Connected Vehicle Data Validation:

How do CV events relate to collision trends?

A Technical Whitepaper by:

Michigan State University

Nischal Gupta

Hisham Jashami, PhD

Peter T. Savolainen, PhD, P.E.

Ford Mobility

Mohammad Abouali

Tim Barrette, PhD

Callahan Coplai, AICP

Wesley Powell

This page intentionally left blank

EXECUTIVE SUMMARY

Traditionally, road agencies have utilized police-reported crash data both for the prioritization of high-risk locations, as well as in the development and implementation of safety projects to address prevailing crash trends. This approach is reactive in nature and can lead to suboptimal investment decisions due to limitations that are inherent in crash data analysis. The use of connected vehicle (CV) data provides a promising means for addressing these limitations as information about CV events can be obtained both at larger scale and in a timelier manner as compared to crash data. To this end, the frequency of engagement in moderate or harsh driving events (e.g., braking, acceleration, cornering) present a promising surrogate measure as a supplement to, or in lieu of, crash data. This white paper examines the viability of using aggregated and de-identified CV data from Ford as a leading indicator for crash trends. Comparisons are made between CV event and crash data to assess the correlation and utility of the event data for predictive and evaluative purposes. Results illustrate the relationships between events and crashes at varying levels of fidelity and suggest such data provide a promising resource for road agencies for the purposes of proactive safety management.

BACKGROUND

Each year, more than 35,000 fatalities occur as a result of motor vehicle crashes in the United States, in addition to more than 5 million injuries (1). For every crash-related fatality, eight people are hospitalized, and 100 are treated and released from hospitals (2). Crashes also incur economic and societal costs, which are equivalent to approximately 1.6% of the US gross domestic product (3). Significant reductions in crashes, injuries, and fatalities have been realized over time due to advances in vehicle safety features, improved roadway design, and the introduction of various policies and programs to address behavioral issues that adversely affect traffic safety. However, these metrics have generally plateaued in recent years, providing motivation for further efforts to address this public health and economic issue (4). In 2020, despite a decrease in vehicle miles traveled due to the pandemic, vehicle-related deaths were up 8% in the U.S.

In response to these broader issues, a diverse range of highway safety stakeholders have adopted the national strategy of 'Towards Zero Deaths', which was initiated by the Federal Highway Administration in 2009. These same stakeholders have developed strategic highway safety plans that outline comprehensive frameworks to help reduce traffic crashes and fatalities on public roads. These plans provide guidance as to the identification of emphasis areas where crash risks are most pronounced, as well as specific strategies that present the greatest potential for near- and long-term improvements in traffic safety.

Historically, the most critical element of these data systems are police-reported crash data. In consideration of resource constraints, it is imperative that agencies are able to proactively identify crash countermeasures and candidate locations that present the greatest opportunities for improvement. To this end, the Highway Safety Manual (5) outlines best practices for data-driven and proactive methods of safety management. These practices are based upon the availability of high-quality, properly maintained, and regularly updated police-reported crash data. These data records are compiled by law enforcement agencies and describe the location, circumstances, persons, and vehicles involved in the crashes. Despite their utility, the use of police-reported crash data for performance monitoring and predictive analytics presents some inherent challenges.

First, crashes are inherently rare and random events. Consequently, there is considerable variability in the frequency of crashes at individual roadway locations (e.g., intersections, segments) on a year-to-year basis. A significant number of crashes go unreported, especially those which involve minimal or no injury (6, 7). There are also differences in the minimum reporting requirements from state to state. For example, all states require a crash to be reported if it resulted in injury or death. However, crashes that do not result in injury are generally reported if minimum levels of property damage occur, ranging from $500 to $2000 on a state-by-state basis (8).

Furthermore, at low-volume and rural locations, numerous years of data are required in order to make meaningful inferences as to where crash risks are overrepresented as compared to locations with similar traffic volumes and geometric characteristics. Police-reported crash data also tend to include relatively limited information as to additional factors that contributed to the crash having occurred. Collectively, these issues limit the ability of road agencies to proactively and quickly respond to emerging road safety issues (9).

To this end, various surrogate measures of road safety have recently emerged as promising alternatives to police-reported crash data (10). These surrogate measures include traffic conflicts and various other types of near-crash events. The advantage of these metrics is that they tend to occur significantly more frequently than crashes, allowing for safety issues to be identified more quickly as compared to reliance on police-reported crash data. Much of the early work in this area focused on facility-level observations, such as monitoring individual road locations through field observation or the use of cameras. Alternately, the observation of traffic over time and space provides an alternative means of network-level analysis. Recent examples include the second Strategic Highway Research Program (SHRP 2) Naturalistic Driving Study (NDS), which included voluntary participation from 3400 drivers using a series of cameras and sensors installed on the vehicles of study participants (11). While more efficient, these methods also tend to be resource-intensive and are difficult to implement at scale.

In contrast, the emergence of connected vehicle (CV) technologies presents opportunities to leverage data for surrogate safety measures using equipment already installed in vehicles on the road today. These CV data can provide information about vehicle location, engine status, speed, and the use of various vehicle systems (12). This data presents a more objective lens than relying on subjective assessment of a crash scene. Moreover, CV event data are more frequently updated, providing significant advantages as compared to police-reported crash data for analysis purposes.

Ford Motor Company (Ford) collaborated with Michigan State University (MSU) in order to assess the potential usefulness of its existing CV data in traffic safety analysis. This paper presents an overview of a pilot project that is using aggregated and de-identified CV event data to demonstrate how these CV data can be used by transportation agencies in developing traffic safety solutions.

Ford Connected Vehicle Data

The vehicle data provided by Ford for this analysis included temporal and spatial information about driving events, including the frequency of acceleration, braking, and cornering at various threshold levels. These data, provided in an aggregate and de-identified format, can provide extensive information regarding traffic patterns and road safety conditions.

Harsh driving events are defined as sudden changes in velocity and/or direction of the vehicle which are usually identified by changes in g-force above "normal" thresholds using an accelerometer (13). These include events such as harsh acceleration, harsh braking, and harsh cornering. These events present a promising surrogate safety measure to supplement police-reported crash data.

Ford has shared a subset of aggregated and de-identified CV event data with MSU in order to assess the utility of leveraging these events in transportation agency roadway safety applications. The research team at MSU assisted with the data visualization and developing statistical models to identify relationships between CV events and crash risk. The idea is to demonstrate how the harsh CV events data can be utilized in lieu of, or in complement to, crash data when assessing crash risk, and also in the identification of high-risk locations.

The primary focus of this research was to examine the relationship between harsh CV events data and crash occurrence. This analysis focused on data from the metro Detroit area, specifically the road network in the seven counties that comprise the Southeast Michigan Council of Governments (SEMCOG) metropolitan planning organization. Ford is headquartered in Dearborn, Michigan and this region presents relatively high levels of Ford CV data coverage compared to others.

To date, CV event data were provided for the six-months period from January 2020 to June 2020. The preliminary analyses have focused primarily on three different event types, namely, harsh acceleration, harsh braking, and harsh cornering. In total, more than 1.9 million of these events were found to occur during this period as shown in Figure 1. Events were significantly less frequent in April and May due, in part, to travel restrictions that were introduced in response to COVID-19. The de-identified data were provided in aggregate three-hour time bins. An additional event-level dataset was provided that included aggregated temporal and spatial (geographic coordinates) information.

For comparison purposes, crash data were obtained from the SEMCOG open data portal for the five-year period from 2015 through 2019. Crash data were aggregated by type (e.g., rear-end, angle) to allow for assessments of the degree to which the CV data are correlated with, or predictive of, various types of crashes. In addition to the crash data, traffic and roadway information were also obtained from the SEMCOG open data portal. These data include information such as the annual average daily traffic (AADT), national functional classification (NFC) of the road, road surface condition, and posted speed limit. Additional roadway inventory data were obtained for the state trunkline system through the Michigan Department of Transportation (MDOT). These data include additional information detailing roadway geometric characteristics, such as the number of lanes by type, as well as the presence of features such as medians, traffic signals, and sidewalks, among others. The Ford CV event data were integrated with the crash and roadway data using geographic information to create a road segment-level database.

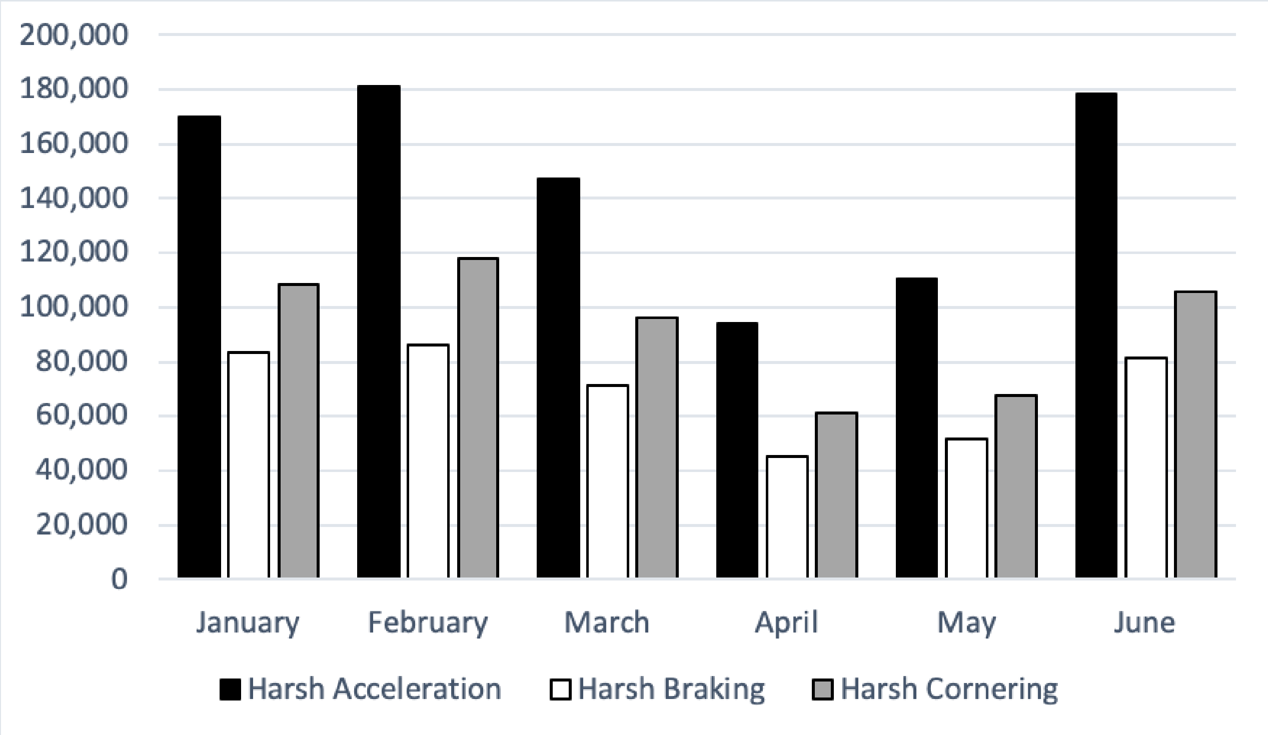

FIGURE 1 DISTRIBUTION OF FORD CV DATA FROM JANUARY-JUNE 2020

Methodology

Using these data, a series of investigations were conducted to assess the value of using CV event data as a supplement or alternative to police-reported crash data. This research involved the following activities:

1. Data visualization – As an initial step, the general relationship between traffic crashes and CV events was examined graphically at various levels of detail. The correlation in crash and CV event data was compared across different geographic areas, roadway environments, and across different subsets of crashes/events.

2. Regression analysis – Regression models were estimated to assess the degree to which CV events were predictive of traffic crashes. Negative binomial models are estimated to examine relationships between the numbers of crashes and CV events on individual road segments while controlling for the effects of other pertinent factors, such as traffic volumes and segment length.

3. Network screening – Historically, transportation agencies have generally prioritized intersections and road segments for safety improvement projects on the basis of historical police-reported crash data. Sites with higher numbers of crashes and/or crash rates are generally viewed as better candidates for such projects. This task involved a comparison of the relative rankings of road segments based upon the frequency of crashes as compared to the frequency of CV events.

4. US-23 Flex Route case study – In November 2017, MDOT opened a Flex Route along US-23 between M-14 and M-36. This project involved widening of the median shoulder, which is used as an alternative travel lane during peak traffic periods or in response to congestion or incidents. The CV event data were integrated for comparative analysis with probe vehicle speed data from the Regional Integrated Transportation Information System (RITIS). RITIS is a data archiving and analytics platform maintained by the University of Maryland, which integrates relevant data from multiple agencies, systems, and the private sector. These data are commercially available to transportation agencies (14) and allowed for an investigation of the relationship between general travel speeds and the frequency of CV events.

Results and Discussions

Data Visualization and Regression Analysis

Figure 2 provides plots of the annual average number of crashes (from 2015-2019) versus the number of CV events (January-June 2020) for segments in the SEMCOG road network. Separate plots are provided for all roads, other principal arterials, minor arterials, and collectors. Principal arterials (e.g., freeways, including the interstate system) were excluded from the analysis due to limited network coverage and sparser events given the relative infrequency of harsh events on such facilities.

Collectively, these plots show strong correlation with R2 values of 0.50, 0.59, and 0.50 for all roads, other principal arterials, and minor arterials. The collector roads showed an increasing trend, but significantly more variability as evidenced by an R2 of only 0.23. This is likely a function of several factors, including smaller road segments, lower traffic volumes, and a lower penetration rate of CV data as compared to the higher-class facilities.

To obtain further insights as to the nature of these relationships, a series of negative binomial regression models are estimated with the results shown in Table 1. Separate models are estimated for total crashes within each of the three functional classes. Traffic volume and speed limit are included as predictor variables, along with the number of CV events that were experienced on these same segments. Both AADT and CV event counts were log-transformed to improve goodness of fit. In examining the model results, crashes were found to increase by approximately 0.2 percent for a 1-percent increase in harsh CV events on both other principal arterials and minor arterials. On the collector roads, this effect was also present but less pronounced (approximately 0.12 percent increase in crashes for a 1-percent increase in harsh CV events). These increases are after controlling for the effects of AADT, speed limit, and segment length.

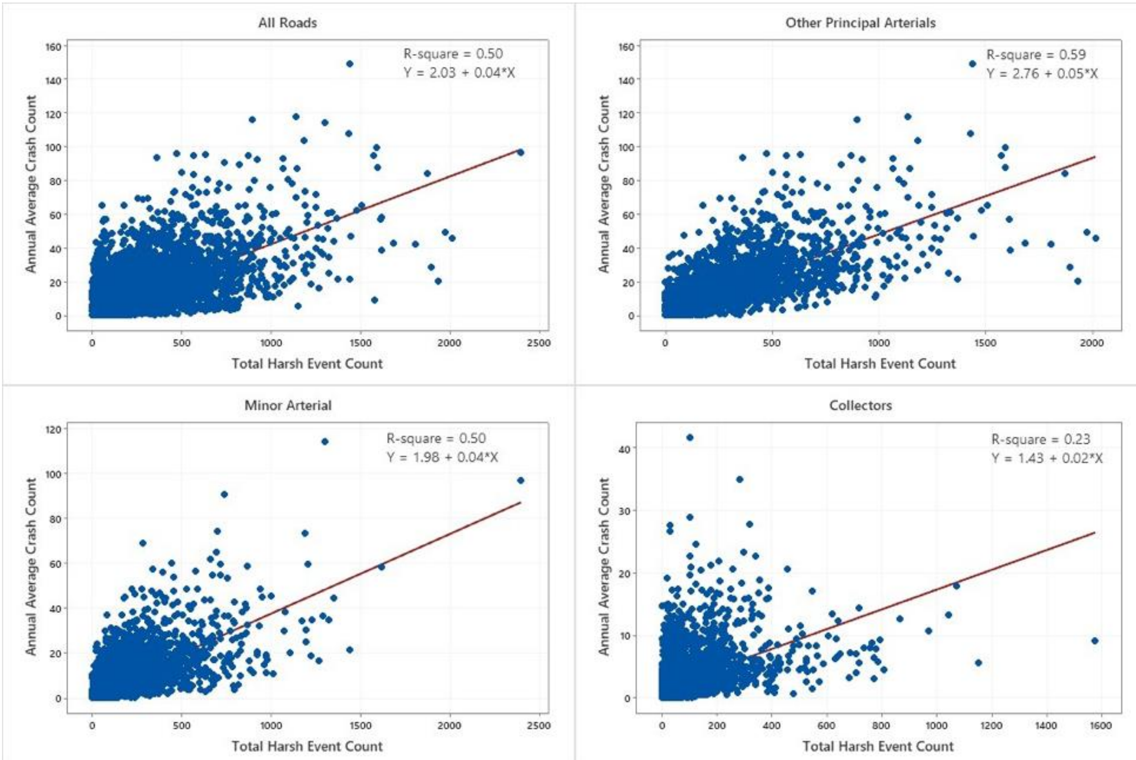

Figure 2 Plot of Crashes vs. CV Events in SEMCOG Region by Functional Class

Table 1 Parameter Estimates for Negative Binomial Model Based on Roadway Class

Response Variable = Total Crash Count (2015-2019)						
` Parameters	Other Principal Arterial (n = 2,956)		Minor Arterial (n = 3,438)		Collectors (n = 3,485)	
	Estimate (Std. Error)	p-value	Estimate (Std. Error)	p-value	Estimate (Std. Error)	p-value
Intercept	-2.39 (0.26)	<0.001	-2.26 (0.20)	<0.001	-1.70 (0.16)	<0.001
Ln(AADT)	0.62 (0.03)	<0.001	0.56 (0.02)	<0.001	0.52 (0.02)	<0.001
Ln(Harsh CV events per mile)	0.20 (0.01)	<0.001	0.21 (0.01)	<0.001	0.12 (0.01)	<0.001
Speed Limit (mph)						
55 mph or more	Base Condition		Base Condition		Base Condition	
40 mph to 50 mph	0.16 (0.08)	0.042	0.21 (0.07)	0.002	0.36 (0.06)	<0.001
35 mph or less	0.35 (0.08)	<0.001	0.44 (0.06)	<0.001	0.43 (0.05)	<0.001

Similar analyses were conducted specifically for the MDOT trunkline network. These analyses leveraged road network inventory files maintained by MSU, which have several advantages as compared to the SEMCOG data, including longer segment lengths (limiting concerns associated with very short segments experienced in the SEMCOG roads file) and more detailed information about roadway geometric characteristics. Separate analyses were conducted for various subsets of the data and samples are illustrated here for multilane (non-freeway) roads.

Figure 3 includes plots for annual average crashes (2015-2019) versus the total number of harsh events and each of the three event subsets (harsh acceleration, harsh braking, and harsh cornering) for multilane roads. Collectively, these plots show very strong linear relationships between crashes and the CV event data (as indicated by R2). This is true for all of the CV event types, though the goodness-of-fit was better for harsh acceleration and harsh braking events (R2 = 0.69 in both cases) as compared to harsh cornering (R2 = 0.53). Collectively, these results suggest significant potential for using the CV event data as a supplement or proxy for crash data.

Table 2 provides results of negative binomial regression models for two-lane and multilane roads. Again, these data show consistent relationships between crashes and harsh CV events, which are statistically significant (p-value < 0.001). The relationship is particularly strong on multilane roads, which may again be attributable to higher traffic volumes and CV penetration rates as compared to the two-lane facilities.

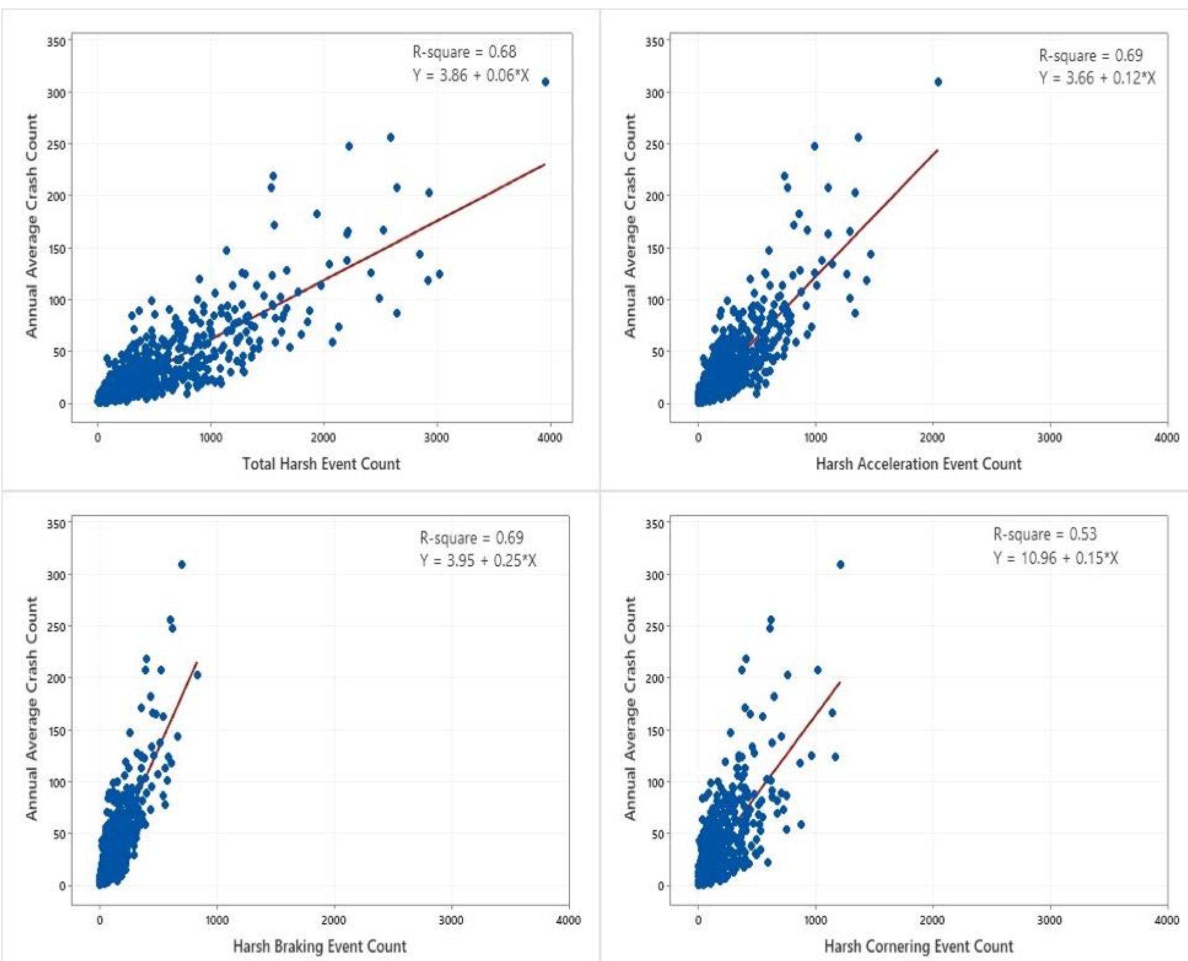

FIGURE 3 PLOTS OF TOTAL CRASHES VS. HARSH CV EVENTS BY TYPE ON MULTILANE STATE TRUNKLINES (NON-FREEWAYS)

TABLE 2 PARAMETER ESTIMATES FOR NEGATIVE BINOMIAL MODEL FOR TOTAL CRASH COUNT

Response Variable = Total Crash Count (2015-2019)				
`	Two-Lane Roads (n = 214)		Multilane Roads (n = 619)	
Parameters	Estimate (Std. Error)	p-value	Estimate (Std. Error)	p-value
Intercept	-2.09 (0.73)	<0.001	-2.87 (0.71)	<0.001
Ln(AADT)	0.53 (0.09)	<0.001	0.52 (0.09)	<0.001
Ln(Harsh CV events per mile)	0.20 (0.05)	<0.001	0.39 (0.05)	<0.001
Speed Limit (mph)				
55 mph or more	Base Condition		Base Condition	
40 mph to 50 mph	0.20 (0.17)	0.016	0.40 (0.13)	0.003
35 mph or less	0.54 (0.23)	0.226	0.93 (0.14)	<0.001

Further investigations were conducted for various subsets of crashes/events of interest. For example, Figure 4 provides plots of rear-end crashes versus harsh braking events. These relationships were consistently strong across facility types and this figure shows separate plots for two-lane highways (R2 = 0.50) and multilane roads (R2 = 0.65), respectively.

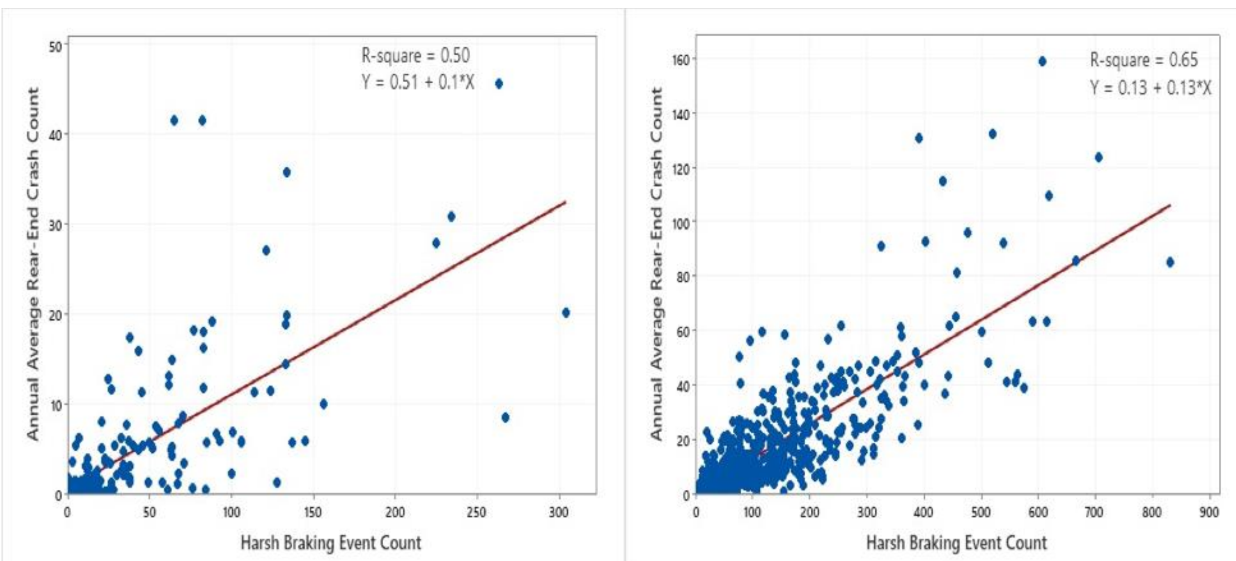

FIGURE 4 PLOTS OF REAR-END CRASHES VS. HARSH BRAKING EVENTS FOR STATE-MAINTAINED TWO-LANE ROADS (LEFT) AND MULTILANE ROADS (RIGHT)

The available data were also used to assess the separate relationship crashes and CV events have with traffic volumes and speed limits. These analyses provide a general sense of the degree to which CV events may serve as a proxy for crash data. For example, Table 3 presents a side-by-side comparison of regression models for crashes and harsh driving events, respectively.

Table 3 Comparison of Negative Binomial Model for Multilane Roads

Multilane Roads (n = 619)				
` Parameters	Response Variable = Total Crash Count		Response Variable = Total Harsh CV Events	
	Estimate (Std. Error)	p-value	Estimate (Std. Error)	p-value
Intercept	-4.10 (0.69)	<0.001	-2.92 (0.76)	<0.001
Ln(AADT)	0.87 (0.07)	<0.001	0.90 (0.08)	<0.001
Speed Limit (mph)				
55 mph or more	Base Condition		Base Condition	
40 mph to 50 mph	0.64 (0.13)	<0.001	0.42 (0.13)	<0.001
35 mph or less	1.23 (0.13)	<0.001	0.49 (0.13)	0.001

These results show very similar relationships with respect to traffic volumes. A one-percent increase in AADT was associated with a 0.87% increase in crashes and a 0.90% increase in harsh driving events. Crashes and CV events were also shown to increase as the speed limit was decreased, which is generally reflective of the interrelationships between speed limits and other segment-specific factors such as access point density, the frequency of signalized and stop-controlled intersections, and the level of roadside development.

In addition to the segment-level analyses presented above, a series of investigations were also conducted for intersections along the MDOT trunkline network. A geographic information system (GIS) shapefile containing locations and characteristics of 4,324 intersections in the metro Detroit region was obtained from MDOT. The dataset included information about the major and minor road traffic volumes, number of legs, type of traffic control, and level of service of safety (LOSS) for each of the intersection. LOSS is a four-class stratification scheme that compares the number of crashes a location experiences to the expected value based on a crash prediction model. LOSS I includes sites experiencing significantly fewer crashes while LOSS IV includes sites experiencing significantly more crashes than expected.

Figure 5 shows the relationship between annual average crashes and harsh CV events across the entire sample of intersections, stratified by LOSS. When segregated into different subsets based upon the type of traffic control and number of approach legs, similar trends emerge.

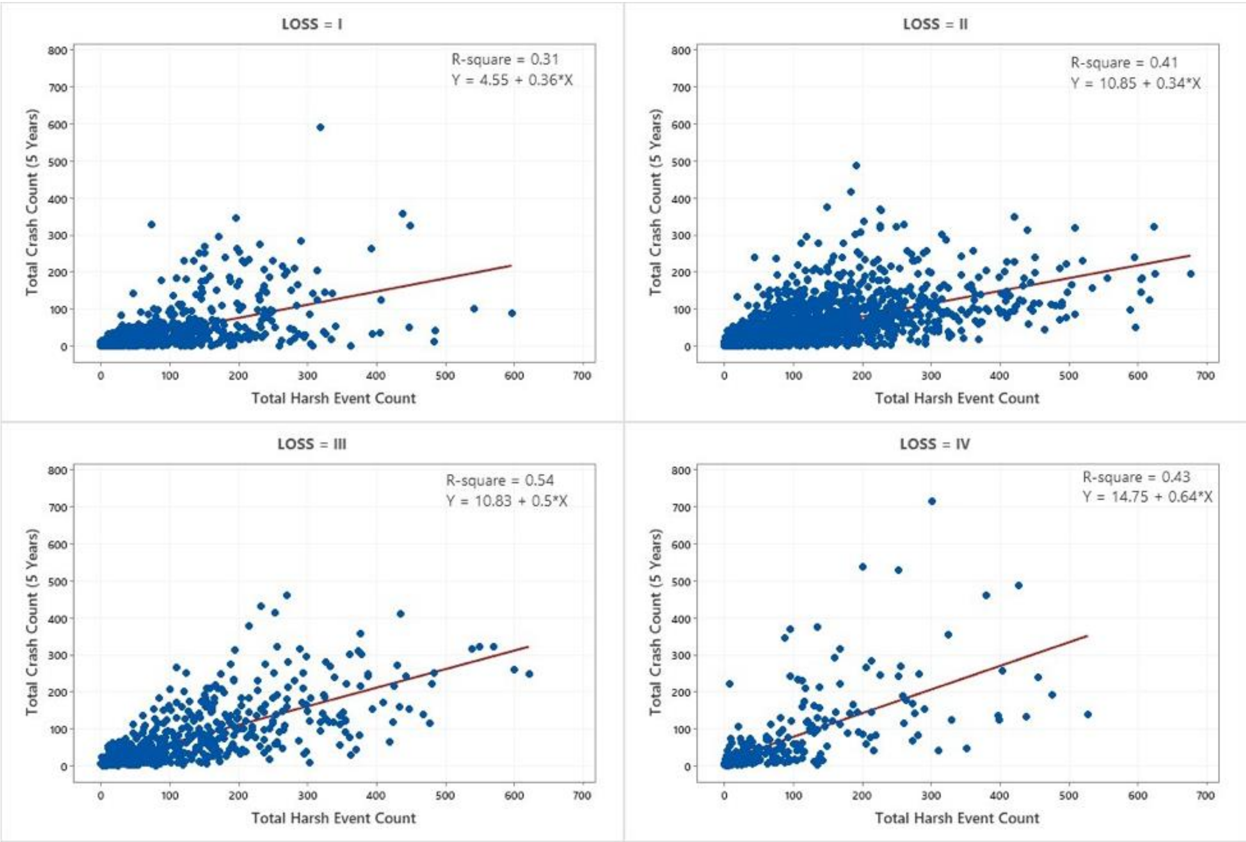

FIGURE 5 RELATIONSHIP BETWEEN CRASHES AND HARSH CV EVENTS AT INTERSECTIONS

Table 4 presents a comparison of negative binomial model results for crash counts and harsh CV events at four-legged signalized intersections. As in the preceding analysis (Table 3), the same predictor variables are used in order to assess the degree to which CV events and crashes vary with respect to major and minor road traffic volumes, as well as the LOSS categories described previously.

Similar trends are again observed between crashes and CV events. CV events and crashes are both found to increase with traffic volumes on the major and minor roads. In both cases, the results are more sensitive with respect to major road volumes as compared to minor road volumes. Interestingly, there is also a consistent relationship between the event data and LOSS. The relationship between LOSS and crash data is expected since these tiers are based upon historical crash data from MDOT. As the CV event data show similar trends (i.e., higher LOSS tiers experience higher number of events), this further reinforces the internal validity of using events as a leading indicator of crashes.

TABLE 4 NEGATIVE BINOMIAL MODELS FOR FOUR-LEGGED SIGNALIZED INTERSECTIONS

` Parameters	Response Variable = Total Crash Count		Response Variable = Total Harsh CV Events	
	Estimate (Std. Error)	p-value	Estimate (Std. Error)	p-value
Intercept	-7.78 (0.36)	<0.001	-3.10 (0.37)	<0.001
Ln(Major Road AADT)	0.80 (0.04)	<0.001	0.51 (0.04)	<0.001
Ln(Minor Road AADT)	0.40 (0.03)	<0.001	0.31 (0.03)	<0.001
LOSS				
I	Base Condition		Base Condition	
II	0.51 (0.06)	<0.001	0.26 (0.06)	<0.001
III	0.97 (0.08)	<0.001	0.48 (0.08)	<0.001
IV	1.50 (0.14)	<0.001	0.78 (0.14)	<0.001

NETWORK SCREENING

The network screening process generally relies heavily on crash data. At the simplest level, many agencies rank locations exclusively on the basis of the annual frequency or rate (per million vehicle-miles traveled or million entering vehicles). Given the nature of crashes, ranking by crash frequency tends to overemphasize high-volume locations (where crashes are most prevalent) while the converse is true when considering crash rates (i.e., low-volume sites tend to receive disproportionate weight).

Given the significantly greater frequency with which CV events occur, such data provide an appealing alternative for ranking locations as to their relative crash risks. To this end, the CV event data was utilized to rank locations based on the number of harsh driving events that were experienced. These rankings were compared to those based on a ranking by annual crash frequency. MDOT-maintained roads in the SEMCOG region were selected for network screening. Segments were included if the AADT was at least 1,000 veh/day and the segment length was at least 0.1 miles.

From this list, the top 50 roadway segments were identified based on both metrics (number of crashes and number of harsh CV events). Figure 11 shows these locations. The segments with the highest annual average crash frequency and highest CV event frequency are shown in red and blue, respectively. The segments that are common in both the lists are shown in green (overlapping segments). Figure 11 shows that the high-crash locations also tended to experience high numbers of CV events. Out of 50 total segments in each of the lists, 29 segments were common in both lists. The rankings were also compared after normalizing the data based on segment length and AADT. In these instances, the rankings began to diverge as 15 common segments were in the top 50 when considering crashes/events per mile and 8 segments were common in terms of crashes/events per million vehicle miles traveled (MVMT). In these instances, segments tended towards the top of the lists if they were either very short in length (i.e., near to 0.1-mi) or very low in AADT (i.e., near 1000 veh/day). There are also likely to be differences in the penetration rate of vehicles generating this CV event data across the SEMCOG region.

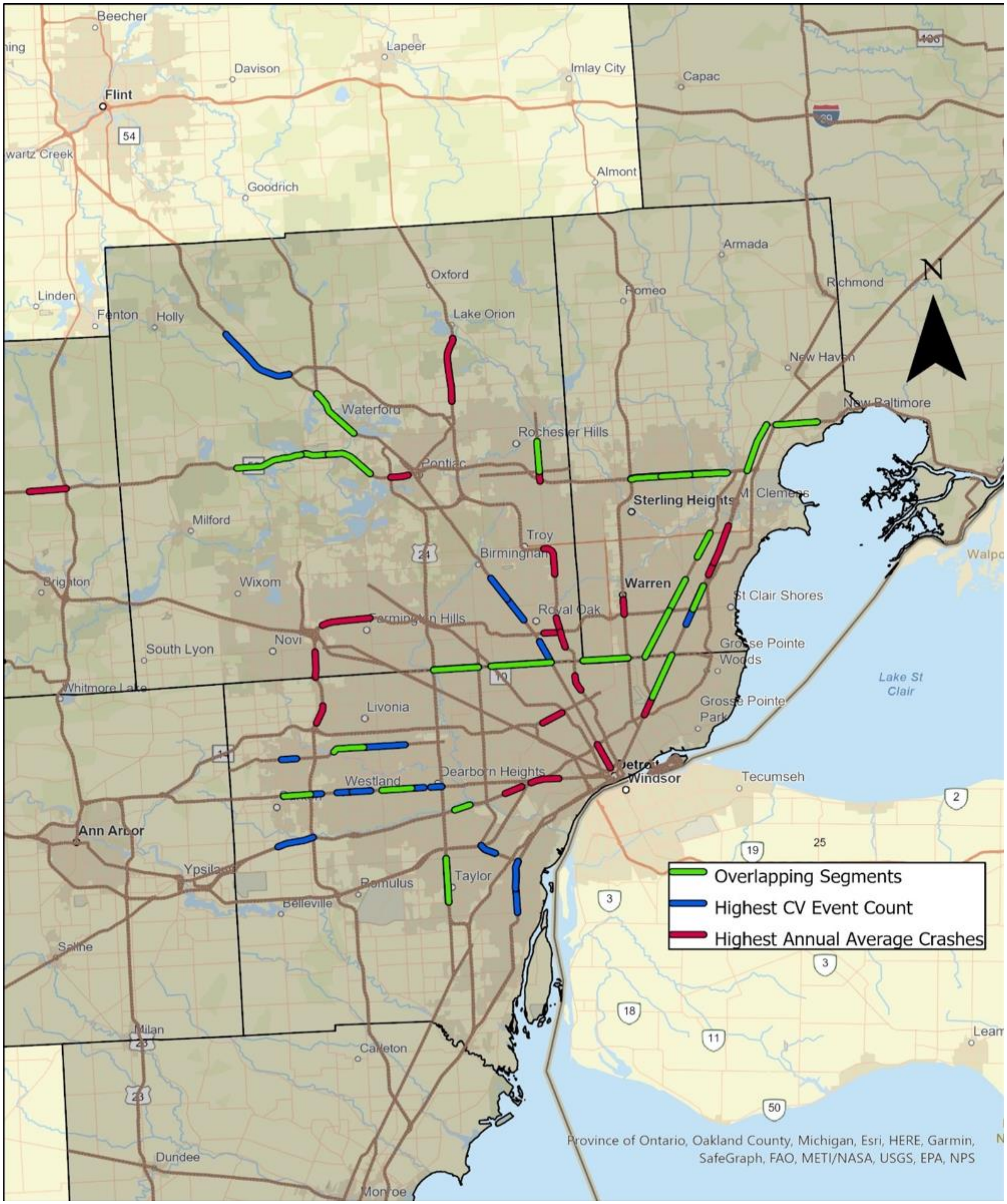

Figure 6 Network Screening Using Average Crash Frequency and CV Event Frequency

US-23 Flex Route

The CV event data were also used to assess relationships with traffic operational data. The US-23 Flex Route was examined as a case study. For visualization purposes, speed data were aggregated at hourly and daily intervals, along with the associated numbers of harsh CV events at these same intervals.

Figure 7 shows a time series plot of average daily speeds on the Flex Route from January 2020 to June 2020. These plots are shown by date (on the x-axis) and mile marker in both the southbound direction (indicated by red lines) and the northbound direction (indicated by blue lines).

These data show correlation in terms of when the CV events occur and when speed drops occur due to traffic congestion and other incidents. CV events were observed on days with lower average speeds. Similar trends are observed when the data is aggregated at one-hour intervals. This suggests the CV event data provide meaningful insights as to traffic operational performance measures even at this relatively high level of aggregation.

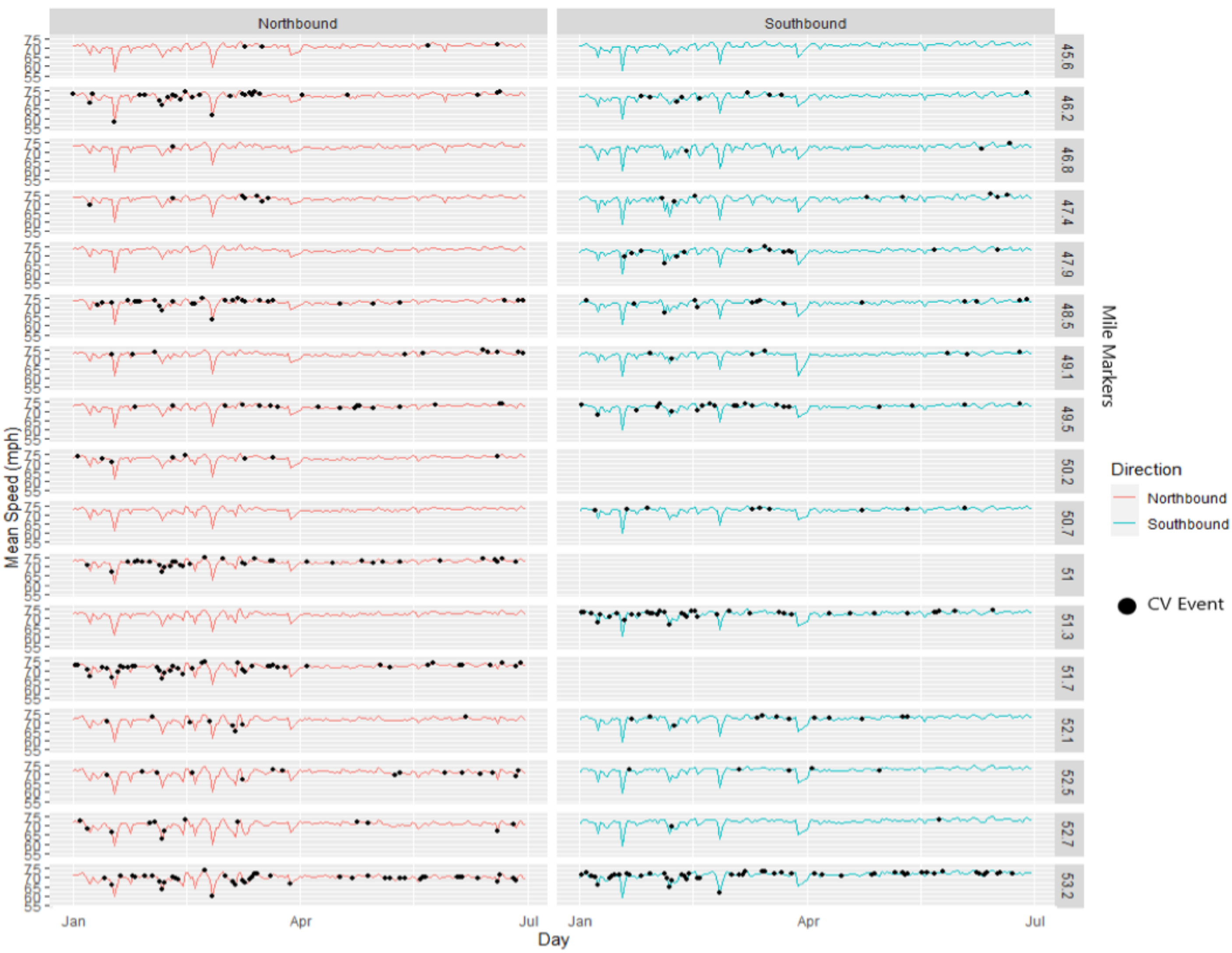

Figure 7 Time Series Plot of Travel Speed on Flex Route at Daily Aggregation

CONCLUSIONS

The Ford CV event data has shown promising results in this initial evaluation of its potential use for safety planning applications. The frequency of harsh CV events (acceleration, braking, cornering) exhibited significant positive correlation with the frequency of crashes. This was true for both road segment and intersection locations, as well as across different site types and in consideration of different subsets of crashes and CV events. The CV event data were also found to exhibit similar relationships with respect to segment-specific traffic volumes, speed limits, and other geometric characteristics. Even when controlling for the effects of these predictors, the CV events show further improvements in goodness-of-fit and increase the reliability of these predictive equations. This performance is likely to improve further with the increased penetration of vehicles generating this data.

With that being said, there are also a few limitations that should be acknowledged. First, the event data is only collected from Ford connected vehicles with certain selected in-vehicle settings, in combination with an enabled FordPass mobile app feature. Consequently, it is unclear how representative this sample of connected vehicles is compared to the general population of vehicles on the road. Continuing on this point, the penetration rate of Ford connected vehicles is not balanced across the SEMCOG region. However, it is unclear how this rate varies spatially, and ongoing work is aimed at further investigating this general issue.

Beyond the potential that has been demonstrated for safety evaluation, the CV event data also were reflective of changes in other traffic conditions based upon average speed data from the US-23 Flex Route. Additional research is proposed to investigate the applicability of these data across similar contexts, such as work zones and in assessing progression through signalized intersections.

Ultimately, this research suggests that CV event data provide a statistically significant surrogate safety measure as a complement to or in lieu of police-reported crash data. Further research will compare these data over the same time periods. Given data availability, this preliminary analysis considered police-reported crash data from 2015 through 2019, and CV event data from January 2020 to June 2020. Consequently, there are significant differences in coverage by time-of-year. Further, much of these CV event data were collected during the COVID-19 pandemic. Consequently, the number of CV events was comparatively low compared to the prior years' crash data. Stronger relationships may be anticipated when aligning the reporting periods for the event and crash data.

REFERENCES

1. National Highway Traffic Safety Administration. Fatality Analysis Reporting System (FARS). http://www-fars.nhtsa.dot.gov//QueryTool/QuerySection/SelectYear.aspx. Accessed March 15, 2021.

2. National Center for Injury Prevention and Control. Motor Vehicle Crash Injuries: Costly but Preventable. Atlanta: Centers for Disease Control and Prevention. www.cdc.gov/vitalsigns/crash-injuries/index.html. Accessed March 15, 2021.

3. Blincoe, L., T. R. Miller, E. Zaloshnja, and B. A. Lawrence. The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised). Washington, D.C., 2015.

4. National Highway Traffic Safety Administration. Traffic Safety Facts (TSF). https://cdan.nhtsa.gov/SASStoredProcess/guest. Accessed June 30, 2021.

5. AASHTO. Highway Safety Manual, First Edition. American Association of State Highway and Transportation Officials. Washington, D.C., 2010.

6. Waldheim, N., E. Wemple, and J. Fish. Applying Safety Data and Analysis to Performance-Based Transportation Planning. U.S. Department of Transportation, Federal Highway Administration, 2015.

7. Hauer, E. The Frequency-Severity Indeterminacy. Accident Analysis and Prevention, Vol. 38, No. 1, 2006, pp. 78–83. https://doi.org/10.1016/j.aap.2005.07.001.

8. Accident Reporting. https://drivinglaws.aaa.com/tag/accident-reporting/#:~:text=Illinois-,Crashes are required to be reported in cases of death,in the crash is uninsured. Accessed March 12, 2021.

9. Arun, A., Haque, M.M. Haque, A. Bhaskar, S. Washington, and T. Sayed. A Systematic Mapping Review of Surrogate Safety Assessment Using Traffic Conflict Techniques. Accident Analysis & Prevention, Vol. 153, 2021, https://doi.org/10.1016/j.aap.2021.106016.

10. Tarko, A.P. Chapter 17. Surrogate Measures of Safety. Safe Mobility: Challenges, Methodoogy and Solutions, 2018.

11. Wang, B., S. Hallmark, P.T. Savolainen, and J. Dong, Crashes and Near-crashes on Horizontal Curves Along Rural Two-lane Highways: Analysis of Naturalistic Driving Data, Journal of Safety Research, Vol. 63, 163-169, 2017.

12. Cohen-Abravanel, D. What Is Connected Car Data? https://otonomo.io/blog/connected-car-data/. Accessed March 15, 2021.

13. Aries, K. Why Monitor Harsh Braking and Acceleration? https://www.verizonconnect.com/resources/article/harsh-braking-acceleration-why-monitor/. Accessed March 12, 2021.

14. RITIS Introduction. https://ritis.org/intro. Accessed April 12, 2021.