Accident Analysis and Prevention 103 (2017) 10–19

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Performance of basic kinematic thresholds in the identification of crash and near-crash events within naturalistic driving data

Miguel A. Perez*, Jeremy D. Sudweeks, Edie Sears, Jonathan Antin, Suzanne Lee, Jonathan M. Hankey, Thomas A. Dingus

Virginia Tech Transportation Institute, 3500, Transportation Research Plaza, Blacksburg, VA 24060, United States

ARTICLE INFO

Article history:
Received 28 October 2016
Received in revised form 27 February 2017
Accepted 5 March 2017
Available online 31 March 2017

Keywords:
Naturalistic driving
Kinematic thresholds
Crash detection

ABSTRACT

Understanding causal factors for traffic safety-critical events (e.g., crashes and near-crashes) is an important step in reducing their frequency and severity. Naturalistic driving data offers unparalleled insight into these factors, but requires identification of situations where crashes are present within large volumes of data. Sensitivity and specificity of these identification approaches are key to minimizing the resources required to validate candidate crash events. This investigation used data from the Second Strategic Highway Research Program Naturalistic Driving Study (SHRP 2 NDS) and the Canada Naturalistic Driving Study (CNDS) to develop and validate different kinematic thresholds that can be used to detect crash events. Results indicate that the sensitivity of many of these approaches can be quite low, but can be improved by selecting particular threshold levels based on detection performance. Additional improvements in these approaches are possible, and may involve leveraging combinations of different detection approaches, including advanced statistical techniques and artificial intelligence approaches, additional parameter modifications, and automation of validation processes.

1. Introduction

Given the thousands of traffic fatalities that occur globally on an annual basis, it is societally-relevant to assess and address causal factors that lead to vehicular safety-critical events (e.g., crashes and near-crashes). Assessment of these factors can be accomplished in a variety of ways, one of which is the analysis of naturalistic driving data (Dingus et al., 2006). One particularly noteworthy naturalistic driving study (NDS) is the Second Strategic Highway Research Program (SHRP 2) NDS, which contains survey data and driving data representing over 4000 participant-years from over 3000 drivers in six different areas of the United States (Dingus et al., 2015).

The assessment of crash and near-crash causal factors using data from studies like the SHRP 2 NDS first requires the identification of the crash and near-crash events within the data. In most cases, this identification takes the form of (1) selecting kinematic thresholds that could indicate involvement in a safety-critical event, and (2) validating, generally by visual inspection of existing video, whether any events that "trigger" those kinematic thresholds are truly a safety-critical event. The kinematic thresholds employed usually use a combination of parameters related to vehicle speed, lateral acceleration, longitudinal acceleration, steering changes, and yaw rate changes. These measures are typically based on either expected kinematic "signatures" on a vehicle that collides with another object or on the "signatures" for the evasive maneuvers that a driver may perform in a crash avoidance attempt (Robertson et al., 1992; af Wåhlberg, 2006; Bagdadi and Várhelyi, 2011; Zaki et al., 2014; Boer et al., 2005).

Previous NDSs have created sets of kinematic thresholds with varying degrees of success. For example, the 100-Car NDS (Dingus et al., 2006) was the first instrumented-vehicle study undertaken with the primary purpose of collecting continuous, large-scale, naturalistic driving data. The study consisted of 78 privately owned vehicles, 22 leased vehicles, and 241 primary and secondary drivers. Candidate crashes in the study were detected post hoc within the continuous driving data when any of the following kinematic thresholds were exceeded:

- Lateral Acceleration: lateral acceleration equal to or greater than 0.7 g.
- Longitudinal Acceleration: longitudinal acceleration or deceleration equal to or greater than 0.6 g.
- Forward Time-to-Collision (TTC): acceleration or deceleration equal to or greater than 0.5 g coupled with a forward TTC of

* Corresponding author.
E-mail address: miperez@vtti.vt.edu (M.A. Perez).

http://dx.doi.org/10.1016/j.aap.2017.03.005

4 s or less; all longitudinal decelerations between 0.4 g and 0.5 g coupled with a forward TTC value of ≤4 s and a corresponding forward range value at a minimum TTC of not greater than 100 ft.
- Rear Time-to-Collision: any rear TTC trigger value of 2 s or less with a corresponding rear range distance of ≤50 ft AND any rear TTC trigger value with the following vehicle's absolute acceleration being greater than 0.3 g.
- Event Button: driver-activated by pressing a button located on the dashboard when an event is deemed critical.
- Yaw Rate: any value greater than or equal to a plus or minus 4° change in heading (i.e., vehicle must return to the same general direction of travel) within a 3 s window of time.

These thresholds have been used and modified in a number of subsequent investigations. For example, the Naturalistic Teenage Driving Study described by Lee et al. (2011) modified the 100-Car NDS thresholds to reflect lessons learned and to account for the different participant pool (i.e., teen vs. adult drivers). The longitudinal and lateral thresholds were raised to 0.65 g and 0.75 g, respectively, as were the deceleration levels to determine the validity of a low TTC threshold (to 0.60 g for TTC of 4 s or less and 0.50 g for TTC of 4 s or less and range of 100 ft or less). Similarly, Malta et al. (2012), proposed a set of kinematic thresholds for the European Field Operational Test on Active Safety Systems (euroFOT) project that added consideration of time headway (≤0.5 s), time to lane crossing (<1.0 s), and jerk (<$-2\,\mathrm{m/s^3}$ when longitudinal acceleration was below $-2\,\mathrm{m/s^2}$ and the conditions persisted for 0.2 s or longer). Fitch et al. (2013) added a geo-location component, considering decelerations in the context of the road type on which they occurred (i.e., longitudinal deceleration above 0.3 g when the vehicle was travelling above 64 km/h in a freeway environment).

Likewise, McGehee et al. (2007) and Carney et al. (2010) used DriveCam, Inc. systems to identify and collect data from potential crashes. In this case, data were only recorded when one or more of the following kinematic thresholds were exceeded:

- Shock trigger threshold: acceleration, in any direction, exceeding 1.50 g.
- Longitudinal trigger threshold: longitudinal acceleration or deceleration exceeding 0.50 g.
- Lateral trigger threshold: lateral acceleration exceeding 0.55 g.

These thresholds for the DriveCam system were relaxed in subsequent studies (Foss and Goodwin, 2014) in order to provide more opportunities for driver behavior assessment.

When available, the activation data from different safety systems in vehicles can also be a useful method to detect potential kinematic events of interest (Benmimoun et al., 2011). The use of a driver's actions and expresssions has also been proposed as a potential means of identifying kinematic events (Dozza and González, 2013).

The sensitivity and specificity of the different kinematic thresholds and other approaches used to find events is important in assessing the usefulness of particular methods. Approaches that result in a large array of event candidates will require time and resources to validate the candidate pool, whereas those that are too selective will run the risk of missing some events. The sensitivity of these approaches, unfortunately, can be low. For example, Victor et al. (2010) reported 19 true events in their assessment of 935 candidate events from the Sweden Michigan Naturalistic Field Operational Test (SeMiFOT) study. Similarly, sensitivity of some of the approaches used in the 100-Car NDS was well under 10% (Dingus et al., 2006). The potential exists, however, for better performance. Dozza and González (2013), for example, reported sensitivity and specificity of 84.2% and 79.7% for a video-based algorithm used in assessing particular event types.

Low sensitivity and specificity have provided some impetus for the initial development of more sophisticated classification models. For instance, a functional yaw rate classifier developed by Sudweeks (2015) retained the majority of safety relevant events (92% of crashes, 81% of near-crashes) while reducing the number of invalid or erroneous yaw rate triggers by 42% for the 100-Car NDS (Dingus et al., 2006). In addition, Dozza et al. (2013) proposed a preprocessing technique called chunking to increase robustness of parameter calculation and statistical sensitivity. The method divided data into equivalent, elementary chunks of data of equivalent duration to enable a robust and consistent calculation of parameters. Wu and Jovanis (2013) proposed iterative screening of candidate events with feedback to minimize video review. Specifically, they recommended initial screening for safety critical events as a first step, followed by classification, then a second screening, and after that, final verification via statistical models. Alternatively, Kluger et al. (2016) presented a discrete Fourier transform of longitudinal acceleration in combination with K-means clustering. This approach detected 78% of crashes and near-crashes within a small subset of the SHRP 2 NDS data, while generating about one false positive for every 2.7 h of driving data (Kluger et al., 2016).

One limitation in characterizing all of these approaches is the size of the sample available for a particular study. While particular approaches may work well in homogeneous datasets, their efficacy may not extend to more variable conditions. In this investigation, we use the SHRP 2 NDS, encompassing over 32 million miles of driving data, to fully describe the sensitivity and specificity of various kinematic thresholds used to identify potential events of interest. These results can be used to determine future tradeoffs in threshold selection as well as to guide improvements in the sophistication of the algorithms used.

2. Materials and methods

As applicable, the data from all the SHRP 2 NDS files were processed through each of the kinematic threshold approaches described in Table 1. These thresholds were selected to be representative of previous approaches, to be executable on a large-scale dataset within the computing constraints of the project, to produce a set of candidate events that could feasibly be validated within available project resources (Hankey et al., 2016), and to encompass a wide variety of kinematic conditions to minimize the events that were missed. Thresholds were run only on the pertinent data when values were not missing. There was no imputation, interpolation, or assumption made when a necessary value was not present. There was no exact calculation of the rates at which each threshold comparison could be executed on particular data, but qualitative assessments based on propensity for data availability are provided within Table 1.The resulting candidate events were validated by trained data coding staff, and were considered valid if they were classified as a crash or near-crash (see Hankey et al., 2016 for definitions). The proportion of valid detections was computed for each threshold as follows:

$$\%\,\mathrm{Valid}_i = \frac{\text{Total number of crashes or near crashes detected by threshold } i}{\text{Total number of candidate events detected by threshold } i} \times 100$$

In addition, each of the kinematic thresholds was examined to determine how often particular thresholds were the only thresholds that detected an event and how often multiple thresholds detected the same event. This was summarized via calculation of the percentage of valid events that were detected by one or more thresholds, as well as a matrix that summarizes which single thresholds and which pairwise threshold combinations yielded the largest proportions of valid events.

$$\text{Unique Detection Proportion}_{i,j} = \frac{\text{Number of type } j \text{ events detected only by threshold } i}{\text{Total number of type } j \text{ events detected only by threshold } i}$$

where j is either a crash or a near-crash.

$$\text{Combined Detection Proportion}_{i,j,k} = \frac{\text{Number of type } j \text{ events detected by both threshold } i \text{ and threshold } k}{\text{Total number of type } j \text{ events detected by threshold } i}$$

The calculation used the precipitating event (i.e., the event that initiated the crash or near-crash sequence) for the crash or near-crash as the anchor point, and considered any kinematic threshold that was crossed within the 20 s following that precipitating event as an instance of detection.

A proportion of false negative observations was calculated for both crashes and near-crashes as follows:

$$\text{False Negative Proportion}_{i,j} = \frac{(\text{Total number of type } j \text{ events} - \text{Number of type } j \text{ events detected by threshold } i)}{\text{Total number of type } j \text{ events}}$$

Receiver Operating Characteristic (ROC) curves, in which crashes and near crashes were combined as valid events, were also generated independently for each kinematic threshold in order to assess their specificity and sensitivity. For this purpose, the subset of crashes and invalid events detected by each kinematic threshold were used as the population of "attainable" events for that threshold. A key parameter of the threshold was then varied systematically based on the empirical distribution of such parameter within the events detected by the threshold. As the parameter was varied, new contingency tables were generated showing the classification of the events detected by the revised threshold. The contingency tables were then used to generate estimates for sensitivity and false positive rate (1 − specificity) that resulted in the ROC curves. Fig. 1 illustrates this process using the longitudinal deceleration kinematic threshold as an example. The initial threshold had a longitudinal deceleration parameter set at 0.65 g and resulted in 48,105 candidate events. Subsequent validation showed that 5289 of these candidate events were either a crash or a near-crash while 42,816 were invalid. This resulted in sensitivity and false positive rates of 1, since all 48,105 events were considered crashes or near-crashes when tested against the 0.65 g parameter. Similar calculations were completed for other potential parameter values. For example, the 70th-percentile of deceleration observed within the candidate events was 0.78 g. If that parameter was used to detect candidate events, a total of 2710 true crashes or near-crashes would be accurately detected, but 2579 would be missed. Similarly, 13,828 invalid events would be classified as crashes or near-crashes, but a total of 28,988 invalid events would be correctly ignored. Sensitivity and false positive rates at the 0.78 g parameter value are 0.51 and 0.32, respectively.

Due to the nature of the sensors and parameters that could be varied for yaw rate, ESC, traction control, and swerve, a significant portion of their original ROC plots were below the 50–50 reference line, indicating that classification performance might be improved by reversing the decision criteria. That is, rather than classifying a candidate event as a crash or near-crash if the observed extreme value exceeded a set parameter value, it was classified as a crash

Table 1
Description of kinematic threshold specifications.

Type	Description	Data availability
Longitudinal deceleration	The level of longitudinal acceleration is less than or equal to −0.65 g. The threshold is exceeded for at least 0.01 s. Multiple triggers are joined into the same potential event as long as the time between successive triggers is less than 2 s.	Very high
Longitudinal acceleration	The level of longitudinal acceleration is greater than or equal to 0.50 g. The threshold is exceeded for at least 0.01 s. Multiple triggers are joined into the same potential event as long as the time between successive triggers is less than 2 s.	Very high
Freeway deceleration	The level of longitudinal acceleration is less than or equal to −0.3 g when the vehicle travels on a freeway segment, as defined by the base map used in the Roadway Information Database. This level of deceleration or higher lasts for at least one timestamp. Multiple triggers are joined into the same potential event as long as the time between successive triggers is less than 2 s.	Moderate (depended on availability of map information)
Lateral acceleration	The lateral acceleration is greater than or equal to 0.75 g or less than or equal to −0.75 g. The threshold is exceeded for at least 0.2 s. Multiple triggers are joined into the same potential event as long as the time between successive triggers is less than 2 s.	Very high
Swerve	The derivative of yaw rate is monitored to find cases where the signal defines one complete cycle of a sine waveform whose minimum and maximum exceed $\pm 15^\circ/s^2$ within 2 s. The minimum speed is 5 m/s (∼11 mph).	High
Yaw rate	Vehicle swerves from ±8° per second to ±8° per second (in the opposite direction) within a window of 0.75 s. The minimum activation speed is 13.4 m/s (30 mph). Multiple triggers are joined into the same potential event as long as the time between successive triggers is less than 2 s. As a noise-reduction measure, the algorithm looks at the first three autocorrelation lag values for the yaw rate of each event. If any of the autocorrelation values is negative, meaning there is a lot of noise, the potential event is discarded. This works well for 10 Hz; a different number of autocorrelation lag values may be necessary at other data collection frequencies.	High
Advanced safety system activation (i.e., ABS, ESC, airbag, traction control)	Some vehicles provided the ability to monitor for activation of advanced safety systems such as anti-lock braking (ABS), traction control, airbag deployment, and electronic stability control (ESC) systems. For these vehicles, different algorithms searched the appropriate network variables to look for system activations.	Very low to Moderate (depending on availability of system/system status for particular vehicles)
Longitudinal jerk	The derivative of longitudinal acceleration was less than −1.0 g/s for 1 s while the vehicle travels at 5 m/s (∼11 mph) or higher speeds.	High
Lateral jerk	The absolute value of the derivative of lateral acceleration is greater than 1.0 g/s for 0.8 s while the vehicle travels at 5 m/s (∼11 mph) or higher speeds.	High

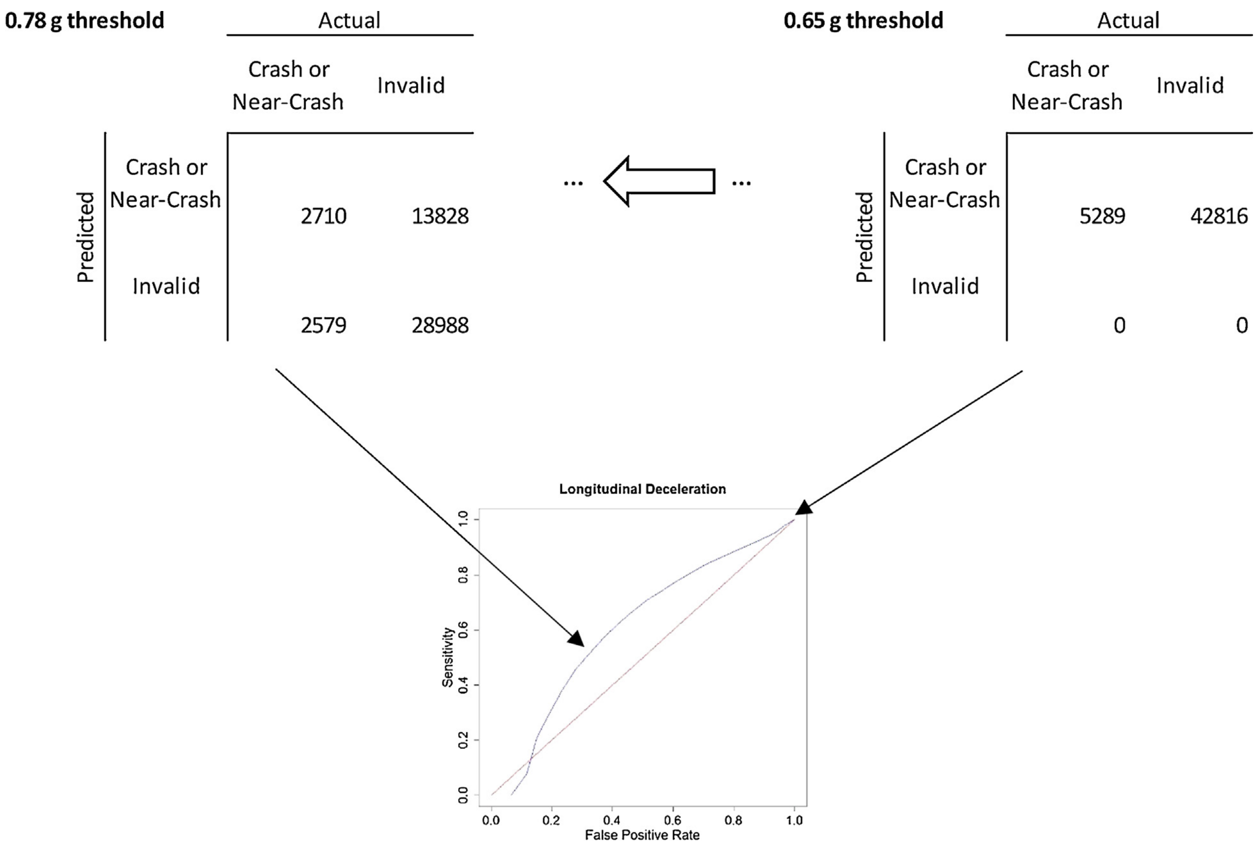

Fig. 1. Illustration of ROC curve generation process for the longitudinal deceleration kinematic threshold. Each point of the ROC relates to a contingency table that was generated as a key parameter for each threshold was varied.

Table 2
Kinematic threshold performance.

Threshold crossing type	Number valid		Candidate events found	Percent valid
	Crash	Near-Crash		
Longitudinal deceleration	682	4607	42,816	12.3%
Yaw rate	31	68	4355	2.6%
ABS	548	1878	133,493	1.8%
Lateral acceleration	229	77	16,773	1.8%
Freeway deceleration	63	2310	150,408	1.6%
Longitudinal acceleration	321	108	53,063	<1%
ESC	28	12	8516	<1%
Traction control	117	38	29,519	<1%
Lateral jerk	413	74	112,678	<1%
Airbag	1	0	283	<1%
Swerve	39	23	66,327	<1%
Longitudinal jerk	0	0	0	N/A

or near-crash real if the extreme value was under the parameter value.

Once developed, the ROC curves were used to suggest potential thresholds for future use via a simple optimization of Euclidean distance between the curve and perfect sensitivity and specificity. These potential thresholds weigh the ability to detect valid events against the possibility of detecting false events. This process weighed the detection of valid events equally against the detection of false events, and was designed based on our preferences and experience in this area. Other researchers may adjust these weights in the future as needed to fulfill their own preferences (e.g., heavily penalizing algorithms for failing to detect events).

As an assessment of the applicability of the suggested potential thresholds on an independent data set, kinematic triggers from the Canada Naturalistic Driving Study (CNDS; Council of Deputy Ministers Responsible for Transportation and Highway Safety, 2016) were evaluated using the updated threshold parameters for triggers executed in that particular data collection.

3. Results

Overall, a total of 555,345 candidate events were identified by the execution of the kinematic thresholds (Table 1) on the totality of the SHRP 2 NDS data. All of these candidate events were manually reviewed by video inspection as part of this investigation. A total of 2036 crashes and 7220 near-crashes were validated from these candidate events. Table 2 shows the performance for the different kinematic thresholds that were tested, along with the number of crashes and near-crashes that each detected. The longitudinal deceleration threshold yielded the highest percentage of valid events, followed by freeway deceleration, ABS, and lateral acceleration thresholds. Most of the thresholds resulted in very low

Table 3
Independent and pair-wise threshold performance in detecting crash events.

	Longitudinal Deceleration	Longitudinal Acceleration	Lateral Acceleration	Yaw Rate	Freeway Deceleration	ABS	Airbag	ESC	Traction Control	Lateral Jerk	Swerve
Longitudinal Deceleration	0.20	0.10	0.15	0.02	0.04	0.23				0.10	0.05
Longitudinal Acceleration	0.22	0.11	0.10	0.02	0.03	0.07				0.06	0.05
Lateral Acceleration	0.45	0.14	0.04	0.10	0.04	0.15			0.05	0.36	0.15
Yaw Rate	0.45	0.16	**0.77**	0.00	0.06	0.16				0.48	0.39
Freeway Deceleration	0.41	0.14	0.14	0.03	0.02	0.33				0.14	0.03
ABS	0.28	0.04	0.06		0.04	0.16		0.02	0.06	0.11	
Airbag	**1.00**	**1.00**	**1.00**			**1.00**	0.00			**1.00**	**1.00**
ESC	0.11	0.04	0.07			0.39		0.01	0.14	0.07	
Traction Control	0.03		0.09			0.28		0.03	0.04	0.06	
Lateral Jerk	0.17	0.05	0.20	0.04	0.02	0.15			0.02	0.13	0.05
Swerve	**0.79**	0.44	**0.90**	0.31	0.05	0.13	0.03			**0.51**	0.00

Table 4
Independent and pair-wise threshold performance in detecting near-crash events.

	Longitudinal Deceleration	Longitudinal Acceleration	Lateral Acceleration	Yaw Rate	Freeway Deceleration	ABS	Airbag	ESC	Traction Control	Lateral Jerk	Swerve
Longitudinal Deceleration	0.42				0.13	0.26					
Longitudinal Acceleration	0.29	0.01	0.03			0.07					0.02
Lateral Acceleration	0.19	0.04	0.00	0.18	0.05	0.10			0.04	0.17	0.14
Yaw Rate	0.12		0.21	0.00	0.12	0.13		0.03		0.24	0.01
Freeway Deceleration	0.27				0.23	0.10					
ABS	**0.64**				0.12	0.08			0.01		
Airbag							0.00				
ESC	0.17			0.17		**0.67**		0.00		0.17	
Traction Control	0.18		0.08			**0.68**			0.00	0.11	
Lateral Jerk	0.08		0.18	0.22	0.03	0.19		0.03	0.05	0.01	0.07
Swerve	0.13	0.09	0.48	0.04		0.13				0.22	0.00

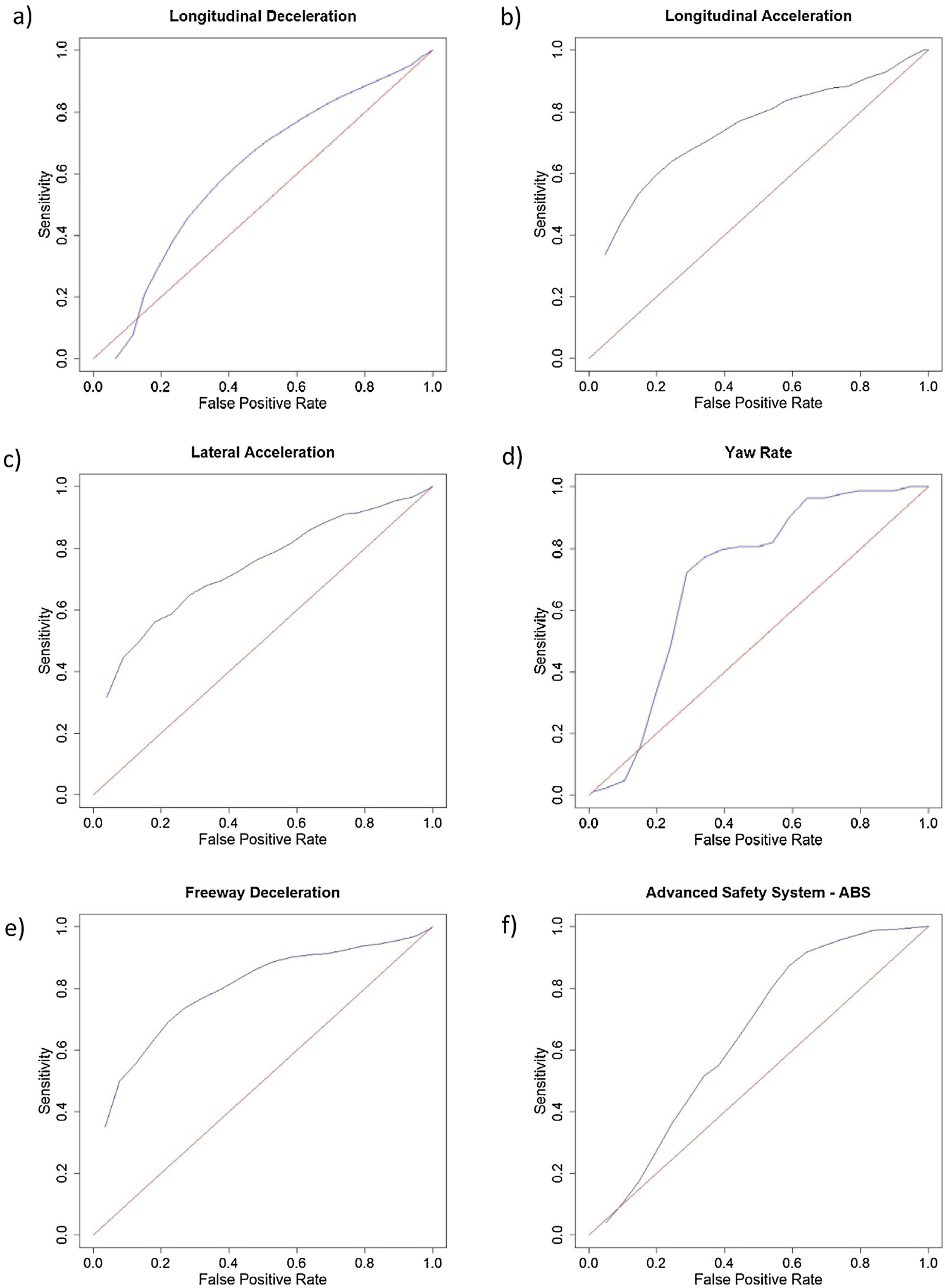

Fig. 2. ROC curves for thresholds of interest. From (a) to (j), respectively: Longitudinal deceleration, Longitudinal acceleration, Lateral acceleration, Yaw rate, Freeway deceleration, ABS, ESC, Traction control, Lateral jerk, Swerve. The diagonal line represents the classification level expected by pure chance.

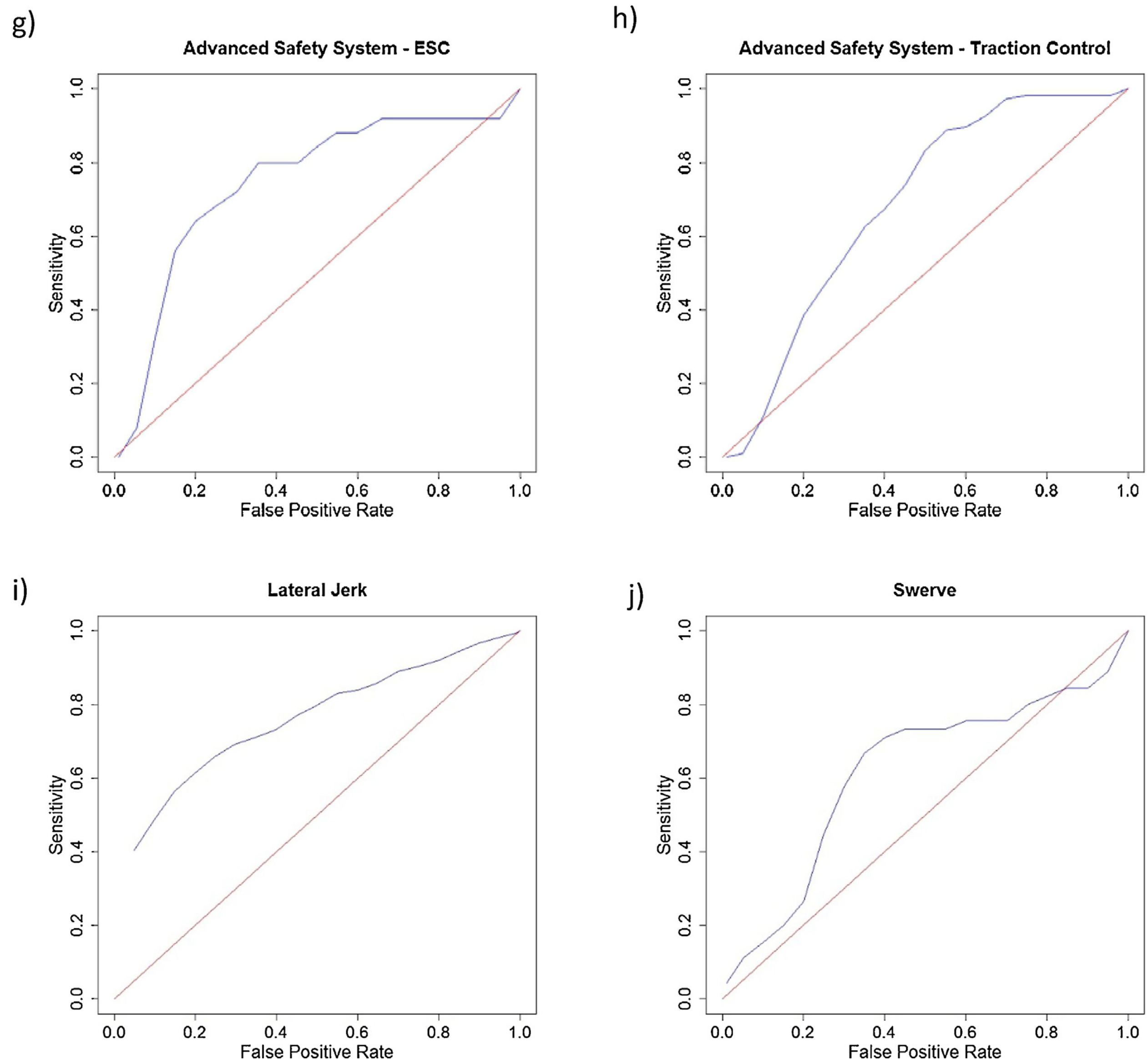

Fig. 2. (*Continued*)

detection rates of crashes and near-crashes. The longitudinal jerk trigger did not identify any candidate events throughout the full dataset.

Note that, in Table 2, crashes and near-crashes listed in each row were not necessarily unique (i.e., the same crash or near-crash may have been identified by several different thresholds, and therefore counted in multiple table rows).

The ability of each threshold to detect valid events in isolation and in combination with other thresholds is presented in Tables 3 and 4. The longitudinal deceleration, longitudinal acceleration, ABS, and lateral jerk thresholds uniquely detected the largest proportions of crashes. Overall, about 70% of crashes were detected using only one of these thresholds. For near-crashes, the percentage increased to nearly 75%. The thresholds contributing most to that percentage were longitudinal deceleration, freeway deceleration, and ABS.

In Table 3, the diagonal cells (shaded) indicate the proportion of all crashes that were uniquely detected by each threshold. The proportions in each non-diagonal cell indicate the proportion of crash events detected by the threshold in each row that were also detected by the threshold in the corresponding column.

In Table 4, the diagonal cells (shaded) indicate the proportion of all near-crashes that were uniquely detected by each threshold. The proportions on each non-diagonal cell indicate the proportion of near-crash events detected by the threshold in each row that were also detected by the threshold in the corresponding column.

For both crashes and near-crashes, there was some independence among the thresholds, but some high proportions of overlap also were observed. The longitudinal deceleration threshold identified the largest proportion of crash and near-crash events, and was also, logically, the most likely to select events that were identified by other thresholds. On average, 39% of crash events and 23% of near-crash events identified by the longitudinal deceleration threshold were also identified by other thresholds. The ABS and lateral jerk thresholds also showed substantial overlap with other thresholds for crashes (on average, 29% of crashes and 26% of near-crashes for ABS; 29% of crashes and 10% of near-crashes for lateral jerk). There were some trends related to the prevailing event kinematics in these findings, as lateral acceleration, yaw rate, and swerve thresholds (which are all related to events requiring lateral movement) showed some degree of overlap, particularly for crashes. In general, the overlap was less frequent for near-crashes,

Table 5
False Negative Proportions.

Threshold crossing type	False negative proportion	
	Crash	Near-Crash
Longitudinal deceleration	0.67	0.36
Longitudinal acceleration	0.84	0.99
Lateral acceleration	0.89	0.99
Yaw rate	0.98	0.99
Freeway deceleration	0.97	0.68
ABS	0.73	0.74
Airbag	1.00	1.00
ESC	0.99	1.00
Traction control	0.94	0.99
Lateral jerk	0.80	0.99
Swerve	0.98	1.00

where the largest overlap was observed between ABS and the other safety-system thresholds (i.e., ESC and traction control).

Most kinematic thresholds exhibited, with their initial parameters, high proportions of "misses" (i.e., false negatives) for both crashes and near-crashes (Table 5). Longitudinal deceleration exhibited the lowest proportion of false negatives for both types of events. ABS and Freeway Deceleration also exhibited some of the lowest false negatives proportions.

Empirical ROC curves were generated for each of the thresholds that generated sufficient valid events (Fig. 2; excludes airbag and longitudinal jerk, which detected almost no valid events). An objective function that assigned equal weight to maximizing sensitivity while minimizing the false positive rate yielded the parameters in Table 6. When convergence was observed, the suggested parameters were stricter than the parameters utilized in this investigation. In addition, Table 6 also shows the sensitivity and false positive rate estimates for the assessment of the suggested parameters using data from the CNDS when such an assessment was possible. Increasing longitudinal trigger thresholds in the CNDS collection resulted in lower sensitivity and false positive rate values than those estimated for the SHRP2 NDS. Classification measures for lateral acceleration were higher for the CNDS than the SHRP2 NDS, but were roughly equivalent for the ABS and lateral jerk thresholds.

4. Discussion

The results of this investigation suggest that a concise set of these kinematic thresholds can reduce the number of candidate events validated while retaining reasonable sensitivity levels. Particularly, use of the following kinematic thresholds is recommended, since other kinematic thresholds employed in this investigation failed to uniquely detect events at high rates.

- Longitudinal deceleration ≤ -0.75 g.
- Longitudinal acceleration ≥ 0.58 g.
- ABS activation lasting longer than 0.74 s.
- ESC activation lasting less than 1.45 s.
- Traction control activation lasting less than 1.20 s.
- Lateral jerk with |absolute value| ≥ 4.5 g/s within any 0.8 s window.
- Decelerations on freeways (as determined by map functional class) ≥ 0.3 g when the vehicle is traveling at ≥ 33 mph.
- Yaw rate oscillations larger than $\pm 8°$/s, smaller than $\pm 40°$/s, and occurring within 0.75 s.
- Swerve maneuvers larger than $15°/s^2$, smaller than $90°/s^2$, and occurring within 2 s.

The advanced safety system activation (i.e., ABS, Airbag, ESC, Traction Control) flags were not available on all the vehicles in the SHRP 2 NDS, and therefore their ability to uniquely detect crashes and near-crashes may have been underestimated. ABS activations, however, managed to uniquely identify 16% of crashes and 8% of near-crashes in spite of this limitation, suggesting that, when available, this particular system should be monitored more carefully in future data collection efforts. Lower than expected rates for the ABS (and similar) sensors were generally due to the implementation of the thresholds, which did not initially consider the activation time. Short activations of systems like ABS occur often, particularly when pavement friction is compromised, without leading to a crash or near-crash event. It was interesting that for some of these systems, however, better performance resulted when a maximum duration threshold was applied. This was most likely due to sensor issues that created longer duration flags that were generally invalid. Monitoring of the airbag flag did not provide useful results, mainly because readings for this particular variable tended to be noisy within the data used in this investigation. That said, it is expected that future noise-free recording of this variable should be very sensitive, particularly to crash events. Combined with the findings for ESC and traction control systems, this suggests that careful signal cleaning may be required before using these sensors to detect crashes and near-crashes.

Similar to the airbag threshold, longitudinal jerk did not detect any safety critical events in this study. However, it may be possible to enhance the performance of this trigger with additional filtering and optimizing the threshold duration. For example, Bagdadi and Várhelyi (2013) recommended a Savitzky–Golay filter for smoothing the original signal. The results of their analysis suggested that for detecting traffic conflicts (TTC $\geq$ 1.5 s) a peak-to-peak jerk threshold value of approximately 1.0 g/s could be used, and for detection of more serious conflicts (TTC < 1.5 s) a threshold of approximately 1.5 g/s for the critical jerk was recommended. The negative jerk value method was implemented by Zaki et al. (2014) who determined that smoothing with a Savitzky–Golay filter and using a minimum jerk value of -8 m/s^3 was sufficient for detecting traffic conflicts (TTC $\leq$ 4 s). Likewise, Pande et al. (2017) describe good correspondence between the presence of jerk events and historical crash rates at some highway locations. Additional filtering and bias correction may also be useful in the future for the yaw rate and swerve thresholds, which relied on a gyroscope and appear to have been subject to sensor noise based on the thresholds obtained from the ROC curves.

Groupings in the overlap of events detected by multiple kinematic thresholds is also informative and can suggest potential threshold improvements. For example, the lateral acceleration, yaw rate, and swerve thresholds, all related to events requiring lateral movement, showed substantial overlap in events detected. This suggests that kinematic thresholds that incorporate these elements may result in improved detection of valid events while reducing the number of invalid detections. This would also be the case, as previously mentioned, for safety system activations. However, there seems to be relative independence between events characterized by longitudinal movement, mainly because in those scenarios there are large kinematic differences between acceleration and deceleration profiles (as opposed to turning left or right in a lateral movement).

This last finding is further evidenced by the relatively large proportions of false negative observations throughout all kinematic thresholds. The lowest initial proportion of false negative observations in this investigation was 36%, and most of the observations were in the 90% range. This finding indicates that using these types of kinematic thresholds independently (e.g., only apply a longitudinal deceleration threshold in a search for crashes and/or near-crashes) comes at the risk of missing a sizeable portion of the existing events. The use of a false negative assessment, when feasible, is also recommended in future investigations that attempt

Table 6
Parameters suggested by optimization of ROC curves for combined selection of crashes and near-crashes (NA – not available for the CNDS collection).

Trigger type	Parameter used	Suggested parameter	Sensitivity/false positive rate	
			SHRP2 NDS	CNDS
Longitudinal deceleration	≤−0.65 g	≤−0.75 g	0.62/0.41	0.44/0.32
Longitudinal acceleration	≥0.50 g	≥0.58 g	0.64/0.24	0.44/0.35
Lateral acceleration	|abs| ≥ 0.75 g	|abs| ≥ 0.92 g	0.65/0.29	1/0.52
Yaw rate	±8°/s within 0.75 s	Between ±8°/s within 0.75 s and ±40°/s within 0.75 s	0.61/0.20	NA
Freeway deceleration	≥0.3 g, no minimum speed	≥0.3 g, ≥33 mph	0.74/0.27	NA
ABS	Flag presence	Flag duration ≥ 0.74 s	0.72/0.48	0.73/0.64
Airbag	Flag presence	No convergence	–	NA
ESC	Flag presence	Flag presence for 1.45 s or less	0.75/0.32	NA
Traction control	Flag presence	Flag presence for 1.20 s or less	0.65/0.37	NA
Lateral jerk	|abs| ≥ 1.0 g/s within 0.8 s	|abs| ≥ 4.5 g/s within 0.8 s	0.66/0.25	0.56/0.29
Swerve	15°/s^2 within 2 s	Between 15°/s^2 within 2 s and 90°/s^2 within 2 s	0.65/0.33	NA
Longitudinal jerk	≤−1.0 g/s within 1.0 s	No convergence	–	NA

to generate more concise approaches for crash and/or near-crash detection.

The success of kinematic thresholds varied in identifying crashes versus near-crashes, indicating that the severity of the events of interest is also a factor in determining which types of kinematic thresholds to employ and what their parameters should be. Investigations focused only on cases where crashes occurred (and perhaps crashes above a certain severity) would likely need very few kinematic thresholds set at high levels to identify the events of interest with minimal invalid events. Indeed, the use of lower severity events (i.e., near-crashes and crash-relevant conflicts) for safety studies has been debated in the literature, with some authors suggesting caution about basing risk estimates on events where no harm occurred (Knipling, 2015). Independent of this discussion, the results in this investigation suggest that research projects where near-crashes and crash-relevant conflicts may be of particular interest would need to relax many of these thresholds and develop ways of minimizing the large number of invalid events detected.

In fact, the large number of candidate events (over half a million) generated by these kinematic thresholds in this investigation suggests that improvements are needed in the sensitivity and specificity of these methods when attempting to detect crashes and near-crashes. Validation of these events is a resource-intensive endeavor, and one that may not be sustainable as naturalistic driving datasets increase in size and complexity. Potential improvements may include the following:

- Combinations of kinematic thresholds that show correlation in their selected events, including the development of models or heuristics that estimate the likelihood or confidence that a candidate event is indeed an actual event.
- Utilization of advanced statistical classifiers (e.g., Sudweeks, 2015; Kluger et al., 2016), ensemble methods that combine multiple classifiers (Dietterich, 2000; Rokach, 2009), or artificial intelligence techniques.
- Additional modification of some of the underlying parameters for each of these thresholds (e.g., duration of flag), which were not directly addressed in this investigation. In particular, the length of time over which multiple similar kinematic thresholds are joined (e.g., see description of "Longitudinal Deceleration" threshold in Table 1) may benefit from sensitivity analyses that characterize whether this time window should be reduced or extended.
- Assessment of combinations of thresholds that are relatively orthogonal in their selection of crash and near-crash events to assess their potential to reduce false negatives while increasing true positives.
- Automatization of the validation processes so that the need for manual review of candidate events is minimized or eliminated.

The SHRP 2 NDS provides a unique dataset to test and develop these recommendations as part of future investigations. As vehicles become more instrumented and smartphones continue to expand their computing power and ability to generate and process data, the availability of ubiquitous data that could identify crash and near-crash events will only continue to increase. Efforts to learn from these crashes and near-crashes will require sensitive and specific methods, such as those investigated here, for identifying true events.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

af Wåhlberg, A.E., 2006. Speed choice versus celeration behavior as traffic accident predictor. J. Saf. Res. 37 (1), 43–51, http://dx.doi.org/10.1016/j.jsr.2005.10.017.
Bagdadi, O., Várhelyi, A., 2011. Jerky driving—an indicator of accident proneness? Accid. Anal. Prev. 43 (4), 1359–1363, http://dx.doi.org/10.1016/j.aap.2011.02.009.
Bagdadi, O., Várhelyi, A., 2013. Development of a method for detecting jerks in safety critical events. Accid. Anal. Prev. 50, 83–91, http://dx.doi.org/10.1016/j.aap.2012.03.032.
Benmimoun, M., Fahrenkrog, F., Zlocki, A., Eckstein, L., 2011. Incident detection based on vehicle CAN-data within the large scale field operational test "euroFOT". Paper presented at the 22nd Enhanced Safety of Vehicles Conference (ESV 2011), Washington DC/USA.
Boer, E.R., Rakauskas, M.E., Ward, N.J., Goodrich, M.A., 2005. Steering entropy revisited. Paper presented at the Proceedings of the 3rd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Rockport, ME.
Carney, C., McGehee, D.V., Lee, J.D., Reyes, M.L., Raby, M., 2010. Using an event-triggered video intervention system to expand the supervised learning of newly licensed adolescent drivers. Am. J. Public Health 100 (6), 1101–1106, http://dx.doi.org/10.2105/AJPH 2009.165829.
Council of Deputy Ministers Responsible for Transportation and Highway Safety, 2016. Canada Naturalistic Driving Study, Retrieved from http://www.canada-nds.net/.
Dietterich, T.G., 2000. Ensemble methods in machine learning. In: Kittler, J., Roll, F. (Eds.), Multiple Classifier Systems: First International Workshop, MCS 2000. , pp. 1–15.
Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J.D., Perez, M.A., Hankey, J., Ramsey, D.J., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., Knipling, R.R., 2006. The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment. U.S. Department of Transportation; National Highway Traffic Safety Administration, Alexandria, VA.
Dingus, T.A., Hankey, J.M., Antin, J.F., Lee, S.E., Eichelberger, L., Stulce, K.E., McGraw, D., Perez, M.A., Stowe, L., 2015. Naturalistic Driving Study: Technical Coordination and Quality Control. SHRP 2 Report S2-S06-RW-1. Transportation Research Board, Washington, D.C.
Dozza, M., González, N.P., 2013. Recognising safety critical events: can automatic video processing improve naturalistic data analyses? Accid. Anal. Prev. 60, 298–304, http://dx.doi.org/10.1016/j.aap.2013.02.014.
Dozza, M., Bärgman, J., Lee, J.D., 2013. Chunking: a procedure to improve naturalistic data analysis. Accid. Anal. Prev. 58, 309–317, http://dx.doi.org/10.1016/j.aap.2012.03.020.

Fitch, G.M., Soccolich, S.A., Guo, F., McClafferty, J., Fang, Y., Olson, R.L., Dingus, T.A., 2013. The Impact of Hand-Held and Hands-Free Cell Phone Use on Driving Performance and Safety-Critical Event Risk. National Highway Traffic Safety Administration, Washington, DC.

Foss, R.D., Goodwin, A.H., 2014. Distracted driver behaviors and distracting conditions among adolescent drivers: findings from a naturalistic driving study. J. Adolesc. Health 54 (5 Suppl), S50–S60, http://dx.doi.org/10.1016/j.jadohealth.2014.01.005.

Hankey, J.M., Perez, M.A., McClafferty, J.A., 2016. Description of the SHRP 2 Naturalistic Database and the Crash, Near-crash, and Baseline Data Sets. Blacksburg, VA, Virginia Tech Transportation Institute, http://hdl.handle.net/10919/70850.

Kluger, R., Smith, B.L., Park, H., Dailey, D.J., 2016. Identification of safety-critical events using kinematic vehicle data and the discrete Fourier transform. Accid. Anal. Prev. 96, 162–168, http://dx.doi.org/10.1016/j.aap.2016.08.006.

Knipling, R.R., 2015. Naturalistic driving events: no harm, no foul, no validity. In: Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Salt Lake City, UT.

Lee, S.E., Simons-Morton, B.G., Klauer, S.E., Ouimet, M.C., Dingus, T.A., 2011. Naturalistic assessment of novice teenage crash experience. Accid. Anal. Prev. 43 (4), 1472–1479, http://dx.doi.org/10.1016/j.aap.2011.02.026.

Malta, L., Ljung Aust, M., Faber, F., Metz, B., Saint Pierre, G., Benmimoun, M., Schäfer, R., Aachen, Germany 2012. Eurofot SP6 D6.4 – Final results: Impacts on traffic safety.

McGehee, D.V., Raby, M., Carney, C., Lee, J.D., Reyes, M.L., 2007. Extending parental mentoring using an event-triggered video intervention in rural teen drivers. J. Saf. Res. 38 (2), 215–227, http://dx.doi.org/10.1016/j.jsr.2007.02.009.

Pande, A., Chand, S., Saxena, N., Dixit, V., Loy, J., Wolshon, B., Kent, J.D., 2017. A preliminary investigation of the relationships between historical crash and naturalistic driving. Accid. Anal. Prev. 101, 107–116, http://dx.doi.org/10.1016/j.aap.2017.01.023.

Robertson, D.I., Winnett, M.A., Herrod, R.T., 1992. Acceleration signatures. Traffic Eng. Control 33 (9), 485–491.

Rokach, L., 2009. Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography. Comput. Stat. Data Anal. 53 (12), 4046–4072.

Sudweeks, J.D., 2015. Using Functional Classification to Enhance Naturalistic Driving Data Crash/Near Crash Algorithms. National Surface Transportation Safety Center for Excellence.

Victor, T., Moeschlin, F., Hjälmdahl, M., Ahlström, C., Fors, C., Steinmetz, E., Morang, M., 2010. Sweden-Michigan Naturalistic Field Operational Test (SeMiFOT) Phase 1: WP 5 Evaluation of Methodology – Final Report.

Wu, K.-F., Jovanis, P., 2013. Screening naturalistic driving study data for safety-critical events. Transp. Res. Rec. 2386, 137–146, http://dx.doi.org/10.3141/2386-16.

Zaki, M., Sayed, T., Shaaban, K., 2014. Use of drivers' jerk profiles in computer vision-based traffic safety evaluations. Transp. Res. Rec. 2434, 103–112, http://dx.doi.org/10.3141/2434-13.