

Jun 29th, 12:00 AM

Threats to Scientific Validity in Truck Driver Hours-of-Service Studies

Ronald R. Knipling

Safety for the Long Haul Inc. Arlington, VA

Follow this and additional works at: <http://ir.uiowa.edu/drivingassessment>

Knipling, Ronald R.. Threats to Scientific Validity in Truck Driver Hours-of-Service Studies. In: Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 26-29, 2017, Manchester Village, Vermont. Iowa City, IA: Public Policy Center, University of Iowa, 2017: 382-388. <https://doi.org/10.17077/drivingassessment.1662>

This Event is brought to you for free and open access by the Public Policy Center at Iowa Research Online. It has been accepted for inclusion in Driving Assessment Conference by an authorized administrator of Iowa Research Online. For more information, please contact lib-ir@uiowa.edu.

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

Ronald R. Knipling
Safety for the Long Haul Inc.
Arlington, Virginia, USA
Email: rknippling@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

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

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

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 (Privitera, 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 <i>Driver Fatigue and Alertness Study</i> (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.

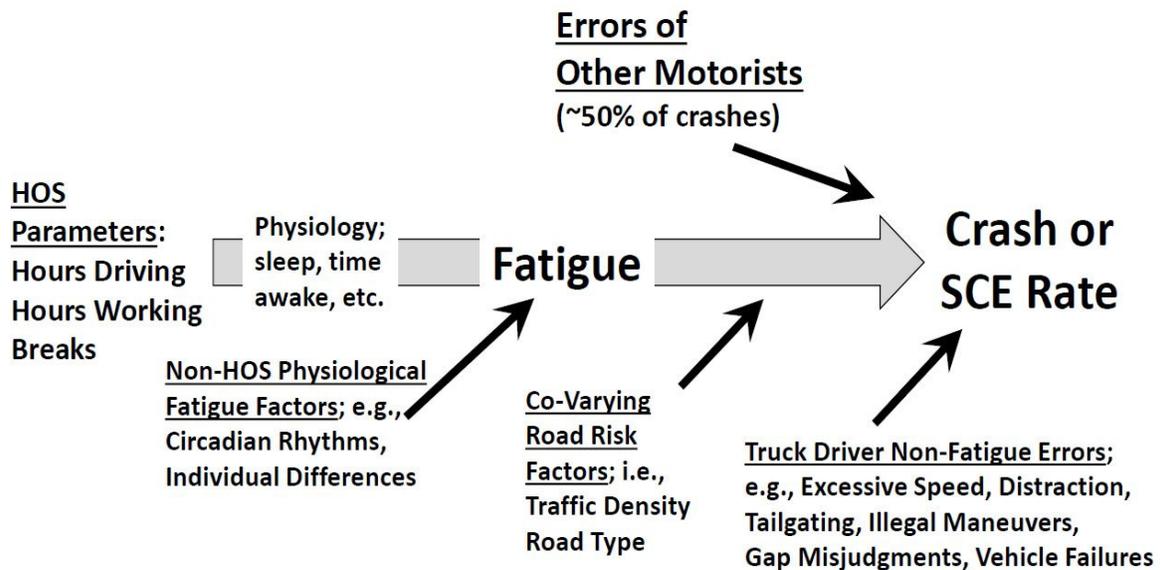


Figure 1. Potential confounds in studies relating HOS parameters to truck crash or SCE rate.

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 (Privitera, 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).

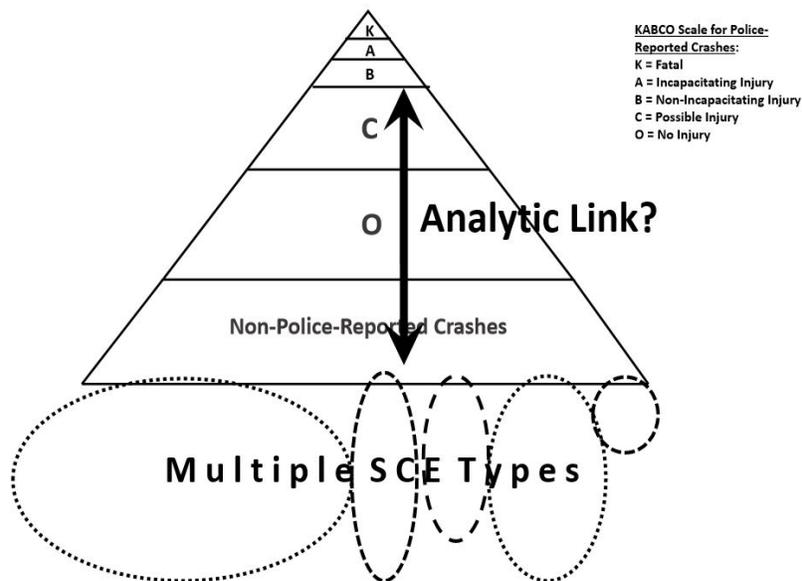


Figure 2. Crash triangle for crashes plus multiple SCE types constituting SCE datasets

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. (2008) 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 <i>inversely</i> 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 at <https://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.