.002)	(0.002)
Average precipitation (inch)	0.019***	0.044***
	(0.005)	(0.005)
Actual speed processed (mph)	−0.011***	−0.010***
	(0.001)	(0.001)
Constant	−21.904***	−20.306***
	(0.470)	(0.447)
Monthly dummies	Y	Y
Yearly dummies	Y	Y
Observations	5,780,112	4,097,184
R^2	0.067	0.104
χ^2	8,051.419*** (df = 27)	12,027.935*** (df = 27)

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01; ”—” means the variable is deleted due to collinearity; ”Y” means the corresponding dummies are controlled in the regression.