



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

Keywords:

Trucks
Naturalistic
Large truck crash causation study
Odds ratio
Following too closely
Tailgating

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

* Corresponding author at: Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg, VA 24061, United States.
E-mail address: jhickman@vtti.vt.edu (J.S. Hickman).

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

factor. The estimate for exposure in this calculation was the crashes in which the truck was not assigned the CR; however, this may not be a typical driving scenario. Naturalistic driving data provides an opportunity to reproduce the relative risk estimate found in the LTCCS for the associative factor, Following Too Closely, by using safety-critical events which did not result in a crash. The CR, Following Too Closely, was selected for this novel comparison between the LTCCS and ND data set given the high relative risk estimate.

2. Method

The NTDS and DDWS FOT data sets were merged to form one ND data set. The LTCCS data set was extracted from data.gov (see reference section for web address) and used the weighted counts. As indicated above, the ND data set used many of the same, or similar, variables and data elements that were employed in the LTCCS, including the associative factor, Following Too Closely. Safety-critical events in the ND data set included (in order of severity) crashes, near-crashes crash-relevant conflicts, and unintentional lane deviations. See [Olson et al. \(2009\)](#) for a detailed operation definition of these safety-critical events in the DDWS FOT and NTDS. Here we summarize the safety-critical events as: (i) crash: any contact with an object, (ii) near-crash: any circumstance that required a rapid, evasive maneuver by the subject vehicle, (iii) crash-relevant conflict: any circumstance that required an avoidance response on the part of the subject vehicle that was less severe than a rapid evasive maneuver (as defined above), and (iv) unintentional lane deviation: subject vehicle crossed over a solid lane line where there was no additional hazard (guardrail, ditch, vehicle, etc.) present.

An odds ratio (OR) estimate was calculated to approximate the risk ratio for the associated factor, Following Too Closely (*i.e.*, tailgating). Typically this analysis involves the use of baseline or control events to illustrate the behavior, tailgating, under normal driving conditions. Over 20,000 baseline epochs [*i.e.*, randomly selected time periods (*e.g.*, 6 s) that did not overlap with a safety-critical event] were included in the ND data set; however, the use of baseline epochs was not possible in this analysis as the sensor suite used to detect safety-critical events in the ND data set would by definition define an instrumented truck following too closely as a safety-critical event. Many of the safety-critical events in the ND data set were not considered high severity (*i.e.*, crashes or near-crashes), but were rather less severe crash-relevant conflicts and unintentional lane deviations. To evaluate the risk ratio of the associative factor, Following Too Closely, “crash-relevant conflicts” were used in place of baseline epochs as a way to estimate risk (*i.e.*, by comparing the odds of a driver tailgating given a high-severity crash in the LTCCS, compared to the odds of tailgating given a crash-relevant conflict in the ND data set). Given the impossibility of following too closely without an additional vehicle, this analysis was performed using only multi-vehicle truck crashes and safety-critical events. More specifically, single vehicle crashes and safety-critical events in the LTCCS and ND data sets, respectively, were excluded. Although unintentional lane deviations were considered the lowest severity events in the ND data set, these safety-critical events were excluded as by definition they involved only one vehicle. Thus, the resulting data set included only multi-vehicle crashes in the LTCCS and multi-vehicle crash-relevant conflicts in the ND data set.

2.1. Analysis approach

An OR estimate is a measure of association commonly employed in the analysis of 2×2 contingency tables ([Agresti, 1996](#)). In the current study, the OR estimates were defined as the probability of event occurrence (*i.e.*, crash in the LTCCS) divided by the probability of non-occurrence (*i.e.*, crash-relevant conflict in the ND data set). These OR estimates were calculated using the presence or absence of the associated factor, Following Too Closely, and a 95% confidence interval was

calculated, including the upper confidence limit (UCL) and lower confidence limit (LCL) ([Schlesselman, 1982](#)). Thus, the data from the LTCCS was used as the numerator with the exposure data in the denominator coming from the ND studies

As the two ND studies used in the analysis employed a retrospective approach, the use of OR estimates rather than relative risk or risk ratios (*i.e.*, ratio of the risk in the exposed divided by the risk in the unexposed) was appropriate. The frequencies obtained from a retrospective and prospective approach are not equal. The prospective relative risk is calculated with the probability of the crash given an at-risk driving behavior; however, the retrospective relative risk is calculated with the conditionality in the other direction (probability of an at-risk driving behavior given a crash). Exposure risk (*i.e.*, risk associated with texting while driving given the driving event is a safety-critical event or a baseline) can be calculated in ND studies as the number of baselines and safety-critical events is known. ORs are not used to approximate the risk ratio in this situations; rather, the OR is used to approximate the rate ratio (*i.e.*, rate of safety-critical events occurring at any given point in time). The current analysis used an OR to approximate the rate ratio as it which does not require the rare event assumption, ([Guo and Hankey, 2009](#)).

3. Results

[Table 1](#) shows the 2×2 contingency table of *Following Too Closely* in multi-vehicle ND crash-relevant conflicts and LTCCS crashes. [Table 1](#) shows there were 66,537 multi-vehicle crashes in the LTCCS (89.5% of the total) and 1644 multi-vehicle crash-relevant conflicts in the ND data set (92.0% of the total) that were not coded with *Following Too Closely*. The remaining 7785 multi-vehicle crashes in the LTCCS (10.5% of the total) and 144 multi-vehicle crash-relevant conflicts in the ND data set (8.0% of the total) were coded with *Following Too Closely*. The OR estimate for the associative factor, Following Too Closely, was 1.34 (with a LCL of 1.12 and an UCL of 1.59). Thus, if drivers were tailgating, or following too closely, they were 1.34 times more likely to be involved in a multi-vehicle crash, than an ND multi-vehicle crash-relevant conflict.

4. Discussion

The synthetic odds ratio for *Following Too Closely* was the first attempt to use the crash frequency data from the LTCCS with exposure data from the ND data set. The comparative analyses above suggested these types of cross comparisons using the LTCCS and ND data sets were feasible. One of the significant limitations in the LTCCS was the lack of baseline data which limits analyses using the LTCCS data set to frequency counts. However, [Craft \(2008\)](#) published “relative risk” statistics regarding the associated factor, Following Too Closely, in the LTCCS. This calculation was based on a comparison between crashes where the truck was assigned the CR to crashes where the truck was not assigned the CR. This analysis found that the associated factor, Following Too Closely, had a relative risk of 22.6. This relative risk estimate can be interpreted as meaning that truck drivers in the LTCCS were 22.6 times more likely to be assigned the CR if they were following too closely than if they were not.

Table 1
 2×2 Contingency Table Used to Calculate the Odds Ratio for Following Too Closely.

Events	Following Too Closely		
	No	Yes	Total
LTCCS Multi-Vehicle Crashes	66,537	7,785	74,322
ND Multi-Vehicle Crash-Relevant Conflicts	1,644	144	1,788
Total	68,182	7,929	76,111

This does not address the risk of *Following Too Closely* as there were no non-crashes in the LTCCS data set (i.e., baseline or control events). The current study assessed the feasibility of combining the LTCCS and ND data to construct a synthetic risk ratio estimate. This type of analysis requires certain assumptions prior to being performed, including: (i) the data collected from both data sets were accurate, (ii) the behavior(s) during baseline epochs in the ND data set were representative of the general driving population, and (iii) any differences in coding between the data sets were reconciled. These estimates should be viewed with an appropriate level of skepticism given that the LTCCS and ND data sets were inherently different. The results of the synthetic odds ratio analysis confirmed that for truck drivers, following too closely significantly increased the odds of being involved in a crash by 1.34 times.

5. Conclusions

With the enormous amounts of data contained in the LTCCS and ND data sets it is important to continue to find methods to extract usable information to answer socially important questions. This analysis proved it was 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. This has important ramifications for crash database analyses that are typically limited to frequency count analyses (as baseline or normal driving behavior is not captured). It appears more sophisticated analyses using crash databases (such as the LTCCS) and ND data sets can yield more precise estimates of environmental conditions and driver behaviors that increase crash risk. This study also demonstrated some of the difficulties that could be encountered in calculating a synthetic risk ratio estimate. Researchers who follow this approach should pay special attention to these issues when designing an approach that compares crash databases and ND data sets.

Although the finding that following too closely was associated with a 1.34 increase in crash risk, rather than the 22.6 reported by Craft (2008), the implications of this should make it clear that CMV drivers should not tailgate. The evidence suggests that practicing good defensive driving techniques is very important, especially with regard to following distance. It appears most CMV safety managers are aware of this issue as a survey of CMV safety managers by Knippling et al. (2003) found that “lack of defensive driving skills” was rated as the fourth most important CMV driver problem area. There are various CMV training approaches that stress “lead time” to avoid a crash likely situation. The data also suggest that forward collision warning systems, which provide an auditory alert when the vehicle is too close to the lead vehicle, may be an effective countermeasure. Battelle (2004) and Fitch et al. (2008) used naturalistic truck data to model the benefits of forward collision warning systems (both found that forward collision warning could reduce approximately 21% of forward collision warning-related crashes). Moreover, (Murray et al., 2009) found that forward collision warning systems provided a significant cost-benefit (i.e., for every dollar spent, carriers get more \$1.33 to \$7.22 back based on different mileages, system efficacies, and purchase prices).

5.1. Limitations

Crash-relevant conflicts in the ND data set were used to calculate the odds ratio. This was necessary as the sensor suite used to detect safety-critical events in the ND data set would by definition define an instrumented truck following too close as a safety-critical event. These

do not represent normative (or baseline) driving; thus, the point estimate and confidence interval in this calculation should be considered a lower limit estimate (as one would expect fewer instances of following too closely under normal or baseline driving conditions, thereby raising the odds ratio). It should also be noted that “following too closely” was defined differently in the LTCCS and ND data sets (subjectively in the LTCCS and objectively in the ND data sets). This may add some noise to the data as it's possible that instances of following too closely were missed in the LTCCS data set.

Acknowledgements

This research was funded by the FMCSA under Contract # DTM75-07-D-00006 (Task Order #8). Mr. Olu Ajayi was the Task Order Manager and Dr. Martin Walker was the Contracting Officer Technical Representative. The authors thank Dr. Walker and Mr. Ajayi for their insight, careful reviews, and prompt attention to administrative matters.

References

- Agresti, A., 1996. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc., New York.
- Battelle, 2004. Phase II Driver Survey Report: Volvo Intelligent Vehicle Initiative Field Operational Test (Contract No. DTFH61-96-C-00077, Task Order 7721). USDOT, Washington, DC.
- Blanco, M., Hickman, J.S., Olson, R.L., Bocanegra, J.L., Hanowski, R.J., Nakata, A., Greening, M., Madison, P., Holbrook, G.T., Bowman, D., 2010. Investigating Critical Incidents, Driver Restart Period, Sleep Quantity, and Crash Countermeasures in Commercial Vehicle Operations Using Naturalistic Data Collection. Federal Motor Carrier Safety Administration, USDOT, Washington DC.
- Craft, R., 2008. Why trucks and buses collide with cars. Presentation at the FMCSA “ART” Forum, TRB 87th Annual Meeting, Washington, DC.
- Data.gov. (n.d.). Large Truck Crash Causation data Extraction. Retrieved from: <http://explore.data.gov/Transportation/Large-Truck-Crash-Causation-Study-LTCCS-/5zg4-dnsb>.
- Federal Motor Carrier Safety Administration, 2006. Report to Congress on the Large Truck Crash Causation Study (LTCCS). MC-R/MC-RRA [No Publication No.].
- Federal Motor Carrier Safety Administration, 2007. The LTCCS. Brochure and Summary. Publication No. FMCSA-RRA-07-017.
- Fitch, G.M., Rakha, H.A., Arafeh, M., Blanco, M., Gupta, S.K., Zimmerman, R.P., Hanowski, R.J., 2008. Safety Benefit Evaluation of a Forward Collision Warning System: Final Report (Report No. DOT-HS-810-910). National Highway Traffic Safety Administration, USDOT, Washington, DC.
- Guo, F., Hankey, J., 2009. Modeling 100-Car Safety Events: A Case-Based Approach for Analyzing Naturalistic Driving Data (Report# 09-UT-006). Retrieved on May 3, 2010 from: National Surface Transportation Safety Center for Excellence, Blacksburg, VA. http://scholar.lib.vt.edu/VTTI/reports/Modeling100CarSafety_Final_041210.pdf.
- Hickman, J.S., Knippling, R.R., Olson, R.L., Fumero, M.C., Blanco, M., & Hanowski, R.J., 2016. *Light Vehicle-Heavy Vehicle Interaction Data Collection and Countermeasure Research Project*. Report# FMCSA-RRR-11-029. Washington, DC: FMCSA, USDOT.
- Knippling, R.R., Hickman, J.S., Bergoffen, 2003. Synthesis 1: Effective Commercial Truck and Bus Safety Management Techniques. Retrieved from: Transportation Research Board, Washington, DC. http://onlinepubs.trb.org/onlinepubs/ctbssp/ctbssp_syn_1.pdf.
- Murray, D., Shackelford, S., Houser, A., 2009. Analysis of Benefits and Costs of Forward Collision Warning Systems for the Trucking Industry (Report No. FMCSA-RRT-09-021). Retrieved from: Federal Motor Carrier Safety Administration, USDOT, Washington, DC. <http://www.fmcsa.dot.gov/facts-research/research-technology/report/09-021-RP-Forward-Collision.pdf>.
- National Highway Traffic Safety Administration, 2017. Traffic Safety Facts 2015: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. Retrieved from: <https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812384>.
- Olson, R.L., Hanowski, R.J., Hickman, J.S., Bocanegra, J., 2009. Driver Distraction in Commercial Vehicle Operations (Report No. FMCSA-RRR-09-042). Retrieved from: USDOT, FMCSA, Washington, DC. <http://www.fmcsa.dot.gov/facts-research/research-technology/report/FMCSA-RRR-09-042.pdf>.
- Schleselman, J.J., 1982. *Case-Control Studies: Design, Conduct, Analysis*. Oxford University Press, Inc., New York, NY.