BEFORE THE

OFFICE OF THE SECRETARY OF TRANSPORTATION

DEPARTMENT OF TRANSPORTATION

COMMENTS OF

SAFETY FOR THE LONG HAUL, INC.

IN RESPONSE TO NOTIFICATION OF REGULATORY REVIEW

Docket Number: DOT-OST-2017-0069

Ronald R. Knipling, Ph.D.
President, *Safety for the Long Haul Inc.*
5059 North 36th Street
Arlington, VA 22207-2946
rknipling@verizon.net
www.safetyforthelonghaul.com

November 28, 2017

Safety for the Long Haul Inc. appreciates the opportunity provided by the Department to bring its attention to regulations that are candidates for repeal, replacement or modification. These comments address the scientific basis of commercial truck driver Hours-of-Service (HOS) regulations promulgated by the Federal Motor Carrier Safety Administration (FMCSA).

Safety for the Long Haul Inc. is the incorporated consultancy of Dr. Ronald R. Knipling. Dr. Knipling has more than 30 years' experience in large truck safety research and development. He is the author of the first and only comprehensive textbook on large truck safety, entitled *Safety for the Long Haul; Large Truck Crash Risk, Causation, & Prevention*. In recognition of the book, he received the International Road Transport Union (IRU) Order of Merit award, the first given to an American scientist. Driver fatigue, HOS, and Naturalistic Driving (ND) have been among the biggest focus areas of Dr. Knipling's research. During six years at NHTSA and also six years as Research Division Chief at FMCSA, Dr. Knipling conceived and managed many of the seminal ND studies on driver fatigue. As Senior Transportation Fellow with Virginia Tech Transportation Institute (VTTI) for seven years, he managed data collection for the *100-Car Naturalistic Driving Study*[1] and directed event coding and analysis for the first large truck study to address crash causation using ND data[2].

It was during these years, however, that Dr. Knipling recognized anomalies in ND methods and data leading him to doubt the validity and representativeness of recorded ND events in relation to crashes causing actual human harm and specifically in relation to truck driver asleep-at-the-wheel crashes. As an independent and primarily self-funded researcher since that time, he has criticized ND HOS research in regard to its internal validity (demonstration of cause-effect relationships), external validity (representativeness in relation to crashes, especially serious crashes), and construct validity (accurate measurement of driver fatigue/performance deterioration).[3] Dr. Knipling's concerns regarding the validity of FMCSA ND research form the primary basis for the comments presented here.

Comments will focus specifically on the commercial driver HOS rule at 49 CFR 395.3(a)(3)(ii). That rule provides that, except for drivers who qualify for certain short-haul exceptions, "driving

[1] Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R.R. *The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment*. NHTSA Report No. DOT HS 810 593, 2006.

[2] Hickman, J.S., Knipling, R.R., Olson, R.L., Fumero, M., Hanowski, R.J., & Blanco, M. *Phase 1 - Preliminary Analysis of Data Collected In The Drowsy Driver Warning System Field Operational Test: Task 5, Phase I Data Analysis*, for the FMCSA under NHTSA Contract DTNH22-00-C-07007, TO #21, September 30, 2005.

[3] Recent publications expressing these concerns include: Knipling, R.R. Threats to scientific validity in truck driver hours-of-service studies. *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Pp. 382-388, Manchester Village VT, June 26-29, 2017 (graph in presentation); also: Knipling, R.R. Toward naturalistic driving crash representativeness. *Tenth SHRPII Safety Data Symposium*, Washington DC, October 6, 2017.

is not permitted if more than 8 hours have passed since the end of the driver's last off-duty or sleeper-berth period of at least 30 minutes." In other words, the rule requires a 30 minute, off-duty rest break for drivers during a maximum 8-hour period of work.

While repeal or reform of the mandatory break rule is the specific objective sought by this submission, the argument herein applies to all HOS regulations promulgated since 2011 based primarily on the Naturalistic Driving (ND) Mixed "Safety-Critical Event" (SCE) methodology. This methodology involves instrumenting vehicles to detect miscellaneous kinematic events such as hard brakings, swerves, lane drifts, and short times-to-collision. ND researchers choose the types of events they consider important and set trigger thresholds at desired levels to achieve their desired proportions of these events in a combined kinematic event database. *Safety for the Long Haul Inc.* believes that this methodology is not true empirical science because event types and proportions are not derived from the actual problem of concern; i.e., serious crashes. Surrogate measures like SCEs are used in various branches of science, but a requirement is that surrogates be validated as representative of the target problem. To be valid, SCE datasets must be shown to be representative of the real-world population of large truck serious crashes and/or the driver state of extreme fatigue. There has been no such validation. On the contrary there is compelling and growing evidence that SCE datasets are *not* representative of either serious crashes or driver fatigue. This submission presents and documents the evidence against SCE dataset validity and asks that HOS rules based primarily on this methodology be suspended or rescinded.

DOT's Notice of Regulatory Review listed 11 potential reasons the public might cite in requesting suspension, repeal, replacement, or modification of a Federal transportation regulation. Reason #7 on Page 45752 of the notice offers the public the opportunity to request that the Department "(7) reconsider regulations that were based on scientific or other information that has been discredited or superseded." This submission requests suspension, repeal, replacement, or modification of the mandatory 30-minute break rule based on this Reason #7. While there are likely other valid reasons for reconsidering the rule, this submission focuses entirely on the issue of scientific validity. We believe that the rule was never based on sound science and that compelling evidence has been produced to discredit the mixed-SCE methodology. Among other reference citations, we will cite a finding of the National Academies of Sciences, Engineering, and Medicine[4] corroborating the view that unfiltered SCE datasets do not validly represent driver fatigue.

The information cited and arguments expressed herein were developed independently by *Safety for the Long Haul Inc.* Its research conducted since 2011 on this topic has been 100% self-

---

[4] National Academies of Sciences, Engineering, and Medicine. 2016. *Commercial Motor Vehicle Driver Fatigue, Long-Term Health, and Highway Safety: Research Needs*. Washington, DC: The National Academies Press. https://doi.org/10.17226/21921.

funded. This extensive work has been motivated solely by concerns about scientific validity and, therefore, the scientific basis of HOS policy decisions. Though self-funded in this endeavor and acting independently, *Safety for the Long Haul Inc.* has conferred with various individuals and organizations in preparing these comments. Most notably, the *Owner-Operator Independent Drivers Association Foundation Inc. (OOFI)* has reviewed these assertions, concurred with them, and is including them in its own commentary to DOT.

The requirement for a mandatory 30-minute break following eight hours of continuous work and other key provisions of the 2011 HOS rulemaking were promulgated primarily on the basis of a report entitled, "The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations by Blanco et al. (2011)[5]. This study employed an ND SCE methodology in long- and line-haul trucking operations. Blanco et al reported associations between hours of driving and SCE rate as well as before- and after changes associated with breaks from driving. However, only 4 of Blanco's 2,197 SCEs (0.2%) were actual crashes; the other 99.8% were non-crash kinematic events such as hard-braking or swerves. Such harmless surrogate events have no *intrinsic* significance; to be significant, they must be validated against actual harmful crashes or against a known hazardous condition such as driver drowsiness. A simple analogy is a political survey; it has no validity unless it is demonstrably representative of an actual voter population. The same requirement should when SCEs are studied in lieu of studying actual crashes or actual cases of driver fatigue.

The following are ten key points indicating that Blanco's SCE dataset was not representative of actual harmful truck crashes or driver fatigue, and therefore was not a sound scientific basis for rulemaking:

1. **SCEs Never Validated**. Neither FMCSA nor its principal contractor, Virginia Tech Transportation Institute (VTTI), has published data showing that SCEs gathered in large truck ND studies are representative of either serious crash risk or of driver drowsiness and fatigue. Most crash harm and most driver fatigue crashes are at the highest severity levels of the "crash triangle." Overwhelming, ND SCEs are harmless kinematic events. No one has shown an analytic link or provided other evidence that SCEs represent asleep-at-the-wheel or other serious truck crashes. This scientific deficiency has been noted in research needs statements of two different Transportation Research Board (TRB) committees, the Committee on Safety Data, Analysis, and Evaluation (ANB20) and the Committee on Truck

[5] Blanco, M., Hanowski, R. J., Olson, R.L., Morgan, J. F., Soccolich, S. A., Wu, S-C, and Guo, F. *The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations*. Report No. FMCSA-RRR-11-017, May 2011.

and Bus Safety (ANB70)[6]. See Figure 1.

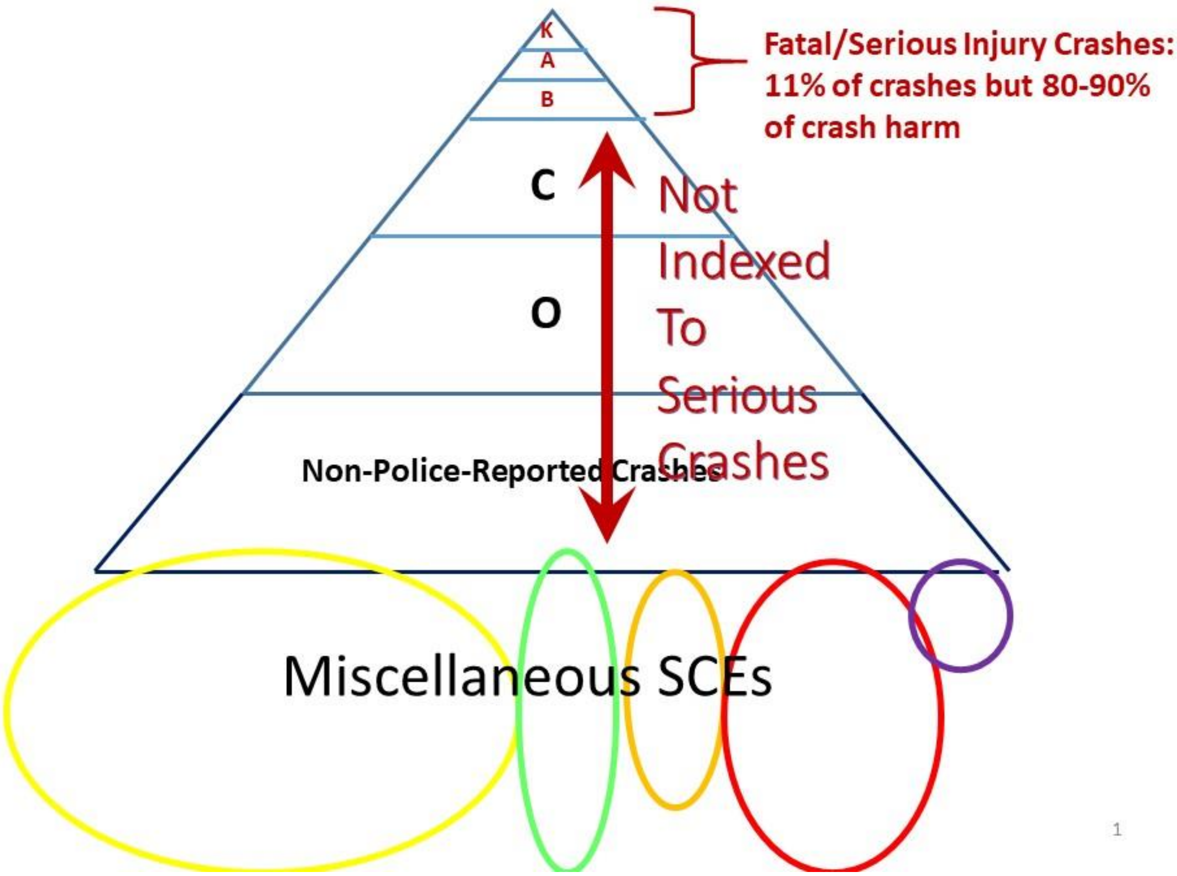


Figure 1. Crash triangle stratified by severity and, below, multiple SCE types constituting SCE datasets, which are overwhelmingly non-crashes. (Knipling, 2017; see Attachment 1).

2. **No Reason to Assume They are Valid**. SCEs are sampled from abrupt but overwhelmingly harmless kinematic events. They are not sampled from the actual target problems – serious crashes and fatigue. There's no *a priori* reason to believe that an arbitrary assortment of SCEs would be representative of any other type of events, outcomes, or condition. As with any other surrogate measure, a relationship with real, important outcomes must be demonstrated with data.

[6] TRB Committee on Safety Data, Analysis, & Evaluation/ANB20. (2016) Indexing Naturalistic Driving Events to Crashes. Research Needs Statement available at https://rns.trb.org/dproject.asp?n=40810; TRB Committee on Truck & Bus Safety/ANB70. (2015) Toward Naturalistic Driving Crash Representativeness (23-2015). Research Needs Statement available at https://rns.trb.org/dproject.asp?n=39354.

3. **SCE Rates Driven by Traffic Interactions**. FMCSA's previous HOS-related ND study in 2008[7] concluded that SCEs primarily reflected "hour-by-hour traffic density variations." The correlation between SCEs rate and average U.S. traffic density across the 24-hour day was +0.83. Thus SCEs primarily reflect external driving conditions faced by drivers, not the alertness states of drivers.

4. **SCEs are Associated with *High* Alertness**. An analysis of truck ND videos conducted at VTTI in 2008 by Wiegand et al[8] found that that SCE involvement was strongly associated with *higher* levels of alertness, not drowsiness. The odds ratio association of high alertness (low drowsiness) with SCE involvement was 1.93 (where 1.0 represents equivalence). The analysis also corroborated the very strong positive association (odds ratio = 7.2) of SCE rate with traffic density, further showing that SCEs mainly reflect interactions of trucks with other traffic.

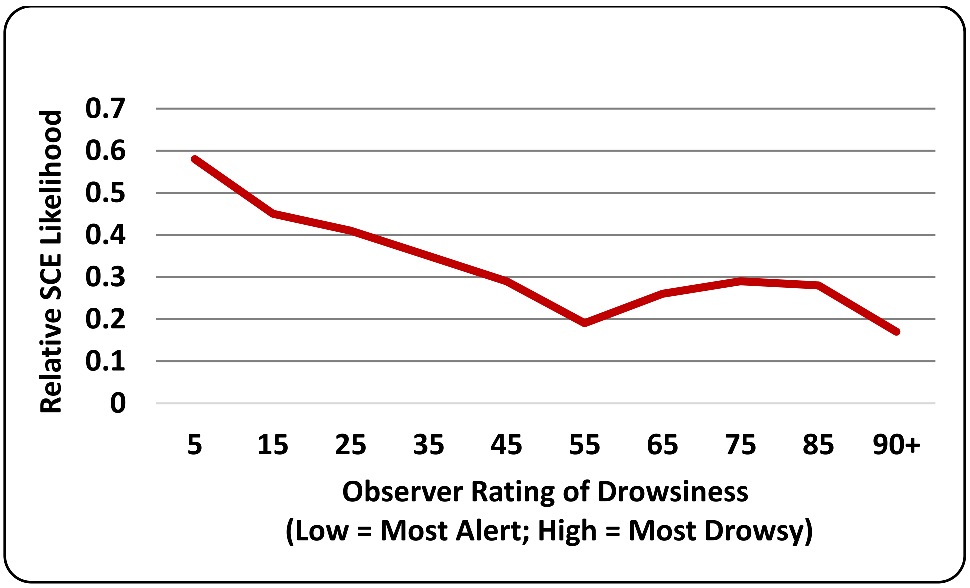


Figure 2. Likelihood of SCEs (versus Control) as a function of observer ratings of driver drowsiness. Alertness was associated with higher, not lower, likelihood of SCEs. The same inverse relationship was seen when eye closure (PERCLOS) was used to measure drowsiness. (Wiegand et al., 2008)

---

[7] Hanowski, R. J., Olson, R. L., Bocanegra, J. and Hickman, J.S.. *Analysis of Risk as a Function of Driving-Hour: Assessment of Driving-Hours 1 Through 11*. Report No. FMCSA-RRR-08-002, January 2008. Also: FMCSA. *Analysis of Risk as a Function of Driving-Hour: Assessment of Driving-Hours 1 Through 11 Final Report.* Report Tech Brief. No. FMCSA-RRR-08-006. 2008. Quotation from Page 3 of Tech Brief.

[8] Wiegand, D.M., Hanowski, R.J., Olson, R., & Melvin, W. *Fatigue Analyses from 16 Months of Naturalistic Commercial Motor Vehicle Driving Data*, 2008, The National Surface Transportation Center for Excellence. Available at: http://scholar.lib.vt.edu/VTTI/reports/FatigueAnalyses_061208.pdf

5. **Key HOS Study Did Not Analyze its Own SCEs**. The key Blanco et al. 2011 study[9] employed SCEs as its dependent variable (i.e., measure of risk) but provided no descriptions or analyses of the SCEs. It did not describe their scenarios, conditions of occurrence, or whether the truck driver was in any way at-fault. It did not examine or report whether SCEs showed any indications of driver fatigue, other hazardous physical condition, or other driver error or failure of any kind. All SCEs, regardless of their characteristics, were treated as fatigue-related events.

   FMCSA's *Share the Road Safely* public awareness campaign[10] urges passenger vehicle drivers to avoid encroaching on large trucks and, among other types of encroachments, to "not cut in close while merging in front of a CMV." Yet this very scenario would have been counted as a truck driver fatigue event in Blanco's study if the truck driver had to brake hard or swerve above threshold levels to avoid the collision.

   The National Academies of Sciences, Engineering, and Medicine, in its critical review of FMCSA's fatigue research methodologies[11], specifically reputiated the indiscriminate use of all SCEs as surrogates for fatigue, as follows (Page 193): "Some of these kinematic events, such as hard-braking and swerving to avoid collisions, may be necessary to avoid a collision that was the fault of other drivers . . . Therefore, these events are not necessarily appropriate surrogate outcomes for studies on fatigued driving."

   While the ND SCE methodology is the main focus of this critique, we'll note that FMCSA principal *crash* study supporting the 2011 rulemaking similarly did not describe its crashes! A study entitled, *Hours of Service and Driver Fatigue: Driver Characteristics Research* (Jovanis et al., 2011[12]) compared truck trips where crashes occurred to non-crash trips. The word "fatigue" was in the report title, abstract, and introduction, but the study itself employed no actual measures of driver fatigue or other causation. It did not describe or analyze its crashes for truck driver fault or any other characteristic relating to causation. In effect, every crash was treated as a truck driver fatigue event even though FMCSA's Large

---

[9] Blanco, M., Hanowski, R. J., Olson, R.L., Morgan, J. F., Soccolich, S. A., Wu, S-C, and Guo, F. *The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations*. Report No. FMCSA-RRR-11-017, May 2011.

[10] FMCSA Share the Road Safely Press Release, June 26, 2017. https://www.fmcsa.dot.gov/newsroom/fmcsa-urges-everyone-share-road-safely-national-safety-campaign

[11] National Academies of Sciences, Engineering, and Medicine. 2016. *Commercial Motor Vehicle Driver Fatigue, Long-Term Health, and Highway Safety: Research Needs*. Washington, DC: The National Academies Press. https://doi.org/10.17226/21921.

[12] Jovanis, P. P. Wu, K-F., Chen, C. *Hours of Service and Driver Fatigue: Driver Characteristics Research*, Report No. FMCSA-RRR-11-018, Contract #19079-425868, Task Order #6, May 2011.

Truck Crash Causation Study[13] found the percentage to be 4% to 13%, depending on the criterion used.

6. **Later Analysis Revealed Lack of Fatigue in Events**. A subsequent report from the same ND data collection published by FMCSA five years later (Blanco et al., 2016[14]) belatedly provided some of the essential event descriptions not provided in 2011. The 2016 report revealed that only 8.9% of the SCEs were attributable to reduced driver alertness with another 0.5% attributable to asleep-at-the-wheel. Thus 90.6% of the SCEs used for the HOS rulemaking were not demonstrably fatigue-related.

   Safety for the Long Haul Inc. believes that dependent variables in HOS research should be demonstrably fatigue-related and/or related to specified relevant populations of harmful crashes caused by commercial drivers. The National Academies, in its same review of FMCSA driver fatigue and HOS research[15], stated the same view (Page 8): "SCEs include incidents that are and are not fatigue-related. For some research purposes, then, only a subset of SCEs is relevant, so methods for identifying the most relevant subset of SCEs for research on CMV driver fatigue need to be determined."

7. **Negative 24-Hour Correlation with Fatigue Crash Distribution**. In the same Blanco et al (2016) data[16], the 24-hour distribution of SCEs correlated *negatively* (i.e., inversely) with the well-known 24-hour distribution of actual fatigue-related crashes. As has been known for decades, driver fatigue peaks in the early pre-dawn circadian low period (e.g., 3-6am). SCEs peak during evening rush hour, the time of greatest traffic density[17]. Again this suggests that SCEs and fatigue are opposites and that a dataset of unfiltered SCEs cannot be representative of the real-world population of fatigue-related crashes. See Figure 3.

---

[13] Starnes, M. *LTCCS: An Initial Overview*. NHTSA National Center for Statistics & Analysis, DOTR HS 810 646, August 2006. Also see: Knipling, R.R. *Safety for the Long Haul; Large Truck Crash Risk, Causation, & Prevention*. American Trucking Associations. ISBN 978-0-692-00073-1, 2009. Available at *www.atabusinesssolutions.com*.

[14] Blanco, M. J. S. Hickman, R. L. Olson, J. Bocanegra, R. J. Hanowski, A. Nakata, M. Greening, P. Madison, G. T. Holbrook, and D. Bowman. (2016). *Investigating Critical Incidents, Driver Restart Period, Sleep Quantity, and Crash Countermeasures in Commercial Vehicle Operations Using ND Data Collection.* FMCSA-RRR-13-017. Wash. DC: DoT.

[15] National Academies of Sciences, Engineering, and Medicine. 2016. *Commercial Motor Vehicle Driver Fatigue, Long-Term Health, and Highway Safety: Research Needs*. Washington, DC: The National Academies Press. https://doi.org/10.17226/21921.

[16] Blanco, M. J. S. Hickman, R. L. Olson, J. Bocanegra, R. J. Hanowski, A. Nakata, M. Greening, P. Madison, G. T. Holbrook, and D. Bowman. (2016). *Investigating Critical Incidents, Driver Restart Period, Sleep Quantity, and Crash Countermeasures in Commercial Vehicle Operations Using ND Data Collection.* FMCSA-RRR-13-017. Wash. DC: DoT.

[17] Seen in the above citation and also: Hanowski, R. J., Olson, R. L., Bocanegra, J. and Hickman, J.S.. *Analysis of Risk as a Function of Driving-Hour: Assessment of Driving-Hours 1 Through 11*. Report No. FMCSA-RRR-08-002, January 2008.

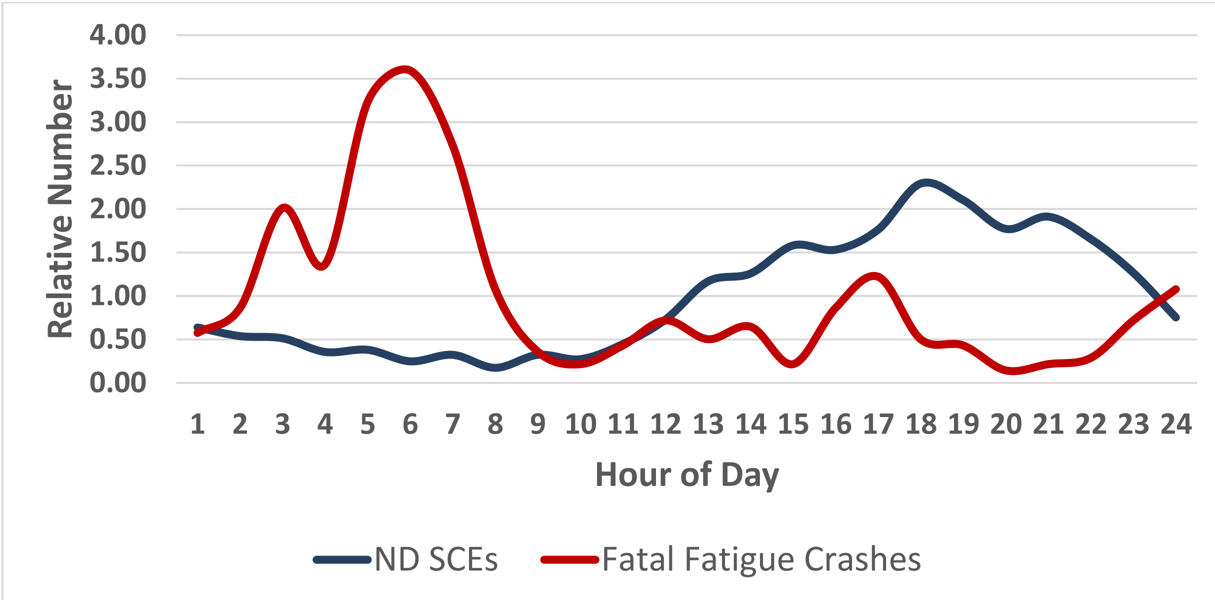


Figure 3. 24-hour distribution of large truck fatal fatigue crashes[18] and ND SCEs in Blanco et al. (2016). These two 24-hour distributions correlate negatively (r = -0.50). Graph and correlation from Knipling (2017[19]).

8. **Same Recent Analysis Revealed Sharp SCE-Crash Dissimilarities**. The same FMCSA Blanco et al. (2016) ND report revealed that 81% of its SCEs were single-vehicle events. In contrast, only about 20% of actual truck crash involvements are single-vehicle[20]. Single-vehicle and multi-vehicle crash involvements are markedly different in their characteristics and causation (see Attachment 2[21]). Such a sharp discrepancy in this key crash characteristic disqualifies the SCE dataset as being representative of actual crash risk. Imagine a political or other opinion survey with this level of unrepresentativeness in relation to its target population.

---

[18]Massie, D.L., Blower, D., & Campbell, K.L. *Short-Haul Trucks and Driver Fatigue*. UMTRI Center for National Truck Statistics. Prepared for FHWA OMC under Contract DTFH61-96-C-00038, Sept., 1997. Note: The hour-of-day data shown above from this report was for long-haul trucks.

[19] Knipling, R.R. Threats to scientific validity in truck driver hours-of-service studies. *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Pp. 382-388, Manchester Village VT, June 26-29, 2017 (graph in presentation); also: Knipling, R.R. Toward naturalistic driving crash representativeness. *Tenth SHRPII Safety Data Symposium*, Washington DC, October 6, 2017.

[20] FMCSA. *2014 Truck and Bus Crash Facts*, FMCSA-RRA-16-001. April 2016.

[21] Knipling, R.R. Crash heterogeneity: implications for naturalistic driving and for understanding crash risks. Paper 17-02225, Session 247, TRB Annual Meeting, Washington DC, 2017. Published in *Transportation Research Record No. 2663*. Available online at http://trrjournalonline.trb.org/doi/abs/10.3141/2663-15.

9. **No Controls for Well-Known Confounding Factors**. The key 2011 Blanco study[22] did not control for any potential confounding variables affecting SCEs and crash risk. Most notably it did not control for time-of-day, traffic density, roadway type (e.g., Interstates vs. undivided arterials and other local roads), or time awake. Previous research, including major FMCSA-sponsored ND studies[23], had demonstrated the importance of these confounding factors. Blanco's drivers could have simply been taking breaks in high-traffic situations and waiting for traffic to subside before resuming their trips. Since the study had no controls, such alternative explanations are as plausible as FMCSA's interpretations.

   Again we will note that FMCSA has made the same scientific errors in its crash studies as in its ND studies. The 2011 crash study supporting HOS rulemaking (Jovanis et al., 2011) compared crash and non-crash trips in regard to hours of driving and other HOS-related parameters, but had no controls for the same well-known confounding factors affecting crash risk and co-varying with driver schedules. Major uncontrolled factors included time-of-day, traffic density, and roadway type. The TRB Truck and Bus Safety Committee (ANB70) has recognized these scientific deficiencies and research needs[24].

10. **Theory Underlying SCE Methodology Has Been Discredited**. The entire ND methodology of combining multiple types of SCEs into a single dataset is based on the 20th century theories of H.W. Heinrich, an industrial safety engineer who never studied traffic crashes. "Heinrich's Law" states that "the predominant causes of no-injury accidents are . . . identical with the predominant causes of major injuries" (P. 31)[25]. This assumption is explicitly embraced in ND studies employing SCEs but has been discredited by the U.S. National Safety Council in regard to industrial accidents and also by peer-reviewed studies in the traffic safety literature (see Attachments 1 & 2). The NSC has published a text[26] debunking Heinrich and lamenting that "his misleading premises continue to be perpetuated" (P. vii). In regard to Heinrich's premise of identical mechanisms, the NSC rejoinder is that, "Causal factors for low-probability, high-consequence events are rarely represented in the analytical data on frequent incidents . . ." (P. 62).

---

[22] Blanco, M., Hanowski, R. J., Olson, R.L., Morgan, J. F., Soccolich, S. A., Wu, S-C, and Guo, F. *The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations*. Report No. FMCSA-RRR-11-017, May 2011.

[23] Reviewed in Knipling, R.R. Threats to scientific validity in truck driver hours-of-service studies. *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Pp. 382-388, Manchester Village VT, June 26-29, 2017 (see Attachment 1, Table 1).

[24] TRB Truck & Bus Safety Committee/ANB70. (2015) Driver Performance and other Causal Mechanisms in Quasi-Experimental HOS Studies. Research Needs Statement available athttps://rns.trb.org/dproject.asp?n=39358.

[25] Heinrich, H.W. (1959). *Industrial Accidents Prevention: A Scientific Approach, 2nd Edition*. McGraw Hill, New York: McGraw-Hill, 1941.

[26] Manuele, F. A. *Heinrich Revisited: Truisms or Myths, 2nd Edition*. National Safety Council, Itasca, IL, 2014.

On the basis of the profound scientific deficiencies documented above, *Safety for the Long Haul Inc.* requests and solicits the U.S. DOT and FMCSA to suspend or rescind the 2011 HOS rule requiring mandatory rest breaks and any other HOS provisions which were based primarily on the ND mixed-SCE methodology and related crash studies containing parallel scientific deficiencies. Stated bluntly, this research has been “junk science.” Future studies in support of HOS rulemaking must be required to demonstrate that its dependent measures validly represent serious truck crashes and/or driver drowsiness/fatigue. In addition, studies of HOS parameters (e.g., hours of driving, taking breaks) must be designed to control for well-known confounding factors such as time-of-day, traffic density, and roadway type. The scientific validity of HOS research extrapolates directly to rule fairness and effectiveness. If research dependent measures are not representative of true risk and/or causal linkages are not real, then more restrictive HOS rules will simply hamstring industry without resulting in true safety improvements.

Attachments

<u>Attachment 1</u>

Knipling, R.R. Threats to scientific validity in truck driver hours-of-service studies. *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Pp. 382-388, Manchester Village VT, June 26-29, 2017.

This peer-reviewed conference paper explains and documents three fundamental scientific deficiencies seen in FMCSA's HOS research over the past ten years. These include deficiencies in:

- *Internal* validity: truthfulness of causal inference (causal relations between HOS parameters and measured outcomes).
- *External* validity: the truthfulness of generalizations from studies to real-world phenomena of importance (i.e., serious crashes).
- *Construct* validity: whether or not driver states such as fatigue/drowsiness are being measured accurately and appropriately based on our knowledge of these states.

<u>Attachment 2</u>

Knipling, R.R. Crash heterogeneity: implications for naturalistic driving and for understanding crash risks. Paper 17-02225, Session 247, Transportation Research Board Annual Meeting, Washington DC, 2017. In press for 2017 publication in *Transportation Research Record No. 2663*.

This paper presents evidence for the causal heterogeneity of crashes involving all motor vehicles and specifically for large trucks. Causal heterogeneity means that findings from one subpopulation of crashes cannot be generalized to other subpopulations or to all crashes. Further, non-crash events such as Naturalistic Driving "Safety-Critical Events" are likely to be even less valid in representing important crashes. The fact of crash heterogeneity disproves a key, explicit assumption of the ND SCE method; i.e., the assumption that all types of crashes and non-crash dynamic events have identical or highly similar causation. The paper concludes that ND estimates of the prevalence and role of specific crash causal factors have little likely validity in relation to fatal and injury crashes where the preponderance of human harm occurs.

Note that both of the above research studies were self-funded by Safety for the Long Haul Inc. and were not financially supported by the trucking industry.

THREATS TO SCIENTIFIC VALIDITY IN TRUCK DRIVER HOURS-OF-SERVICE STUDIES

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

Summary: Commercial truck driver Hours-of-Service (HOS) rules are periodically revised to reduce driver fatigue and improve driver health in cost-efficient ways. HOS research must demonstrate *causal* relationships between HOS parameters and *important* safety outcomes. Thus, two scientific requirements are *internal* validity (demonstration of true cause-effect relationships) and *external* validity (generalizability to important real-world consequences). HOS rules ostensibly act by mitigating driver fatigue; thus, dependent measures in most HOS studies must verifiably capture and measure alertness/fatigue. That is, dependent measures must have *construct* validity. This paper examines these basic scientific validity requirements and finds significant threats to them within the designs of major U.S. HOS studies. Lessons learned apply to many other areas of behavioral research. Improved designs and compensatory methods are suggested for addressing validity threats and thereby increasing internal, external, and construct validity. Improving scientific validity would in turn raise the likelihood that HOS changes based on research would be safety-effective in the real world of truck transport on our nation's highways.

INTRODUCTION

Principal Acronyms
HOS – Hours-of-Service
CR – Critical Reason
ND – Naturalistic Driving
SCE – Safety-Critical Event (in ND)

Hours-of-Service (HOS) rules support commercial truck safety and driver health by setting legal limits to driver schedules. These include minimum daily off-duty hours, maximum daily driving hours, required breaks from driving, and weekly maximum work hours. The Federal Motor Carrier Safety Administration (FMCSA) of the U.S. Department of Transportation conducts on-road research to demonstrate causal linkages between HOS parameters (e.g., hours of driving) and important safety outcomes. HOS-outcome linkages must be *causal*; if they are not, rule changes will not have expected effects. Further, effects must be *important*, resulting in meaningful reductions in human harm. Such concerns are not unique to HOS research; they arise widely in science as questions of *internal* and *external validity* (Privatera, 2014). Internal validity is the truthfulness of causal inference. External validity is the truthfulness of generalizations from studies to real-world phenomena of importance. Another type of validity related to both is *construct* validity. A *construct* is an underlying factor known to exist but which cannot be directly observed. "Fatigue" is a classic example. We may say and think we are measuring fatigue in a study, but can we prove that? Without clear evidence of internal, external, and construct validity, the effects of HOS rule changes cannot be predicted with confidence before those changes are implemented nationwide

Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Manchester Village Vermont, June 26-29, 2017, Pp.382-388.

across the $600 billion U.S. trucking industry. Scientific rigor is thus essential to ensure public safety and transportation efficiency.

The same scientific requirements apply generically across many areas of behavioral research. If one were testing a pill to increase happiness, one would need a design rigorous enough to show causality (internal validity), laboratory measures consistent with established indicators of subject happiness (construct validity), and measures predictive of happiness indicators in real life (external validity).

This paper focuses on HOS research designs and their scientific vulnerabilities. Methodological dissection isn't just an exercise in criticism – it reveals important considerations about sound practices in behavioral research, about human fatigue and performance, about regulatory effects on safety, and about the nature of crash risk.

THREATS TO INTERNAL VALIDITY (CAUSAL INFERENCE)

The *experimental* method uniquely demonstrates causality between a condition and a dependent measure. Required elements of an experiment include (a) manipulation (not merely observation) of conditions, (b) randomized assignments to conditions, and (c) one or more comparison groups (Privatera, 2014). Key U.S. HOS studies (notably Hanowski et al., 2008; Blanco et al., 2011, and Jovanis et al., 2011) have been *quasi-experiments*, also called *pseudo-experiments*. These studies "look like" experiments. They had nominal independent variables (e.g., hours-of-driving), dependent measures (ND SCEs or crashes), and were statistically analyzed like experiments. But they lacked required experimental elements (a) and (b). They recorded ND SCEs or crashes occurring under different HOS conditions (most notably hours-of-driving) but did not manipulate conditions or randomly assign drivers to them. Such designs are essentially correlational, showing associations but not causality (Knipling, 2015a). They are subject to *confounding variables*, variables not accounted for but which could be causing or partially causing observed changes. Table 1 presents evidence of the strong confounding effects of four such factors. With such strong confounds, interpretation becomes tenuous. For example, if risk increases at dawn at the end of overnight driving shifts, is this due to time-on-task driving, time awake, the early morning circadian trough, the incipient rise in traffic with morning rush, or to changeover from freeways onto more risky local roads? One cannot know without controlled analysis.

Figure 1 illustrates internal validity concerns regarding HOS quasi-experiments. HOS parameters are nominal independent variables, presumed to affect risk by way of the construct "fatigue." Yet multiple confounding variables threaten the validity of causal inference. Some confounds create systematic bias while others act randomly to add error. The relative strengths of various factors operating in such designs cannot be inferred without analyses controlling *post hoc* for potential confounding variables. Such analyses were not performed in the HOS studies cited. A more rigorous approach would enlist a large fleet with flexible operations (e.g., a private fleet delivering to its own outlets) to manipulate trips per HOS parameters and potential confounding variables. Driver performance could be observed over hours of driving on standard routes while varying and counterbalancing start times (and therefore their circadian and traffic conditions). Driver assignments could be random or counterbalanced. Event rates could be disaggregated by roadway type and traffic density to see if these strong factors were distorting results.

Table 1. Major Potential Confounding Variables Threatening Valid HOS Causal Inference

Confounding Variable	Relevance to Fatigue and/or Risk	How It May Co-vary with Schedules
Time-of-Day	Time-of-day was "the strongest and most consistent factor influencing driver fatigue and alertness" in the *Driver Fatigue and Alertness Study* (Wylie et al., 1996), an effect attributed to daily circadian rhythms. In the Large Truck Crash Causation Study (LTCCS), 62% of truck driver asleep-at-the-wheel crashes occurred between 4:01 and 6:00 a.m. (Knipling, 2009). In Hanowski et al., 2008, SCE involvement rates varied more than 7-fold across the 24-hour day.	The lowest circadian trough and period of greatest fatigue risk occurs near the end of overnight driving schedules.
Traffic Density	Traffic density changes sharply and predictably over driving schedules. Hanowski et al. (2008) found a +0.83 correlation between SCE rate and traffic density patterns. Wiegand et al., 2008 found an SCE-to-baseline odds ratio of 7.2 for high-traffic conditions.	A disproportionate number of CMV driving schedules end during periods of high traffic; i.e., during morning or evening rush hours.
Roadway Type	Interstate fatal crash rates are about one-half those of arterial roads and one-third those of local roads. In one truck study, the SCE-to-baseline odds ratio for undivided versus divided roads was 5.3 (Knipling, 2009).	Long-haul trip terminations generally involve a change from Interstates/freeways to arterial or local roads.
Time Awake	Time awake (especially >16 hours) is a strong independent factor in the biological sleep-wake homeostat (Knipling, 2015). Its operation is largely independent of the level and type of physical activity.	Time awake is a continuous but "hidden" co-variate of hours of driving and hours of work.

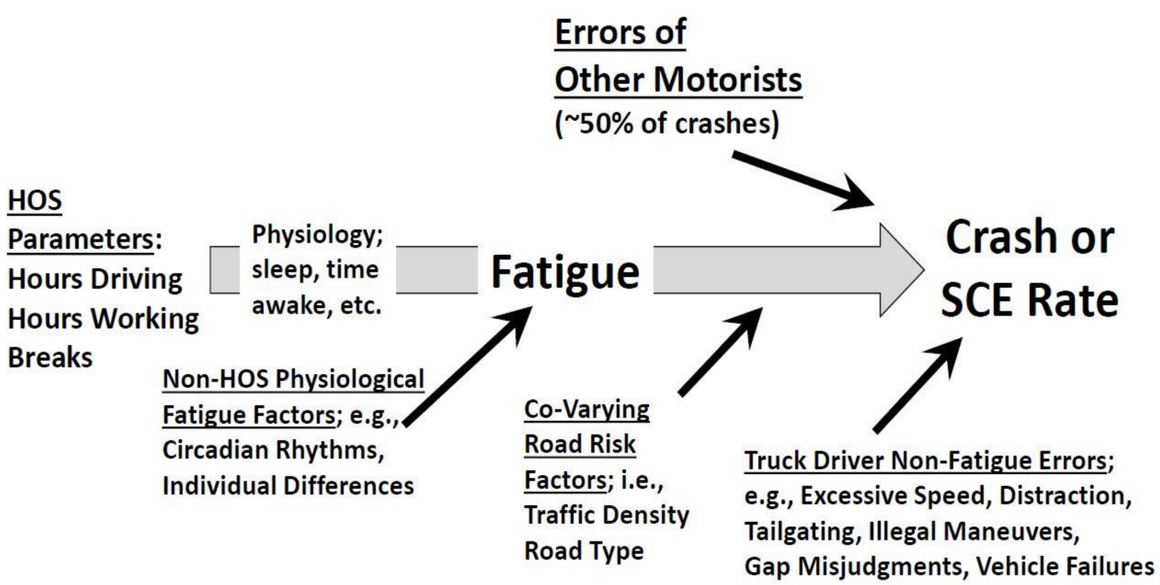


Figure 1. Potential confounds in studies relating HOS parameters to truck crash or SCE rate.

The causal gauntlet shown in Figure 1 represents more than just a scientific challenge. It is a real-world limitation to the likely effects that HOS rules can have on truck crash risk. Physiological fatigue factors not addressable by HOS rules (e.g., sleep quality, circadian rhythms), external traffic/road vagaries, and non-fatigue driver errors (many by other motorists on the roadways) combine to likely outweigh and largely mask the effects of various specific HOS schedule parameters.

THREATS TO EXTERNAL VALIDITY (TRUE, MEANINGFUL GENERALIZATIONS)

External validity is the extent to which study data represent phenomena of real-world importance beyond the confines of a study (Privatera, 2014). Generalizations are often problematic in crash studies because crashes are heterogeneous, both "horizontally" and "vertically" (Knipling, 2015b, 2017; Kidd and McCartt, 2015). Horizontal heterogeneity refers to the variety of scenarios within any crash severity level. In the LTCCS (all serious crashes), truck driver asleep-at-the-wheel was the Critical Reason (CR) for 19% of road departures, but 1% or less of rear-end, sideswipe, and opposite direction involvements (Knipling, 2009). Vertically, crash profiles can differ sharply by severity level. For example, the known causal role of fatigue is about five times greater in fatal than in property damage truck crashes (FMCSA, 2014). One cannot simply assume crash generalizability (e.g., as in Blanco et al., 2016), either horizontally or vertically. In Figure 2, layers of the triangle represent levels of police-reported crash severity plus non-police-reported crashes. Layers K, A, and B are fatal/injury crashes representing about 11% of police-reported crashes but 80-90% of harm (Zaloshnja and Miller, 2007). ND SCEs of multiple types are shown almost entirely beneath the triangle since they overwhelmingly involve no impact. An analytic link could be perhaps created, however, through mathematical indexing of ND events to serious crashes just as some unrepresentative survey samples are indexed to their target populations to improve representativeness (TRB ANB20, 2016; Knipling, 2017).

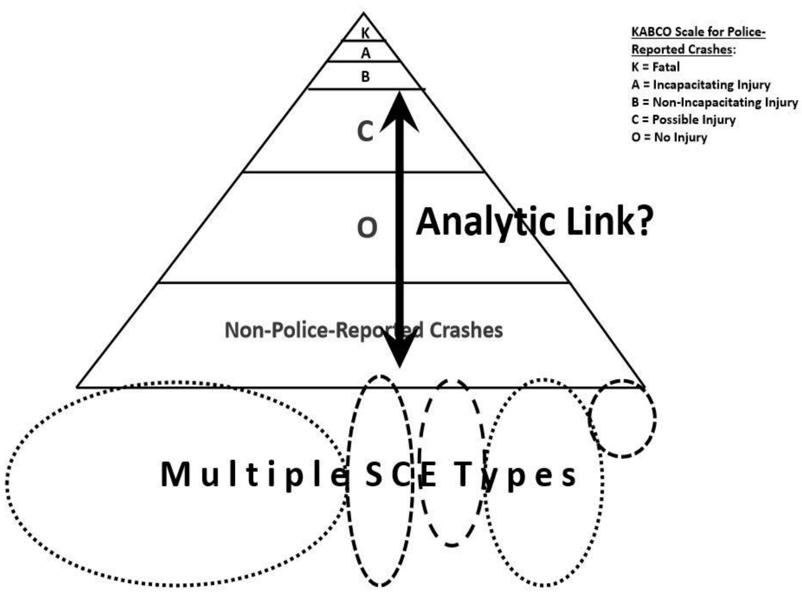


Figure 2. Crash triangle for crashes plus multiple SCE types constituting SCE datasets

No large truck ND dataset has included a sufficient number of crashes for causal analysis of any kind. Blanco et al.'s 2011 truck ND dataset included just four (4) crashes out of 2,197 SCEs (0.2%), with "crash" defined as "any contact." Passenger car statistics from the Strategic Highway Research Program (SHRP2) ND dataset suggest that most ND physical contacts are extremely minor and would not qualify as even non-police-reported crashes using U.S. DOT criteria (Blincoe et al., 2015; Kidd and McCartt, 2015; Knipling, 2017). Marked differences in crash type distributions, and therefore causal profiles, were seen vertically between levels of SHRP2 low-severity impacts (Kidd and McCartt, 2015). They concluded that, “Researchers must take into account how the crash populations in the SHRP2 data differ from national crash databases in order to describe results precisely and use due caution in generalizing results to the population of U.S. crashes.” Truck driver fatigue-related crashes tend to be even more severe than truck crashes in general (Knipling, 2009). Truck ND SCEs are clearly not representative of these serious fatigue-related crashes or of any other important, definable truck crash population. Findings relating to SCE causal genesis, including HOS-related factors, cannot be generalized to the serious truck crashes concerning to industry and to society.

THREATS TO CONSTRUCT VALIDITY (KNOWING WHAT WE ARE MEASURING)

HOS rules are instituted to mitigate fatigue. Thus, driver alertness/fatigue is the logical conceptual dependent variable for most HOS studies. Demonstrating explicitly that alertness/fatigue is indeed being measured would be one way to allay both internal and external validity concerns. However, key HOS reports have not addressed the construct validity of their dependent measures and not even attempted to specifically measure fatigue. In a major non-ND study, Jovanis et al. (2011) employed relative crash rate as their dependent measure, but used no filters to identify fatigue involvement or even whether the truck driver was at-fault in any respect. All crashes were treated as fatigue events. Yet most large truck crashes are *not* known to be fatigue-related (Starnes, 2006; Knipling, 2015). Thus, unfiltered crash rate has no construct validity as a measure of fatigue. ND HOS studies (e.g., Hanowski et al., 2008, Blanco et al., 2011) have used SCE rate as a fatigue surrogate, but with no identification of actual fatigue or any other cause of the events. A subsequent report (Blanco et al., 2016) from the same source revealed that only 8.9% of events were attributable to reduced alertness with another 0.5% attributable to asleep-at-the-wheel. ND SCEs and fatigue-related crashes are more opposite than similar (Table 2). Such divergence refutes SCE construct validity in relation to fatigue. Neither crashes nor SCEs should be treated as fatigue events without event-specific indications (e.g., from interviews or videos) that fatigue was actually involved.

Table 2. SCEs and Driver Fatigue-Related Crashes: Notable Contrasts

ND Safety-Critical Events (SCEs)	Fatigue-Related Crashes
Lowest rate in early morning (Blanco et al., 2016)	Highest rate in early morning
Most likely in heavy urban traffic	Most likely on low traffic rural roads
Most likely on undivided roads	Most likely on divided highways
Driver is active, usually distracted (Barr et al., 2011)	Driver is passive with tunnel vision and relinquishing vehicle control (Barr et al., 2011).
AATW % of CRs = 0.1% (Knipling 2009); 0.5% in Blanco et al., (2016)	AATW % of CRs = 3.8% (Knipling 2009)
Risk *inversely* related to PERCLOS (Percent Eye Closure). Eyes were more open in SCEs than during normal driving control events (Weigand et al., 2008)	Risk strongly indicated by PERCLOS; e.g., lane tracking deteriorates as eyes close (Knipling, 2009)

Note: Knipling (2015a) provides additional citations for these statements. AATW = Asleep-at-the-wheel.

At this writing, FMCSA has pending the publication of a new ND HOS study focusing on driver weekly “restart” rest periods. In the published study plan (FMCSA 2015), SCE rate was stated as the study’s measure of “safety impacts.” Recorded SCEs comprising this dependent measure included "electronically-recorded hard brakes, hard accelerations, swerves, contact with other objects, and driving in excess of posted speed limits." It is unclear how such a mixture of dynamic events could operationally represent fatigue or any other discrete safety construct. Certainly SCEs cannot be assumed, without evidence, to represent serious crashes resulting in human harm. Greater scientific rigor in the selection and specification of measurements seems called for given the millions of dollars invested in these studies and the national safety and economic ramifications of truck driver HOS rules.

CONCLUSIONS

Scientific terms like *internal validity*, *external validity*, and *construct validity* are not found in U.S. HOS reports. No major HOS report has directly addressed validity threats. Yet HOS research validity likely extrapolates directly to rule effectiveness. If causal linkages are not real and/or dependent measures not representative of true risk, then changes to HOS rules will likely not result in true safety improvements. Two Transportation Research Board (TRB) committees have recognized these concerns (ANB20, 2016; ANB70, 2015a, 2015b). Methods are available to address and improve scientific validity and thereby avoid policy errors and unintended societal consequences. Such methods also promise greater knowledge about fatigue and how best to manage commercial driver schedules.

DISCLAIMERS

The views expressed are solely those of the author. Safety for the Long Haul Inc. received no external funding for this work and has no vested interest in any HOS rule change.

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.

Blanco, M. J. S. Hickman, R. L. Olson, J. Bocanegra, R. J. Hanowski, A. Nakata, M. Greening, P. Madison, G. T. Holbrook, and D. Bowman. (2016). *Investigating Critical Incidents, Driver Restart Period, Sleep Quantity, and Crash Countermeasures in Commercial Vehicle Operations Using ND Data Collection.* FMCSA-RRR-13-017. Wash. DC: DoT.

Blincoe, L. J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015) *The economic and societal impact of motor vehicle crashes, 2010 (revised).* DOT HS 812 013, NHTSA.

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

FMCSA. (2015) *Commercial motor vehicle driver restart study plan*. U. S. 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.

Jovanis, P. P. Wu, K-F., Chen, C. (2011) *Hours of service and driver fatigue: driver characteristics research*, Report No. FMCSA-RRR-11-018, U.S. DOT.

Kidd, D. G. & McCartt, A. T., Insurance Institute for Highway Safety. The Relevance of Crash Type and Severity when Estimating Crash Risk Using the SHRP2 Naturalistic Driving Data. 4th International Driver Distraction and Inattention Conference, Sydney, New South Wales, 2015.

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

Knipling, R.R. (2015a) *Critical Review of Driver Fatigue & HOS-Related Research Methodologies*. Commissioned paper for National Academy of Sciences Commercial Driver Fatigue Panel.

Knipling, R.R. (2015b) Naturalistic driving events: no harm, no foul, no validity. Driving Assessment 2015, Paper #30.

Knipling, R.R. (2017) Crash heterogeneity: implications for naturalistic driving and for understanding crash risks. Paper 17-02225, TRB Annual Meeting, Washington DC.

Privitera, G. J. (2014) *Research methods for the behavioral sciences,* Sage Publications, Inc.

Rosekind, M.R. (2005) Managing safety, alertness and performance through federal hos regulations: opportunities and challenges. Alertness Solutions. Rulemaking docket #FMCSA-2004-19608.

Starnes, M. (2006) *Large Truck Crash Causation Study (LTCCS): an initial overview*. NHTSA National Center for Statistics & Analysis, DOTR HS 810 646, August 2006.

TRB Committee on Safety Data, Analysis, & Evaluation/ANB20. (2016) Indexing Naturalistic Driving Events to Crashes. Research Needs Statement available at https://rns.trb.org/dproject.asp?n=40810.

TRB Truck & Bus Safety Committee/ANB70. (2015a) Driver Performance and other Causal Mechanisms in Quasi-Experimental HOS Studies. Research Needs Statement available athttps://rns.trb.org/dproject.asp?n=39358.

TRB Committee on Truck & Bus Safety/ANB70. (2015b) Toward Naturalistic Driving Crash Representativeness (23-2015). Research Needs Statement available at https://rns.trb.org/dproject.asp?n=39354.

Wiegand, D.M., R. J. Hanowski, R. L. Olson, and W. Melvin. (2008) *Fatigue analyses from 16 months of naturalistic commercial motor vehicle driving data*, The National Surface Transportation Center for Excellence. Available at: http://scholar.lib.vt.edu/VTTI/reports/FatigueAnalyses_061208.pdf.

Wylie, C.D., T. Shultz, J. C. Miller, M. M. Mitler, and R. R. Mackie. (1996) *Commercial motor vehicle driver fatigue and alertness study*, FHWA, U.S. DOT.

Zaloshnja, E. and T. Miller. (2007) *Unit costs of medium & heavy truck crashes*. Final Report, Pacific Institute for Research & Evaluation for FMCSA, U.S. DOT.

CRASH HETEROGENEITY: IMPLICATIONS FOR NATURALISTIC DRIVING STUDIES AND FOR UNDERSTANDING CRASH RISKS

Ronald R. Knipling
Safety for the Long Haul Inc.
5059 36th Street North
Arlington, Virginia 22207-2946
United States
Phone: 1 703 533 2895
FAX: 1 703 533 2895
Email: rknipling@verizon.net

Final Revision Submitted: March 5, 2017

Word count: 6,247 text + 2 figures (250 each) + 3 tables (250 each) = 7,497

ABSTRACT
Motor vehicle crashes are heterogeneous in their conditions of occurrence, risk factors, and causal scenarios. Horizontal heterogeneity refers to the many distinct scenarios within any crash severity level. Vertical heterogeneity is seen in the different proportions of characteristics at different severity levels. This paper presents evidence for the causal heterogeneity of crashes involving all motor vehicles and also specifically for large trucks. If horizontally or vertically defined crash subsets are not representative of other subsets, then findings from them cannot be validly generalized to other populations. Further, crash heterogeneity contradicts a key assumption of the "Heinrich Triangle," the assumption that crashes within the triangle have identical or highly similar causal factors regardless of outcome severity. The Heinrich assumption is explicit in Naturalistic Driving Studies (NDS) capturing mainly non-crash dynamic events and minor crashes below conventional reporting thresholds. NDS causal prevalence estimates have little likely validity in relation to fatal and injury crashes where the preponderance of human harm occurs. NDS external validity could perhaps be improved by *post hoc* mathematical indexing of captured events to the objective profiles of target populations.

Knipling, R.R. Crash heterogeneity: implications for naturalistic driving and for understanding crash risks. Paper 17-02225, Session 247, TRB Annual Meeting, Washington DC, 2017. Published in *Transportation Research Record No. 2663*. Available online at http://trrjournalonline.trb.org/doi/abs/10.3141/2663-15.

1. INTRODUCTION

Heterogeneity is seen pervasively in motor vehicle crashes. Crashes have many different scenarios and physical configurations, and each vehicle in a multi-vehicle crash plays a distinct role. Conditions of occurrence and risk factors vary across involvement types, and drivers make many different types of errors leading to crashes. At the same time, crash scenario profiles vary in predictable ways across different levels of outcome severity. This paper presents evidence of "horizontal" heterogeneity (how crashes occur) and "vertical" heterogeneity (as a function of severity) and explores their implications for understanding crash risk and causation.

Most studies of crash occurrence and causation involve a sample of events meant to represent, implicitly or explicitly, a target population of crashes. Crash heterogeneity means that various target populations vary significantly from one another, and therefore that generalizations from part-to-part or part-to-whole may be spurious. Prevalence estimates of crash factors and other characteristics may be erroneous unless they are linked empirically or analytically to a specific target population.

Selected Acronyms

NMVCCS – National Motor Vehicle Crash Causation Survey
LTCCS – Large Truck Crash Causation Study
PDO – Property Damage Only
CR – Critical Reason
NDS – Naturalistic Driving Studies
SCE – Safety-Critical Event
GES – General Estimates System

External validity is the extent to which any study generalizes beyond its specific conditions to phenomena of broader importance. An external validity challenge exists in many studies involving crashes. Does the study reveal truth about crashes causing significant harm? That same challenge looms even larger when one seeks understanding based on samples of non-crashes or of crashes too minor to cause societal concern.

Naturalistic Driving Studies (NDS) collect primarily non-crash and minor crash events seeking generalizable causation insights. Following its presentation on crash heterogeneity, this paper examines the construction and underlying assumptions of NDS event datasets. It presents evidence challenging these assumptions and articulates concerns regarding NDS external validity. The paper concludes with suggestions for improving validity.

2. METHOD

This paper presents both newly reported and previously reported statistics. In many cases, previously reported statistics have been aggregated and/or further analysed to highlight specific points. Newly reported crash statistics are from data retrievals performed at the direction of the author. This includes statistics from two major in-depth crash causation investigations, both of which employed stratified sampling and case weighing to generate nationally representative prevalence estimates:

- The National Motor Vehicle Crash Causation Survey (NMVCCS) was conducted by NHTSA between 2005 and 2007. Trained researchers investigated 5,471 crashes, each of which involved a light passenger vehicle (*1*). NMVCCS included all five police-reported "KABCO" severity levels: K = Killed; A = Incapacitating injury; B = Non-incapacitating injury; C = Possible injury; O = No injury (also known as Property Damage Only). Thousands of specific characteristics could be recorded for each crash and each involved person or vehicle. Weighed NMVCCS statistics represented 2,189,166 crashes with 4,031,226 involved vehicles.

- The 2001-2003 Large Truck Crash Causation Study (LTCCS) preceded NMVCCS and employed a similar methodology (*2*). It reported data on 963 large truck crashes, each of which resulted in a K, A, or B injury. These top three levels represented just 11% of truck police-reported crashes, but a high majority of human harm from truck crashes *(6)*. Weighted LTCCS statistics represented 119,417 crashes involving 141,200 trucks and 99,828 other vehicles.

Of the hundreds of NMVCCS and LTCCS variables coded for each case, perhaps the most pivotal and heuristic was the *Critical Reason* (CR). The CR is the “immediate reason” for the destabilizing event or collision path which became the crash (*1, 2, 3*). Notable coding choices included falling asleep, driver inattention (including distraction), inadequate surveillance, excessive speed for conditions, following too closely, and illegal maneuvers. Vehicle failures and extreme environmental or roadway conditions were also cited as CRs. Only one CR was assigned in each NMVCCS and LTCCS crash, thus making the CR largely equivalent with “fault.” Fault attribution is admittedly simplistic and misleading in some individual cases, but it is useful for sorting crash involvements into major categories corresponding to the location and types of driver errors or other failures triggering crashes.

A key distinction in crash analysis is between *crashes* and *crash involvements*. When two vehicles collide, there is one crash but two involvements. One of the involvements would be assigned a CR, while the other has none. In single-vehicle crashes, the CR is virtually always assigned to the sole vehicle/driver. Thus, the number of CR assignments equals the number of sampled crashes. Most new NMVCCS and LTCCS statistics reported here are *involvement* statistics disaggregated by multi- vs. single-vehicle and by CR assignment (yes vs. no, i.e., at-fault vs. not-at-fault).

This paper also reports one set of involvement statistics newly accessed from the second Strategic Highway Research Program (SHRP2) NDS via its InSight website.

Statistical tests of significance are not employed in this paper because its assertions are based on macro patterns of data (mostly nationally representative estimates) rather than on paired comparisons. Statistical significance would be too low a criterion for assessing the validity and implications of study findings.

3. HORIZONTAL CRASH HETEROGENEITY

Crash involvements can occur in many different ways. Profiles of why crashes occur are strongly associated with profiles of where, when, and how they occur. “Horizontal” heterogeneity refers to the variety of types, roles, and higher categories seen within any population of crashes or involvements.

3.1 Heterogeneity of Involvement Types

Numerous examples may be cited. In the LTCCS, truck driver asleep-at-the-wheel was the CR in 19% of road departures but 1% or less of rear-end, sideswipe, and intersection/crossing path involvements. Excessive truck speed (for curve/turn or for responding to other vehicles) as the CR was 21% of road departures, 32% of head-on strikes where the truck was the encroaching vehicle, 25% of rear-end strikes into moving vehicles, but just 6% of rear-end strikes into stopped vehicles.

Figure 1 depicts the distributions of seven CR types and other factors for four crash involvement types from NMVCCS. The first five factors are CRs assigned to these vehicles/drivers. The last two (alcohol and vehicle factor) are associated factors which likely contributed to the involvement. Each of the four types shows a distinctive pattern of causal and contributing factors. Even the two rear-end-striking subtypes have notable differences. Conversely, the prevalence of the seven factors varies across the four involvement types.

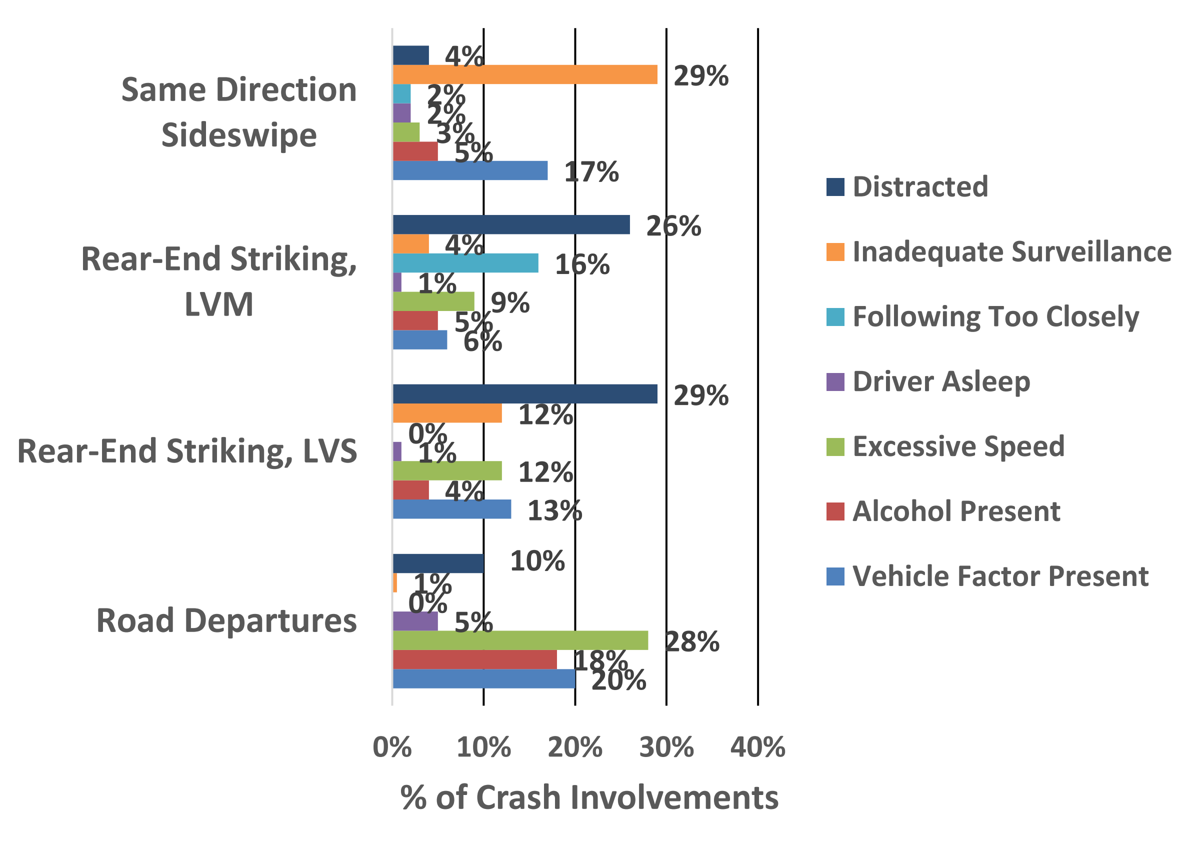


FIGURE 1 Weighted Distributions of Selected CRs for Four Crash Involvement Types in NMVCCS. LVM = Lead Vehicle Moving, LVS = Lead Vehicle Stopped. Distraction includes two types: internal and external. Excessive speed includes three: excessive speed for conditions, to be able to respond, and for a curve/turn. Alcohol and vehicle percentages exclude cases coded as unknown or unreported.

3.2 Fault Categories: the "Good," the "Bad," and the "Ugly"

The pivotal nature and heuristic value of the CR variable is demonstrated in this section. From a causal perspective, crash involvements may be divided into three largely distinct categories, deemed here the "good," the "bad," and the "ugly." While no crash is good in its consequences, involvements are causally "good' if the subject driver and vehicle are not at-fault; i.e., not assigned the CR in causation studies. "Bad" involvements are those involving two or more vehicles where the CR *is* assigned to the subject vehicle. “Ugly” involvements are those involving just one vehicle. Single-vehicle involvements are deemed “ugly” because they are often more serious and much more likely to involve driver impairment, extreme misbehavior, and/or vehicle maintenance failures than are “bad” involvements. Thus, the three causal involvement categories are:

- Multi-vehicle, no CR assigned (“good”)
- Multi-vehicle, CR assigned (“bad”)
- Single-vehicle (“ugly”).

These three involvement categories have markedly different characteristics. Table 1 provides examples from the LTCCS, wherein subject vehicles were all large combination-unit or single-unit trucks. The value in any taxonomy lies in its ability to elucidate differences. Here one sees many, suggesting that generalizations across the three categories or from one category to the whole could be perilous. Being assigned the CR is associated with a variety of driver and vehicle deficiencies compared to “good” involvements. Even more striking are the numerous differences between the two truck at-fault categories. At-fault multi-vehicle (“bad”) involvements are associated mostly with traffic factors, whereas single-vehicle involvements (“ugly”) are associated with curves, driver misbehaviors, impairment, and vehicle maintenance deficiencies.

Table 1 Three Involvement Categories: Weighted Percentages of LTCCS Trucks Assigned Attribute

LTCCS Variable	Attribute (or Attribute Aggregation)	“Good” (MV, No CR)	“Bad” (MV, CR)	“Ugly” (Single Vehicle)
Road Alignment	Curve (Left + Right)	19%	22%	60%
Pre-Event Movement	Truck negotiating a curve	9%	12%	46%
Attempted Avoidance Maneuver	Braking, steering, and/or accelerating (% of knowns)	47%	64%	59%
Driver Seat Belt Use	None used or not indicated*	6%	8%	23%
Driver Roadway Familiarity	Truck driver rarely/never drove road before*	17%	29%	38%
Vehicle Factor (Truck)	Present (any inspection deficiency)	21%	50%	62%
Driver Fatigue Present	Truck driver fatigued*	3%	14%	30%
Critical Reason (CR)	Driver Asleep	NA	1%	13%
Hours of Last Sleep	< 6 hours last main sleep*	10%	15%	29%
Critical Reason (CR)	Heart attack/other physical impairment	NA	2%	6%
Relation to Junction	Intersection	14%	23%	9%
Trafficway Class	Urban (6 different roadway types)	53%	65%	38%
Traffic Factor	Ambient traffic present at time of crash	31%	42%	6%
Critical Reason (CR)	Too fast for conditions or curve/turn	NA	13%	30%

* % of knowns. Table percentages are LTCCS estimates for all U.S. large trucks (combination- plus single-unit) involved in serious (KAB) crashes.

Table 2 presents CR assignment percentages per equivalent involvement categories for both studies. Recall that light passenger vehicle involvements dominated NMVCCS since every NMVCCS crash involved one. The overall category percentages for NMVCCS were

44% "good," 39% "bad," and 17% "ugly." The LTCCS involvements shown are all large trucks. The truck involvement distribution was 44% "good," 29% "bad," and 27% "ugly."

Table 2 Three Involvement Categories: Weighted Critical Reason Percentages

Study/Category:	Vehicles in NMVCCS			Large Trucks in LTCCS		
Critical Reason (CR):	MV, No CR ("Good")	MV, CR ("Bad")	Single Vehicle ("Ugly")	MV, No CR ("Good")	MV, CR ("Bad")	Single Vehicle ("Ugly")
Non-Performance; e.g., Asleep, Medical Crisis	NA	3%	14%	NA	3%	21%
Inadequate Surveillance; e.g., Looked But Did Not See	NA	28%	2%	NA	20%	4%
Other Recognition Failure; e.g., Distraction	NA	21%	14%	NA	20%	12%
Too Fast (for Conditions or Curve/Turn)	NA	6%	27%	NA	13%	30%
Other Decision Errors; e.g., Misjudged Gap	NA	16%	1%	NA	24%	1%
Maneuver Execution Error; e.g., Overcorrected	NA	4%	29%	NA	3%	9%
Vehicle Failure; e.g., Brakes, Tires, Cargo Shift	NA	1%	4%	NA	7%	13%
Environmental or Roadway Condition	NA	2%	3%	NA	1%	2%
Other CRs Not Shown	NA	19%	6%	NA	9%	8%
Total	NA	100%	100%	NA	100%	100%

As discussed, CRs in both studies were assigned to only one vehicle and thus were not applicable to "good" involvements. "Good" involvements are relevant to this discussion, however, since they constituted 44% of involvements in both studies. A study not representative of its target crash involvement population in regard to *all three* categories would over- or under-estimate the prevalence of various driver errors and other failures precipitating crashes.

Note the dissimilarities between the "bad" (multi-vehicle CR) and "ugly" (single-vehicle) CR distributions in Table 2. In both studies, "bad" involvements most often resulted from driver inadequate surveillance, other recognition failures (principally distraction), and decision errors such as tailgating, misjudging gaps, or false assumptions. "Ugly" involvements usually resulted from physical non-performance (asleep or ill), excessive speed, maneuver execution errors (especially in NMVCCS), or vehicle failures (especially in the LTCCS). In fact, the within-study Pearson **r** correlations between the "bad" and "ugly" CR distributions in Table 2 are *negative*: -0.35 for NMVCCS and -0.15 for LTCCS. In other words, the distributions are more dissimilar than similar. This further reinforces the need for representative sampling; any sampling bias would likely distort prevalence estimates of causal factors.

In contrast, Table 2 correlations across studies but within involvement categories are positive: +0.75 for multi-vehicle involvements and +0.67 for single-vehicle involvements.

These high correlations are found even though NMVCCS and LTCCS were separate studies involving different vehicle types, driver demographics, and data collection periods. Such concordance demonstrates the robustness of causal processes operating within crash involvement categories.

4. VERTICAL CRASH HETEROGENEITY

4.1 Systematic Differences Associated with Severity

Vertical heterogeneity refers to trends seen in crash compositions across different severity levels. For example, an Australian study (*4*) found that 66% of fatal car crashes involved extreme or illegal behaviors, versus just 18% of non-fatal crashes. Reported U.S. truck driver fatigue is about five times greater in fatal crashes than in all police-reported crashes (*3*, *5*). Fatal truck crashes are twice as likely to occur at night, 62% more likely to occur on undivided roads, and 62% more likely to be frontal impacts than are truck property damage only (PDO) crashes (*5*).

CR assignment and other indicators of driver error or fault in crashes involving both trucks and passenger vehicles shift with increasing crash severity. Driver error appears more-or-less evenly split in minor crashes, but shifts heavily toward passenger vehicle drivers in more serious crashes. This trend was seen even across the three adjacent severity levels of the LTCCS. The CR was assigned to the truck in 46% of truck "B" involvements, 37% of "A" involvements, and just 23% of "K" involvements.

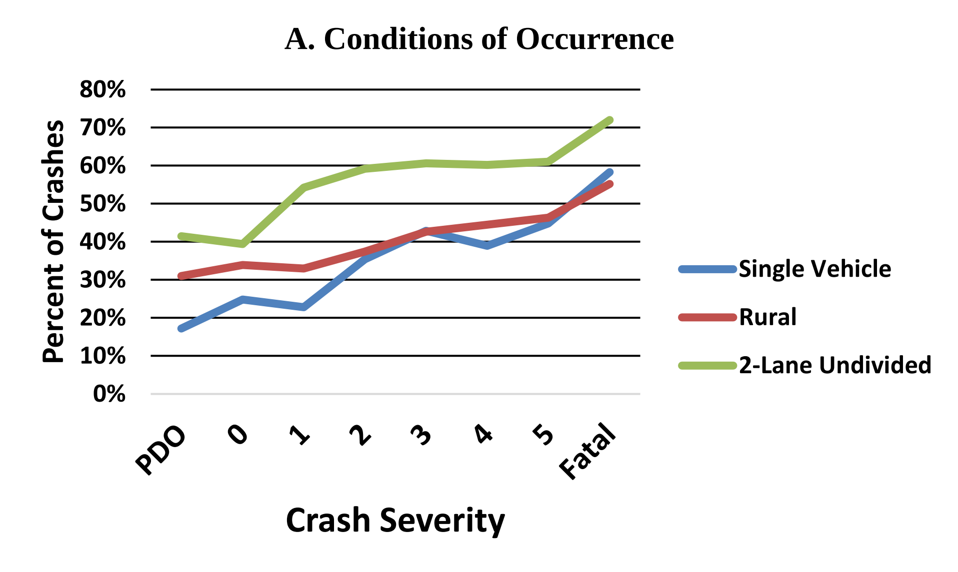
Recent AAA Safety Foundation (*6*) estimates for the percent of drowsy drivers in passenger vehicle crashes were:

- 3% of drivers involved in no-injury crashes
- 8% of drivers involved in injury crashes
- 15% of drivers involved in fatal crashes.

A 2015 NHTSA report by Blincoe et al. (*7*) shows the varying presence of six factors in crashes (including both police-reported and non-police-reported) of different severities. NHTSA estimated factor presence for the years 2008-2010 based on its National Automotive Sampling System (NASS) and related extrapolations. The severity scale used was the Maximum Abbreviated Injury Scale (MAIS), which ranges from PDO up to MAIS 6 (fatal) based on the most severe injury in the crash. Figure 2A shows percentages of three crash conditions of occurrence as a function of crash severity. Only 17% of PDO crashes were single-vehicle, but this percentage increased to 58% for fatal crashes. Similarly, the rural crash percentage rose from 31% to 72%. Of those rural crashes, the percentage occurring on two-lane roads increased from 42% to 72%.

Figure 2B shows the same NHTSA estimates for three causation-related factors: alcohol, excessive speed, and distraction. Alcohol was estimated to be present in 14% of PDO crashes versus 40% of fatals. The speed-related percentage rose from 16% to 32%. Driver distraction followed a different pattern, however. Here the percentage generally declined slightly with increasing severity, and then more sharply for fatal crashes.

These differences demonstrate that crashes cannot be accurately characterized without reference to specific severity levels. Quantitative statements about crash risks and the prevalence of causal factors (e.g., distraction, alcohol) require a stated target crash population to be meaningful.



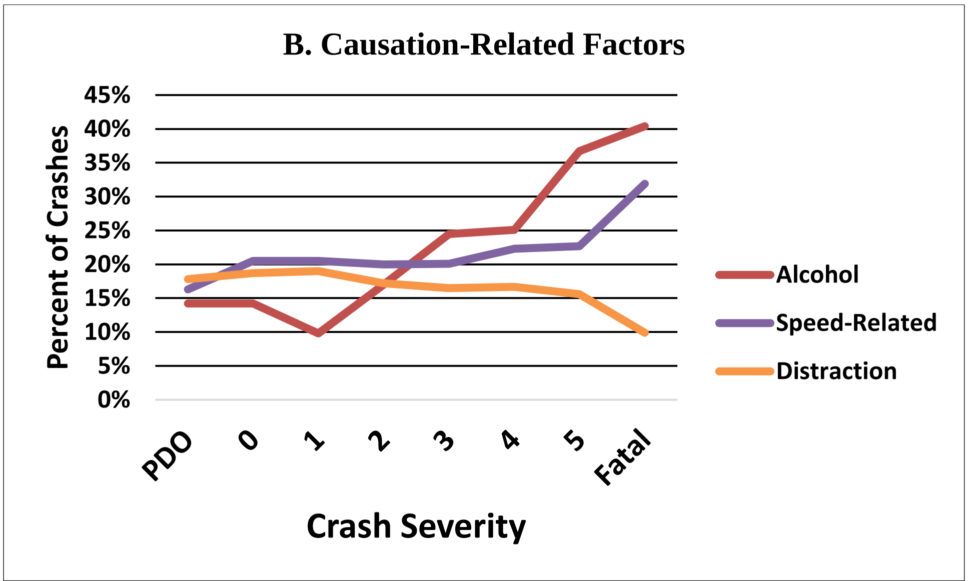


FIGURE 2 Trends in Crash (A) Conditions of Occurrence and (B) Causation-Related Factors with Increasing Severity (*7*). Note: Data for 2-lane undivided includes rural crashes only.

4.2 Is "Heinrich's Law" Invalidated?

Based on studies of industrial accidents (not traffic crashes) beginning in the 1920s, H. W. Heinrich (*8*), formulated theories regarding accident causation and prevention. The Heinrich Triangle is a well-known schematic of accident severity and frequency. As one ascends the triangle, severity increases but frequency decreases. Per Heinrich, for every 300 human errors or other uncontrolled events resulting in no injury, 29 others result in minor injury and one results in major injury.

Heinrich saw accident occurrence as a forward, linear chain where human errors and other causes were distinct and separate from their consequences. The genesis of accidents was *identical* up and down the triangle. He wrote, "The predominant causes of no-injury accidents are, in average cases, identical with the predominant causes of major injuries, and incidentally of minor injuries as well" (*8*). For this to be true of crashes, they would need to be causally homogeneous throughout the triangle.

This and some other Heinrich theories have been challenged in recent decades. The National Safety Council (*9*) has questioned the empirical basis of the work and noted that no detailed records of Heinrich's original methods or data can be found. The NSC adds that there were no established practices for rigorous behavioral observations at the time of the work. In regard to the premise of identical mechanisms, the NSC view is that, "Causal factors for low-probability, high-consequence events are rarely represented in the analytical data on frequent incidents . . ."

Given a lack of crash data supporting Heinrich and the positive evidence of vertical crash heterogeneity presented above, it appears that Heinrich's theory of identical causal mechanisms cannot be validly applied to motor vehicle crashes. Yet, the Heinrich assumption continues to be an explicit premise in major causation studies *(10, 11)*.

4.3 Crash Severity and Total Crash Harm

The main purpose of crash safety research is to identify ways to reduce the material and human consequences of crashes. The most important of these occur at the highest levels of severity. An analysis (*12*) of three years of large truck crashes in the General Estimates System (GES) found that serious (fatal + injury) crashes were about 11% of all police-reported large truck crashes but caused 78% of crash costs, 91% of reduced quality-of-life years, and 92% of lost productivity. Relevance and representativeness in relation to serious crashes seem essential for any dataset claiming high safety significance.

Table 3 presents NHTSA estimates (*7*) of 2010 U.S. all-vehicle crash numbers, economic cost, and total societal harm. NHTSA's estimates are shown here in three severity categories: fatal, injury, and PDO. Economic cost elements include damage, traffic congestion, medical, legal, insurance, and lost income. Total societal harm includes economic loss but adds monetary valuations of lost quality-of-life. In 2010, 22.1% of crashes resulted in fatalities or other injuries, but they accounted for 70.5% of economic cost and 91.4% of total societal loss.

Table 3 encompasses both police-reported and non-police-reported crashes, but the two were disaggregated by NHTSA. Police-reported estimates were based on public records, while non-police-reported estimates came from surveys in which respondents recalled crashes experienced in the previous year. Only 44.8% of 13.6 million 2010 crashes were reported, but they accounted for 83% of economic costs and 89% of total societal harm (*7*).

Table 3 Severity Percent Distributions of 2010 U.S. Crash Numbers, Economic Cost, and Total Societal Harm (*7*)

Metric: Severity:	Number of Crashes	Total Economic Cost	Total Societal Harm
Fatal	**0.2%**	**19.1%**	**36.1%**
Injury	**21.9%**	**51.4%**	**55.4%**
Property Damage Only	**77.9%**	**29.5%**	**8.6%**

Clearly, crash costs and harm reside primarily in the top tiers of the crash triangle. Studies primarily of unreported and non-injury crashes address just a small portion of the societal problem, and are not likely to be causally representative of crashes causing principal harm.

5. IMPLICATIONS FOR NATURALISTIC DRIVING STUDIES (NDSs)

5.1 Creation of Mixed-SCE Naturalistic Driving Datasets

NDSs continuously record driver behavior and road events using videos and other sensors. NDS researchers create mixed Safety-Critical Event (SCE) datasets by selecting and combining multiple dynamic events such as those recorded during avoidance maneuvers. The Federal Motor Carrier Safety Administration's (FMCSA's) current NDS on truck driver Hours-of-Service rules (*10*) measures HOS "safety impacts" by compiling five SCE types: hard brakings, hard accelerations, swerves, contacts with other objects, and driving in excess of posted speed limits. Previous HOS studies (*13*, *14*) have employed the first four of these, as well as short times-to-collision, other events chosen by researchers, and events selected by drivers themselves by activating a critical incident button. For the 2011 study (*14*), unintentional lane deviations were added to the mix as a "reliable indicator of fatigue" (P.30). They were 51% of that SCE dataset (1,118 of 2,197), though no crash-based rationale was stated for their number or the resulting proportions. Only four (0.2%) SCEs were crashes, with "crash" defined as "any contact."

NDS researchers select their SCE types based on their judgments of event importance, and on instrumentation installed on study vehicles. Researchers set unique, desired trigger thresholds for each event type, which in turn determines numbers and proportions. Thus, NDS researchers themselves control the composition of their SCE datasets.

SCEs are not sampled from crash populations. Their characteristics differ from crashes, often markedly. A recent large truck NDS *(11)* reported that 81% of its SCEs involved only the truck, whereas the truck-only proportion for police-reported crash involvements is around 20% (3). The scenario "rear-end, truck struck" (had a crash actually occurred) was seen just once in 2,899 SCEs (0.0%), compared to 10% of serious crash involvements in the LTCCS. Truck driver asleep-at-the-wheel was the CR in 14 of the 2,899 SCEs (0.5%). The LTCCS crash percentage was eight times higher, 3.8% (2). Trucks made avoidance maneuvers in 99% of SCEs versus 62% (of knowns) in the LTCCS.

In the same truck NDS study *(11)*, ”good” (not-at-fault, multi-vehicle) truck involvements were just 7.5% of SCEs compared to 44% of LTCCS involvements. Under-sampling of ”good” involvements in non-crash SCEs (compared to crashes) is almost inevitable since they are least likely to involve an avoidance maneuver or kinetic change. This was seen in Table 2 and is especially true in “good” roles such as struck vehicles in rear-end conflicts. NDS does not capture such events well.

Though one would not expect perfect SCE-crash concordance, the clear differences in origin and characteristics suggest that mixed-SCE datasets should not be taken as representing crash involvement populations.

5.2 Mixed-SCE Datasets and Heinrich’s Triangle

From its inception, the NDS SCE methodology has been based on the Heinrich theory of identical mechanisms. This was expressed in a validation study (*15*) of the NHTSA *100-Car Naturalistic Driving Study* (*16*):

- “The underlying assumption of Heinrich’s Triangle is that the unsafe acts, minor injuries, and major injuries all share the same underlying causal mechanism” (P.4).
- “For [NDSs], a surrogate measure should have the following properties:
    - The causal mechanism for surrogates . . . and crashes are the same or similar.
    - There is a strong association between the frequency of surrogate measures and crashes under different settings” (P.4).
- “One key requirement for using near-crashes as a surrogate measure is that they possess the same causal mechanism as crashes (the only difference between a crash and an appropriate near crash surrogate is the severity of the safety outputs)” (*15*, P.16).

The validation study (*15*) compared *100-Car* crashes (1% of its dataset) to near-crashes (8%). There were no comparisons to crashes in public records or to the lowest-intensity SCEs constituting 91% of study data. Various high similarities between SCE crashes and near-crashes were reported, particularly in conditions of occurrence like weather. Yet calculations from their data performed by this author (*17*) found dissimilarities:

- The crash and near-crash profiles for precipitating factors (e.g., object in roadway, other vehicle crossing straight across path, etc.) correlated only +0.18 (R^2 = .03).
- Conflict type (e.g., single-vehicle, conflict with lead vehicle, conflict with vehicle in adjacent lane, etc.) profiles correlated only +0.44 (R^2 = .20).
- Drivers reacted to crash threats in 45/68 crashes (66%) versus 723/760 near-crashes (95%). The latter percentage was near 100% because avoidance maneuvers were the principal means of SCE detection.
- Single-vehicle scenarios (including object/obstacle and parked vehicle) were 37 of 69 crashes (54%), versus 59 of 761 near-crashes (8%).

Few *100-Car* crashes were reported to police. In all there were 69 SCEs (0.8% of all 9,125) resulting in “any measureable dissipation or transfer of energy” (*15*, P. xxxvi). Only five of these caused injuries while another seven were police-reported PDO. The total police-reported percentage was 17% of detected crashes and 0.1% of all SCEs (*15, 16*).

Note also the sampling discontinuities inherent within NDS event datasets. NDS crashes are detected from impact forces, whereas NDS non-crashes are detected from various dynamic triggers. There are no consistent sampling rules across events.

5.3 All-Crash SCE Datasets

The small number of crashes within dynamically defined SCEs and the observed differences between crash and non-crash SCEs has prompted NDS researchers to limit some analyses to crashes only. The SHRP2 study was by far the largest NDS and the first to capture enough crashes for reliable analysis. Dingus et al. (*18*) reported causation-related statistics based on 905 crash involvements. The SHRP2 severity categories were (1) Airbag/injury/rollover/high delta-V crash (virtually all police reported), (2) Police-reportable, (3) Physical contact with another object, and (4) Low-risk tire strikes. The report, however, did not provide numeric distributions or disaggregations of findings by severity. Nor did it report crash involvement types.

Dingus et al. did report that 74% of its involvements had some type of associated driver error. Estimates of prevalence and elevated risk compared to "model driving" (alert, attentive, and sober) were provided for specific driver errors and risky behaviors/states. Results were extrapolated directly to the estimated U.S. crash (police-reported plus non-reported) without discussion of the concordance between SHRP2 crash severities and those in public records (e.g., per MAIS or KABCO). In its discussion, the paper asserted that 4 million (36%) of 11 million annual U.S. crashes could be avoided if no driver distraction were present. However, NHTSA's finding (*7*) that the role of distraction decreases with increasing crash severity (see Figure 2B) and broader evidence presented here suggests that NDS findings should not be extrapolated directly to the national picture.

The Insurance Institute for Highway Safety (*19*) analysed a sample of crash involvements from the same SHRP2 source. IIHS did provide a numeric distribution per the same severity categories:

1. 98 Airbag/injury/rollover/high delta-V (7%)
2. 150 Police-reportable (10%)
3. 597 Minor non-reportable contacts (41%)
4. 620 Low-risk tire strikes (42%).

Excluding tire strikes, the ratio of non-reportable to reportable events was 597:(98+150) = 597:248 = 2.41. NHTSA's estimated 2010 unreported-to-reported ratio was 7.51M:6.09M = 1.23 (*7*). From this ratio difference and based on the methodologies used, it seems clear that SHRP2 used a lower damage threshold for "crash" than that used by NHTSA to develop its estimates. The two unreported crash samples cannot be assumed to be equivalent. NDS crash thresholds detected via onboard vehicle sensors are likely to be considerably lower than survey respondents' thresholds for recalling and reporting a crash experienced over the prior year. A small animal strike, for example, could have sufficient associated kinematic change for NDS detection and classification as a minor contact (Category 3 above) but would not likely be considered or remembered by a driver as a crash.

IIHS reported the two predominant SHRP2 crash types to be rear-end and road departure, with the proportions of these two types varying sharply by event severity. For example, rear-end scenarios were 54% of the 98 most severe events, but only 12% of the 597 minor contacts. They are about one-third of police-reported crashes nationally (*19*).

Risks associated with various behaviors such as cell phone use also varied significantly with severity. IIHS concluded that the SHRP2 crash severity and crash type distributions differed substantially from crashes reported to police and forming the basis of national assessments of the motor vehicle crash problem. They concluded that, "Researchers must take into account how the crash populations in the SHRP2 data differ from national crash databases in order to describe results precisely and use due caution in generalizing results to the population of U.S. crashes."

IIHS provided this author with SHRP2 crash statistics related to their analysis and to discussions here of horizontal heterogeneity. SHRP2 analysts did not assign CRs but they did assign fault to vehicles/drivers. Version 2.0 of the SHRP2 dataset was accessed through the InSight website. Statistics on fault for motorists (plus a very small number of non-motorists) were accessed for 765 of the same 845 Category 1-3 crashes/contacts in the IIHS report. Classifications were:

- 138 (18%) multi-vehicle, not-at-fault ("good")
- 165 (22%) multi-vehicle, at-fault ("bad")
- 462 (60%) single-vehicle and at-fault ("ugly").

Recall that the NMVCCS involvement distribution was 44% "good," 39% "bad," and 17% "ugly." Such large discrepancies between the two datasets are difficult to unravel. Compared to NMVCCS, the SHRP2 crash dataset contained a far higher percentage of single-vehicle events. Also, of SHRP2 multi-vehicle involvements, 54% (165/303) were "bad." In NMVCCS, 47% were "bad," consistent with the fact that one vehicle/driver in a multi-vehicle crash was at-fault, but that some of these crashes involved 3+ vehicles and thus more than one vehicle/driver designated as "good." Given this possibility, the SHRP2 "bad" percentage of more than 50% in multi-vehicle crashes suggests a sampling bias, either involving study subjects or criterion events.

Thus it appears that neither all-crash nor mixed SCE datasets adequately represent the horizontal profiles of established target crash involvement populations. One way to make NDS data correspond more clearly to conventional crash (as distinct from *involvement*) statistics would be to use only at-fault events in analyses. Allowing that every crash has one at-fault vehicle/driver, using only at-fault involvements eliminates concerns about undersampling not-at-fault ("good") involvements. This manipulation would not address other sampling concerns, however.

6. IMPROVING EXTERNAL VALIDITY

External validity is the extent to which observations made in any study generalize beyond its specifics to phenomena of broader importance (*20*). In crash studies, this ordinarily equates to the representativeness of a sample in relation to a specified national target crash population. The LTCCS, for example, was a stratified random sample of U.S. fatal and injury crashes involving a large truck (*2*, *3*). Other target crash populations may be chosen based on study goals. Various crash strata are commonly targeted, including all fatal crashes, all serious (fatal/injury) crashes, and all police-reported crashes. None of these populations represents all physical impacts experienced by motorists, but they do represent crash populations designated and analyzed as public health threats.

Naturalistic driving presents a more daunting external validity problem, however. NDS datasets are not sampled or derived from any crash population unless those datasets are limited to crashes. The vast majority of SCEs in most NDS datasets are *not* crashes. Researcher-chosen dynamic thresholds (e.g., for deceleration in hard braking) determine the numbers and relative proportions of diverse event types within the dataset. Lower thresholds increase event-type proportions while higher thresholds decrease them. This process produces a mixed-event dataset intended to elucidate crash genesis but not one that is analytically or empirically linked to any crash population. Thus there is no firm basis for believing that mixed-SCE datasets are externally valid; i.e., that their estimates of the prevalence of various crash risks generalize beyond their samples to important, defined crash populations.

A potential partial solution would retain the use of non-crash SCEs but index them to a target crash population such as police-reported crashes. The SCE-crash gap could be reduced by differentially weighting SCEs *post hoc* to better match the profiles of target crashes from datasets such as the GES and the Fatality Analysis Reporting System (FARS). Indexing could be based on objective crash and SCE characteristics; i.e., descriptors already standard in GES and FARS. A key NDS advantage (the ability to replay event videos and other data) would be retained while a key disadvantage (lack of external representativeness) would be reduced somewhat.

Indexing is used to make unrepresentative political, social science, and other survey samples more representative of their target populations (*21*). A common technique assigns an adjustment weight to each survey respondent. Persons or cases in under-represented groups get a weight larger than 1 while those in over-represented groups get a weight smaller than 1. This modeling technique requires *auxiliary variables* with known population distributions. Familiar auxiliary variables in surveys are objective characteristics like gender, age, marital status, and region of the country.

GES and other national crash datasets regularly classify crash involvements using potential auxiliary variables. These include attributes describing the “who,” “when,” “where,” and “how” of crashes:

- “Who” -- driver age and sex
- “When” -- hour and day of crash
- “Where” – roadway type variables
- “How” – number of vehicles involved, crash type.

Indexing SCEs to crashes would take full advantage of NDS’s unique “instant replay” capabilities while reducing sample unrepresentativeness. Conclusions drawn about the prevalence of risk factors and about overall risk would be more realistic and relevant to target crashes. Two TRB committees (*22, 23*) have recognized the potential value of this approach in research needs statements available on the TRB website.

Restricting NDSs to crashes can achieve external validity in relation to all physical contacts assuming that driver and vehicle samples are representative. This does not make them externally valid in relation to societally designated critical crash populations, however.

A high majority of physical contacts do not qualify as police-reportable and certainly do not result in serious injuries or death. Thus even all-crash NDS datasets are not externally valid in relation to crash populations causing most societal harm. Given the marked differences between minor and serious crashes, one could not draw valid causal conclusions about our highest priority crash populations without even larger NDS crash samples with higher thresholds and/or *post hoc* indexing of these samples to key populations.

7. CONCLUSIONS

This paper asserts the following:

- Crashes are heterogeneous "horizontally" in regard to when, where, and how they happen. CRs and other cause-related factors vary for different involvement types and categories.
- An instructive taxonomy separates crash involvements by CR assignment and number of involved vehicles into three categories labeled here as "the good, the bad, and the ugly."
- Crashes are vertically heterogeneous. The prevalence of their causal factors and other characteristics varies markedly by severity.
- Therefore, the historic "Heinrich Law" positing identical causal mechanisms across accidents of different severities is not true in regard to traffic crashes.
- Abstract phrases like "crash risk" have no definite meaning without an accompanying crash population referent.
- Narrow (e.g., within a crash type) or otherwise limited extrapolations are likely to be more valid than broader extrapolations.
- Only about 20% of U.S. crashes cause injuries or fatalities, but these serious crashes account for about 70% of economic costs and more than 90% of total societal harm.
- Therefore, an important criterion for judging the value of crash-related research is its validity in relation to serious crashes.
- NDS event datasets combine data from disparate sources: impact forces for crashes and miscellaneous dynamic triggers for non-crashes.
- Mixed non-crash NDS SCE datasets are not likely to be externally valid in relation to crash populations because they are not derived from those populations. *Post hoc* numeric indexing might improve validity somewhat.
- Important prevalence differences may be seen even between adjacent crash and near-crash subsets of the same NDS dataset.
- All-crash NDS SCE datasets potentially capture samples representative of the universe of physical contacts experienced by motorists. This does not, however, make their prevalence estimates applicable to crashes deemed important by society.
- The SHRP2 crash dataset should not be taken as representing U.S. crash populations documented in public records and major crash research databases.
- "Minor non-reportable contacts" within SHRP2 should not be taken as representing non-reported crashes as defined by NHTSA.
- NDS is unchallenged in its capabilities to provide "why" answers for individual events captured and recorded by its sensors. Direct and accurate observation of individual

events does not in itself make prevalence extrapolations accurate, however. Comparing NDS event data to the "who," "when," "where," and "how" distributions of important crashes and crash involvements might provide more meaningful insights from this innovative research technology.

ACKNOWLEDGEMENTS

Thanks to the staff of the NHTSA National Center for Statistics and Analysis, including Seymour Stern, Adam Toth, and Lyn Cianflocco, and to Joseph Bocanegra, then of the Virginia Tech Transportation Institute, for retrieving crash statistics per specifications provided by author. Thanks also to David Kidd of IIHS for providing SHRP2 statistics supplemental to its cited report.

REFERENCES

(1) NHTSA. *National Motor Vehicle Crash Causation Study: Report to Congress.* DOT HS 811 059. U.S. DOT, 2008.

(2) Starnes, M. *LTCCS: An Initial Overview*. DOT HS 810 646. NHTSA, U.S. DOT, 2006.

(3) Knipling, R.R. *Safety for the Long Haul; Large Truck Crash Risk, Causation, & Prevention*. American Trucking Associations, Arlington, VA, 2009.

(4) Wundersitz, L. and M. Baldock. *The Relative Contribution of System Failures and Extreme Behaviour in South Australian Crashes,* Publication CASR092, Centre for Automotive Safety Research, University of Adelaide, 2011.

(5) FMCSA. *Large Truck Crash Facts 2012*. FMCSA-RRA-14-004. U.S. DOT, 2014.

(6) Tefft BC. Prevalence of Motor Vehicle Crashes Involving Drowsy Drivers, United States, 1999-2008. *Accident Analysis & Prevention*, Vol. 45, No. 1, 2012, pp.180-186.

(7) Blincoe, L. J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. *The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised).* DOT HS 812 013, NHTSA, 2015.

(8) Heinrich, H.W. (1959). *Industrial Accidents Prevention: A Scientific Approach, 2nd Edition*. McGraw Hill, New York: McGraw-Hill, 1941.

(9) Manuele, F. A. *Heinrich Revisited: Truisms or Myths, 2nd Edition.* National Safety Council, Itasca, IL, 2014.

(10) FMCSA. *Commercial Motor Vehicle Driver Restart Study Plan*. U. S. DOT, February 12, 2015.

(11) Blanco, M. J. S. Hickman, R. L. Olson, J. Bocanegra, R. J. Hanowski, A. Nakata, M. Greening, P. Madison, G. T. Holbrook, and D. Bowman. *Investigating Critical Incidents, Driver Restart Period, Sleep Quantity, and Crash Countermeasures in Commercial Vehicle Operations Using ND Data Collection.* FMCSA-RRR-13-017. Wash. DC: DoT., November 2016.

(12) Zaloshnja, E. and T. Miller. *Unit Costs of Medium & Heavy Truck Crashes*. Final Report, Pacific Institute for Research & Evaluation for FMCSA, U.S. DOT, 2007.

(13) Hanowski, R. J., R. L. Olson, J. Bocanegra, and J. S. Hickman. *Analysis of Risk as a Function of Driving-Hour: Assessment of Driving-Hours 1 Through 11*. FMCSA-RRR-08-002, FMCSA, 2008.

(14) Blanco, M., R. J. Hanowski, R. L. Olson, J. F. Morgan, S. A. Soccolich, S-C Wu, and F. Guo. *The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations*. FMCSA-RRR-11-017, FMCSA, 2011.

(15) Guo, F., S. G. Klauer, M. T. McGill, and T. A. Dingus. *Evaluating the Relationship between Near-Crashes and Crashes: Can Near-Crashes Serve as a Surrogate Safety Metric for Crashes?* DOT HS 811 382, NHTSA, October 2010.

(16) Dingus, T. A., S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R. R. Knipling. *The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment*. DOT HS 810 593, NHTSA, 2006.

(17) Knipling, R.R. Naturalistic Driving Events: No Harm, No Foul, No Validity. *Driving Assessment 2015*, Paper #30, 2015.

(18) Dingus, T.A., F. Guo, S. Lee, J.F. Antina, M. Pereza, M. Buchanan-Kinga, and J. Hankeya. Driver Crash Risk Factors and Prevalence Evaluation Using Naturalistic Driving Data. *Proceedings of the National Academy of Science*, Vol. 113, No. 10, 2016, 2636-2641.

(19) Kidd, D. G. & McCartt, A. T., Insurance Institute for Highway Safety. The relevance of Crash Type and Severity when Estimating Crash Risk Using the SHRP2 Naturalistic Driving Data. *4th International Driver Distraction and Inattention Conference*, Sydney, New South Wales, 2015.

(20) Privitera, G. J. *Research Methods for the Behavioral Sciences,* Sage Publications, Inc., Los Angeles, 2014.

(21) National Institutes of Health (NIH) Office of Behavioral and Social Sciences Research. Sample Surveys. *e-Source*. http://www.esourceresearch.org/tabid/374/default.aspx, Accessed February 7, 2016.

(22) TRB Committee on Safety Data, Analysis, & Evaluation (ANB20). Indexing Naturalistic Driving Events to Crashes. Research Needs Statement available at *https://rns.trb.org/dproject.asp?n=40810*. Accessed July 22, 2016.

(23) TRB Committee on Truck & Bus Safety (ANB70). Toward Naturalistic Driving Crash Representativeness. Research Needs Statements available at *https://rns.trb.org/dproject.asp?n=39354*. Accessed July 22, 2016.