#7 - Wearable DSRC Devices for Workers

Final Research Report

PI: Burcu Akinci, Ph.D.

Sean Qian, Ph.D.

Depart of Civil and Environmental Engineering

Carnegie Mellon University

Disclaimer

The Problem

Deteriorating road network and increasing travel demands for transportation facilities bring increasing construction activities and consequently a large number of work zones (Yang et al., 2014). There are several safety issues associated with construction in work zones where active traffic exists nearby. According to Federal Highway Administration (FHWA), a work zone crash occurred once every 5.4 minutes in the U.S., and 96,626 accidents took place in work zones in 2015, an increase of 7.8% over 2014 (FHWA, 2017). It is important to investigate the relationships between work zones and crash frequency and severity (i.e. crash metrics) and understand what aspects of a work zone impact possibility and severity of crashes. Despite many research studies on work zone related crash analysis (Garber and Woo, 1990; Hall and Lorenz, 1989; Jin et al., 2008; O Ozturk et al., 2014; Theofilatos et al., 2017), a thorough investigation of possible relationships between work zone characteristics and crash severity and frequency is still lacking.

There are three main challenges in exploring crash causality of work zone. Difference on roadway characteristics will bring pre-selection bias on work zones' crash causality effect, and usually these roadway characteristics are hard to quantify and control (such as the sight range of the road segment) (Abdel-Aty and Pemmanaboina, 2006). Dynamic traffic flow can also affect the crash rates, so there is a need to control this factor when researching the work zone's crash causality using a high-temporal-resolution traffic flow data (Theofilatos and Yannis, 2014). Various work zone settings (like closure setting, total length, etc.) have different effects on work zone's crash causality, so it is necessary to consider different work zone settings separately (Theofilatos et al., 2017). Failing to consider these three aspects will result in biased estimation on work zone's crash causality, and the results to be obtained will lack robustness. Here, robustness refers to the estimated effect on work zones being significantly different from temporal neighboring events and spatial neighboring places. In this study, we discussed ways to address these challenges in the developed analysis framework. This framework is designed to quantify work zone crash causality in a rigorous way, and check the robustness of the causality effect.

The first challenge is the pre-selection bias brought by unobserved roadway characteristics. Road segments selected for work zones usually have different characteristics compared with other road segments. For example, if a road segment is deteriorated and lack of maintenance, it may induce more crashes, while it has a higher possibility to be selected for retrofitting than other well-maintained roads(Anastasopoulos and Mannering, 2009). Hence, we may observe more collisions during roadway maintenance on this type of road segments compared with other road segments, but this may be in part due to road characteristics, not necessarily work zones. These pre-selection bias may lead to estimation bias of work zone effects. To consider this effect, there is a need for a larger sample set of work zones. Through examining crash data within a relatively small work zone sample size (such as focus on only 60 work zones), it is not possible to identify whether crashes are caused by work zones or characteristics of the roadway specified by our sample size (Chen and Tarko, 2014; Ozgur Ozturk et al., 2014). In this research, we processed far

more work zone cases compared with the previous study (i.e. 10k vs. 60) and developed a Regression Discontinuity (RD) method to quantify the causality effect.

In addition to roadway characteristics, dynamic traffic flow can also affect work zone safety, such as traffic speed. Those factors may be coupled with the presence of work zones to induce crashes, which can be challenging to separate respective effects if crash data are not carefully examined. For example, roadway construction projects may perform during daytime, when the traffic going on the work zone segment is more than what it would be during night time. To consider this effect, high-temporal-resolution traffic flow data is needed to yield the requirement of "dynamic". If only provided aggregated speed information (such as daily traffic volume), it could be challenging to identify whether the crashes with the presence of a work zone is caused by the work zone or high traffic volume. Hence, actual time-varying observation speed is needed to perform this analysis. From the view of method, high-traffic-speed zones usually bring the unbalanced possibility of scheduled work zones (higher speed limit represents higher road class), and this high-traffic-speed also contributes to higher crash occurrence possibility (Theofilatos and Yannis, 2014). To exclude the effect of high-traffic speed when testing the effect of work zones on crash frequency or severity, it is necessary to limit the traffic speed into a homogeneous group so that the consequences of crash frequency or crash severity in this subgroup would get rid of the influence of traffic speed. By this way, we can exclude the confounding effect of traffic speed and get a pure causality effect of work zone on crash metrics.

Except for the stable road characteristics and the dynamic outside environment, the third part that could affect the work zone effect is the settings of the work zone. Previous researchers have done abundant exploration on building a statistical model of crash frequency or severity in the work zone (Theofilatos et al., 2017). However, only a few of them focused on comparing no-work-zone conditions and with-work-zone conditions. Ozturk et al. (2014) realized this research gap and compared the number of crashes between pre-work zone and during-work zone crashes. Their assumption was that the pre-work zone conditions should be similar to the during-work zone condition, and the crash numbers should be unchanged. We relaxed their assumption to continuous temporal trend of road conditions and crash metrics by using a time-series-observations and Regression Discontinuity (RD) analysis. Then, we discussed the crash causalities of different work zones under this analysis framework.

Our approach

Data Preprocessing As illustrated in the first section, one of the challenges that previous researchers did not fully incorporate is utilizing multiple-source, high-resolution databases to infer the crash causality of work zones. In this research, we used PennDOT open source roadway network (Pennshare, 2018) as the most-up-to-date roadway data. The crash data (crashes occurred and reported in Pennsylvania in 2013) is obtained from PennDOT(PennDOT, 2018). Crash occurrence time, severity (fatal/non-fatal) and the weather information when the crash occurred are extracted from this database. PennDOT provides Road Condition Reporting System

(RCRS) data to the public (Commonwealth Pennsylvania, 2018). Real-time information for traffic incidents, roadwork, and other events are provided in this database. We selected the road events, which is labeled as "roadwork", as work zone events. This database provided us the start time and end time of each incident, the position of the start point and end point, and the closure situation in each roadway activities. INRIX company released real-time monitored speed information for major roads in Pennsylvania (Kim and Coifman, 2014). The data is reported every 5 min on major roadways. They provide actual observed historical speed (real-time speed), historical mean speed and free flow speed. We use MultiNet to geocode the speed information to the base map, PennShare Roads. MultiNet is a highly accurate map provided by TomTom company (TomTom, n.d.). They marked the road segments with INRIX code, which produced high accuracy to geocode INRIX speed data onto real road networks. The MultiNet road segments with INRIX speed code are subsets of the PennShare Roads data. Also, the two road networks could be joined to match the speed data with work zones. A summary of utilized data sources is shown in Table 1, a map of the work zones and the roadway network is displayed in Figure 1.

Table 1 Data source

Code	Full name	Role	Provided information
PennShare Roads	PennDOT open source roadway network (Pennshare, 2018)	Base map	AADT, number of intersections ($Intersections_i$), number of ramps ($Ramps_i$), road class (RoC_i), and total number of lanes ($Lane_i$).
PennCrash	PennDOT Crash Data (PennDOT, 2018)	Crash source	crash occurrence time, severity (fatal/non-fatal) and the weather information when the crash occurred
RCRS	Road Condition Reporting System (Commonwealth Pennsylvania, 2018)	Work zones	start time and end time of each event, the position of start point and end point, and the closure situation in each roadway activities
INRIX	INRIX speed (Kim and Coifman, 2014)	Speed source	grounded-truth historical speed (real-time speed), historical mean speed and free flow speed
MultiNet	TomTom MultiNet (TomTom, n.d.)	Speed marker	Geocode INRIX speed data onto PennShare Roads
Milepost	NHS Milepost (Federal Highway Administration, 2017)	Validation	Used to validate work zone length

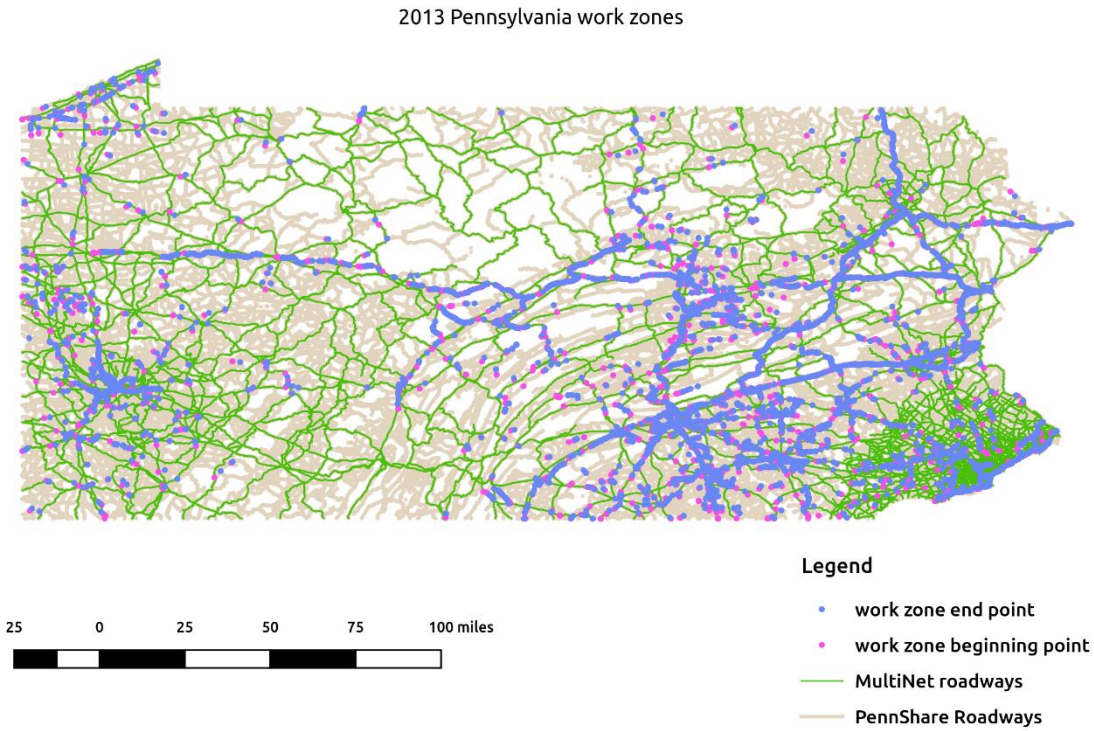

Figure 1 Map of work zones in PA, 2013

An overview of data processing utilized in this research is shown in Figure 2. The main steps are extracting dependent variables, explanatory variables, and built an observation panel. As J. Dan Turner (2017) pointed out, work zone related crashes includes accidents occurred at the approach or exit of a work zone. However, the crash database may have under-reporting crashes due to many factors, like less severe crashes may not be reported as work zone related (Lord and Mannering, 2010). To eliminate these pre-selective biases, we identified the crashes occurred related with work zone by utilizing spatial-temporal relationships between them. More specifically, through spatiotemporal reasoning, we identified crashes that occurred within a work zone and during a roadway project. The way to identify the upstream/downstream section is based on a tracing algorithm developed based on the NetworkX python package (https://networkx.github.io/). The roadway network is converted to a node-edge based abstract network. For the upstream section, we traced the predecessors of the node in the work zone. For the downstream section, we traced the successors of the node in the work zone. By using this node-edge based network, the accuracy is improved as the road vertices.

The explanatory variables include two parts; inventory controls, and dynamic controls. The inventory controls are variables that are related to work zone position, while the dynamic controls are related to observations at different times. The inventory controls include number of ramps, intersections, total lanes count, and road class. The dynamic controls include the historical

ground truth information about traffic. In the crash likelihood study, the work zone information is mainly extracted from base map, such as lane counts, road classes and lane closures. Number of intersections and ramps are geometrically calculated from base map. The number of intersections is calculated by number of intersected roads within a work zone. The number of ramps is calculated by the number of intersected roads whose road name contains "ramp". Speed information is extracted from INRIX speed database. In the crash level crash severity study, the crash database provides well-format control variables. Lane counts, road class, and weather information are provided at the PennCrash database. Speed information is similarly extracted from INRIX database.

The observations are obtained by the same time window between 6 weeks before the work zones to the date that roadway activities performed, and 6 weeks after the roadway activities. We divided the work zones into two groups; one is the short-duration work zones (shorter than $\Delta t = 0.5\ h$), the other is the long-duration work zones (longer than $\Delta t = 0.5\ h$). The long-duration work zones are divided as time slots with a duration equal to $\Delta t = 0.5\ h$. The last time slots may be shorter than $\Delta t = 0.5\ h$. Each time slot will be a new observation.

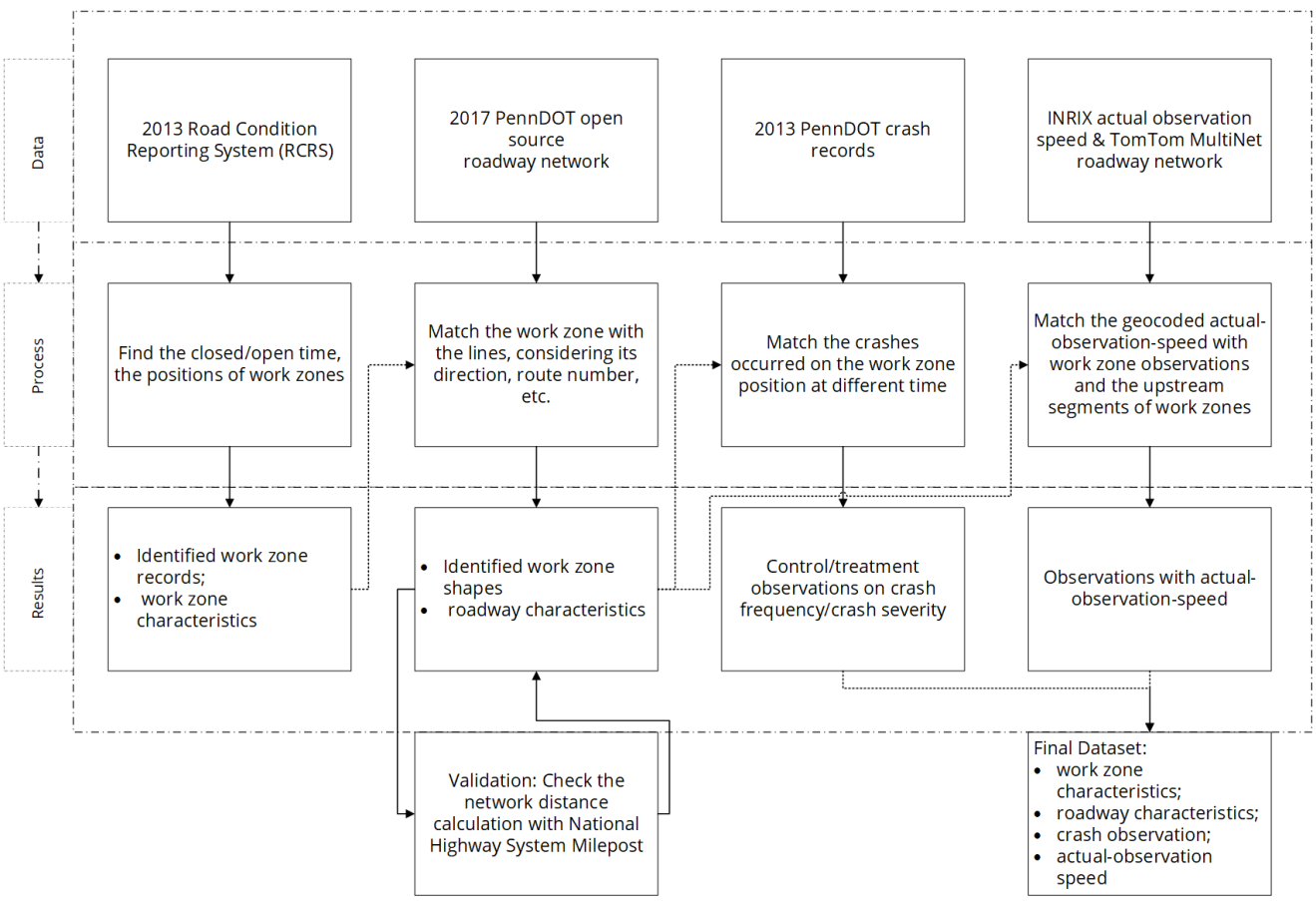

Figure 2 Data processing summary

Regression Discontinuity Design. The core analysis method in this study is Regression Discontinuity (RD) design. For the work zone related crash study, as far as the authors reviewed, none of the current research has tried using this matched-case control design. Regression Discontinuity is a conventional research design when investigating a temporal shock on time series data. In this study, we assumed the roadway characteristics before work zone and after work zone to be similar. Hence, the temporal trend of the number of crashes should change smoothly during the observation period. With this assumption, by extracting crashes occurred at the same location and including a time period six weeks before and six weeks after the roadway projects, the only time-varying "shock" is a roadway project during the work zone period. Thus, we created a counterfactual observation of "there were no roadway projects" on the work zone position. Comparing the counterfactual observation and the actual occurred crashes at the work zone position, we could conclude the causality of roadway projects on traffic safety.

Many social studies faced with the same problem as work zone crash studies, which is the fact that it is not possible to perform experiments with control of the independent variables. Hence, economists provided such methods like Regression discontinuity (RD) and Difference in Difference (DD) to perform the causal inference. Besides RD, Difference in Difference is another possible way to investigate the causality of the work zone on crash metrics. However, to perform DD, researchers need to find control road segments whether there is no roadway project performed while the experimental group is experiencing roadway project. Besides, the control group needs to have the same trend of all factors that affect crash metrics compared with the experimental group. The second assumption is too strong to satisfy in the work zone safety. If the control group is too far away from the experimental group in terms of spatial distance, we cannot make sure that they are experiencing the same trend of related factors. For example, if a control group and experimental group are located too close to each other, they will be affected by similar traffic flow. So, this method is not suitable for our work zone crash study. Looking back to the RD method, it assumes the temporal trend of the studied road segments is stable or is linear with time. This assumption is satisfied when the temporal trend is continuous. Our linear model serves as a Taylor approximation for any forms of the continuous trend. We do not need to consider special events (like big sports games), because we are using the crash sample that distributed all over the state and all over the year relatively randomly. Thus the effect of special events is diluted by random assignments.

Analysis under small temporal scale. The temporal scale of crash analysis determines the variable scale, the model choice, and the analysis method. In a long observation period (e.g. weekly), there tend to be more crashes, so people use crash counts to describe crash frequency (Lord and Mannering, 2010), while in a shorter period (e.g. hourly), crash likelihood (i.e., whether there is a crash occurrence) is more suitable for describing the crash frequency(Xu et al., 2013). In this paper, the crash likelihood is used to represent the crash frequency in a short observation period.

In a long observation period (e.g. weekly), crash counts are usually over-desperation (i.e., the variance is more significant than the mean of crash counts). When modeling the crash frequency

data, previous studies found common Poisson regression model may induce biased and inconsistent estimation because the over-desperation data violated the assumption of Poisson distribution(Lord and Mannering, 2010). Therefore, previous researchers modified and developed a series of model forms like Negative binomial(Srinivasan et al., 2011), Poisson-gamma(Lord, 2006), and Poisson-lognormal(Lord and Miranda-Moreno, 2008). However, these models suffered from problems like low sample-mean and small sample size, which reduce the model reliability(Lord and Mannering, 2010). In this paper, we divided the roadway project period into small time intervals to decay the crash counts over a more extended period into crash likelihood in a shorter period. With that, it is possible to model it as a classification problem, which can be solved by conventional classification algorithm like logistics regression without losing useful information. Then, the problem collapsed as a classification problem, thus will not suffer from the count variable problems.

When using the time-varying speed information, we noticed that the work zone treatment is highly related with speed on the roadway. This indicates an association between the work zone and high-speed. As we illustrated above, it is a confounding problem (Jager et al., 2008). Usually, researchers could use methods like randomization, restriction, matching, and stratification to solve the confounding problem. Here, since we cannot directly perform the experiment, and speed is not a category variable, we cannot use randomization or matching method. But, restriction or stratification is suitable for our study. Here, we restricted the high speed as a subgroup in our analysis. The criteria are illustrated in the finding section. By analyzing this subgroup, we made sure the estimated power is pure and robust.

Findings

Effects of work zone on crash likelihood. Using i to denote the list of positions, using t to denote the time of each observation, the regression model used in shown in Equation (1). The explained variable is the crash likelihood (LL) per time slot (1=crash occurrences; 0 = crash does not occur).

$$LLs_{i,t} \sim WZ_{i,t} + Control_{i,t} + ToD_i + ToW_i + RoC_i + Intersections_i + Ramps_i + LaneCounts_i + SoP_{i,t} + (Historical\ Mean\ Speed_{i,t}) + e_{i,t} \quad (1).$$

Where $WZ_{i,t}(treatment)$ denotes whether there is a roadway project performed at the studied position during the studied period. $Control_{i,t}$ represents time series indicator of each observation (). ToD_i denotes dummies of time of day (6AM to 6PM as day time, other time as night time). ToW_i denotes a dummy of weekday indicator ($ToW_i = 1$ when it is weekdays, $ToW_i = 0$ when it is weekends). $RoC_i(NHS\_IND\_Y)$ denotes road class ($RoC_i = 1$ when the road is National Highway System recognized roads, otherwise $RoC_i = 0$).

The results of work zone causality on crash likelihood are shown in Table 2. The first model is a univariate model only focuses on the treatment (roadway project) variable, reporting crash occurrence odds ratio with work zone is related to $exp(0.313) = 1.368$ times higher compared to crash occurrence odds ratio without work zone. The second model assumes that there exists

a linear time trend from the six weeks before the work zone to the six weeks after the work zone. The third model introduces some control variables as illustrated in the method section. The fourth model adds the historical mean speed into the model, which makes the coefficient of treatment change to insignificant. In the fifth and sixth model, we restricted the historical mean speed to larger than 53 km/h. The threshold is picked by checking when the effect of high speed is diluted by the work zone treatment. Note that this number is not exact and numbers around this threshold can also cause similar effects

Using this stratification method, we found a positive causality effect of work zone on crash likelihood. In the road segments whose historic mean speed is high, work zone has a positive effect on crash likelihood. The average crash occurrence odds ratio with work zone is related to $exp(0.544) = 1.723$ times higher compared with crash occurrence odds ratio without work zone.

Table 2 Effects of work zone on crash likelihood

VARIABLES	(1) No time changes	(2) Linear time changes	(3) Add controls	(4) Add speed reference	(5) Speed stratification	(6) Speed stratification & controls
treatment	0.313***	0.314***	0.365***	-0.159	1.764***	0.544***
	(0.0963)	(0.0964)	(0.0906)	(0.199)	(0.645)	(0.105)
Control		0.0219*	0.0214*	-0.00427	-0.197**	0.0293**
		(0.0114)	(0.0114)	(0.0216)	(0.0997)	(0.0133)
log_netlength			0.333***	0.424***	0.166	0.266***
			(0.0599)	(0.0988)	(0.759)	(0.0700)
LANE_COUNT_1			0.796***	-1.529		1.068***
			(0.175)	(0.977)		(0.193)
log_duration			-0.102*	-0.171	-0.632**	-0.112*
			(0.0602)	(0.143)	(0.280)	(0.0656)
log_numinters			0.435***	0.241**	-2.988***	0.482***
			(0.0624)	(0.122)	(0.723)	(0.0745)
log_numramps			-0.00566	0.148	4.311***	-0.0341
			(0.0606)	(0.121)	(0.934)	(0.0769)
DaytimeofDay			0.871***	1.036***		0.833***
			(0.108)	(0.201)		(0.128)
WeekdayofWeek			0.274*	0.957***	1.762	0.0363
			(0.163)	(0.228)	(1.238)	(0.205)
NHS_IND_Y			0.661***	0.756***		0.690***
			(0.118)	(0.262)		(0.138)
Sequence_0			0.283**	0.0341	-0.567	0.350**
			(0.130)	(0.218)	(0.766)	(0.156)
speeds_in_historical_mean				-0.0348***	-0.127	
				(0.0105)	(0.0867)	
Constant	-8.676***	-8.680***	-12.15***	-10.51***	1.616	-11.64***
	(0.0557)	(0.0560)	(0.763)	(1.718)	(4.893)	(0.827)
Observations	2,846,337	2,846,337	2,846,337	870,532	39,955	2,075,541

Speed range	>53	>53

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Effects of the work zone on crash severity. The model used in the crash severity analysis is similar to the model used in the crash likelihood analysis. The difference is that the dependent variable now is the likelihood of crashes that are labeled as "Fatal or major injury" in PennCrash database.

$$LL\ of\ Severe\ Crashes_{i,t} \sim WZ_{i,t} + Control_{i,t} + ToD_i + ToW_i + RoC_i + + Intersections_i + Ramps_i + LaneCounts_i + SoP_{i,t} + (Historical\ Mean\ Speed_{i,t}) + e_{i,t} \quad (2).$$

Using this model, we did not find significant effects ($p < 0.05$) of the work zone on severe crash likelihood. When we tried using the same speed stratification setting as in the crash likelihood analysis, the model failed to run because of singular problem. So, we concluded that the occurrence likelihood of severe crash during roadway projects did not change.

Table 3 Crash severity analysis: work zone level analysis

VARIABLES	(1) No time changes	(2) Linear time changes	(3) Add controls	(4) Add speed reference	(5) Speed stratification & controls
treatment	-1.241*	-1.241*	-1.282*	0.170	
	(0.748)	(0.747)	(0.750)	(0.918)	
Control		-0.0102	-0.00953	-0.122	0.0594
		(0.0649)	(0.0648)	(0.0882)	(0.0920)
log_netlength			0.199	0.549	-0.0754
			(0.326)	(0.543)	(0.305)
log_duration			-0.554***	-0.747***	-0.509***
			(0.0982)	(0.162)	(0.175)
log_numinters			-0.158	-1.547***	0.543*
			(0.344)	(0.469)	(0.296)
log_numramps			0.530	0.921**	0.471
			(0.428)	(0.451)	(0.503)
DaytimeofDay			-0.872*	-2.354**	-0.119
			(0.525)	(0.948)	(0.683)
NHS_IND_Y			0.160	-0.120	0.193
			(0.553)	(0.968)	(0.686)
Sequence_0			-1.113**	-1.627*	-0.981
			(0.484)	(0.832)	(0.608)
speeds_in_historical_mean				-0.128**	
				(0.0591)	
Constant	-11.69***	-11.69***	-6.549***	0.0446	-6.363***
	(0.242)	(0.242)	(2.286)	(4.078)	(2.126)
Observations	2,846,337	2,846,337	2,498,808	743,457	1,287,669
Speed range					>53

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Placebo tests. For the crash severity analysis, we performed a placebo test analyzing whether the severe possibilities are different between crashes occurred in a work zone or out of the work zone. The results indicated that there is no significant difference. This conclusion solidified our finding that work zone did not bring much difference on the crash severity.

For the crash likelihood analysis, we performed a temporal test analyzing whether the crash likelihood changed before and after the roadway project. The result indicates that there are no changes during the preceding and succeeding weeks. Another spatial test includes analyzing whether the crash likelihood changed near the work zone. The results indicated that there are no changes at the neighborhood area and that work zone area is temporally and spatially different on the crash likelihood compared with the spatial neighboring road segments and temporal neighboring observation windows.

Role of work zone characteristics. We also analyzed the effects of work zone settings based on Equation (1). We tested the interaction term and the stratum of different work zone settings. The log odds-ratio of crash occurrence likelihood during "shoulder closed" crash is 1.167 larger than no work zones, and 0.708 larger than other kinds of work zones. Closure situation "lane restriction" has no significant (p>0.05) smaller positive effect on crash likelihood. Besides, more extended work zone (i.e. > 600meter) has larger crash occurrence likelihood. The log odds ratio of crash occurrence likelihood could increase 0.452. The longer duration of the work zone, the less crash occurrence likelihood in each 0.5-hour observation. The log odds ratio of crash occurred in work zones longer than 6 hours is 0.324 smaller than that of work zones shorter than 6 hours.

Conclusions

The deteriorating transportation infrastructure systems maintained with work zones unavoidably brought safety issue attracting people's concern. In this paper, we investigated the causality of work zones on crash frequency and severity. In addition, we studied possible relationships between work zone characteristics and crash safety metrics.

Compared with previous studies, this study built an analysis framework employing a causality inference method to quantify the causal effect of work zone on crash likelihood and severity. We solved three problems in this framework. To solve the road inherent characteristics issues, we performed a Regression Discontinuity analysis to enhance the what-if counter-factual analysis. To perform the analysis in a high temporal-spatial resolution, we emphasized the significance of data source and hence the modeling method. Besides, confounding problems are discussed with the method of stratification and restriction. The detailed placebo tests also enhance the robustness of the model conclusions.

In summary, the results indicated the crash likelihood would increase 72.3% for high-historical-mean speed road segments related with roadway projects, while the crash severity will not significantly change. Work zone characteristics like shorter work zone duration, longer work zone length, and “shoulder closure” will bring more risk for the crash occurrence.

This study only focused on the crashes in 2013. Future research could target performing a time-series analysis with different years’ data. Besides, the development of Intelligent Transportation System (ITS) provides abundant high-temporal-resolution and high-spatial-resolution data. The conclusion drew from the high-temporal-resolution and high-spatial-resolution data may challenge those drew by previous research methods based on aggregated data. Similar to our research framework, researchers could perform causality studies on more high-resolution and large-scale data as well as carefully robustness check to get a more rigorous understanding of the activities on transportation infrastructure. Such as snowy weather, traffic jams, etc. With these understanding, researchers and policymakers could take more pertinent approaches to improve traffic safety.

Reference:

Abdel-Aty, M.A., Pemmanaboina, R., 2006. Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data. IEEE Trans. Intell. Transp. Syst. 7 2 , 167–174. doi:10.1109/TITS.2006.874710

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accid. Anal. Prev. 41 1 , 153–159. doi:10.1016/J.AAP.2008.10.005

Chen, E., Tarko, A.P., 2014. Modeling safety of highway work zones with random parameters and random effects models. Anal. Methods Accid. Res. 1, 86–95. doi:10.1016/j.amar.2013.10.003

Commonwealth Pennsylvania, 2018. RCRS_Event_Data Service [WWW Document]. URL http://www.penndot.gov/Doing-Business/OnlineServices/Pages/Developer-Resources-DocumentationAPI.aspx (accessed 6.10.18).

Federal highway Administration, 2017. National Highway System - Planning - FHWA [WWW Document]. URL https://www.fhwa.dot.gov/planning/national_highway_system/ (accessed 6.11.18).

FHWA, 2017. Work Zone Facts and Statistics. Washington, DC .

Garber, N.J., Woo, T.-S.H., 1990. Accident Characteristics at Construction and Maintenance Zones in Urban Areas. Virginia.

Hall, J.W., Lorenz, M., 1989. Characteristics of Construction-Zone Accidents. Transp. Res. Rec. J. Transp. Res. Board 1163, 4–14.

J. Dan Turner, 2017. Public Roads - What's A Work Zone? [WWW Document]. URL https://www.fhwa.dot.gov/publications/publicroads/99mayjun/workzone.cfm (accessed 7.26.18).

Jager, K.J., Zoccali, C., MacLeod, A., Dekker, F.W., 2008. Confounding: What it is and how to deal with it. Kidney Int. 73 3 , 256–260. doi:10.1038/SJ.KI.5002650

Jin, T.G., Saito, M., Saito, J., 2008. Spatial and Temporal Analyses of Crashes At Two Work Zones With Different Traffic Control Devices: Barrels and Jersey Barriers.

Kim, S., Coifman, B., 2014. Comparing INRIX speed data against concurrent loop detector stations over several months. Transp. Res. Part C Emerg. Technol. 49, 59–72. doi:10.1016/J.TRC.2014.10.002

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accid. Anal. Prev. 38 4 , 751–766. doi:10.1016/J.AAP.2006.02.001

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44 5 , 291–305. doi:10.1016/J.TRA.2010.02.001

Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. Saf. Sci. 46 5 , 751–770. doi:10.1016/J.SSCI.2007.03.005

Ozturk, O., Ozbay, K., Yang, H., 2014. Estimating the impact of work zones on highway safety.

Ozturk, O., Ozbay, K., Yang, H., 2014. Estimating the Impact of Work Zones on Highway Safety. Transp.

Res. Board 93rd Annu. Meet. 1–16.

PennDOT, 2018. Crash Download Map [WWW Document]. URL https://pennshare.maps.arcgis.com/apps/webappviewer/index.html?id=8fdbf046e36e41649bbfd9d7dd7c7e7e (accessed 6.10.18).

Pennshare, 2018. RMSSEG (State Roads) | [WWW Document]. URL https://data-pennshare.opendata.arcgis.com/datasets/rmsseg-state-roads (accessed 6.10.18).

Srinivasan, R., Ullman, G., Finley, M., Council, F., 2011. Use of Empirical Bayesian Methods to Estimate Crash Modification Factors for Daytime Versus Nighttime Work Zones. Transp. Res. Rec. J. Transp. Res. Board 2241 1 , 29–38. doi:10.3141/2241-04

Theofilatos, A., Yannis, G., 2014. A review of the effect of traffic and weather characteristics on road safety. Accid. Anal. Prev. 72, 244–256. doi:10.1016/J.AAP.2014.06.017

Theofilatos, A., Ziakopoulos, A., Papadimitriou, E., Yannis, G., Diamandouros, K., 2017. Meta-analysis of the effect of road work zones on crash occurrence. Accid. Anal. Prev. 108, 1–8. doi:10.1016/J.AAP.2017.07.024

TomTom, n.d. TomTom MultiNet | ADCi | TomTom GIS Data [WWW Document]. URL https://www.adci.com/tomtom/gis/ (accessed 6.11.18).

Xu, C., Tarko, A.P., Wang, W., Liu, P., 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accid. Anal. Prev. 57, 30–39. doi:10.1016/J.AAP.2013.03.035

Yang, H., Ozturk, O., Ozbay, K., Xie, K., 2014. Work Zone Safety Analysis and Modeling: A State-of-the-Art Review. Transp. Res. Board 93rd Annu. Meet. January .