Accident Analysis and Prevention 96 (2016) 198–207

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Identify sequence of events likely to result in severe crash outcomes

Kun-Feng Wu[a,*], Craig P. Thor[b], Muhammad Nashir Ardiansyah (Ph.D. Student)[a]

[a] Department of Transportation and Logistics Management, National Chiao Tung University
[b] Turner-Fairbank Highway Research Center, Federal Highway Administration

ARTICLE INFO

Article history:
Received 13 December 2015
Received in revised form 5 August 2016
Accepted 6 August 2016

Keywords:
Crash sequence
Crash characterization
FARS
Crash severity

ABSTRACT

The current practice of crash characterization in highway engineering reduces multiple dimensions of crash contributing factors and their relative sequential connections, crash sequences, into broad definitions, resulting in crash categories such as head-on, sideswipe, rear-end, angle, and fixed-object. As a result, crashes that are classified in the same category may contain many different crash sequences. This makes it difficult to develop effective countermeasures because these crash categorizations are based on the outcomes rather than the preceding events. Consequently, the efficacy of a countermeasure designed for a specific type of crash may not be appropriate due to different pre-crash sequences. This research seeks to explore the use of event sequence to characterize crashes. Additionally, this research seeks to identify crash sequences that are likely to result in severe crash outcomes so that researchers can develop effective countermeasures to reduce severe crashes. This study utilizes the sequence of events from roadway departure crashes in the Fatality Analysis Reporting System (FARS), and converts the information to form a new categorization called "crash sequences." The similarity distance between each pair of crash sequences were calculated using the Optimal Matching approach. Cluster analysis was applied to group crash sequences that are etiologically similar in terms of the similarity distance. A hybrid model was constructed to mitigate the potential sample selection bias of FARS data, which is biased toward more severe crashes. The major findings include: (1) in terms of a roadway departure crash, the crash sequences that are most likely to result in high crash severity include a vehicle that first crosses the median or centerline, runs-off-road on the left, and then collides with a roadside fixed-object; (2) seat-belt and airbag usage reduces the probability of dying in a roadway departure crash by 90%; and (3) occupants who are seated on the side of the vehicle that experience a direct impact are 2.6 times more likely to die in a roadway departure crash than those not seated on the same side of the vehicle where the impact occurs.

1. Introduction

To effectively target and treat severe crashes, there is a need to first characterize crashes in a way that supports the development of effective countermeasures. This is often referred to as crash type analysis. However, one of the challenges in developing effective countermeasures to improve traffic safety is to characterize crashes in ways that capture the factors that are associated with crash risk while also accounting for their contribution to the crash progression, referred to as the crash sequence in this study. Current crash categorization methods such as head-on, sideswipe, rear-end, angle, and hit fixed-object lacks sufficient details for countermeasure development as it only provides limited information on crash severity (e.g. head-on crashes usually have the most severe outcomes and rear-end crashes usually have the least severe outcomes), and provides almost no information on opportunities to intervene in the crash sequence.

1.1. Crash characterization

Previous research suggests that crash characterization analysis is critical not only for studying how and why crashes occur but for developing targeted countermeasures (e.g. Snyder and Knoblauch, 1971; Knoblauch, 1977; Cross and Fisher, 1977; Retting et al., 1995; Preusser et al., 1995). Consider, as an example, a single vehicle run-off-road crash where driver fatigue or distraction is present, which results in the vehicle leaving the roadway and ultimately striking a roadside object or resulting in a rollover. This scenario will often end in a severe crash outcome. If the crash is defined based on this sequence of events, it makes it possible to consider the countermeasures that may have prevented the final, likely severe

* Corresponding author.
E-mail addresses: kfwu@nctu.edu.tw (K.-F. Wu), Craig.Thor@dot.gov (C.P. Thor), nashirard.tem03g@nctu.edu.tw (M.N. Ardiansyah).

http://dx.doi.org/10.1016/j.aap.2016.08.009
0001-4575/

outcome. For instance, shoulder rumble strips are generally considered a proven, effective safety countermeasure as this treatment provides drivers with auditory or tactile vibrations that could facilitate recovery from roadway departure events and possibly prevent this scenario from resulting in a severe crash. Alternatively, a more forgiving roadside design may prevent the driver from hitting roadside objects or prevent a rollover and hence reduce crash severity. As such, to develop effective countermeasures that may reduce crash occurrence or severity, crash characterization is of the utmost importance.

Preusser et al. (1995) carefully analyzed police accident reports for crash typing and summarized that crash characterization involves pre-crash characteristics, driver behavior, roadway situations, and vehicle movement. Developing crash types is conventionally based on three steps (Preusser et al., 1995): (1) analyzing police crash reports and classifying crashes based on common pre-crash behavior and/or situations; (2) reading additional reports to test the integrity of the preliminary classification; and (3) developing crash type definitions for each of the identified crash groups. For the Fatality Analysis Reporting System (FARS) and National Automotive Sampling System-General Estimates System (NASS-GES) database, the variable of "Crash Type" is used to characterize "the pre-crash circumstances, vehicle's trajectory, driver maneuvers to avoid, and the event that resulted in the most severe injury or, if no injury, the greatest property damage involving this motor vehicle" (NHTSA, 2015). Unfortunately, the categorization presented as the Crash Type is limited in its usefulness when developing effective countermeasures.

The conventional practice of crash characterization in highway engineering combines multiple dimensions of crash contributing factors and categorizes crashes into broad, post-crash categories such as head-on, sideswipe, rear-end, angle, and hit fixed-object. Take motorcycle left-turn oncoming crashes as an example. These crashes often result in severe outcomes and may be characterized as follows: involve multiple vehicle, typically intersection or driveway related; on the roadway; one vehicle turning left and the other going straight; impact points indicate that the vehicle were coming from opposite directions; and a common driver factor: failure to yield right-of-way (Preusser et al., 1995). In a State Department of Transportation's crash inventory data these crashes are usually categorized as either a head-on, sideswipe opposite direction, or angle crash. As a result, there may be crashes that are etiologically similar to the left-turn crashes, but are classified into different categories. This lack of specificity creates a problem when targeting a certain type of crash for safety improvement. As most of crash characterizations do not include the information regarding how the events were connected sequentially, there is only little information available for a safety researcher to identify countermeasures, or worse, the information available for identifying countermeasures may be confusing or even misleading.

1.2. Sequence of events

In a letter of Safety Recommendation the National Transportation Safety Board (NTSB) sent to the National Highway Safety Administration (NHTSA) and Governors Highway Safety Association (GHSA), the NTSB suggested that the use of sequence of events in the FARS data can provide a more direct characterization of types of crashes (NTSB, 2011). As an example, although cross-median crashes are one of the most serious crash types, the NTSB raised concerns about identifying these crashes from the Fatality Analysis Reporting System (FARS). Prior to 2010, there is only one data element used to identify cross-median maneuver in FARS—Cross Median/Centerline, which lacks a clear definition.

The NTSB pointed out questionable examples that may fall within this category but lack significant detail concerning the actual crash event such as a vehicle that departs the roadway, crosses the median, and collides with a tree when compared to a vehicle that crosses the median and, while not hitting another vehicle, causes another vehicle to lose control. The examples above illustrates that a lack information on the sequence of events within a category, crossing the median in this case, would lead to ambiguity and make it difficult to understand the extent of the cross-median crash problem and to develop countermeasures to address them. Therefore, the NTSB suggested that NHTSA provide a more direct characterization of cross-median crashes in the Sequence of Events of the vehicle data element in FARS.

Table 1
Sample sequences.

Sequence	Stage						
	1	2	3	4	5	6	7
Sequence 1	A	A	A	B	B	C	C
Sequence 2	A	A	B	B	B	B	C

Why is dissecting the crash sequence important? Not only because complicated crash causations can be disentangled, but also because effective prevention strategies can be identified. Consider a real-world crash, retrieved from the National Highway Traffic Safety Administration's (NHTSA) National Motor Vehicle Crash Causation Study (NMVCCS) study database, as an example. The crash was on a rural two-lane highway, where a non-contact truck approached from the opposite direction and subject driver steered right. The subject driver's vehicle continued off the right side of the road and ultimately struck a tree. In this crash example, a physical median or center rumble strips may have prevented the approaching truck from encroaching into the center of the lane at the start of the process. More forgiving roadside design or in-vehicle safety device may have helped the driver recover earlier in the crash stage.

As discussed above, each crash contains a set of events in a sequence. This paper intends to make the argument that the sequence of events is just as important as the ultimate crash configuration. In other words, there is a need to identify the sequential character of a crash without reducing it to an unordered set of single events. A sequence is defined as an ordered list of elements, where an element can be a certain event, as shown in Table 1. Here A, B, and C are event types, and they have been marked on a time line. Any partially ordered sets of elements are referred to as an episode, for example "A B B."

Recently, interest in knowledge discovery from sequential data has increased (Pearson and Lipman, 1988; Abbott and Hrycak, 1990; Abbott and Tsay, 2000; Wu, 2000). Sequence data analysis has been widely applied to many scientific fields, in particular, Deoxyribonucleic acid (DNA) sequencing analysis in biology (e.g. Pearson and Lipman, 1988). In the pursuit to map the human genome, researchers identify relationships between episodes, any partially ordered sets of events, which constitute the DNA, and certain outcomes of interest. Conceptually, the DNA sequence analysis is similar to traffic crash sequence analysis. The DNA sequencing analysis focuses on identifying the relationship between certain DNA and a disease or disease predisposition. Similarly, crash sequence analysis focuses on the relationship between certain crash sequences and crash occurrence or severity. By identifying these relationships it becomes possible to gain insights on the development and implementation effectiveness of specific countermeasures. Nevertheless, including the sequential information poses challenges by adding to the complexity of analysis, which grows exponentially with an increase in the number of possible sequences of events involved and the length of sequence. Take the sequence one in Table 1 as an example, suppose there are 10 possible events at each stage, theoretically there would be 10^8