Jun 24th, 12:00 AM

Naturalistic Driving Events: No Harm, No Foul, No Validity

Ronald R. Knipling

Safety for the Long Haul Inc., Arlington, VA

Knipling, Ronald R.. Naturalistic Driving Events: No Harm, No Foul, No Validity. In: Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 22-25, 2015, Salt Lake City, Utah. Iowa City, IA: Public Policy Center, University of Iowa, 2015: 197-203. https://doi.org/10.17077/drivingassessment.1571

NATURALISTIC DRIVING EVENTS:
NO HARM, NO FOUL, NO VALIDITY

Ronald R. Knipling
Safety for the Long Haul Inc.
Arlington, Virginia, USA
rknipling@verizon.net

Summary: This paper challenges the validity of vehicle-based Naturalistic Driving (ND) Safety Critical Events (SCEs) in relation to injury and fatal crashes. It asserts that mixed SCE datasets have no known or likely representativeness in relation to serious crashes and are likely invalid in regard to their causal factors. This argument is made in the context of ND attempts to associate truck driver Hours-of-Service parameters and safety. But the argument generally applies to other mixed SCE datasets. In part, the challenge is to a monolithic "Heinrich Triangle." Crashes are heterogeneous, both "horizontally" within any severity strata and "vertically" across strata. Serious crashes account for the vast majority of human harm, and are very different from minor crashes. Yet all crashes have, and are defined by, tangible external consequences. In contrast, SCEs are defined by driver maneuvers. Their datasets contain almost no crashes, let alone harm. As such, they are not properly part of the "triangle." Mixed SCE datasets are collections of multiple, disparate driver maneuvers chosen and defined by researchers. They are thus contrived, not analytically derived from the phenomenon of importance, serious crashes. No valid quantitative inferences about the genesis of crash harm can be made from such datasets. This deficiency does not invalidate all ND applications, however. And SCE and real crash datasets could be linked by systematic sampling and case weighting based on objective crash characteristics.

## CRASH HARM AND CRASH HETEROGENEITY

The purpose of traffic safety research is to understand causes and identify interventions relevant to crash harm; i.e., fatalities, injuries, and damage resulting from crashes. Research on phenomena not verifiably linked to crash harm has questionable value. The term *crash harm* has intrinsic meaning but is also defined technically. *Harm* is a quantitative measure of combined human and material loss based on monetary valuation of crashes and injuries of various severities. Crash harm studies (e.g., Zaloshnja and Miller, 2007) tabulate all the property damage, injuries, lost income, and other consequences of crashes of different severities to derive unified measures of crash consequences. Some scales include "pain and suffering," reduced quality-of-life, and related extra-economic consequences. Using harm as a metric permits objective yet human-centered comparisons across different categories of crashes.

| Acronyms |
|---|
| **ND** – Naturalistic Driving |
| **SCE** – Safety-Critical Event |
| **KABCO** – Severity scale for police reports |
| **HOS** – Hours-of-Service |
| **LTCCS** – Large Truck Crash Causation Study |
| **CR** – Critical Reason ("proximal cause") |

Figure 1 shows a "Heinrich Triangle" adapted for crashes, stratified by the five police-reported "KABCO" severity levels coded from the highest crash injury. The levels are: K = Killed; A = Incapacitating injury; B = Non-incapacitating injury; C = Possible injury; O = No injury (also known as Property Damage Only). The Large Truck Crash Causation Study (LTCCS) investigated nationally representative fatal and injury crashes (K, A, and B). Based on statistics in Zaloshnja and Miller (2007), LTCCS-eligible crashes represented just 11% of police-reported large truck crashes but 80-90% of known truck crash harm. Specifically, KAB crashes were 78% of crash costs, 91% of reduced quality-of-life years, and 92% of lost productivity. Relevance to KAB crashes seems required for any study claiming safety significance.
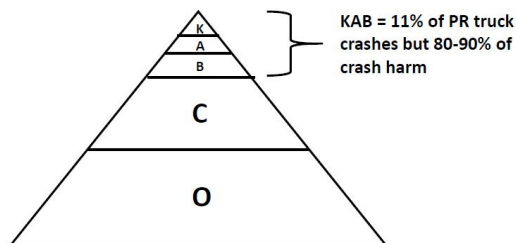
Figure 1. Heinrich's triangle for police-reported crashes

Beneath KAB crashes are possible injury (C) and no injury (O) crashes. These are almost 90% of police-report crashes but only 10-20% of known harm. Beneath police-reported crashes are the many *non*-police-reported crashes. These result in damage and other tangible consequences, but few effects to human wellness. Threshold criteria might include reported to insurance, or requiring vehicle repair. Figure 2 expands the crash triangle to include them.
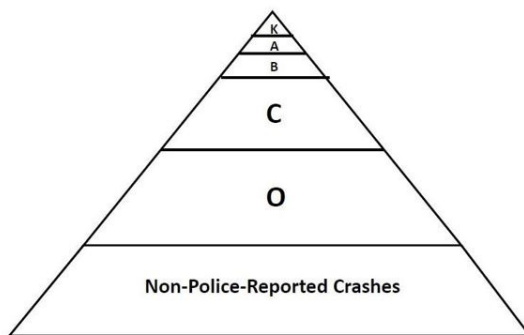
Figure 2. Heinrich's triangle for all crashes (non-police-reported added)

All actual crashes, however, have two pertinent characteristics. First, they have real negative consequences and thus intrinsic importance. Secondly, they are defined and classified by those external consequences. Regardless of their reliability, crash classifications are independent of researcher judgment and are not artifacts of the methodology used to study them.

Underlying Heinrich's Triangle is the assumption that major injuries, minor injuries, and unsafe acts all share the same underlying causal mechanisms (Knipling, 2009, Guo et al, 2010). Yet motor vehicle crashes are heterogeneous, both "horizontally" and "vertically." Horizontal heterogeneity refers to the variety of scenarios seen within any crash severity strata. Road departure, rear-end, lane change/merge, crossing path, and other scenario types are all different from each other. They can be classified using common descriptors and there is overlap among

them, but each major crash type has a distinct profile in regard to when, where, how, and why the crash occurred. Crash heterogeneity is further increased by the simple fact that, for multi-vehicle crashes, the different drivers usually play different roles; i.e., "at-fault," or "not-at-fault."

An example of horizontal crash heterogeneity is provided by Critical Reason (CR) profiles for different crash types. In the LTCCS, truck driver asleep-at-the-wheel was the CR (proximal cause) of 19% of road departures, but 1% or less of rear-end, sideswipe, and opposite direction involvements. Recognition failures (failures to perceive threats) were CRs in 65% of same-direction sideswipes, but only 14% of road departures. Three fundamental crash involvement types are truck single-vehicle (27% in the LTCCS), multi-vehicle truck at-fault (29%), and multi-vehicle truck not-at-fault (44%). The first two of these are both "at-fault," but otherwise they are largely different in regard to conditions of occurrence and causal factors. Their LTCCS CR percentage profiles correlated only +0.18 across 34 CRs (Knipling, 2009a). The third type, truck multi-vehicle not-at-fault involvements, are even more different from the truck perspective since the truck generally plays a passive role. A dataset not capturing these categories proportionately will be skewed in relation to real crash causation.

Vertically, crash profiles can differ sharply by severity level. Table 1 presents seven objective comparisons between 2012 large truck Property Damage Only and fatal crashes (FMCSA, 2014). In addition to comparisons shown, a reliable finding is that the non-truck causal contribution becomes greater as crash severity increases. This is seen even within the three adjacent severity categories of the LTCCS. In truck multi-vehicle involvements, the other motorist was assigned the CR (i.e., at-fault) in 54% of B, 63% of A, and 77% of K involvements (Knipling, 2009b). Driver impairment also increases with crash severity. Police-reported driver fatigue is about five times greater in fatal truck crashes than in all police-reported crashes (FMCSA, 2014).

Table 1. Assorted contrasting characteristics of 2012 Property Damage Only and fatal large truck crashes

Characteristic:	PDO	Fatal
Nighttime (6pm to 6am)	**18.0%**	**35.9%**
Two-way, undivided road	**32.7%**	**53.0%**
Crash occurred in intersection	**12.8%**	**20.4%**
Intersection-related [but not in intersection]	**22.9%**	**5.5%**
Front (of truck) impact	**36.1%**	**58.6%**
Head-on crash, passenger vehicle crossed road center	**0.2%**	**17.7%**
Pedestrian/Pedalcyclist	**<0.1%**	**9.1%**

NATURALISTIC DRIVING EVENTS AND CRASHES

Naturalistic Driving (ND) Safety-Critical Events (SCEs) are mostly driver evasive maneuvers or other dynamic events captured using videos and other instrumentation. A large ND study (Blanco et al., 2011) focusing on Hours-of-Service (HOS) related factors captured 2,197 dynamically triggered SCEs. Yet only four (4) of the 2,197 SCEs (0.2%) were crashes, defined as events with "any contact." In the *100-Car Naturalistic Driving Study*, 69 of 9,125 SCEs (0.8%) resulted in "any impact." Of these, five (5) caused injuries and another seven (7) were police-reported PDO, making the total police-reported percentage 0.1% (Guo et al., 2010). If crashes themselves vary greatly in regard to when, where, how, and why they happen, what is the likely representativeness of events that are not even crashes?

Crash severity categories are defined by their common external consequences. In contrast, SCE datasets are collections of disparate dynamic events, each type with its own unique, *researcher-selected* criterion. Such datasets are "synthetic." They are *contrived* by researchers rather than *derived* from crash characteristics, probability, or harm. In a large HOS-related study (Hanowski et al., 2008) the principal SCE triggers were hard-braking, short times-to-crash, and swerves. In the follow-on 2011 Blanco et al study, unintentional lane deviations were added as an event type. Figure 3 illustrates SCEs in relation to the Heinrich Triangle for crashes. The five ovals represent five possible SCE types, each individually defined and then mixed in a dataset. There's no analytic link to serious (or even minor) crashes. How would we know, and why should we believe, that SCEs are representative of the distal events at the top of the triangle?
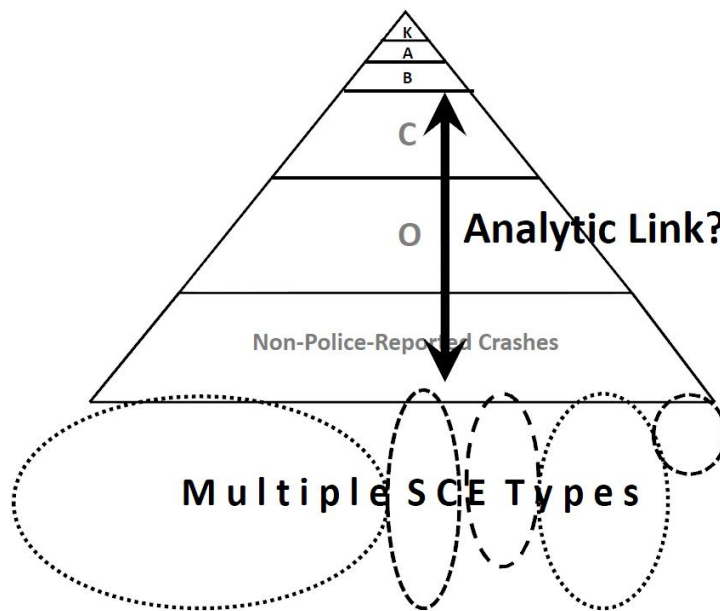

Figure 3. Heinrich's triangle for all crashes plus multiple SCE types constituting SCE datasets

Hickman et al. (2005) coded 915 combination-unit truck ND SCEs using variables similar or identical to those in the LTCCS. There were many similarities to LTCCS profiles, but also stark differences, as shown in Table 2. Even a few major discrepancies invalidate assumptions one might make about overall representativeness. How would you know which conclusions drawn from SCEs have external validity in relation to crashes, and which do not?

Table 2. Contrasting characteristics of Combination-Unit Truck SCEs and LTCCS crash involvements

Truck/Truck Driver Characteristic:	SCE	KAB Crashes
No avoidance maneuver by truck	1.5%	43.8%
Rear-end event, truck would have been/was striking	43.1%	12.3%
Rear-end event, truck would have been/was struck	0.5%	5.7%
Opposite direction event, other vehicle encroached	0.4%	5.6%
Truck "at-fault" (assigned CR, % of all events)	80.9%	55.3%
Of CRs assigned to trucks:		
Truck driver critical non-performance (e.g., asleep, ill)	1.6%	6.5%
Inadequate surveillance (e.g., looked but did not see)	4.2%	6.6%
Vehicle-related failure	0.0%	4.7%

Major ND studies (e.g., (Hanowski et al., 2008, Blanco et al., 2011) have employed SCE rate as a surrogate measure of driver fatigue. It's not. In the Hickman et al. (2005) truck ND study, only *one* of 915 SCEs (0.1%) involved truck driver asleep-at-the-wheel as the CR. That's compared to 3.8% in the LTCCS, a 38-fold difference. How can one reliably measure fatigue effects and factors in such a dilute testbed? Further, SCE occurrence seems to indicate *driver alertness*. Wiegand et al. (2008) observed 3,270 truck SCEs and found an *inverse* relationship between SCE occurrence and drowsiness. Relative SCE risk was 1.9 times *greater* when drivers were rated *less* fatigued. Relative SCE risk was also 1.7 times greater when percent eye closure (PERCLOS) was *below* the fatigue threshold. In Hanowski et al. (2008), SCE rates were lowest in pre-dawn hours when fatigue and asleep-at-the-wheel crashes are greatest.

To increase the fatigue content of their truck SCE dataset, Blanco et al (2011) added 1,118 unintentional lane deviations to an unspecified mix of 1,075 other SCEs. Two scientific issues arise; first, there was no crash basis for the number of lane deviations added or the overall mix. This is like cooking without a recipe. Secondly, though lane deviations can be "a reliable indicator of fatigue," as the report states, it is not a *specific* indicator. A prior ND study (Olson et al., 2009) found 77.5% of lane deviations to be *distraction*-related; i.e., drivers were performing tertiary tasks. Barr et al. (2011) viewed SCEs to compare drowsiness and distraction, and found them more opposite than alike. Distracted drivers are generally alert, not drowsy. Blanco et al. could have *tested* the role of fatigue in their SCEs by describing and analyzing them. SCE analysis by time-of-day, divided vs. undivided highway, driver maneuvers, single-vehicle vs. multi-vehicle, "fault," specific CR (i.e., asleep-at-the-wheel), observer ratings of drowsiness, PERCLOS, or other classifications would have captured and quantified fatigue. Yet unfiltered SCE rate was the only dependent variable used.

Few studies have critically examined the validity of SCEs in relation to crashes. A comparison of near-crashes and crashes in the 100-Car Naturalistic Driving Study (Guo et al., 2010) was an internal consistency check, not an external validity check. Yet its findings are relevant. The study compared 761 near-crashes to 69 crashes within the overall dataset of 9,125 SCEs. "Any impact" defined the 69 crashes, of which only 12 (0.1% of all SCEs) were police-reported. All the SCEs here were gathered in the same study from the same drivers and using the same dynamic parameters. Near-crashes and crashes were adjacent categories, all making this perhaps the easiest test of SCE realism imaginable. Guo et al found near-crash/crash profiles to be highly similar for many descriptive variables, including driver gender, age, lighting condition, road alignment, surface condition, and weather. For example, for six codable weather conditions, the distributions of near-crashes and crashes correlated +0.99. Several key causation-related variables did not show high correlations, however. Most notably, the two profiles for 54 precipitating factors (e.g., object in roadway, other vehicle crossing straight across path, etc.) correlated only +0.18. The correlation between the two distributions for conflict type (e.g., single-vehicle, conflict with lead vehicle, conflict with vehicle in adjacent lane, etc.) was +0.44. Single-vehicle scenarios (here including object/obstacle and parked vehicle) were 37 of 69 crashes (54%), versus just 59 of 761 near-crashes (7.8%). A much higher percentage of crashes (41/69 = 59%) than near-crashes (244/761 = 32%) occurred under low-traffic conditions.

Guo et al reported that *100-Car* drivers reacted to crash threats in only 45 of 68 crashes (66%), versus 723 of 760 near-crashes (95%). The report (P. 23) stated that, "this difference shall not be considered as evidence against the identical causal mechanism . . . [as it] . . . is considered as the

risk factors that trigger the safety events, not the driver's last response to avoid a crash." Yet, in the classic Indiana Tri-Level causal taxonomy, three of the four *main categories* of driver error are defined by faulty or absent driver reactions. These include critical non-performance (e.g., asleep), recognition failures (which delay or fail to evoke a driver reaction) and performance/ response execution errors (e.g., steering overcompensation). These three categories, all involving absent or faulty evasive maneuvers, were 65% of known driver CRs in the LTCCS (Starnes, 2006), and 63% in the National Motor Vehicle Crash Causation Survey (NHTSA, 2008). Are we to believe that "crash causal mechanisms" do not include driver reactions to threat, or the lack thereof? Or is it that SCEs and crashes may have different causal mechanisms?

Though their report emphasized crash/near-crash similarities, Guo et al (P.48) concluded the following: "There is no debate that crashes and near-crashes are two different types of events. This is not only true by operational definition but several results in this report demonstrate that the two cannot be completely identical. However, this does not eliminate using near-crashes as crash surrogates for a specific purpose." This author agrees! Specific SCEs may be modeled against specific, tightly defined crashes; e.g., hard braking and forward collisions in traffic. But the lack of a common criterion or link to crashes across a mixed SCE dataset means that its characteristics (i.e., profiles of when, where, how, and why) are not likely to match up *generally* with those of serious crashes. Associations between causal factors (e.g., HOS parameters) and mixed SCE datasets cannot be extrapolated to serious crash datasets. If Guo's near-crashes and "any impact" crashes from the same SCE dataset were "different types of events," then how much more different would be the population of injury and fatal crashes?

SCE rates can be predictive of individual driver crash rates, particularly crashes of minimal severity. For *100-Car Study* drivers, Guo et al reported a modest but statistically significant crash/near-crash correlation of +0.21. Prediction of individual crash rates has both scientific and practical importance, but that predictiveness does not imply concordance across crash causes. A single driver trait like tendency to speed could account for all or most of the relationship. SCEs reflect driver "style." Driver "style" affects safety, but the factors underlying many serious crashes are different, more extreme, and not necessarily apparent from everyday driving.

GETTING REAL

The blacktop slang phrase "No Harm, No Foul!" suggests that behaviors without consequences have no significance. A playground scientist might add, "No Harm, *and No Verified Relation to Harm*, No Foul!" There's a chasm between SCEs and serious crashes. If ND studies must use non-crash SCEs, they should at least follow standard best practices of sampling and validation. Define, specifically, an external target crash population. Select and refine SCE types and triggers based on target crash objective characteristics. Differentially weight SCEs until their profiles match crashes. Cross-validate across multiple benchmarks, just as you would validate a survey sample against its target population. Ensure that operational measures (i.e., of fatigue) have construct validity. Consider alternatives to SCEs, such as driver performance measures taken during the lonely highway driving periods when drivers actually fall asleep and crash. The onus should not be on skeptics to fact-check every SCE-based inference. It should be on ND researchers to prove the meaningfulness of their datasets in relation to actual target crash populations and to real harm.

ACKNOWLEDGMENTS

REFERENCES

Barr, L.C., Yang, D., Hanowski, R. J., and Olson, R. (2011) *An assessment of driver drowsiness, distraction, and performance in a naturalistic setting*. FMCSA-RRR-11-010, Washington DC: Department of Transportation (DoT).

Blanco, M., Hanowski, R. J., Olson, R.L., Morgan, J. F., Soccolich, S. A., Wu, S-C, and Guo, F. (2011) *The impact of driving, non-driving work, and rest breaks on driving performance in commercial motor vehicle operations*. Report No. FMCSA-RRR-11-017. Wash. DC: DoT.

FMCSA. (2014) *Large truck crash facts 2012*. FMCSA-RRA-14-004. Wash., DC: DoT.

Guo, F., Klauer,S.G., McGill, M.T., and Dingus, T.A. (2010) *Evaluating the relationship between near-crashes and crashes: can near-crashes serve as a surrogate safety metric for crashes?* NHTSA Report DOT HS 811 382. Wash., DC: DoT.

Hanowski, R. J., Olson, R. L., Bocanegra, J. and Hickman, J.S. (2008) *Analysis of risk as a function of driving-hour: assessment of driving-hours 1 through 11*. Report No. FMCSA-RRR-08-002, Wash., DC: DoT.

Hickman, J.S., Knipling, R.R., Olson, R.L., Fumero, M., Hanowski, R.J., & Blanco, M. (2005) *Phase 1 - preliminary analysis of data collected in the drowsy driver warning system field operational test: task 5, phase i data analysis*, NHTSA Contract DTNH22-00-C-07007, TO #21. Wash., DC: DoT.

Knipling, R.R. (2009a) Three large truck crash categories: what they tell us about crash causation. *Proceedings of the Driving Assessment 2009 Conference*, Big Sky, MT, 31-37.

Knipling, R.R. (2009b) *Safety for the long haul; large truck crash risk, causation, & prevention*. Arlington, Virginia: American Trucking Associations. ISBN 978-0-692-00073-1.

NHTSA. (2008) *National Motor Vehicle Crash Causation Survey (Report to Congress, July 2008)*. DOT HS 811 059, Washington, DC: Department of Transportation.

NHTSA. (2014) *Traffic safety facts: 2012 data*. DOT HS 812 016, 2014.

Olson, R.L., Hanowski, R.J., Hickman, J.S., & Bocanegra, J. (2009). *Driver distraction in commercial vehicle operations*. FMCSA-RRR-09-042. Wash. DC: DoT.

Starnes, M. (2006) *LTCCS: an initial overview*. NHTSA Publication DOT HS 810 646, NHTSA, Washington, DC: Department of Transportation.

Zaloshnja, E. and Miller, T. (2007) *Unit Costs of Medium & Heavy Truck Crashes*. FMCSA Report, http://ai.volpe.dot.gov/carrierresearchresults/pdfs/crash%20costs%202006.pdf, Wash., DC: DoT.