Accident Analysis and Prevention 153 (2021) 106010

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

The Impact of driver distraction and secondary tasks with and without other co-occurring driving behaviors on the level of road traffic crashes

Ali Jazayeri [a,*], John Ray B. Martinez [a], Helen S. Loeb [b], Christopher C. Yang [a]

[a] College of Computing & Informatics, Drexel University, Philadelphia, PA 19104, USA
[b] Center of Injury Prevention and Research, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

ARTICLE INFO

Keywords:
Distracted driving
Secondary tasks
Co-occurring behaviors
Driving behaviors

ABSTRACT

Driving safety is typically affected by concurrent non-driving tasks. These activities might negatively impact the trips' outcome and cause near-crash or crash incidents and accidents. The crashes impose a tremendous social and economic cost to society and might affect the involving individuals' quality of life. As it stands, road injuries are ranked among top-ten leading causes of death by the World Health Organization. Distracted driving is defined as an attention diversion of the driver toward a competing activity. It was shown in numerous studies that distracted driving increase the probability of near-crash or crash events. By leveraging the statistical power of the large SHRP2 naturalistic data, we are able to quantify the preponderance of specific distractions during daily trips and confirm the causality factor of an ubiquitous non-driving task in the crash event. We show that, except for phone usage which happens more frequently in near-crash and crash categories than in baseline trips, both distracted driving and secondary tasks occur almost uniformly in different types of trips. In this study, we investigate the impact of the co-occurrence of distracted driving with other driving behaviors and secondary tasks. It is found that the co-occurrence of distracted driving with other driving behaviors or secondary tasks increase the chance of near-crash and crash events. This study's findings can inform the design and development of more precise and reliable driving assistance and warning systems.

1. Introduction

According to NHTSA, 25 percent of the police-reported crashes are due to driver inattention defined as "insufficient or no attention to activities critical for safe driving" (Regan et al., 2011). The most substantial form of driver inattention is distracted driving. Distracted driving is defined as events or activities within or outside the vehicle (Young et al., 2007) that negatively affect a driver's ability to process information that is necessary to operate a vehicle safely (Regan and Hallett, 2011; Regan et al., 2008, 2011). This includes talking on cell phones, texting, eating, drinking, and other non-driving activities, called secondary tasks (Regan and Hallett, 2011). Distracted driving accounts for approximately 16 percent of economic loss and 15 percent of societal harm. In addition, 10 percent of fatal and 18 percent of injury crashes have been reported as "distraction-affected crashes" (Blincoe et al., 2015). These numbers represent driving trips that ended in non-fatal injuries or deaths. However, it is shown that as much as 16.1% of driving time gets affected by inattention (Stutts et al., 2003). Moreover, distracted driving has adverse impacts on traffic operation due to greater fluctuation in speed and significant lane deviation (Stavrinos et al., 2013). Therefore, numerous research studies have focused on the definition, theoretical foundation, formulation, prediction, and prevention of distraction and distracted driving to inform the development of the technological, behavioral, and infrastructure mitigating measures to enhance driving safety.

Classically, studies that focus on distracted driving use different combinations of data collection and analytical approaches. For example, the data used for examining distracted driving may be collected from human-in-the-loop simulation studies for retrospective (Jin et al., 2012; Ameyoe et al., 2015; Stavrinos et al., 2013) and real-time analysis or prediction (Wang et al., 2015; Liang et al., 2007). Another common approach is collecting naturalistic driving data using an instrumented vehicle for retrospective (Dukic et al., 2013; Jenkins et al., 2017; Aksan et al., 2013; Li et al., 2018) or real-time analysis and prediction (Liu et al., 2016; Deshmukh and Dehzangi, 2017; Kircher and Ahlstrom, 2010; Botta et al., 2019). Another approach is adopting qualitative techniques for data collection, such as interviews (Bakiri et al., 2013). Different methods are also used for analysis, modeling, and prediction of

* Corresponding author.
E-mail address: aj629@drexel.edu (A. Jazayeri).

https://doi.org/10.1016/j.aap.2021.106010
Received 11 September 2020; Received in revised form 15 December 2020; Accepted 25 January 2021
Available online 18 February 2021

distraction, such as driver modeling including perceptual and motor components (Hermannstädter and Yang, 2013; Ameyoe et al., 2015; Li et al., 2018), statistical analysis (Bakiri et al., 2013; Dukic et al., 2013), and machine learning algorithms such as classification and regression (Jin et al., 2012; Liu et al., 2016; Jenkins et al., 2017; Deshmukh and Dehzangi, 2017; Wang et al., 2015; Kircher and Ahlstrom, 2010; Liang et al., 2007; Botta et al., 2019).

An avenue of research on distracted driving now highlights the effects of secondary tasks. According to the Second Strategic Highway Research Program (SHRP 2) Researcher Dictionary for Video Data Reduction (VTTI, 2015), the secondary task is defined as any distraction that includes non-driving related glances away from the direction of vehicle movements such as radio adjustments, seatbelt adjustments, window adjustments, visor adjustment, and other non-critical tasks. It does not include tasks that are critical to the driving, such as speedometer checks, blind spot checks, activating wipers/headlights, and other critical tasks.

As a specific data collection approach gets adopted, different sets of variables get generated. These variables can be grouped into three categories. The first category includes variables related to the driver (such as age and prior experience, and visual, motor, and cognitive capabilities) or variables measured to collect the level of distraction or inattention of driver (such as physiological changes in the driver state, eye movement patterns, and brain activity measures). The second category includes variables collected from the instrumented vehicle or simulator dynamics (such as lateral and longitudinal speed and acceleration, lateral deviation, and steering angle over the course of driving). Then the third category is composed of variables associated with the environment. The latter category characterizes the sources of internal or external distractions such as cell-phone and billboards or time and physical characteristics of the environment, such as traffic signs, the surrounding vehicle dynamics, or road curvature.

The findings of the above studies can be summarized as follows. In the studies conducted to measure the impacts of distracted driving, it is shown that distracted driving significantly and adversely impact the performance of drivers. Besides, distracted drivers experience changes in their physiological and brain state, functionality, and performance. These changes are meaningful enough to be used for prediction purposes and for the design and development of warning systems.

However, the impact of distraction co-occurring with other driving behaviors are studied only in a few studies. One study shows the relationship between driving drowsy and distracted driving (Anderson and Horne, 2013), while another study investigates the distractive effects of cell phone use on safe driving (Unknown, 2003). None of the previously reported studies explicitly considered the data-driven co-occurrence of driving behaviors and secondary tasks to crash risk. In this study, we categorize driving epochs based on their outcome: (i) epochs ending in a crash, (ii) epochs with a near-crash incident (but no crash), and (iii) baseline epochs without any near-crash or crash incidents. We adopt a data-driven approach to identify co-occurring behaviors in a repository of driving behaviors. The data used in this study is collected in a naturalistic driving experiment. The objectives of this study are to identify: frequent driving behaviors and their co-occurrences, frequent secondary tasks and their co-occurrences with driving behaviors, and impacts and frequency of distraction or secondary tasks with and without other driving behaviors among different outcome categories.

To meet the objectives of this study, we mine frequent driving behaviors and secondary tasks in a data set. The association rule mining technique is utilized to identify frequent driving behaviors. This approach has been adopted for mining of co-occurring patterns in other applications in previous studies as well. For example, in Brossette et al. (1998), a data set of health surveillance data is mined to reveal unknown patterns. Or, Abdullah et al. (2008) uses a data set of medical billing data to identify frequent associations between diagnosis codes and treatment procedures. A similar approach is adopted in Shan et al. (2008), Kareem et al. (2017) for the identification of suspicious claims and potentially fraudulent individuals from billing records. Other applications are prediction and forecasting of cardiovascular diseases and heart attacks (Ordonez, 2006; Jabbar et al., 2011; Khare and Gupta, 2016), location-wise and time-wise mining of frequent diseases (Ilayaraja and Meyyappan, 2013), identification of associations among environmental exposure to different chemical compounds and adverse health outcomes (Bell and Edwards, 2014), and identification of patterns which can inform the diagnosis of asthma in pediatric using a sequential version of this method (Campbell et al., 2020). For a survey on the applications of association rule mining techniques in healthcare applications, please refer to Altaf et al. (2017).

Using the association rule mining technique, this study aims to identify frequent co-occurring behaviors. In other words, we would like to investigate and identify the non-driving behaviors, more specifically, distraction, and related secondary tasks, that are commonly observed in different types of epochs. At the same time, their co-occurrences are more frequent in near-crash or crash epochs. Overlooking potential differences in impacts of behaviors when they occur individually or co-occur with other behaviors might result in higher false positive and false negative rates. Therefore, identifying such sets of behaviors can inform more accurate predictions of epochs' outcomes and, consequently, the design and development of more reliable warning systems.

2. Material and methods

2.1. The SHRP 2 dataset

This study uses a naturalistic driving data set collected under the Second Strategic Highway Research Program (SHRP 2) conducted by the Virginia Transportation Technology Institute (VTTI) (Hankey et al., 2016; Transportation Research Board of the National Academy of Sciences, 2013). The program involves multiple partner organizations such as the Federal Highway Administration (FHWA), the American Association of State Highway and Transportation Officials (AASHTO), and Transportation Research Board (TRB) that interact with states and other organizations. The data set includes the driving data of more than 3,000 volunteer drivers, which amounts to some 5.5 million trips for a period of 4–24 months. The data is recorded by vehicles equipped with a comprehensive data acquisition system (DAS) from six site locations in the United States, namely New York, Florida, Washington, Indiana, Pennsylvania, and North Carolina. The DAS collects data in the form of recorded videos from driver and roadway, including the driver's face and hands, and forward and rear roadway. Other static information (e.g., road infrastructure) and dynamic information (e.g., sensor and GPS data) are also collected. The sequences of drivers' actions and manners are manually annotated in the epochs from seconds prior to the crash and near-crash events until the conflicting condition ends. Crash events are defined as any contact between the subject vehicle with either a moving or fixed object. Near-crash events are considered as circumstances which make drivers have a rapid evasive maneuver. The baseline events or epochs are randomly sampled from normal driving periods in a way that both i) prevalence analysis of factors under typical driving and ii) relative risk analysis among crash, near-crash, and baseline events would be possible (Hankey et al., 2016). From the recorded videos, the driver behaviors are manually annotated throughout driving epochs. In the collected data, up to three behaviors of drivers and three secondary

Table 1
The frequency of different epoch outcomes included in the study.

Epoch type	Frequency
Baseline	19,998
Near-Crash	6,914
Crash	1,843
total	28,775

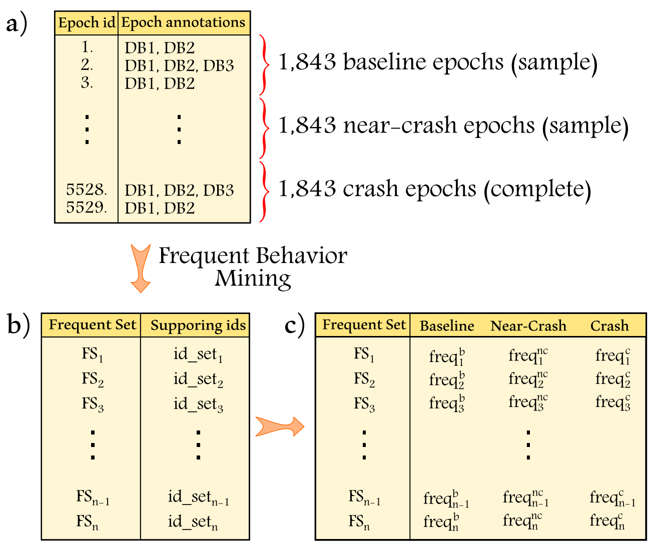


Fig. 1. The frequent behavior mining process. (a) data set creation for one iteration of implementation. Each row might include up to three driving behaviors (DB) and up to three secondary tasks. (b) frequent behavior mining implementation. The frequent sets of behaviors (FS) and the list of epochs that support each FS (id_set) are identified in this step. (c) frequency computation of FS based on the number of epochs supporting each FS (b: baseline, nc: near-crash, c: crash).

tasks are annotated. The manual annotation of driving behaviors and secondary tasks and events is performed by trained annotators in the SHRP 2 project. The annotators are employed based on their background and after reference checks, conducting interviews, proficiency tests, and other considerations. Besides, the annotation is undertaken under a standard quality assurance and quality control workflow. The reliability of annotations is further evaluated by intra-rater and expert-rater tests. For further details about the annotation process, refer to Hankey et al. (2016).

2.2. Data pre-processing

The data set used in this study is a subset of the SHRP2 dataset related to 3,542 drivers. The epochs are annotated with up to three behaviors. The total number of unique behavior annotations in the SHRP2 dataset used in this study is 57 driving behaviors. For the sake of generalizability, we grouped similar driving behaviors into one category in this study. For example, there are six driving behaviors that are related to improper turn. These six behaviors are only different in the direction of turn or other aspects. Therefore, we represent all these six behaviors with one behavior category, improper turn. Therefore, we categorized the 57 behaviors into 13 behavior categories. We follow the same approach for secondary tasks. There are 63 secondary tasks in the SHRP 2 data set. We grouped similar tasks and categorized them into seven secondary task categories. The complete lists of the post hoc mapping for both driving behaviors and secondary tasks are provided in the Appendix.

In this data set, the epoch outcomes are summarized through multiple variables. One such variables is the epoch category or type which can take three values: baseline, near-crash, and crash. In the baseline epochs, the driver has experienced neither crash nor any near-crash events. A ranking of crash severity is provided for the epochs that ended in a crash. The crash severity variable is based on the vehicle dynamics, the presumed level of damage, and the level of potential risks for others in the road. This ranking categorized the crashes into “I - Most Severe”, “II - Police-reportable Crash”, “III - Minor Crash”, and “IV - Low-risk Tire Strike”. In this study, we focused on the baseline, near-crash, and crash categorization of epochs and represented different severity

Table 2
A sample from three epoch outcomes with their associated categories of driving behaviors and secondary tasks. The event narratives related to a near-crash epoch and a crash epoch are provided in the footnote.

Epoch outcome	Behavior 1	Behavior 2	Behavior 3	Secondary behavior 1	Secondary behavior 2	Secondary behavior 3
Baseline	distracted	-	-	food/beverage	-	-
Baseline	distracted	-	-	interaction	external	-
Baseline	vehicle signal error	-	-	-	-	-
Near-crash	improper passing/neighbor lane conflict	-	-	-	-	-
Near-crash[1]	distracted	drowsiness	vehicle signal error	internal	-	-
Near-crash	distracted	-	-	interaction	-	-
Crash	driving fast	inexperience/ unfamiliarity with environment	distracted	phone	-	-
Crash[2]	distracted	right of way error	signal/sign violation	grooming	-	-
Crash	distracted	-	-	internal	-	-

[1] Event narrative: “Subject vehicle (SV) is accelerating after making a right turn without signaling into an aisle of a school parking lot at more than 75% occupancy during daylight hours. SV’s driver is drowsy and becomes distracted by adjusting the climate controls when a golf cart (V2) makes a left turn into the aisle from the intersecting parking lot roadway ahead. When SV looks up, she sees that V2 is cutting the corner on his turn and heading straight toward her. She brakes hard and steers to the left to avoid a head-on collision. The work vehicle brakes and steers to the right to evade SV and proceeds past SV. SV continues down the aisle.”

[2] Event narrative: “Subject vehicle is traveling through a commercial/residential area on an undivided two-way road. There is no lead traffic. Subject driver is flossing throughout part of the event, which distracts him from the driving task. Subject vehicle is traveling below fifteen miles per hour. Subject begins to decelerate at an intersection with a stop sign, while continuing to floss. He stops flossing and glances in both directions, evidently intending to execute a rolling stop. While proceeding through the intersection, the subject is struck by a vehicle (V2) from the perpendicular direction of travel to the left. The camera angles make it impossible to see whether V2 attempted an evasive maneuver.”

of crashes by one crash category. Table 1 shows the total number of epochs and the frequency of each outcome category in this study.

2.3. Frequent behavior mining

In this paper, we used the *Apriori* algorithm (Agrawal and Srikant, 1994) implemented by the arules library (Hahsler et al., 2005) in the R software. Table 1 shows that the number of records related to each outcome category is different. Therefore, the baseline and near-crash epochs are sampled to create a balanced data set. Considering that the lowest number of epoch categories belongs to the epochs associated with crash outcome, the number of samples drawn from each of the baseline

input :
- data set := driving data composed of driving behaviors for the three epoch outcome categories
- *intra-supp* := user-defined support threshold for considering a set of behaviors frequent in each iteration
- *inter-supp* := user-defined support threshold for considering a set of behaviors frequent in all the iteration
- sample_size := # of crash epochs
$\text{ITER} = \left\lceil \frac{\#\ of\ baseline\ epochs}{\#\ of\ crash\ epochs} \times \frac{\#\ of\ near-crash\ epochs}{\#\ of\ crash\ epochs} \right\rceil$

output: complete list of frequent sets (FSs) of behaviors for epoch outcome categories, and the mean and SD of their frequencies

begin
  agg_table = [];
  for *iter = 1 to ITER* **do**
    baseline_sample = sample(baseline data set, sample_size);
    near_crash_sample = sample(near_crash data set, sample_size);
    crash_sample = sample(crash data set, sample_size);
    data_set = row_bind(baseline_sample, near_crash_sample, and crash_sample);
    frequent_FSs = *Apriori*(data_set, *intra-supp*);
    epoch_outcome_FSs = summarizing the frequency of each FS for each epoch outcome;
    append the epoch_outcome_FSs to agg_table;
  end
  use agg_table to create the unique list of FSs, and compute the mean and SD of their frequencies;
  //FSs are included which appeared in more than *inter-supp* of iterations
end

and near-crash epochs is 1,843 (Fig. 1-(a)). In other words, we randomly down-sample the baseline and near-crash epochs and set the number of samples to 1,843 (the number of crash epochs). Because down-sampling might cause losing some of the data records, to minimize any biases due to sampling, we implement the frequent behavior mining for $\left\lceil \frac{\#\text{ of baseline epochs}}{\#\text{ of crash epochs}} \times \frac{\#\text{ of near-crash epochs}}{\#\text{ of crash epochs}} \right\rceil = 41$ times. Table 2 shows a small sample from different epoch outcomes. In this table, the SHRP 2 data set annotations are replaced with the corresponding categories of behaviors and secondary tasks (refer to Appendix).

Also, one needs to note that each epoch might have up to three different annotated behaviors contributing to the epoch outcome. As a final tally, we have 13 types of categorization of 57 behaviors. Each iteration of frequent behavior mining might produce a different list of frequent behaviors with varying values of frequency. Also, this approach mines the frequent behaviors independently from epoch outcomes (Fig. 1-(b)). Therefore, after each iteration, the frequent behaviors at the epoch outcome level are extracted and recorded in a table (Fig. 1-(c)). And at the end of the implementation of frequent behavior mining, the recorded tables are aggregated, and the mean and standard deviation (SD) of frequencies are calculated for each outcome category. A set of behaviors is considered frequent at each iteration if it is occurring in a more than a user-specified percentage of epochs (*intra-supp* threshold). Besides, the same set of behaviors is ultimately reported in the aggregated table if it is found to be frequent in at least 5% of iterations (*inter-supp* threshold). Fig. 1 visualizes the process while Algorithm 1 shows the pseudo-code of the implementation of frequent behavior mining in this study.

For the secondary tasks, we follow the same approach for mining frequent co-occurring secondary tasks and driving behaviors. However, we replace the distracted driving behavior with the secondary tasks recorded for the corresponding epoch. Considering that we are replacing distracted driving behavior with the secondary tasks, each epoch might have up to two driving behaviors and up to three secondary tasks. We present in the next section the results for both analyses: (i) mining frequent individual and co-occurring driving behaviors, and (ii) mining frequent individual and co-occurring secondary tasks and driving behaviors.

3. Results

The association rule mining technique adopted to mine frequent driving behaviors and secondary tasks requires a user-specified threshold (*intra-supp* threshold). This threshold is used to identify frequent driving behaviors and secondary tasks. In this study, we use the support threshold of 0.5% to consider a set of behaviors or secondary tasks frequent. It means that a set of behaviors or tasks is deemed frequent if this set is observed in at least 0.5% of the epochs. Because we consider 5,529 (1,843 × 3) epochs in each implementation of the frequent behavior mining, this threshold translates to about 28 epochs.

Table 3
The average (SD) of frequency of different individual behaviors in 1843 epochs of each outcome category.

Driving behavior	Baseline	Near-crash	Crash
Aggressive driving	2.4 (1.4)	112 (8.1)	21 (0)
Distracted	862.8 (19.1)	921.3 (19.2)	721 (0)
Driving fast	52.6 (6.2)	153.7 (8.8)	181 (0)
Drowsiness	28.9 (5.2)	23.7 (4.5)	48 (0)
Improper backing	0 (0)	5.6 (2.1)	131 (0)
Improper passing/ neighbor lane conflict	4.2 (2.1)	63.7 (6.9)	14 (0)
Improper turn	9 (3.4)	35.2 (4.6)	673 (0)
Inexperience/unfamiliarity with environment	1.6 (1.1)	19.2 (3.6)	97 (0)
Lane changing error	0.4 (0.6)	65.4 (7.4)	9 (0)
Other	32.4 (5.6)	50.5 (5.3)	88 (0)
Right of way error	0.1 (0.3)	61.5 (7.1)	25 (0)
Signal/sign violation	23 (4.1)	68.8 (7.3)	116 (0)
Vehicle signal error	41.5 (5.4)	67.5 (7.5)	76 (0)

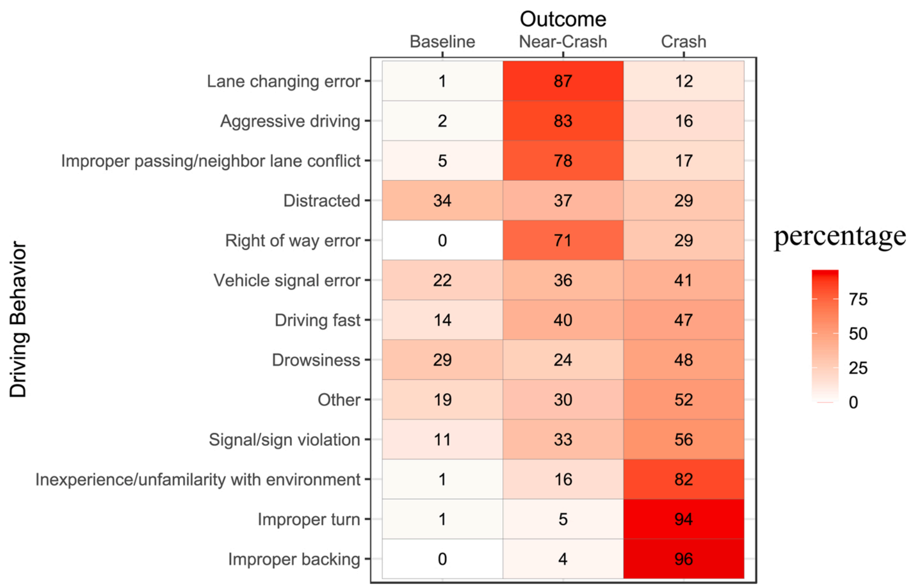


Fig. 2. The frequent behaviors at the support level of 0.5%. The numeric values represent the row-wise normalized values of the average frequencies for each driving behavior. The behaviors are sorted based on their crash percentages.

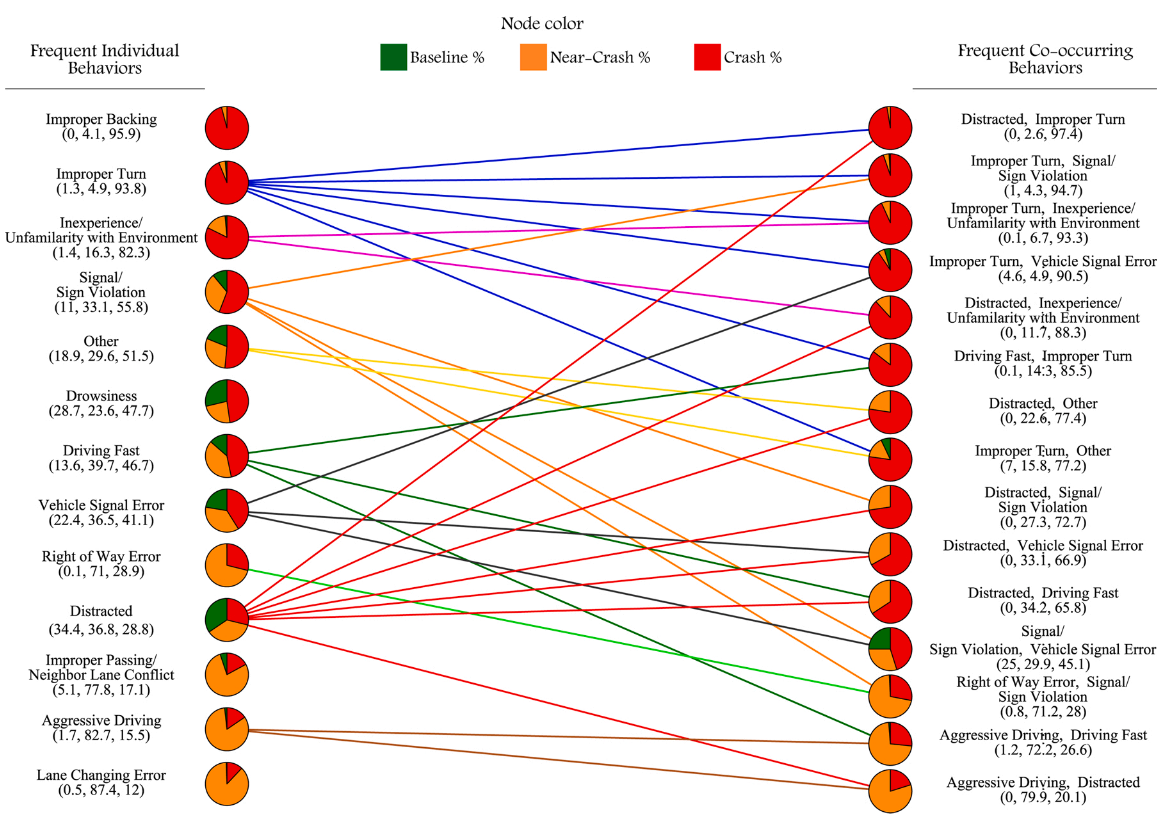


Fig. 3. The network of co-occurring driving behaviors. Each node has a tuple representing the percentage of baseline, near-crash, and crash epochs, respectively. The links from the same behavior are shown in the same color. The frequent sets of two co-occurring behaviors are composed of two frequent individual behaviors co-occurring together.

Therefore, we consider any subset of driving behaviors and secondary tasks frequent if it appears in at least 28 epochs. At this threshold, none of the individual driving behaviors and secondary tasks are excluded (the results are shown in Tables 3 and 4). However, only some of the co-occurring behaviors are frequent (the results are visualized in Figs. 3 and 4). Adopting lower values for threshold results in a large number of co-occurring behaviors recorded only in a few epochs. On the other hand, increasing the threshold results in excluding some of the individual driving behaviors and secondary tasks and, consequently, their co-occurrences. After trying multiple thresholds, we learned that the threshold of 0.5% provides acceptable interpretability of results. In the following, the results of the implementation of the association rule mining technique are presented.

Table 4
The average (SD) of frequency of different secondary tasks in 1843 epochs for each outcome category.

Secondary task	Baseline	Near-crash	Crash
External	198.8 (12.2)	210.6 (12)	228 (0)
Food and beverage	60.8 (7.6)	44.5 (5.6)	50 (0)
Grooming	86.5 (8.2)	115.7 (8.7)	107 (0)
Interaction	445.1 (17)	374 (14.2)	428 (0)
Internal	103.7 (8.7)	454.5 (16.1)	310 (0)
Phone	153.1 (9.5)	265.1 (10.5)	226 (0)
Unknown	73.4 (8.3)	51.3 (6.6)	77 (0)

3.1. Frequent driving behaviors

At the 0.5% support threshold, the frequent sets are either composed of one behavior or two behaviors. No set of three behaviors are found to be frequent. Table 3 lists the frequent sets consisting of one behavior. This table shows that all the 13 behavior categories considered are frequent. The numeric values are the average number of frequencies of each behavior over the 41 iterations of association rule mining implementation. The numeric values in parentheses are standard deviation (SD) of these frequencies. Because the same set of crash epochs is used in different iterations, the frequencies of behaviors in the crash epochs are identical; therefore, the associated SD would be zero. The absolute frequency values for behaviors are row-wise normalized and shown in Fig. 2.

Fig. 3 shows individual behaviors that contribute to different frequent sets of co-occurring behaviors. The individual behaviors are shown in the left column of nodes. And, the frequent sets composed of two behaviors are shown in the right column. Each node on the right column (frequent co-occurring behaviors) is created from the combination of two nodes from the left column (frequent individual behaviors). The two individual behaviors contributing to a frequent co-occurring behavior are connected to the corresponding co-occurring behavior with two edges. For example, the set of {distracted, improper turn} is frequent. Therefore, the corresponding node on the right column is connected to the distracted node and improper turn node on the left column. All the nodes on the right column have two edges as they are created from two individual behaviors. However, nodes on the left column might have multiple edges. A node with a higher number of edges on the left column implies that the corresponding individual behavior co-occurs more frequently with other behaviors. Besides, each node represents a pie-chart showing the relative percentage of different epoch categories corresponding to each node. The nodes are descendingly ordered in two columns based on the crash percentages. For each node, a tuple of numeric values is provided that represents the relative percentage of observation of each behavior set in baseline, near-crash, and crash categories, respectively. For example, the first node of the right column is associated with a frequent set composed of co-occurring distracted driving and improper turn. The corresponding tuple shows this co-occurrence results in near-crashes in 2.6% of times and in crash 97.4% of times.

3.2. Frequent secondary tasks and driving behaviors

We use the the same support threshold of 0.5% for mining data set of secondary tasks and driving behaviors. The frequent sets identified after implementing the association rule mining on this data set are composed of maximum two components: only one driving behavior, only one secondary task, or combination of each co-occurring together. Table 4 shows the average and SD of the frequency of different secondary tasks in 1843 epochs for each outcome category. At the support threshold

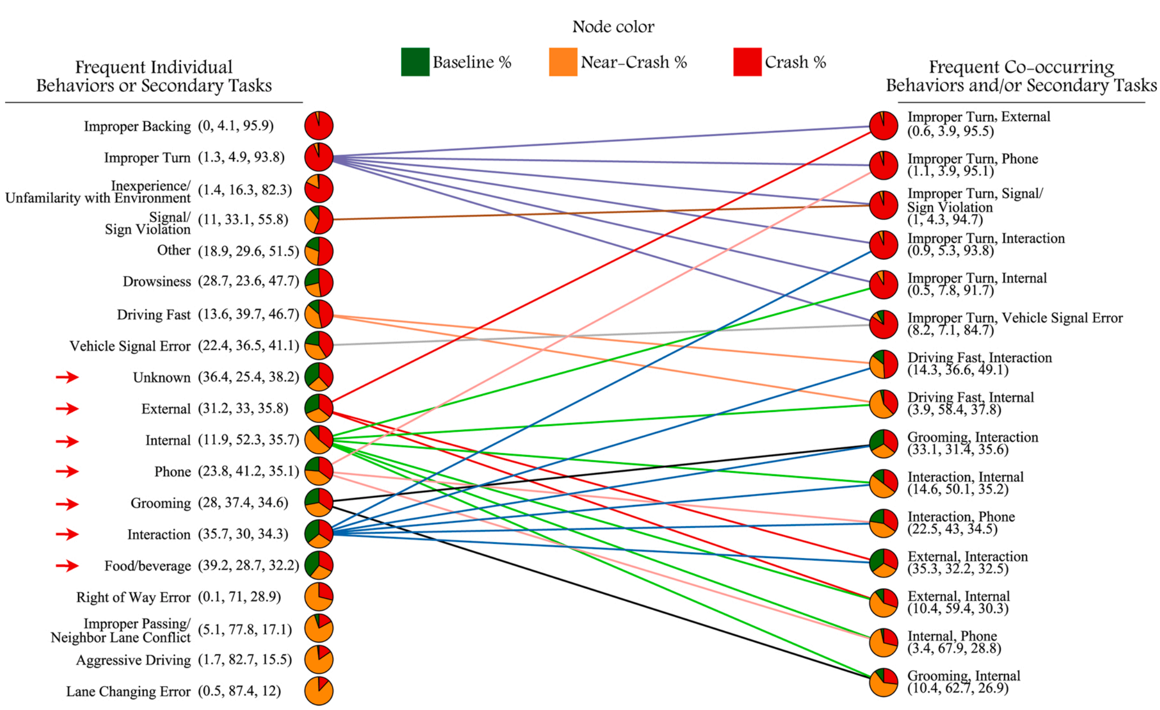
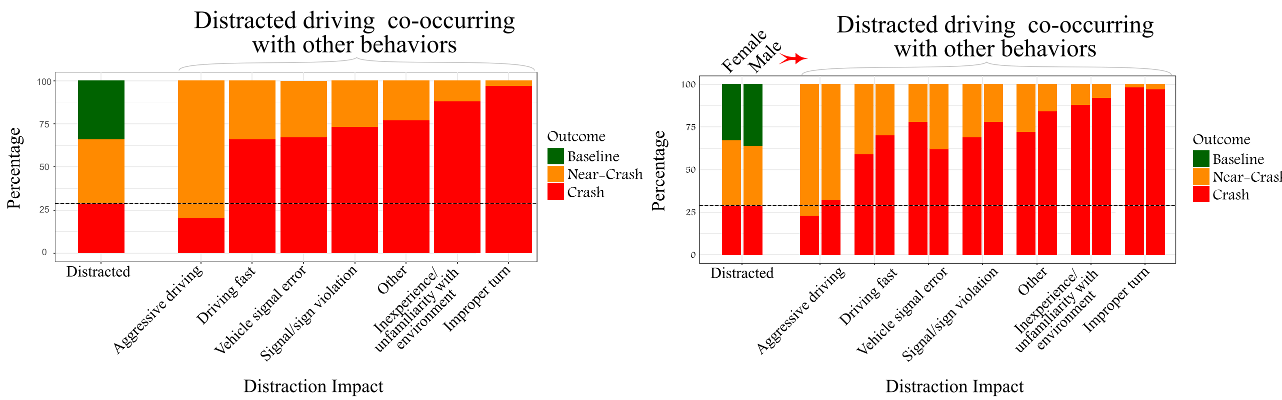


Fig. 4. Co-occurrence driving behavior and secondary task network. Each frequent set of behavior or secondary task has a tuple representing the average percentage of occurrences of the frequent set in the baseline, near-crash, crash categories, and the total data set. The secondary tasks are shown with a red arrow in the left column of nodes. The links originated from the same behavior or secondary task are shown with the same color (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article). Each frequent set on the right columns of behaviors or secondary tasks is composed of two frequent behaviors or two secondary tasks or combinations of each co-occurring together. The nodes are shown as pie-charts representing the normalized percentage of occurrences in different outcome categories.



(a) The percentages of different outcome categories shown for the population of study.

(b) The percentages of different outcome categories in female and male drivers.

Fig. 5. The percentages of different outcome categories involving distracted driving in general and with other co-occurring driving behaviors.

adopted, all the secondary tasks are frequent. Similar to the previous implementation of mining tasks for driving behaviors, the frequencies of secondary tasks in crash epochs are identical. Therefore the SD would be zero since the same set of crash epochs is used in different iterations. Fig. 4 depicts a network for the individual driving behaviors and secondary tasks contributing to the frequent sets of co-occurring behaviors. As shown in the figure, the left group of nodes represents the individual driving behaviors or secondary tasks, while the right column shows the frequent sets composed of two behaviors. Each node on the right column has two edges originated from the individual behaviors and/or secondary tasks on the left column. Similar to Fig. 3, each node on the right column has exactly two edges. However, the number of edges originated from the left column nodes might differ (none to many). A node with a large number of edges co-occur more frequently with other driving behaviors or secondary tasks. A node with no edges shows that the corresponding behavior has not been co-occurred frequently with other behaviors or secondary tasks in the data set. The nodes are shown as pie-charts in this figure and represent the relative percentages of different epoch categories corresponding to each node.

It should be noted that the behaviors considered in this study are not independent, and the probabilities of observing co-occurrences of various behaviors with a given behavior might be different. For example, drivers driving fast might show aggressive driving or improper passing more frequently than other behaviors such as improper backing. This study focuses more on the correlations or co-occurrences of behaviors instead of performing a cause and effect analysis among different behaviors. Therefore, we consider all the behaviors at the same level initially and look at the co-occurrences of other behavior with each specific behavior by mining frequent behaviors. In the following section, the results of this study and potential future works are discussed.

4. Discussion

The results of the adopted approach can be summarized as follows. Table 3 shows that distracted driving is the most frequent behavior. This behavior is observed in almost half of the epochs of different outcome categories. The total number of epochs in each category is 1,843. Table 3 shows that distracted driving is happening in about 47%, 50%, and 39% of baseline, near-crash, and crash epochs, respectively. Fig. 2 visualizes the normalized values for these percentages.

Although not as frequent as distracted driving, based on the normalized values among the three types of epochs, the vehicle signal error is the second behavior uniformly observed among the three types of epochs (Fig. 2). It means that distracted driving and vehicle signal error are observed independently from the outcome in all the epoch categories. On the other hand, in Fig. 3, the impacts of co-occurrences of these two behaviors are visually presented. Considering these two observations (their individual versus their simultaneous occurrences), Fig. 3 shows that the co-occurrences of distracted driving and vehicle signal error change the outcome of the epochs entirely in comparison with their individual occurrences.

Fig. 2 shows that some of the behaviors are observed in specific epoch categories more frequently than others. For example, the behaviors related to lane changing error, aggressive driving, improper passing/neighbor lane conflict, and right of away error are most observed in epochs with a near-crash incidence. And, inexperience/unfamiliarity with the environment, improper turn, and improper backing are more frequently observed in epochs ending in a crash than other epoch categories.

Fig. 3 shows that distracted driving and improper turn are the behaviors most frequently co-occur with other behaviors. The distracted driving co-occurs with seven other, and improper turn co-occurs with six other driving behaviors frequently. Although distracted driving is the most common behavior among different outcome categories, the improper turn is the second most dangerous individual behavior (after improper backing). The co-occurrences of these two behaviors with other behaviors increase the probability of both crash and near-crash occurrences. Although distracted driving behavior is common in different epoch categories, its co-occurrence with other behaviors might change the epoch outcome. Fig. 5(a) shows that the combination of distracted driving co-occurring with the other frequent seven behaviors almost always are observed in near-crash or crash epochs. This figure shows that the co-occurrence of other behaviors with distracted driving increases the probability of crash occurrence more than two times (except for aggressive driving). This probability would be more than three times when distracted driving is observed with inexperienced/unfamiliar with the environment and improper turn driving behaviors. The co-occurrence of aggressive driving might slightly decrease crash probability. However, the combination of distracted and aggressive driving increases the likelihood of near-crash incidents significantly. Fig. 5(b) shows the percentages of different outcome categories involving distracted driving in general and with other co-occurring driving behaviors for different genders. The frequencies of distracted driving are almost the same among male and female drivers. However,

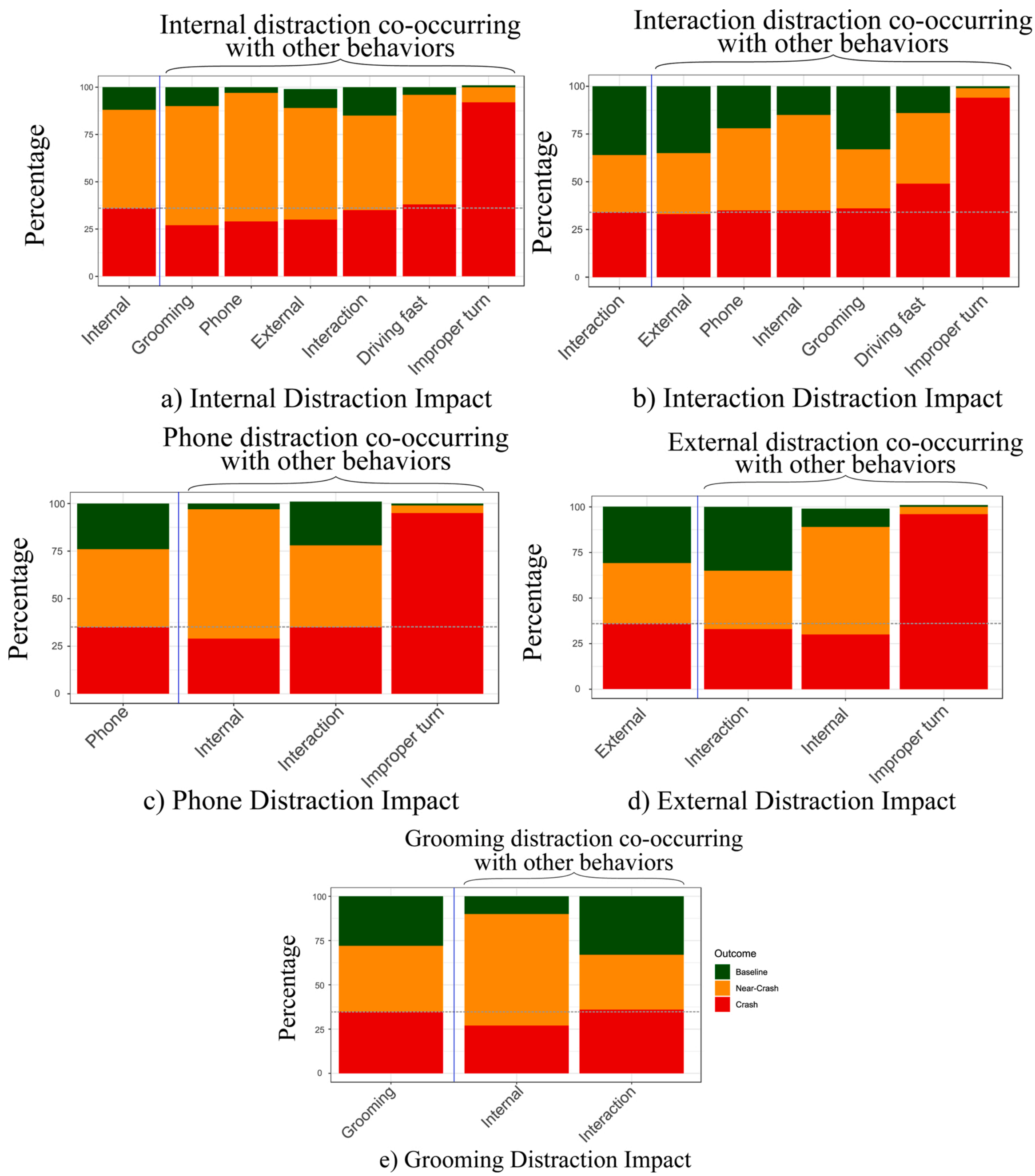


Fig. 6. The changes in percentages of different outcome categories involving secondary tasks (left stacked bar chart in each sub-figure) and with other co-occurring behaviors (stacked bar charts at the right-side of each sub-figure).

the percentages of crashes associated with distracted driving and its co-occurrences with other behaviors are higher for males than females. The higher rates of crashes in male drivers compared to female drivers have been extensively studied and shown in the literature (Massie et al., 1997; Regev et al., 2018). The only exception we found is that female drivers are slightly more involved in crash events caused by the co-occurrences of distracted driving and vehicle signal error.

The same observation can be made for the improper turn. For the frequent behaviors co-occurring with the improper turn, the percentages of near-crash and crash incidents and accidents increase, and the number of baseline epochs consisting of these two behaviors decreases. Furthermore, the most common pair of co-occurring behaviors is composed of distracted driving and improper turn. The drivers of this data set showed this pair of behaviors in 2.5% of epochs on average. And remarkably, on average, in the 7.4% of epochs ended in a crash, this co-occurrence was observed in five seconds before the crash. The pie-chart of this co-occurrence in Fig. 3 shows this co-occurrence results in crash 97.4% of times (and in near-crashes in 2.6% of times), implying that this co-occurrence is the most frequent and most dangerous pair of behaviors simultaneously.

Fig. 3 shows that there is at least one frequent co-occurring behavior for most of the individual behaviors. However, there are a few frequent individual behaviors not frequently observed to co-occur with other behaviors. These behaviors are improper backing, drowsiness, improper passing/neighbor lane conflict, and lane changing error. Fig. 3 shows that there are no outgoing links for these behaviors. Improper backing is the most dangerous individual behavior. On the other hand, drowsiness is among the behaviors commonly observed in all three types of epochs, similar to distracted driving and vehicle signal error. The improper passing/neighbor lane conflict and lane changing error are behaviors with relatively high percentages for near-crash incidents. The variability in the outcome of these behaviors implies that co-occurring patterns depend on the nature of behaviors. In other words, when drivers engage in some of these behaviors (e.g., drowsiness and improper backing), it is not possible for drivers (or at least not frequently observed) to simultaneously engage in other driving behaviors (e.g., driving fast).

Among the secondary tasks, the internal cause of distractions and interactions co-occur more frequently with other behaviors; both co-occur with six other behaviors and secondary tasks (Fig. 4). At the next level are phone and external causes of distraction. However, the occurrences of two food/beverage and unknown type of secondary tasks with other behaviors are not frequent at the support threshold adopted. The columns of nodes in this figure are descendingly ordered based on the crash percentages. It can be seen that secondary tasks are not among the behaviors with the highest rate of crash (secondary tasks in the left column are shown with a red arrow). Table 4 shows that although the frequencies of secondary tasks are different, they almost uniformly occur in different categories of epochs. The main exception for this case is phone usage, which happens more frequently in near-crash and crash categories than in baseline epochs (with about 73% and 42% higher chance of occurrence in near-crash and crash epochs than in baseline epochs, respectively). Based on statistical analysis, another study shows that observable distraction has the highest baseline prevalence but the lowest overall odds ratio among observable impairment, driver momentary judgment error, and observable distraction. However, among the contributing factors to distraction, using handheld electronic devices is among the riskiest secondary tasks with handheld phone dialing having the highest odds ratio among all sources of distraction (Dingus et al., 2016). Another study shows that not the distraction or secondary tasks, but their duration is among the top ten features important for predicting safety-critical events, namely near-crash or crash events (Monselise et al., 2019).

Fig. 6 visualizes the impact of secondary tasks on the epoch outcome when they co-occur with other behaviors or secondary tasks. This figure shows that when each of the secondary tasks frequently occurs with other behaviors or secondary tasks, the chance of more severe outcomes (either near-crash and crash) for the epochs almost always increases.

The results of this study can inform the design and improvement of Advanced driver-assistance and warning systems and countermeasures such as Autonomous Emergency Braking (AEB), Smart Cruise Control, and Lane Departure Warning based on the co-occurring behaviors of drivers. These advanced driver-assistance systems (ADAS) provide the safety margins needed to help avoid or reduce crash risks when the system, for example, detects that driver gets distracted or involved in risky co-occurring behaviors.

5. Conclusion and future works

In this work, we studied the frequent individual and co-occurring behaviors and secondary tasks using a data set created based on a naturalistic driving experiment. The findings show that all the 13 behavior and seven secondary task categories are individually frequent. However, only some of their combinations might be frequent. Among all the 13 driving behaviors, distracted driving is the most frequent behavior. Also, it co-occurs with other behaviors more often than other behaviors. Distracted driving is observed almost uniformly among the baseline, near-crash, and crash epochs.

There are different definitions for distraction and distracted driving in the literature. In this study, we referred to the definition provided in Regan et al. (2011). This definition considers distraction as the attention diversion of the driver toward a competing activity. The attention diversion might originate from different events and sources internal or external to the vehicle. Besides, different types of distraction can arise due to various behavioral and cognitive inattention mechanisms (for example, refer to taxonomies provided in Regan et al. (2011), Regan and Hallett (2011)). To analyze distraction at a more granular level, we studied secondary tasks resulting in distraction. The secondary tasks considered are interactive activities (e.g., interacting with other passengers or pet-related activities), internal distraction (e.g., adjusting/monitoring devices integral to a vehicle, reaching for an object, reading), external distraction (e.g., looking at previous crashes, animals, construction sites), phone-related activities (e.g., talking, texting, and browsing), eating and drinking, and grooming (e.g., shaving and applying make-up). The distraction occurring due to some of these secondary tasks, such as cell-phone use (Unknown, 2003; Lesch and Hancock, 2004; Oscar Oviedo-Trespalacios et al., 2016) and advertising billboards (Crundall et al., 2006; Dukic et al., 2013), have been studied previously. However, we adopted a data-driven approach in this study. Therefore, instead of focusing on specific driving behaviors or secondary tasks, we used association rule mining to identify the frequent individual and co-occurring behaviors and secondary tasks and their impacts on drivers' performance.

Nonetheless, there are some limitations associated with this study. First, the annotations represent the behaviors considered as most critical directly contributing to the epoch outcome. However, the temporal sequences of occurrences of behaviors are not included in the analysis. Although the most critical behaviors are recorded, consideration of their sequential occurrences might be beneficial as well. Besides, this study is limited to the behaviors of drivers. Other relevant data, such as data collected from sensors, surrounding environment, and demographic, physiological, and psychological characteristics of drivers are not considered. An avenue of future research is using individual and co-occurring driving behaviors for prediction of epochs' outcome taking into account the demographic data and physiological and psychological states of drivers. This idea is based on this fact that some of the behaviors are frequent in a specific category of epochs. For example, behaviors related to lane-changing errors are frequent in epochs involving a near-crash incident, or improper backing or turn are frequent in epochs ending in a crash. Therefore, these activities should be good predictors of the corresponding trip categories. The next step, therefore, would be to test and confirm these hypotheses.

CRediT Author statement

Ali Jazayeri: Conceptualization, Methodology, Software, Data Curation, Visualization, Writing – Original Draft

John Ray B. Martinez: Software, Data Curation, Visualization, Writing – Original Draft

Helen Loeb: Supervision, Writing – Review & Editing, Funding acquisition

Christopher C. Yang: Conceptualization, Methodology, Supervision, Writing – Review & Editing, Project administration, Funding acquisition

Acknowledgments

This work was supported in part by the National Science Foundation under the GrantNSF-1741306,IIS-1650531, andDIBBs-1443019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the views of the VTTI, SHRP 2, the Transportation Research Board, or the National Academy of Sciences.

Conflict of interest: The authors declare no conflict of interest.

Appendix A

For the mapping of driving behaviors to the 13 behavior categories and secondary tasks to the seven secondary task categories, please refer to Table 5 and 6 , respectively.

Table 5
The mapping of the 57 driving behaviors into the 13 behavior categories used in the study.

#	Driver behavior	Behavior category
1	Aggressive driving, other	Aggressive driving
2	Aggressive driving, specific, directed menacing actions	Aggressive driving
3	Following too closely	Aggressive driving
4	Distracted	Distracted
5	Exceeded safe speed but not speed limit	Driving fast
6	Exceeded speed limit	Driving fast
7	Speeding or other unsafe actions in work zone	Driving fast
8	Drowsy, sleepy, asleep, fatigued	Drowsiness
9	Improper backing, did not see	Improper backing
10	Improper backing, other	Improper backing
11	Driving in other vehicle's blind zone	Improper passing/neighbor lane conflict
12	Illegal passing	Improper passing/neighbor lane conflict
13	Other improper or unsafe passing	Improper passing/neighbor lane conflict
14	Passing on right	Improper passing/neighbor lane conflict
15	Improper turn, cut corner on left	Improper turn
16	Improper turn, cut corner on right	Improper turn
17	Improper turn, other	Improper turn
18	Improper turn, wide left turn	Improper turn
19	Improper turn, wide right turn	Improper turn
20	Making turn from wrong lane	Improper turn
21	Apparent general inexperience driving	Inexperience/unfamiliarity with environment
22	Apparent unfamiliarity with roadway	Inexperience/unfamiliarity with environment
23	Apparent unfamiliarity with vehicle	Inexperience/unfamiliarity with environment
24	Cutting in, too close behind other vehicle	Lane changing error
25	Cutting in, too close in front of another vehicle	Lane changing error
26	Did not see another vehicle during lane change or merge	Lane changing error
27	Avoiding animal	Other
28	Avoiding other vehicle	Other
29	Avoiding pedestrian	Other
30	Driving slowly in relation to other traffic: not below speed limit	Other
31	Driving slowly: below speed limit	Other
32	Driving without lights or with insufficient lights	Other
33	Improper start from parked position	Other
34	Non-signed crossing violation	Other
35	Other	Other
36	Parking in improper or dangerous location	Other
37	Sudden or improper braking	Other
38	Sudden or improper stopping on roadway	Other
39	Unknown	Other
40	Use of cruise control contributed to late braking	Other
41	Wrong side of road, not overtaking	Other
42	Right-of-way error in relation to other vehicle or person, apparent decision failure	Right of way error
43	Right-of-way error in relation to other vehicle or person, apparent recognition failure	Right of way error
44	Right-of-way error in relation to other vehicle or person, other or unknown cause	Right of way error
45	Disregarded officer or watchman	Signal/sign violation
46	Other sign (e.g., Yield) violation, apparently did not see sign	Signal/sign violation
47	Other sign (e.g., Yield) violation, intentionally disregarded	Signal/sign violation
48	Other sign violation	Signal/sign violation
49	Signal violation, apparently did not see signal	Signal/sign violation
50	Signal violation, intentionally disregarded signal	Signal/sign violation
51	Signal violation, tried to beat signal change	Signal/sign violation
52	Stop sign violation, "rolling stop"	Signal/sign violation
53	Stop sign violation, apparently did not see stop sign	Signal/sign violation
54	Stop sign violation, intentionally ran stop sign at speed	Signal/sign violation
55	Failed to signal	Vehicle signal error
56	Failure to dim headlights	Vehicle signal error
57	Improper signal	Vehicle signal error

Table 6
The mapping of the 63 secondary tasks into the 7 task categories used in the study.

#	Secondary task	Task category
1	Distracted by construction	External
2	Looking at animal	External
3	Looking at an object external to the vehicle	External
4	Looking at pedestrian	External
5	Looking at previous crash or incident	External
6	Other external distraction	External
7	Drinking from open container	Food/ beverage
8	Drinking with lid and straw	Food/ beverage
9	Drinking with lid, no straw	Food/ beverage
10	Drinking with straw, no lid	Food/ beverage
11	Eating without utensils	Food/ beverage
12	Eating with utensils	Food/ beverage
13	Reaching for food-related or drink-related item	Food/ beverage
14	Applying make-up	Grooming
15	Biting nails/cuticles	Grooming
16	Brushing/flossing teeth	Grooming
17	Combing/brushing/fixing hair	Grooming
18	Extinguishing cigar/cigarette	Grooming
19	Lighting cigar/cigarette	Grooming
20	Other personal hygiene	Grooming
21	Reaching for cigar/cigarette	Grooming
22	Removing/adjusting clothing	Grooming
23	Removing/adjusting jewelry	Grooming
24	Removing/inserting/ adjusting contact lenses or glasses	Grooming
25	Shaving	Grooming
26	Smoking cigar/cigarette	Grooming
27	Child in adjacent seat - interaction	Interaction
28	Child in rear seat - interaction	Interaction
29	Dancing	Interaction
30	Passenger in adjacent seat - interaction	Interaction
31	Passenger in rear seat	Interaction
32	Pet in vehicle	Interaction
33	Talking/singing, audience unknown	Interaction
34	Adjusting/monitoring climate control	Internal
35	Adjusting/monitoring other devices integral to vehicle	Internal
36	Adjusting/monitoring radio	Internal
37	Insect in vehicle	Internal
38	Inserting/retrieving CD (or similar)	Internal
39	Moving object in vehicle	Internal
40	Object dropped by driver	Internal
41	Object in vehicle, other	Internal
42	Reaching for object, other	Internal
43	Reaching for personal body-related item	Internal
44	Reading	Internal
45	Writing	Internal
46	Cell phone, Browsing	Phone
47	Cell phone, Dialing hand-held	Phone
48	Cell phone, Dialing hand-held using quick keys	Phone
49	Cell phone, Dialing hands-free using voice-activated software	Phone
50	Cell phone, holding	Phone
51	Cell phone, Holding	Phone
52	Cell phone, Locating/reaching/answering	Phone
53	Cell phone, other	Phone

(continued on next page)

Table 6 (*continued*)

#	Secondary task	Task category
54	Cell phone, Talking/listening, hand-held	Phone
55	Cell phone, Texting	Phone
56	Tablet device, Locating/reaching	Phone
57	Tablet device, Operating	Phone
58	Tablet device, Other	Phone
59	Tablet device, Viewing	Phone
60	Other known secondary task	Unknown
61	Other non-specific internal eye glance	Unknown
62	Unknown type (secondary task present)	Unknown
63	Unknown	Unknown

References

The distraction effects of phone use during a crucial driving maneuver. Accident Analysis & Prevention 354. Anal. Prev. 35 (4), 2003, 501–514. https://doi.org/10.1016/S0001-4575(02)00028-3, 2003.

Abdullah, Umair, Ahmad, Jamil, Ahmed, Aftab, 2008. Analysis of effectiveness of apriori algorithm in medical billing data mining. 2008 4th International Conference on Emerging Technologies 327–331.

Agrawal, Rakesh, Srikant, Ramakrishnan, 1994. Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB `94). Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, pp. 487–499.

Aksan, Nazan, Dawson, Jeffrey D., Emerson, Jamie L., Yu, Lixi, Uc, Ergun Y., Anderson, Steven W., Rizzo, Matthew, 2013. Naturalistic Distraction and Driving Safety in Older DriversHumandistraction and driving safety in older drivers. Factors 55 (4), 841–853. https://doi.org/10.1177/0018720812465769 (2013). PMID: 23964422.

Altaf, Wasif, Shahbaz, Muhammad, Guergachi, Aziz, 2017. Applications of association rule mining in health informatics: a survey. Artificial Intelligence Review. Intell. Rev. 47 (3), 313–340 (2017).

Ameyoe, Ablamvi, Mars, Franck, Chevrel, Philippe, Le-Carpentier, Eric, Illy, Hervé, 2015. Estimation of Driver Distraction Using the Prediction Error of a Cybernetic Driver Model. DSC, Simulation Design and Architecture, Germany (2015).

Anderson, Clare, Horne, James A., 2013. Driving drowsy also worsens driver distraction. Sleep Medicine. 14 (5), 466–468. https://doi.org/10.1016/j.sleep.2012.11.014 (2013).

Bakiri, Sarah, Galéra, Cédric, Lagarde, Emmanuel, Laborey, Magali, Contrand, Benjamin, Ribéreau-Gayon, Régis, Salmi, Louis-Rachid, Gabaude, Catherine, Fort, Alexandra, Maury, Bertrand, Lemercier, Céline, Cours, Maurice, Bouvard, Manuel-Pierre, Orriols, Ludivine, 2013. Distraction and driving: Results from a case-control responsibility study of traffic crash injured drivers interviewed at the emergency room. Accident Analysis & Prevention. Anal. Prev. 59, 588–592. https://doi.org/10.1016/j.aap.2013.06.004 (2013).

Bell, Shannon M., Edwards, Stephen W., 2014. Building Associations between Markers of Environmental Stressors and Adverse Human Health Impacts Using Frequent Itemset Mining, pp. 551–559. https://doi.org/10.1137/1.9781611973440.64 arXiv:https://epubs.siam.org/doi/pdf/. doi: 10.1137/1.9781611973440.64.

Blincoe, Lawrence, Miller, Ted R., Zaloshnja, Eduard, Lawrence, Bruce A., 2015. The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised). Technical Report.

Botta, Marco, Cancelliere, Rossella, Ghignone, Leo, Tango, Fabio, Gallinari, Patrick, Luison, Clara, 2019. Real-time detection of driver distraction: random projections for pseudo-inversion-based neural training. Knowledge and Information Systems. Inf. Syst. 60 (3), 1549–1564. https://doi.org/10.1007/s10115-019-01339-0 (2019).

Brossette, Stephen E., Sprague, Alan P., Michael Hardin, J., Waites, Ken B., Jones, Warren T., Moser, Stephen A., 1998. Association Rules and Data Mining in Hospital Infection Control and Public Health SurveillanceJournal of the American Medical Informatics Associationrules and data mining in hospital infection control and public health surveillance. J. Am. Med. Inform. Assoc. 5 (4), 373–381. https://doi.org/10.1136/jamia.1998.0050373 (1998 07). arXiv:https://academic.oup.com/jamia/article-pdf/5/4/373/2175027/5-4-373.pdf.

Campbell, Elizabeth A., Bass, Ellen J., Masino, Aaron J., 2020. Temporal condition pattern mining in large, sparse electronic health record data: A case study in characterizing pediatric asthma. Journal of the American Medical Informatics Association. Am. Med. Inform. Assoc. 27 (4), 558–566. https://doi.org/10.1093/jamia/ocaa005 (2020 02). arXiv:https://academic.oup.com/jamia/article-pdf/27/4/558/32917026/ocaa005.pdf.

Crundall, David, Editha [Van Loon], Underwood, Geoffrey, 2006. Attraction and distraction of attention with roadside advertisements. Accident Analysis & Prevention. Anal. Prev. 38 (4), 671–677. https://doi.org/10.1016/j.aap.2005.12.012 (2006).

Deshmukh, Shantanu, Dehzangi, Omid, 2017. Identification of real-time driver distraction using optimal subband detection powered by Wavelet Packet Transform. 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN) 9–12. https://doi.org/10.1109/BSN.2017.7935995.

Dingus, Thomas A., Guo, Feng, Lee, Suzie, Antin, Jonathan F., Perez, Miguel, Buchanan-King, Mindy, Hankey, Jonathan, 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proceedings of the National Academy of Sciences. Natl. Acad. Sci. 113 (10), 2636–2641. https://doi.org/10.1073/pnas.1513271113 (2016).

Dukic, Tania, Ahlstrom, Christer, Patten, Christopher, Kettwich, Carmen, Kircher, Katja, 2013. Effects of Electronic Billboards on Driver Distraction. Traffic Injury Prevention. Prev. 14 (5), 469–476. https://doi.org/10.1080/15389588.2012.731546 (2013). PMID: 23682577.

Hahsler, Michael, Grün, Bettina, Hornik, Kurt, 2005. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets, Journal of Statistical SoftwareArticlesa computational environment for mining association rules and frequent item sets. J. Stat. Softw. 14 (15), 1–25. https://doi.org/10.18637/jss.v014.i15 (2005).

Hankey, Jonathan M., Perez, Miguel A., McClafferty, Julie A., 2016. Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets. Virginia Tech Transportation Institute. Technical Report.

Hermannstädter, Peter, Yang, Bin, 2013. Driver distraction assessment using driver modeling. 2013 IEEE International Conference on Systems, Man, and Cybernetics 3693–3698.

Ilayaraja, M., Meyyappan, Thiru, 2013. Mining medical data to identify frequent diseases using Apriori algorithm. 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering 194–199.

Jabbar, M.A., Chandra, Priti, Deekshatulu, B.L., 2011. Cluster based association rule mining for heart attack prediction. Journal of Theoretical and Applied Information Technology. Theor. Appl. Inf. Technol. 32 (2), 196–201 (2011).

Jenkins, Syndney, Codjoe, Julius, Alecsandru, Ciprian, Ishak, Sherif, 2017. Exploration of the SHRP 2 NDS: development of a distracted driving prediction model. Advances in Human Aspects of Transportation. Springer, pp. 231–242.

Jin, Lisheng, Niu, Qingning, Hou, Haijing, Xian, Huacai, Wang, Yali, Shi, Dongdong, 2012. Driver Cognitive Distraction Detection Using Driving Performance MeasuresDiscrete Dynamics in Nature and Societycognitive distraction detection using driving performance measures. Discrete Dyn. Nat. Soc. 2012 https://doi.org/10.1155/2012/432634 (2012 11).

Kareem, Saba, Ahmad, Rohiza Binti, Sarlan, Aliza Binit, 2017. Framework for the identification of fraudulent health insurance claims using association rule mining. 2017 IEEE Conference on Big Data and Analytics (ICBDA) 99–104.

Khare, Sangita, Gupta, Deepa, 2016. Association rule analysis in cardiovascular disease. 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP) 1–6.

Kircher, Katja, Ahlstrom, Christer, 2010. Predicting visual distraction using driving performance data. Annals of advances in automotive medicine. Association for the Advancement of Automotive Medicine. Adv. Automot. Med. 54, 333–342 (2010).

Lesch, Mary F., Hancock, Peter A., 2004. Driving performance during concurrent cell-phone use: are drivers aware of their performance decrements? Accident Analysis & Prevention. Anal. Prev. 36 (3), 471–480. https://doi.org/10.1016/S0001-4575(03)00042-3 (2004).

Li, Zhaojian, Bao, Shan, Kolmanovsky, Ilya V., Yin, Xiang, 2018. Visual-Manual Distraction Detection Using Driving Performance Indicators With Naturalistic Driving DataIEEE Transactions on Intelligent Transportation Systemsmanual distraction detection using driving performance indicators with naturalistic driving data. IEEE Trans. Intell. Transp. Syst. 19 (8), 2528–2535 (2018).

Liang, Yulan, Reyes, Michelle L., Lee, John D., 2007. Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines. IEEE Transactions on Intelligent Transportation Systems. Intell. Transp. Syst. 8 (2), 340–350 (2007).

Liu, T., Yang, Y., Huang, G., Yeo, Y.K., Lin, Z., 2016. Driver Distraction Detection Using Semi-Supervised Machine Learning. IEEE Transactions on Intelligent Transportation Systems. Intell. Transp. Syst. 17 (4), 1108–1120 (2016).

Massie, Dawn L., Green, Paul E., Campbell, Kenneth L., 1997. Crash involvement rates by driver gender and the role of average annual mileage. Accident Analysis & Prevention. Anal. Prev. 29 (5), 675–685. https://doi.org/10.1016/S0001-4575(97)00037-7 (1997).

Monselise, Michal, Liang, Ou Stella, Yang, Christopher C., 2019. Identifying important risk factors associated with vehicle injuries using driving behavior data and predictive analytics. 2019 IEEE International Conference on Healthcare Informatics (ICHI) 1–10.

Transportation Research Board of the National Academy of Sciences, 2013. The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset, Technical Report. Available from the SHRP 2 NDS InSight Data Dissemination web site: https://insight.shrp2nds.us.

Ordonez, Carlos, 2006. Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology in Biomedicine. Inf. Technol. Biomed. 10 (2), 334–343 (2006).

Oscar Oviedo-Trespalacios, Md., Haque, Mazharul, King, Mark, Washington, Simon, 2016. Understanding the impacts of mobile phone distraction on driving performance: A systematic review. Transportation Research Part C: Emerging Technologies. Res. Part C. Emerg. Technol. 72, 360–380. https://doi.org/10.1016/j.trc.2016.10.006 (2016).

Regan, Michael A., Hallett, Charlene, 2011. Driver distraction: definition, mechanisms, effects, and mitigation. In: Porter, Bryan E. (Ed.), Handbook of Traffic Psychology. Academic Press, San Diego, pp. 275–286. https://doi.org/10.1016/B978-0-12-381984-0.10020-7.

Regan, Michael A., Hallett, Charlene, Gordon, Craig P., 2011. Driver distraction and driver inattention: Definition, relationship and taxonomy. Accident Analysis & Prevention. Anal. Prev. 43 (5), 1771–1781. https://doi.org/10.1016/j.aap.2011.04.008 (2011).

Regan, Michael A., Lee, John D., Young, Kristie, 2008. Driver Distraction: Theory, Effects, and Mitigation. CRC Press.

Regev, Shirley, Rolison, Jonathan J., Moutari, Salissou, 2018. Crash risk by driver age, gender, and time of day using a new exposure methodology. Journal of Safety Research. Safety Res. 66, 131–140. https://doi.org/10.1016/j.jsr.2018.07.002 (2018).

Shan, Yin, Jeacocke, David, Wayne Murray, D., Sutinen, Alison, 2008. Mining medical specialist billing patterns for health service management. Proceedings of the 7th Australasian Data Mining Conference - Volume 87 (Glenelg, Australia) (AusDM '08). Australian Computer Society, Inc, AUS, pp. 105–110.

Stavrinos, Despina, Jones, Jennifer L., Garner, Annie A., Griffin, Russell, Franklin, Crystal A., Ball, David, Welburn, Sharon C., Ball, Karlene K., Sisiopiku, Virginia P., Fine, Philip R., 2013. Impact of distracted driving on safety and traffic flow. Accident Analysis & Prevention. Anal. Prev. 61, 63–70. https://doi.org/10.1016/j.aap.2013.02.003 (2013). Emerging Research Methods and Their Application to Road Safety Emerging Issues in Safe and Sustainable Mobility for Older Persons The Candrive/Ozcandrive Prospective Older Driver Study: Methodology and Early Study Findings.

Stutts, Jane, Feaganes, John, Rodgman, Eric, Hamlett, Charles, Reinfurt, Donald, Gish, Ken, Mercadante, Michael, Staplin, Loren, 2003. The causes and consequences of distraction in everyday driving. Annual Proceedings /Association for the Advancement of Automotive Medicine. Association for the Advancement of Automotive Medicine 47, 235–251 (2003 02).

Wang, Shouyi, Zhang, Yiqi, Wu, Changxu, Darvas, Felix, Chaovalitwongse, Wanpracha Art, 2015. Online Prediction of Driver Distraction Based on Brain Activity Patterns. IEEE Transactions on Intelligent Transportation Systems. Intell. Transp. Syst. 16 (1), 136–150 (2015).

Young, Kristie, Regan, Michael, Hammer, M., 2007. Driver distraction: a review of the literature. In: Stevenson, M., Brown, J., Porter, A., Faulks, I.J., Regan, M., Irwin, J.D. (Eds.), Distracted Driving, vol. 2007. CiteSeer, pp. 379–405.