# A synthetic approach to compare the large truck crash causation study and naturalistic driving data

Jeffrey S. Hickman[a,*], Richard J. Hanowski[a], Joseph Bocanegra[b]

[a] Virginia Tech Transportation Institute, United States
[b] National Grid, United Kingdom

## ARTICLE INFO

## ABSTRACT

Truck crashes represent a significant problem on our nation's highways. There is a great opportunity to learn about crash causation by analyzing and comparing the Large Truck Crash Causation Study (LTCCS) and naturalistic driving (ND) data. These data sets provide in-depth information, but have contrasting strengths and weaknesses. The LTCCS contains information on high-severity crashes (crashes and fatal crashes), but relied on data collected during crash investigations. The LTCCS identified principal driver errors in the crash, such as the Critical Reason, but not detailed behaviors or scenario sequences. The ND data sets relate primarily to non-crashes that are detectable from dynamic vehicle events, such as hard braking, swerve, etc., provide direct video observations of the driver and the surrounding driving scene and precise information on driver inputs (kinematics) and captured events, and provide certain types of exposure data that cannot easily be obtained using crash reconstruction data. The ND data are collected continuously, thereby capturing both safety-critical events and normative driving (i.e., baseline). The current project evaluated large-truck crash data from the LTCCS and two large-truck ND data sets, the Naturalistic Truck Driving Study and the Drowsy Driver Warning System Field Operational Test. A synthetic risk ratio analysis on the associated factor, Following Too Closely, indicated that truck drivers in the LTCCS were 1.34 times more likely to be involved in a crash, than an ND crash-relevant conflict, if they were following too closely (i.e., tailgating). Given several caveats noted in the paper, this study suggests it's possible to use the ND data set to calculate the exposure of a given behavior and use the LTCCS data set to calculate the crash exposure to the same behavior.

## 1. Introduction

Large truck and bus (gross weight greater than 10,000 lbs) crashes are a concern for all road users. In 2015, 4.3% of all registered vehicles were large trucks and buses; however, they accounted for 12.4% of all vehicular fatalities (National Highway Traffic Safety Administration, 2017). A better understanding of crash genesis, which can inform regulations, technology development, and safety management techniques, is necessary to prevent and mitigate future large truck and bus crashes. One of the primary aims of the Large Truck Crash Causation Study (LTCCS) was to understand the reasons for serious large truck and bus crashes. The LTCCS included a nationally representative sample of 963 truck crashes that resulted in a serious injury and/or fatality. The data collected in the LTCCS included a detailed description of the events (e.g., driver behaviors and vehicle performance) prior to and during each truck crash, along with an unprecedented amount of information about the vehicles, drivers, and trucking companies involved in the crash, and weather and roadway conditions (FMCSA, 2006). The LTCCS

data relate to post-hoc crash reconstructions based on physical reconstruction, vehicle inspection, and interviews with drivers and witnesses. This type of post-hoc crash investigation, no matter how thorough and in-depth, has the inherent limitation of being an "after-the-fact" reconstruction rather than a direct observation of the crash and the circumstances surrounding the crash.

On the opposite end of the spectrum from crash reconstruction is a data collection approach termed naturalistic driving (ND). ND data collection is a proactive approach which involves data collection while drivers carry out their "day-to-day" operations in vehicles instrumented with sensors and video cameras. In comparison to the traditional approach in generating crash statistics through police accident reports (PARs), ND data have the advantage of recording exactly what happened in the instrumented vehicle prior to, during, and after the safety-critical event. Rather than only being able to collect information at the time of the event, ND data allow for estimations of exposure to various environmental conditions and control conditions (i.e., normative driving incidence of various driver actions and behaviors). These

exposure data (also called baselines) can then be used for rate ratio calculations. Crash databases do not lend themselves as readily to such exposure-based risk analysis; thus, they are limited to frequency counts rather than estimates of risk.

### 1.1. Overview of the Large Truck Crash Causation Study

The LTCCS was conducted jointly by the FMCSA and NHTSA and based on 33 months of data collection. Two-person teams consisting of researchers from NHTSA's National Automotive Sampling System and state truck inspectors assessed over 1000 variables related to the contributing factors in these large-truck crashes. These research teams were located in 24 sites across 17 states. Each truck crash investigated in the LTCCS had at least one large truck (gross wt. greater than 10,000 lbs) and one fatality and/or injury. The LTCCS was based on nationally representative crash counts involving 963 truck crashes from 1123 different large trucks. These 963 truck crashes were assigned sample weights that allowed the LTCCS to estimate the total number of fatal or injury-causing large-truck crashes that occurred during the study (an estimated 120,000 truck crashes; FMCSA, 2006).

Data collection in the LTCCS began as soon as possible after the crash occurred and included: (1) interviews with drivers, passengers, and eyewitnesses, (2) inspections on the truck, drivers' log books, and other pertinent documents, and (3) review of the police, hospital, and coroners' reports. Since the goal of the LTCCS was to determine the contributing factors in large-truck crashes, the data collection focused on pre-crash events. The core of the LTCCS database was the variables relating to the crash event, crash characteristics, conditions of occurrence, and associated factors. One of the key variables of interest related to crash causation in the LTCCS was the critical reason (CR), defined as follows (FMCSA, 2006): "Establishes the critical reason for the occurrence of the critical event. The critical reason is the immediate reason for this event and is often the last failure in the causal chain (*i.e.*, closest in time to the critical pre-crash event). Although the critical reason is an important part of the description of crash events, it is not the cause of the crash nor does it imply the assignment of fault." (p.51).

### 1.2. Overview of the Naturalistic Driving Data Sets

The large-truck ND data sets included two separate studies sponsored by the FMCSA: the Drowsy Driver Warning System Field Operational Test (DDWS FOT) and the Naturalistic Truck Driving Study (NTDS). See Hickman et al. (2016) and Blanco et al. (2010) for a detailed description of the methodologies used to collect data in the DDWS FOT and NTDS, respectively. In total, approximately 3 million miles of driving data and 250,000 h of actigraphy data were collected in these two ND studies. The DDWS FOT was the largest ND commercial vehicle study ever conducted by the U.S. Department of Transportation with over 12 terabytes of kinematic and video data. The DDWS FOT involved three fleet companies across eight locations and 103 drivers. The study used continuous data collection in the 46 trucks which were instrumented to gather kinematic and video data. Each driver in the study was also asked to wear an actigraphy device in order to collect sleep quantity and quality data. The resulting database contains approximately 2.3 million miles traveled and over 8000 days worth of actigraphy data (Hickman et al., 2016).

The NTDS was another ND study using instrumented heavy trucks that collected over 4 terabytes of kinematic and video data. The NTDS involved four fleet companies across seven locations and 100 drivers. As in the DDWS FOT, the NTDS collected continuous driving data from nine instrumented trucks (including kinematic and video data). However, unlike the DDWS FOT, an additional channel of video was collected that allowed a view over the driver's shoulder. Actigraphy devices were also worn by participants in the NTDS. The resulting NTDS database contained approximately 735,000 miles of driving data and 65,000 h of actigraphy data (Blanco et al., 2010). To ensure future

comparisons with the LTCCS, the DDWS FOT and NTDS included many of identical variables in the LTCCS relating to the crash event, crash characteristics, conditions of occurrence, and associated factors.

### 1.3. Project overview

There is a great opportunity to learn about crash causation by analyzing and comparing the LTCCS and ND data. Both of these data sets provide in-depth information, but have contrasting strengths and weaknesses. The LTCCS contains information on high-severity crashes (crashes and fatal crashes), but relied on data collected during crash investigations. The study identified principal driver errors (*e.g.*, CR), but not detailed behaviors or scenario sequences. The ND data sets relate primarily to non-crashes that are detectable from dynamic vehicle events, such as hard braking, swerve, etc. These data provide direct video observations of the driver and the surrounding driving scene and precise information on driver inputs (kinematics) and captured safety-critical events. The ND data sets provide an "instant replay," allowing for the identification of critical driver behaviors leading to traffic conflicts. These studies also provide certain types of exposure data that cannot easily be obtained using crash reconstruction data. More specifically, ND data are collected continuously, thereby capturing both safety-critical events and normative driving (*i.e.*, baseline).

Although the descriptive analyses described in the LTCCS illustrate the prevalence of certain variables, this data set is unable to evaluate which of those variables increases (or decreases) risk of involvement in a crash. This information is critical in identifying which contributing factors and/or associative factors increase crash risk. Thus, those conditions can be targeted by enforcement, safety management techniques, technologies, etc. to avoid the genesis of crashes. The current project evaluated large-truck crash data from the LTCCS and two large-truck ND data sets (*i.e.*, the DDWS FOT and the NTDS). One exploratory *synthetic* odds ratio analysis was performed combining the LTCCS data with ND exposure data. This comparison involved combining the LTCCS crash data as the numerator with ND exposure data as the denominator. Obviously, a major caveat in this comparison was the different sampling frames in the two data sets. The rationale for conducting this comparison was that the exposure data in the DDWS FOT and NTDS were defined in the same manner as the LTCCS data and drawn from the same vehicle type and freight operations. Given this limitation, the synthetic odds ratio may be the best practical approach to estimate crash risk associated with certain environmental situations and driver behaviors in which no other reliable exposure data exist (as in the LTCCS data set).

The FMCSA published "relative risk" statistics regarding the associated factor, *Following Too Closely*, in the LTCCS; these were based on comparisons between crashes where the truck was assigned the CR to crashes where the truck was not assigned the CR. Only one CR was selected in any truck crash. There were 963 truck crashes that were investigated; thus, a CR was assigned to one of these trucks or to the other vehicle, if another vehicle was involved (Craft, 2008; FMCSA, 2007).

An associative factor was a driver behavior, environmental condition, etc. that contributed to the large-truck crash but was not necessarily assigned as the CR. *Following Too Closely* was defined as the vehicle following too closely to the lead vehicle to be able to respond to the actions of the lead vehicle (FMCSA, 2007). Craft (2008) found that the associated factor, Following Too Closely, had a relative risk of 22.6 (*i.e.*, heavy-truck drivers were 22.6 times more likely to be assigned the CR if they were following too closely than if they were not). In itself, this is extremely alarming; however, the methodology used to calculate this statistic had a serious limitation. The relative risk estimate in this case was not an estimate of the increase in risk of a heavy-truck driver following too closely, but rather the increase in risk of being assigned the CR if the truck driver were following too closely. An accurate risk ratio estimate needs to have a reliable estimate for exposure to a risk