NRC/CNSTAT Commercial Driver Fatigue Panel

COMMISSIONED PAPER

REVIEW OF COMMERCIAL DRIVER FATIGUE RESEARCH METHODOLOGIES

By

Ronald R. Knipling
Safety for the Long Haul, Inc.
(703) 533-2895, rknipling@verizon.net
www.safetyforthelonghaul.com

Submitted to:
National Research Council (NRC) Committee on National Statistics (CNSTAT)

April 9, 2015

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

The author expresses particular thanks to panel member Gerald P. Krueger for his detailed review of the manuscript along with extensive comments, corrections, and contributions of supportive material.

ACRONYMS

Acronym	Term
AATW	Asleep at the wheel
ANOVA	Analysis of Variance
ATRI	American Transportation Research Institute
CDL	Commercial Drivers License
CDS	Crashworthiness Data System (passenger vehicle crash database)
CFR	Code of Federal Regulations
CI	Confidence Interval
CMV	Commercial motor vehicle
CNSTAT	Committee on National Statistics
CV	Controlled variable (held constant)
CR	Critical Reason
CTT	Critical Tracking Task
CUT	Combination-Unit Truck (Tractor Semi-Trailer)
CVSA	Commercial Vehicle Safety Alliance (enforcement professional association)
DFAS	Driver Fatigue & Alertness Study (Wylie et al., 1996)
DOT	Department of Transportation (Federal, unless otherwise specified)
DOW	Day of week
DSST	Digit Symbol Substitution Test
DV	Dependent variable (measure of driver performance and/or safety)
EEG	Electroencephalograph
EOG	Electrooculograph
FARS	Fatality Analysis Reporting System (census of fatal crashes)
FMCSA	Federal Motor Carrier Safety Administration
FMCSR	Federal Motor Carrier Safety Regulation
GES	General Estimates System (sampling system for all police-reported crashes)
GVWR	Gross Vehicle Weight Rating
HOS	Hours-of-Service
IIHS	Insurance Institute for Highway Safety
IV	Independent variable
“KABCO”	Five severity levels of police-reported crashes in most states
KSS	Karolinska Sleepiness Scale (subjective self-assessment)
LCV	Longer combination vehicle (e.g., double-trailer)
LOS	Level-of-service (measure of traffic density)
L/SH	Local/short haul (trucking operation)
LTCCS	Large Truck Crash Causation Study
LTL	Less-than truckload (trucking operation)
MVR	Motor Vehicle Record
NAS	National Academy of Science
NASS	National Automotive Sampling System (NHTSA crash databases)
ND	Naturalistic Driving
NHTSA	National Highway Traffic Safety Administration

NIOSH	National Institute for Occupational Safety & Health
NMVCCS	National Motor Vehicle Crash Causation Survey
NRC	National Research Council
NTSB	National Transportation Safety Board
OOS	Out-of-service
ORD	Observer Rating of Drowsiness
OSA	Obstructive Sleep Apnea
PAR	Police Accident Report
PERCLOS	Percent eye closure (alertness measure)
PDO	Property damage only (crash)
PSG	Polysomnographic
PSU	Penn State University
PVT	Psychomotor Vigilance Test
SCE	Safety-Critical Event (in Naturalistic Driving)
SDLP	Standard deviation of lane position (driver performance measure)
SPM	Sleep performance model
SSS	Stanford Sleepiness Scale
SUT	Single-Unit Truck (Straight Truck)
SV	Single-vehicle [crash]
TIFA	Trucks in Fatal Accidents database
TOD	Time-of-day
TOT	Time-on-task
TL	Truckload (trucking operation)
TRB	Transportation Research Board
TTC	Time-to-collision
UCV	Uncontrolled variable
ULD	Unintended lane deviation
VMT	Vehicle Miles Traveled
VTTI	Virginia Tech Transportation Institute

SUMMARY

This paper reviews methodologies applied (and which could be applied) to the study of commercial motor vehicle (CMV) driver fatigue and driving safety. This includes methodologies to quantify and describe the role of fatigue and methodologies to quantify and characterize factors affecting fatigue. Typically these factors are HOS parameters (e.g., hours driving) or are otherwise closely related to HOS concerns. After a general review of crash causation and the fatigue problem, the paper overviews basic scientific concepts and principles which should be applied in evaluating past work and planning future studies. Methodological topics include scientific variables (both manipulated and measured), sampling from populations, and research designs. "Fatigue" is a construct which cannot be observed directly. Therefore, the validity of fatigue measures is problematic. There are multiple measures of both alertness/fatigue and safety/risk which can be validated, but none should be accepted without scrutiny. Representative sampling from driver populations is conceptually straightforward but in practice is extremely difficult in relation to CMV drivers because of their heterogeneity. Few studies have claimed to generate statistically representative samples. Research designs are classified as nonexperimental, experimental, and quasi-experimental. Only experimental designs can demonstrate cause-effect relationships unequivocally. Yet most HOS research has not been experimental. Many studies have used quasi-experimental designs, so defined because they lack one or more key element of experimental control. Specifically, they involve non-random assignments, pre-existing (vs. manipulated) factors, and/or lack comparison/control groups. These shortcomings compromise the validity of findings, including the extent to which they accurately predict real-world driver alertness and safety.

Following the review of principles, the paper reviews 20 studies from the perspective of research methodology and validity. Their key specific findings are cited, and notable study limitations are discussed. Study descriptions and discussions address:

- Overview and primary study purpose.
- Study design.
- Subjects and sample frame.
- Predictors; i.e., independent variables (IVs) and quasi-IVs.
- Dependent variables (DVs).
- Notable controlled variables (CVs).
- Notable uncontrolled variables (UCVs).
- Principal study findings.
- Study limitations & potential improvements.
- Citation.

The 20 studies are subdivided into two groups based on general purpose, though there is some overlap between the groups. The first group includes studies designed primarily to quantify and describe the driver fatigue problem and/or crash causes in general. These include:

1. Safety Study: Fatigue, Alcohol, Other Drugs, and Medical Factors in Fatal-to-the-Driver Heavy Truck Crashes (National Transportation Safety Board, 1990)
2. Large Truck Crash Causation Study (FMCSA, 2006; Starnes, 2006; other reports)
3. Fatigue Analyses from 16 Months of Naturalistic Commercial Motor Vehicle Driving Data (Wiegand et al., 2008)
4. Near-Crashes as Surrogate Safety Metric for Crashes (Guo et al., 2010)
5. An Assessment of Driver Drowsiness, Distraction, and Performance in a Naturalistic Setting (Barr et al., 2011; Hanowski et al., 2000)
6. Prevalence of Fatigue-Related Crashes Estimated from Multiple Imputation of Crashworthiness Data System (CDS) Unknowns (Tefft, 2014; Tefft, 2012).

The second and larger group includes studies examining factors affecting fatigue, most notably HOS-related parameters such as hours working and hours off-duty:

1. Case-Control Studies of Large Truck Crashes (Jones and Stein, 1987, 1989; Teoh et al., 2015)
2. Driver Fatigue & Alertness Study (DFAS; Wylie et al., 1996)
3. Effects of Operating Practices on Commercial Driver Alertness (O'Neill et al., 1999)
4. Effects of Sleep Schedules on CMV Driver Performance: (Balkin et al., 2000)
    a. (1) Actigraphic Assessment of Sleep of CMV Drivers over 20 Days
    b. (2) Sleep Dose/Response Study
5. Stress and Fatigue Effects of Driving Longer Combination Vehicles (FMCSA, 2000)
6. HOS & Fatigue-Related Survey of Long-Distance Truck Drivers (McCartt et al., 2005, 2008)
7. Analysis of Risk as a Function of Driving-Hour: Assessment of Driving-Hours 1 Through 11 (Hanowski et al., 2008)
8. The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations (Blanco et al., 2011)
9. Hours of Service and Driver Fatigue: Driver Characteristics Research (Jovanis et al., 2011)
10. Motorcoach Driver Fatigue Study 2011 (Belenky et al., 2012)
11. Investigation of the Effects of Split Sleep Schedules on Commercial Vehicle Driver Safety and Health (Belenky et al., 2012)
12. Laboratory Study of the Efficacy of the 34-Hour Restart (Van Dongen & Belenky, 2010)
13. Field Study of the Efficacy of the New Restart Provision for Hours of Service (Van Dongen & Mollicone, 2013)
14. Effect of Circadian Rhythms and Driving Duration on Fatigue Level and Driving Performance of Professional Drivers (Zhang et al., 2014).

This paper's scope is largely defined by the Federal Motor Carrier Safety Administration's (FMCSA's) role in establishing HOS rules for interstate commercial vehicle transport and in promoting related countermeasures to driver fatigue-related crashes. However, this paper does *not* address various important fatigue issues outside its scope. Topics not addressed include driver medical qualifications (e.g., relating to Obstructive Sleep Apnea, a major sleep disorder), the long-term health consequences of fatigue, and HOS enforcement methods (e.g., paper logs vs. Electronic Logging Devices).

This paper does not attempt to reach firm conclusions in regard to fatigue's causes and characteristics, except insofar as they affect methodology. Conclusions are drawn in regard to recommended methods and conspicuous research needs. The final section of the paper suggests 16 best practices for future research and articulates 13 research/development needs. The emphasis in this discussion is on major studies with implications HOS rules, other government policies, or major fatigue countermeasures.

1. INTRODUCTION

1.1 Overview

This paper has been written in support of a National Research Council (NRC) Committee on National Statistics (CNSTAT) Panel on Research Methodologies and Statistical Approaches to Understanding Driver Fatigue Factors in Motor Carrier Safety and Driver Health. The NRC/CNSTAT Commercial Driver Fatigue Panel is reviewing the relationship between HOS regulations, driver fatigue, and truck (and bus/motorcoach) accident frequency, as well as the longer term health implications of truck and bus driving. Fatigue factors addressed include hours of driving, hours on duty, breaks, time-of-day, and periods of rest. The committee sponsor is the Federal Motor Carrier Safety Administration (FMCSA).

Commercial driver fatigue should be understood within the more general frameworks of crash causation and the general commercial motor vehicle (CMV) crash picture. Numerous interacting factors affect driver alertness and fatigue, many of which are not easily addressable by HOS rules or other government regulations. This limits potential safety impacts, and also greatly complicates the scientific process as it relates to rule development. These concepts and challenges are reviewed below.

1.2 Concepts of Crash Causation

Efforts to reduce fatigue-related crashes require an understanding of the causal mechanisms by which fatigue operates. Stated more simply, we ought to know *how* fatigue causes crashes. This section briefly presents two conceptual models of crash causation. Neither model has been fully validated, but both provide heuristic frameworks. The two models present different perspectives but are not necessarily incompatible. Future research may define and elaborate how fatigue operates within these or other models.

1.2.1 Risk-Cause Model

Figure 1 shows a simplistic, conceptual crash timeline encompassing two types of causal factors: predisposing risk factors and proximal causes (Knipling, 2009). In the model, risk factors set up a probability that driver errors or other proximal failures occur or have greater consequences. Proximal causes are seen as discrete triggering behaviors or other events, as opposed to preexisting driver, vehicle, or environmental risk factors.

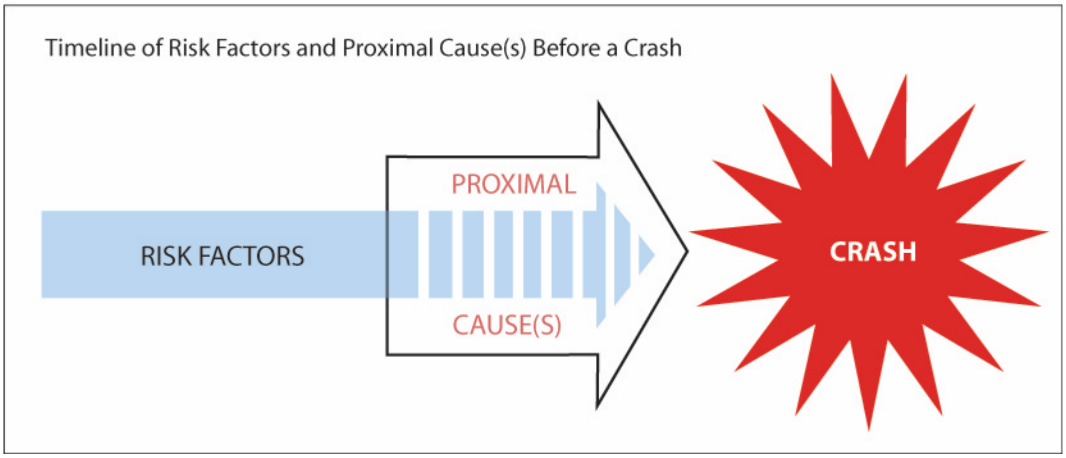

***Figure 1. Timeline of risk factors and proximal cause(s) before a crash.* Reprinted from Knipling (2009).**

There are numerous categories of crash risk factors, and many different discernible factors may be operating simultaneously to raise or lower risk. Risk factor categories include:

- Enduring driver factors; e.g., gender, personality, medical conditions, age, experience
- Temporary driver factors; e.g., mood, recent sleep, time-of-day, drug use, road familiarity
- Vehicle; e.g., mechanical condition, safety features & technologies
- Roadway and environmental; e.g., divided vs. undivided, traffic density
- Carrier operations & management; e.g., fleet-based driver training, driver performance monitoring & evaluation
- Government policies & practices; e.g., driver licensing, HOS rules, enforcement practices.

Proximal causes also fall into multiple categories. The Large Truck Crash Causation Study (LTCCS) performed in-depth investigations of 963 large truck crashes (see Section 3.2 below). The LTCCS classified proximal causes (termed Critical Reasons or CRs) into six main categories, four of which were types of driver errors. Just one CR was designated and assigned to one involved vehicle. The percentages below are from the LTCCS for truck at-fault crashes (Starnes, 2006).

- Driver physical factor; e.g., medical crisis, asleep-at-the-wheel (12%)
- Recognition failure; e.g., inattention, daydreaming, distraction, looked but did not see (30%)
- Decision error; willful unsafe behavior (e.g., speeding, tailgating, illegal maneuver) or misjudgment, such as misjudging the speed of another vehicle (40%)
- Performance or response execution error; e.g., poorly executed turn, overcompensation after avoidance maneuver (6%).
- Vehicle failures; e.g., brakes, tires, cargo shifts (10%)
- Roadway/environmental factors; e.g., missing signs, extreme weather (2%).

Data on driver fatigue from the LTCCS, and related caveats, will be discussed in Section 3.2 below. Fatigue was coded in two different ways in the LTCCS, corresponding partially to the two factor types in the above model.

The LTCCS designated just one CR per crash, but other studies have permitted more than one. The National Transportation Safety Board's 1990 study of fatal-to-the-driver truck crashes (see Section 3.1 below) also identified proximal "contributing" causes, but many of their 182 crashes were attributed to more than one cause.

Comparisons of crashes and non-crashes (time periods or points in time where crashes did not occur) can potentially show the strength of the association of risk factors with crashes. This is the approach taken in a series of case-control studies described in Section 3.1. In Naturalistic Driving (ND), comparisons can be made between Safety-Critical Event (SCEs) and randomly-selected "baseline" time periods (e.g., see Wiegand et al., 2008, Section 3.3).

1.2.2 "Swiss Cheese" Model

British human error theorist James Reason formulated a "Swiss Cheese" accident model (Reason, 1990). Reason visualized multiple layers of error prevention termed "defenses." Conceptually, the defense layers are like slices of Swiss cheese, with holes where there is no defense. A crash or other accident occurs when there is alignment of the holes or, stated in another way, a convergence of risk factors. Short et al. (2007) conceptualized Reason's Swiss cheese model in relation to motor carrier safety management practices like driver training and vehicle maintenance. Figure 2, from Knipling (2009), conceptualizes it in the context of driver behavior, attention, and traffic conditions.

In the Swiss Cheese conception, fatigue would function as one of the holes in the Attention layer. The fatigue hole would enlarge or contract with changes in alertness. Other, non-fatigue factors (e.g., speed, road conditions, vehicle features) would also modulate overall crash risk. Individual risk layers would not need to be primary causes in order to directly affect risk. An increase or decrease in the "holes" of any layer would have the same proportional effect on overall risk. For example, if the holes in the attention layer doubled in size, twice as many "arrows" would make it through to cause a crash.

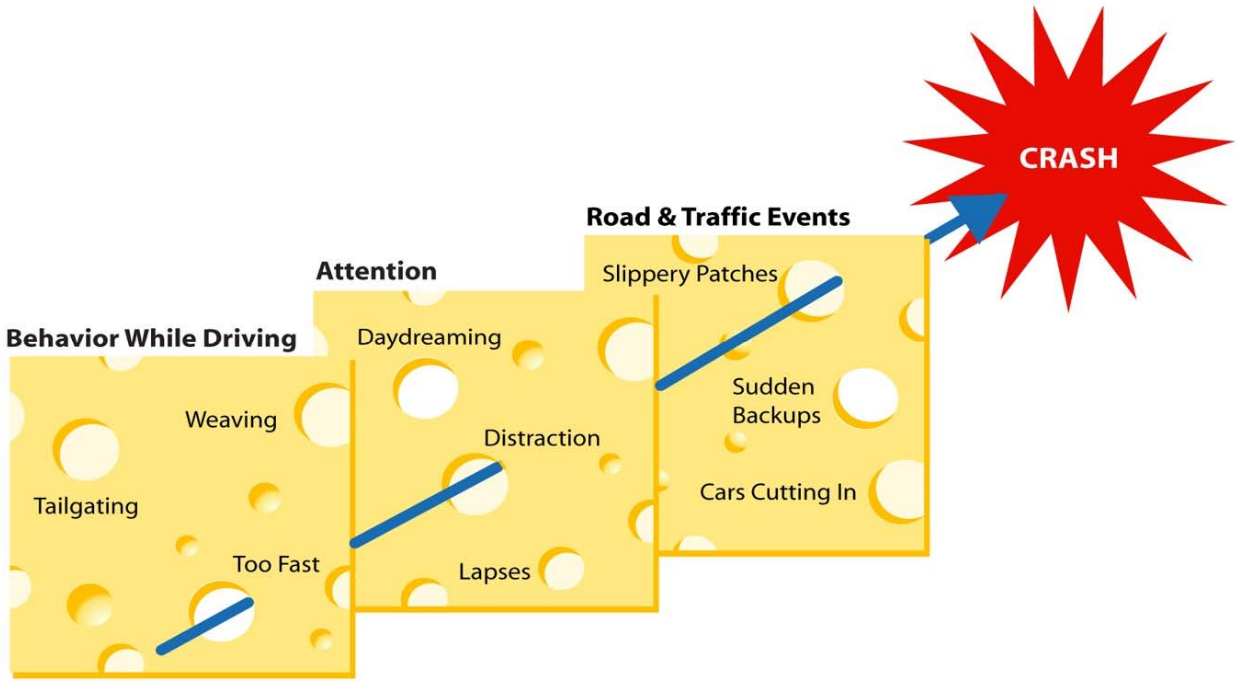

Figure 2. "Swiss Cheese" crash causation model. **Reprinted from Knipling (2009). Adapted from Reason (1990) and Short et al., 2007.**

The Swiss Cheese model has intuitive appeal but has not been validated in regard to fatigue or other factors. Multi-variate analyses showing combined effects of two or more fatigue-related factors (e.g., early morning and lack of sleep) are supportive of the concept, but they are also consistent with the risk-cause model. For the model to function as shown, the different layers would have to be independent of each other. This is clearly not the case for some crash factors. For example, busy traffic and drowsiness both increase crash risk, but driving in busy traffic seems to reduce observed drowsiness.

Future fatigue research could elucidate and validate one or both of the above models. For the Risk-Cause model, such research could explain how fatigue operates as a risk factor and how it precipitates crashes. For the Swiss Cheese model, research could elucidate the nature of risk increase ("hole" enlargements) and decreases ("hole" contractions).

1.3 The Large Truck & Bus Crash Picture

Commercial vehicles include large trucks and buses (motorcoaches). Large trucks are those with gross vehicle weight ratings (GVWR) of greater than 10,000 pounds, but a high majority of large truck crashes involve trucks with GVWRs of greater than 26,000 pounds. The two major large truck configurations are combination-unit trucks (typically tractor-semitrailers) and single-unit trucks (also called straight trucks). Combination-unit trucks (CUTs) typically operate in long-haul service whereas most single-unit trucks (SUTs) are short-haul. The greater CUT mileage means greater exposure to crash risk. In 2008, CUTs were 25% of registered trucks, compiled 63% of truck Vehicle Miles Traveled (VMT), and were 74% of trucks involved in fatal crashes (Craft, 2010).

Large trucks (CUTs + SUTs) far outnumber buses in number of vehicles, mileage, and crash involvements. In the U.S. in 2012 there were 3,802 trucks involved in fatal crashes, versus 251 buses (FMCSA, 2014). From a statistical perspective, the much larger number of truck-related crashes means that statistics on them are more robust and can be analyzed in more detail. Thus, many of the crash statistics presented here and in many other crash reports are truck-only.

In 2012, 4,183 people were killed in 3,702 fatal crashes involving large trucks and buses. This was 12.5% of the 33,561 total traffic crash fatalities for the year. About 1.0% of crashes involving a large truck or bus were fatal, versus 0.5% of crashes involving passenger vehicles. The last four decades have seen impressive declines in fatal crash involvement rates for most vehicle types. Between 1975 and 2012, the large truck fatal crash involvement rate per vehicle miles traveled (VMT) declined by 71% while that for passenger vehicles declined 65%. The truck rate still exceeds the passenger vehicle rate, however, due primarily to truck size. In 2012, there were 1.42 truck and 1.33 passenger vehicle fatal crash involvements per 100M VMT (FMCSA, 2014).

Although fatal crash rates are persistently higher for large trucks than for passenger vehicles, the opposite is true for less severe crashes. For example, the 2012 large truck injury crash involvement rate was 28.6 involvements per 100M VMT, versus 104.0 for passenger vehicles. One safety advantage trucks and buses have over cars is that a much larger percentage of their mileage is on Interstates and other divided highways with relatively low crash risks.

The human and economic cost of commercial vehicle crashes is significant. Zaloshnja & Miller (2007) calculated the average comprehensive cost of a police-reported crash involving a large truck to be $91,112 in 2005 dollars. These costs encompassed tangible economic human and material consequences, including medical and emergency services, property damage, and lost productivity. They also included the monetized value of pain, suffering, and quality-of-life reduction. An earlier study (Zaloshnja and Miller, 2002) estimated the annual total comprehensive U.S. costs for large truck crashes to be $20 billion annually in 2000 dollars.

CMV drivers make many of the same kinds of driving errors as do light vehicle drivers, but their crashes are less likely to involve extreme unsafe driving acts such as reckless driving and alcohol use (Knipling, 2009; Starnes, 2006). Among all crashes involving a truck and a lighter vehicle, principal fault seems to be more-or-less evenly divided (Council et al., 2003). For more severe crashes, however, principal fault (i.e., the critical driver error or other failure precipitating the crash) shifts strongly toward light vehicle drivers (Blower, 1999; FHWA OMC, 1999). In the LTCCS, trucks were at-fault (assigned the CR) in 40% of their multi-vehicle crash involvements. This percentage varied greatly depending on crash severity, as follows:

- "B" (non-incapacitating injury): truck 46%, other vehicle 54%
- "A" (incapacitating injury): truck 37 percent, other vehicle 63 percent
- "K" (fatal injury): truck 23 percent, other vehicle 77 percent.

Embedded in truck crash statistics is a paradox. By many measures, large trucks are driven more safely than are passenger vehicles. Their overall crash rates are less than half those of passenger vehicles. Egregious traffic violations like reckless driving and DUI are far less common among truck drivers. A high majority of fatal truck-car crashes are precipitated by the car driver. Yet trucks remain as much higher-risk vehicles because of their large size and high mileage exposure. In 2012, each individual truck was, on average, more than twice as *likely* to be involved in a fatal crash than was each individual car. Fatal crash likelihood was 0.36 per 1,000 trucks versus 0.15 per 1,000 cars (FMCSA 2014). The truck-car disparity is even greater if one focuses on CUTs in relation to cars. This "paradox of large truck safety" seems inherent in trucks and their use. The upside of the same coin, however, is that there are greater potential benefits from truck safety investments when they are viewed from the perspective of individual vehicles or drivers (Knipling, 2009). From a return-on-investment perspective, society can afford to invest more in the safety of one truck driver than it can in one car driver.

1.4 Factors Affecting Driver Alertness & Fatigue

Driver fatigue involves decreased alertness, decreased vigilance, reduced performance, reduced motivation, impaired judgment, and feelings of drowsiness. Falling asleep-at-the-wheel (AATW) is the greatest known fatigue-related crash risk. Two general categories of fatigue causes are internal physiological factors and task-related factors (Thiffault, 2011). Prominent physiological causes include the following:

- Individual differences in fatigue susceptibility, which may be related to sleep disorders, other medical conditions, or physiological variability.
- Circadian rhythms, with early morning (e.g., 4:00 to 7:00am) as the highest risk time.
- Hours of recent sleep, including primary sleep periods and naps.

- Sleep recency; i.e., grogginess (technically termed sleep inertia) experienced upon awakening.
- Hours awake since last principal sleep; especially at 16+ hours, and independently of work or specific work activities.
- General health and wellness and recent related behaviors; i.e., diet and exercise.
- Caffeine intake.
- Prescription and over-the-counter drug use.
- Light/dark.

Much of the daily variation in human alertness can be modeled based on three main factors: recent sleep, time awake, and circadian status. In a 2005 white paper, current National Highway Traffic Safety Administration (NHTSA) Administrator Mark Rosekind highlighted three factors, as follows (from Page 12):

> While there are a variety of complex factors that can affect fatigue, there are three primary physiological factors that have been scientifically demonstrated to affect alertness, performance and safety. These three factors are: a) sleep (specifically acute sleep loss and cumulative sleep debt), b) hours of continuous wakefulness, and c) circadian rhythms (time of day effects on sleep, alertness and performance).

There are a variety of sleep-performance models which predict alertness based on physiological factors (Balkin et al., 2000; Dawson et al., 2011; FMCSA, 2009). These biomathematical models attempt to quantify and predict the effects of circadian and sleep/wake processes on alertness. Prior sleep and circadian status are the major predictive factors. Various models may also use time awake and sleep recency (sleep inertia) in their computations.

Task-related fatigue factors include time-on-task (such as hours driving), task complexity, and task monotony (Thiffault, 2011). Task-related performance deterioration is most striking for highly demanding tasks, but may also be seen in less demanding tasks like driving. Time-on-task is of particular interest in regard to HOS because two primary parameters of HOS rules are time driving and time working. Several studies assessing time-on-task associations with alertness are reviewed in this paper (see Sections 4.7 through 4.9).

Time awake is well established as a physiological factor in alertness, and is an element in many Sleep Performance Models (Krueger, 2004). In almost any driving schedule, driving hours and work hours co-vary with time awake to a high degree. Few driving studies have clearly distinguished time awake effects from time-on-task effects, but it is likely that time awake is the more operative factor. For most people on most days, the steepest decline in daily alertness occurs after about 16 hours of wakefulness, and relatively independently of driving or other specific activities (Rosekind, 2005, Dawson et al., 2011).

Moore-Ede (1993) lists a number of other "alertness switches," including most of those listed above. This includes ambient temperature, sounds and noises, and certain aromas. Any of these factors might be prominent at any particular time, but are generally not relevant to HOS rules.

1.5 HOS Rules & Crashes: Challenges to Causal Inference

Commercial driver HOS rules contain numerous specific provisions relating to driver schedules. These include minimum daily off-duty hours, maximum daily driving hours, maximum tour-of-duty (which, for truck drivers, limit total work hours), schedule regularity (not regulated directly, but rather as a product of the above), weekly maximum work hours, restart (i.e., 34-hour restart) provisions after time off, required breaks from driving, and sleeper berth use (including "split sleep" provisions). Driver medical qualifications are not HOS rules *per se* but support the rules by screening out drivers with clinical levels of alertness-related conditions such as heart disease, Obstructive Sleep Apnea (OSA), and alcohol/drug abuse. Driver alcohol and drug testing further support driver alertness.

FMCSA bases its HOS provisions primarily on factors affecting driver fatigue and alertness, but there are some inherent differences between the profile of factors affecting alertness and the profile of HOS parameters. Table 1 below presents two lists. The first column shows various physiological and task-related factors that can affect driver alertness and performance. The second column lists HOS parameters. In some cases, there are clear and direct linkages; e.g., time-on-task and maximum daily driving hours. In other cases, the relationship is clear but indirect. For example, recent sleep is a prime physiological fatigue factor, but cannot be regulated directly by HOS rules. Rather, the minimum time-off provisions are designed to afford the opportunity for sufficient sleep. Some major fatigue causes are not addressed, or only partially addressed, by HOS rules. There are large individual differences in fatigue susceptibility, even among healthy individuals (e.g., Wylie et al., 1996; Dinges et al., 1998, Van Dongen et al., 2004). Yet all CMV drivers are governed by the same HOS rules. Time-of-day has a pronounced effect on human alertness, but, with one exception, is not factored into HOS rules. An exception is the current requirement that 34-hour restart periods include two off-duty periods encompassing the overnight four-hour period from 1:00am to 5:00am.

The imperfect alignment of human fatigue factors and HOS parameters means that not all studies show significant effects of HOS parameters on driver alertness. For example, the Driver Fatigue and Alertness Study (Wylie et al., 1996, see Section 4.2) found significant alertness effects from amount of sleep and time awake, but not from time-on-task (hours of driving).

Table 1. Human Alertness/Fatigue Factors and HOS Parameters

Factors in Alertness and Fatigue	HOS Parameters
Individual differences in fatigue susceptibility Circadian status Recent sleep Sleep recency (sleep inertia) Time awake General health and wellness Caffeine (or other stimulant) intake. Prescription and over-the-counter drug use Alcohol and other recreational drugs Light/dark Time-on-task (hours driving or working) Task complexity Task monotony Ambient temperature Sounds and noises Social interaction Certain aromas	Minimum daily off-duty hours Maximum daily driving hours Maximum tour-of-duty Maximum daily work hours Schedule regularity (a product of compliance with other provisions) Weekly maximum work hours Restart (i.e., 34-hour restart) Breaks from driving Sleeper berth use (including "split sleep" provisions)

Logically, one would expect HOS parameters to have their greatest effects on fatigue-specific dependent variables (DVs) but lesser effects on DVs known to be affected by factors other than fatigue. Motor vehicle crash rates are known to reflect numerous interacting factors, most of which are not discernably related to fatigue. Different studies show different quantitative roles of fatigue in crashes, but no study suggests that a *majority* of crashes are fatigue-related. The LTCCS (see Section 3.2) is considered by this reviewer to the best single information source on truck crash causation. Only 4% of truck crash involvements in the LTCCS involved truck driver asleep-at-the-wheel as the critical reason (CR). Thirteen percent (13%) were reported to involve fatigue as an associated factor, defined as the presence of fatigue (Starnes, 2006). Below are prominent crash causes and risk factors not known to be significantly related to fatigue. For some (e.g., inattention, misjudgments), fatigue relevance is possible but not demonstrated categorically, or quantifiably, for crashes. For most, a discernable connection to fatigue seems unlikely.

- Errors of other motorists or other failures (e.g., vehicle) associated with them.
- Truck driver traffic violations or other misbehaviors (e.g., speeding, tailgating).
- Awake inattention, ranging from transient distractions associated with the driving task to egregious inattention associated with cell phone use or other non-driving behaviors. Barr et al. (2011; see Section 3.5) have shown that drowsiness and distraction are in many ways opposites.

- Information processing errors and misjudgments, such as misjudgment of cross traffic closing distances.
- Errors executing specific driving maneuvers, such as merges and turns. Two particularly difficult maneuvers for large trucks are merges/lane changes and 90^{o} turns.
- Environmental and roadway factors, most notably adverse weather and roadway design factors (e.g., sharp curves and ramps).
- Vehicle deficiencies or defects
- Reduced driver alertness due to factors other than fatigue; i.e., illness.

The complexity of motor vehicle crashes and the many different factors affecting them means that scientific rigor is critical for studies attempting to show HOS effects on safety outcomes. The next section of this paper addresses three areas where rigor is needed: scientific variables, sampling, and research designs.

2. RELEVANT RESEARCH CONCEPTS & METHODS

The research design and methodological concepts relevant to CMV driver fatigue are fundamentally the same as for many other behavioral science questions. Accordingly, this chapter is structured to be consistent with standard behavioral science practice and usage, but with examples relating to driving safety and driver fatigue. The chapter draws heavily from *Research Methods for the Behavioral Sciences* by Gregory J. Privitera (2014, Sage Publications, Inc.). The terminology presented is generally non-technical and may not be universally or consistently used by all researchers. Nevertheless, the terminology and concepts provide a basis for understanding the structure of most driver fatigue and HOS studies. They also provide a basis for articulating the strengths and weaknesses of various methodologies, and for identifying potential improvements.

2.1 Scientific Variables

2.1.1 Scientific Variables: Core Concepts

Readers are referred to the Glossary for definitions of basic terminology relating to scientific variables. These terms are used throughout this paper. They include the following:

- Variable
- Independent variable
- Dependent variable
- Controlled variable
- Uncontrolled variable
- Construct (aka hypothetical construct)
- Operational definition
- Reliability
- Internal consistency
- Validity

A few of these concepts and terms are especially critical in the discussions below, or they may be used in specific contexts in this paper. Therefore, they are also reviewed and discussed here:

- **Independent variable (IV)** – The variable manipulated in an experiment. IVs are often called "treatments" and are seen as the cause in any cause-effect relationship identified through experimentation. In this paper, *the term IV is used only for variables actually manipulated in an experiment*, not for other predictor variables such as "quasi-IVs" in quasi-experiments (to be discussed below). The general term *predictor* encompasses IVs, quasi-IVs, and other variables treated as potential causes or antecedents.

- **Dependent variable (DV)** – The variable believed to change in the presence of the IV or other predictor. It is the response shown by humans or other subjects, and the presumed effect in a cause-effect relationship. DVs are usually the measurable performance indicators collected by researchers as "data."
- **Construct (aka hypothetical construct)** – A conceptual variable known (or assumed) to exist but which cannot be directly observed. Fatigue, however defined, is a prime example. "Safety" might also be considered a construct since there may be multiple measures of it.
- **Validity** – The extent to which a measurement of a variable or construct actually measures what is purports to measure. Four types are important and relevant:
    - *Face validity*. Does the measure *appear* to measure the construct?
    - *Construct validity*. Does the measure actually measure the construct?
    - *Criterion*-related validity. Does the measure predict or correlate with an expected outcome?
    - *Content validity*. Do the contents of the measure represent the features of the construct?

2.1.2 HOS- and Other Schedule-Related Predictors

Most HOS-related studies treat schedule or driver experiential parameters as predictors (IVs or quasi-IVs) and fatigue as a DV. As a construct, fatigue is not measured directly but rather defined and measured operationally. In other words, fatigue is inferred from a measured DV. Overall safety (e.g., crash rate) is another common DV. Principal predictors include individual traits, time off-duty, sleep duration, time-of-day, time awake, tour-of-duty (time transpired from start of work), time-on-task (hours working and/or hours driving), task characteristics (e.g., monotonous vs. busy driving), breaks, days working, and recovery periods. Most of these factors correspond directly or indirectly to HOS parameters. For example, one may study effects of sleep duration on performance because sleep duration is related to daily off-duty time requirements.

2.1.3 Dependent Fatigue and Safety Measures

The following are types of measures which may be captured in studies relating to driver fatigue. A given study may employ multiple types of measures. The first ones listed are mostly general measures of safety while later ones are more closely related to fatigue *per se*.

Crashes. Crashes are almost always defined in relation to specified damage/injury threshold criteria. Common criteria include police-reported, DOT-reported (towaway vehicle or injury), serious injury, and fatal. In most states, the police classify the severity of the crashes they report by the "KABCO" system based on the most serious injury in the crash. The levels are: K =

Killed; A = Incapacitating injury; B = Non-incapacitating injury; C = Possible injury; O = No injury (also known as Property Damage Only or PDO). Crash characteristics vary widely by crash severity level, so the reporting threshold is an important characteristic of any crash dataset. Crashes are usually analyzed in one or more of the following ways:

- Counts; Number of crash involvements. Examples include Penn State fleet studies (see Section 4.9).
- Crash Characteristics; descriptions, conditions of occurrence. Examples include the NHTSA General Estimates System (GES), Fatality Analysis Reporting System (FARS), and Trucks in Fatal Accidents (TIFA).
- Crash Causal Scenarios. Causal scenarios are broken down into a series of coded variables describing the crash sequence. This includes critical events, critical reasons, and, in lay terms, "fault." Examples include the LTCCS (Section 3.2), National Motor Vehicle Crash Causation Survey (NMVCCS), and NTSB studies.

Different large truck target crash groups have widely different concentrations of driver fatigue. In 2012, police-reported driver fatigue was about five times greater in fatal truck crashes than in all police-reported crashes (FMCSA, 2014). The study by Tefft (2014; see Section 3.6) illustrates fatigue variation by crash severity. His estimates for the percent of drowsy drivers in Crashworthiness Data System passenger vehicle crashes are:

  - 3% of drivers involved in crashes resulting in no injuries
  - 8% of drivers involved in crashes resulting in a person being admitted to a hospital
  - 15% of drivers involved in fatal crashes.

Fatigue differences among crash subsets are further illustrated in Figure 3, based on LTCCS truck crash involvements (statistics from Knipling and Bocanegra, 2008). All truck involvements can be classified as either single-vehicle, multi-vehicle "at-fault" (i.e., the truck was assigned the critical reason or CR), or multi-vehicle "not-at-fault." The two LTCCS fatigue indicators were truck driver asleep-at-the-wheel as the critical reason (i.e., primary proximal cause, 4% of involvements) and truck driver fatigue as an associated factor (13%). The criterion for the latter was simply the identified presence of fatigue. By both indicators, the causal importance of fatigue varies greatly depending on what target group of truck crashes are chosen for study. A study of single-vehicle truck crashes would have a relatively high fatigue involvement, whereas a study of not-at-fault crashes would have a low involvement (and no asleep-at-the-wheel, since that would make the truck driver at-fault). Studies of crash groups with low fatigue content are highly vulnerable to confounding and misinterpretations due to various non-fatigue causal factors. Thus, clear identification of target crashes and understanding of the likely role of fatigue in those target crashes are critical for accurate causal inference.

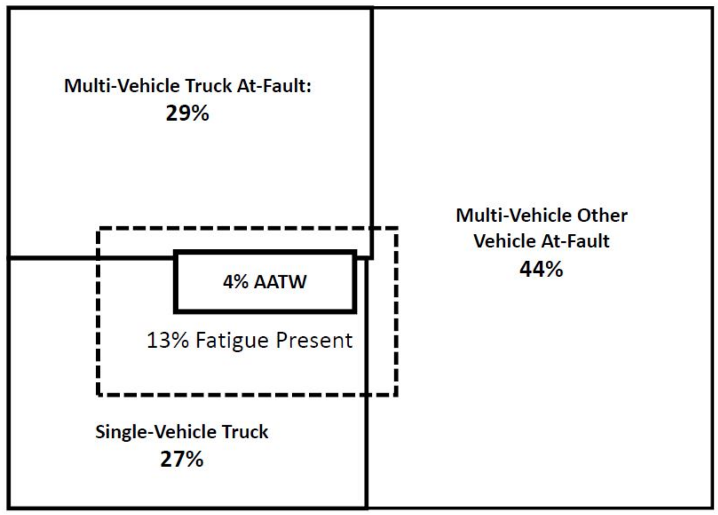

Figure 3. Large Truck "Crash Space" with two fatigue measures superimposed. **Based on truck involvements in the LTCCS (Knipling and Bocanegra, 2008).**

As with almost any kind of assay, a concentrated sample is more likely to yield true results than a dilute sample. Figure 3 illustrates that DVs such as "all crashes" or "overall crash rate" are dilute in regard to driver fatigue. The DV "all single-vehicle crashes" would be a more robust measure, those still not a concentrated measure. Most robust would be DVs incorporating a fatigue requirement such as the two LTCCS fatigue measures shown in the figure.

Harm. *Crash harm* is a quantitative measure of the combined human and material loss from traffic crashes based on economic valuation of crashes and injuries of various severities. Crash harm studies (e.g., Zaloshnja and Miller, 2007) tabulate all the property damage and injuries of different severities in target crashes and, based on crash cost data, derive a single measure of crash consequences. Using harm as a metric permits objective comparisons across different vehicle types, crash types, crash severity levels, and causal factors. Crash harm is a more sensitive and comprehensive measure than crash counts or maximum crash severity because it includes all the injured parties and tabulates a single quantitative, ratio-scale measure.

Among all truck crashes, most harm is concentrated at the top of the KABCO scale. Statistics from Zaloshnja and Miller (2007) show that the top three categories combined (i.e., KAB) constitute about 11% of police-reported large truck crashes but 80-90% of known truck crash harm. Specifically, KAB crashes were 78% of crash costs, 91% of reduced quality-of-life years, and 92% of lost productivity. Harm measures might be particularly appropriate for studies of fatigue crashes since the role of fatigue varies directly with crash severity or, stated in another way, fatigue-related crashes tend to be more severe than most other crashes (Knipling, 2009a). The percentage of fatigue-related crash harm resulting from KAB crashes has not been reported, but it is likely well over 90%.

Non-Crash Surrogate Events; e.g., "Safety-Critical Events" (SCEs). SCEs are mostly dynamic non-crash events captured using full naturalistic driving (ND) instrumentation or simpler in-cab camera systems (e.g., DriveCam® or similar video event recorders). SCE triggers include hard braking, proximity to other vehicles (short "times-to-collision"), and swerves. Possible SCE DVs include:

- SCE Counts; Number of SCEs. Examples include two Virginia Tech Transportation Institute (VTTI) HOS-related studies reviewed in this paper: Hanowski et al., (2008) and Blanco et al. (2011)
- SCE Characteristics; descriptions, conditions of occurrence. Examples include earlier VTTI naturalistic driving studies (e.g., Hickman et al., 2005).
- SCE Causal Scenarios; critical events & reasons leading to SCE. Examples include earlier VTTI naturalistic driving studies (e.g., Hickman et al., 2005).

ND studies can gather huge amounts of data. Vehicle instrumentation suites collect data on dozens of kinematic and driver-related variables concurrently and continuously. The 2011 VTTI study (Blanco et al.) was based on 735,000 miles of data recordings and captured 2,197 dynamically triggered SCEs. Since SCEs are far more numerous than crashes, they can be studied quantitatively with far more precision and statistical power.

ND SCEs contain few crashes and virtually no serious crashes, however. A touted strength of ND is that it captures normal driving, yet the other side of the same coin is that normal driving "suffers" from a paucity of crashes, especially serious crashes. In the Blanco study, only four (4) of the 2,197 SCEs (0.2%) were crashes, and the criterion for a "crash" was "any contact." The paucity of real-world consequences in SCEs raises the question of whether SCEs are representative of serious crashes in regard to crash causal factors such as driver fatigue and HOS parameters. As noted in the Introduction, large truck crashes are heterogeneous both "horizontally" (within any severity level) and "vertically" (across different severity levels). Serious crashes and SCEs are at opposite ends of the severity dimension. There is no *a priori* reason why SCE datasets should be representative of serious crash populations, and there is positive evidence against representativeness for some variables. For example, of 915 combination-unit truck ND SCEs in Hickman et al. (2005), 43.1% were rear-end crash scenarios in which the truck would struck another vehicle had a crash occurred. In only 0.5% of the events, the truck would have been struck in the rear. In the LTCCS, the corresponding percentages for combination-unit trucks were 12.3% and 5.7%. Even a few sharp discrepancies such as these would seem to invalidate use of SCE datasets for assessing causal factors, since those causal factors vary markedly across different crash scenarios.

Figure 4 illustrates this concern. The layers of the triangle represent five levels of police-reported crash severity (K, A, B, C, O) while the bottom layer of the triangle represents non-police-reported crashes. The top three layers (K, A, B) represent crashes with fatalities or known injuries. These fatal and injury crashes are about 11% of police-reported crashes but represent 80-90% of known crash harm (Zaloshnja and Miller, 2007; Knipling, 2009). SCEs are of multiple types but are almost entirely "below the triangle" since they involve no impact. The schematic shows that a few (0.2% in Blanco) are actual collisions. Of those impacts, a minority would be police-reported crashes classified per KABCO. Because the number is so small, no attempt is made to show them in the figure. The scientific concern is whether a mixed dataset of various SCE types can be representative of harmful crashes given the severity disparity between them and the fact that most SCEs are captured and defined based on driver reactions whereas crashes are defined by consequences. None of the studies reviewed in this paper explicitly address this question of representativeness in relation to serious, externally-defined crashes, although Guo et al. (2010; see Section 3.4) does provide comparisons of SCE near-crashes to SCE crashes.

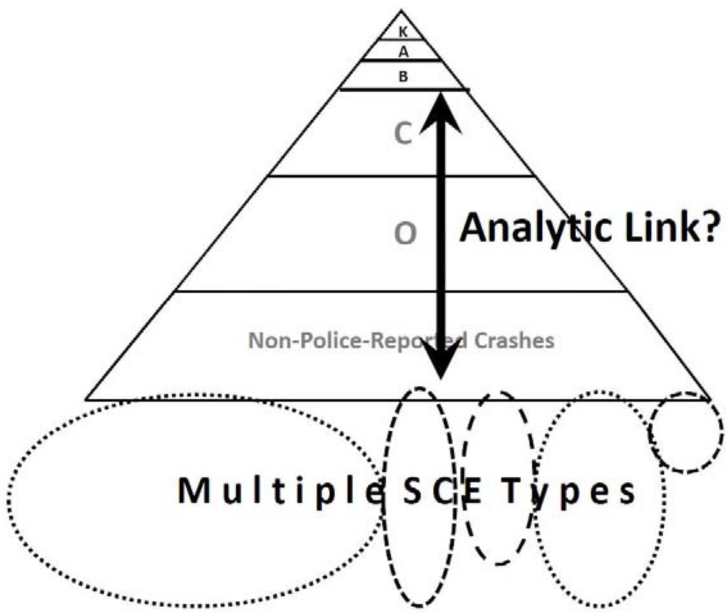

Figure 4. Heinrich's triangle for crashes plus multiple SCE types constituting SCE datasets (Knipling, 2015).

Unfiltered SCEs certainly cannot be considered a valid surrogate for driver fatigue. In the only large truck ND study to record asleep-at-the-wheel (AATW) as a CR (Hickman et al., 2005), only *one* of 915 SCEs (0.11%) was assigned that CR. The LTCCS percentage for serious crashes was 3.8%, about 35 times higher. Wiegand et al. (2008; see Section 3.3) observed 1,271 truck SCEs and found an *inverse* relationship between SCE occurrence and drowsiness per two different measures of drowsiness. Table 2 summarizes several sharp contrasts between unfiltered SCEs and fatigue-related crashes.

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

ND Safety-Critical Events (SCEs)	Fatigue-Related Crashes
Lowest rate in early morning (Hanowski et al., 2008)	Highest rate in early morning (Massie et al., 1997; Knipling, 2009)
Most likely in heavy urban traffic (Hanowski et al., 2008; Hickman et al., 2005)	Most likely on low traffic rural roads (Wiegand et al., 2008; Knipling and Wang, 1994)
Most likely on undivided roads (Hickman et al., 2005)	Most likely on divided highways (Wiegand et al., 2008; Knipling and Wang, 1994)
Mostly multi-vehicle (Hanowski et al., 2008)	Mostly single-vehicle (Starnes, 2006)
Driver is active, usually distracted (Barr et al., 2008; Olson et al., 2009).	Driver is passive with tunnel vision (Barr et al., 2008) and relinquishing vehicle control (Knipling and Wang, 1994; NTSB, 1990).
AATW % of CRs = 0.1% (Hickman et al., 2005)	AATW % of CRs = 3.8% (Starnes, 2006)
Risk *inversely* related to PERCLOS (Percent Eye Closure; Wiegand et al., 2008)	Risk strongly indicated by PERCLOS (Wierwille, 1999; Dinges et al., 1998; Krueger, 2004; Miller, 2014)

Driver Performance. In this paper, the phrase "driver performance" is reserved for measures of driver actions, behaviors, and responses. Thus, crashes and SCEs are not measures of driver performance but rather are outcomes which may or may not reflect driver performance. Most notably, not-at-fault crash and SCE involvements cannot reasonably be considered as indicative of driver performance. At any level of practicality, most (though certainly not all) not-at-fault crashes are unavoidable. Driver performance measures may be continuous or episodic. Examples of continuous measures include lane tracking (several measures, in particular Standard Deviation of Lane Position), steering patterns, speed maintenance, and vehicle following. Driver performance may also be measured in responses to driving events; examples include decision choices and reaction times for avoidance maneuvers in response to crash threats. Driver performance may be measured in real driving or in driving simulators.

Computer-Based or Other Dynamic Non-Driving Performance. Subject alertness and performance may be measured in computer-based or other laboratory testing. Examples include the Psychomotor Vigilance Test (PVT), Critical Tracking Task (CTT), and the Digit-Symbol Substitution Test (DSST). Extensive research has shown that these tests capture lapses of attention and that they are sensitive to prior sleep and other fatigue factors.

Percent Eye Closure (PERCLOS). PERCLOS is the proportion of time that the eyes are 80-100% closed. It is a measure of slow eyelid closure not inclusive of eye blinks. PERCLOS is well-validated as a continuous measure of alertness. Correlations of +0.8 to +0.9 with lane tracking deterioration (Wierwille, 1999) and PVT lapses (Dinges et al., 1998) have been reported. PERCLOS may be measured via manual measuring of video frames, or using various video image processing devices. Some of these are marketed commercially as in-vehicle safety technologies.

Other Physiological Measures of Alertness (or Sleep) State. Other physiological measures relating to alertness include brain electroencephalogram (EEG), electrooculogram (EOG), heart rate variability (Vagal Tone), measures of body activity (e.g., from wrist-worn activity monitors/recorders), and sleep latency (time to fall asleep when given opportunity). These measures are employed mainly in monitoring sleep, but some may be used to monitor states of wakefulness (Miller, 2014). The DFAS (Wylie et al., 1996; see Section 4.2) and other studies have used them. Applications to real driving are limited, however, because the measures are obtrusive and also because they can be highly variable both within and between subjects.

Self- or Observer-Ratings of Alertness. Numeric self-rating scales include the Karolinska Sleepiness Scale (KSS) and the Stanford Sleepiness Scale (SSS). Most used has been the KSS, which obtains self-ratings on a 9-point semantic differential scale from 1 (extremely alert) to 9 (extremely sleepy). Observer Rating of Drowsiness (ORD; Wierwille and Ellsworth, 1994) is a scale in which a trained observer rates subjects' alertness states. Several studies reviewed (e.g., Wylie et al., 1996; Van Dongen and Belenky, 2012) have reported that subjective self-measures like the KSS do not correlate well with objective measures of alertness, such as the PVT or eye closure measures.

Driver History Self-Report Questionnaire Responses. Information (e.g., personal history, opinions) obtained through interviews or written questionnaires. Examples include Insurance Institute for Highway Safety (IIHS) survey studies in which drivers are asked to report drowsy driving episodes over the past week or month of driving (e.g., see Section 4.6).

2.2 Sampling from Populations

A few fundamental terms and principles of scientific sampling are relevant to considerations of past and future HOS/fatigue studies. Most driver fatigue studies involve specific measures of individual driver subjects, making sample size and representativeness important concerns in regard to research validity.

2.2.1 Core Sampling Concepts

Readers are referred to the Glossary for definitions of basic terminology relating to scientific sampling. These terms are used throughout this paper. They include the following:

- Target population
- Sampling frame (accessible population)
- Representative sample

- Probability sampling
- Convenience sampling
- Stratified random sampling
- Sampling error
- Sampling (selection) bias
- Nonresponse bias.

2.2.2 Sampling Issues Relevant to CMV Driver Fatigue Studies

Reviews of individual fatigue studies later in this paper will note the sampling limitations of most studies. Most driver fatigue studies, even those whose results affect millions of drivers via HOS rule changes, have involved 100 or fewer CMV driver subjects recruited from a few companies at a few geographic locations. In December 2013 there were approximately 5.6 Million CDL holders working for 539,000 motor carriers (FMCSA, 2014). Fleet size varies widely and is associated with marked variations in carrier safety management practices (Knipling and Nelson, 2011). To be representative, a driver sample would need to be huge and stratified to accommodate multiple dimensions of driver, vehicle, and motor carrier characteristics. Truck and bus operations are different from each other in many respects, and each is complex in its own right. The trucking industry is highly differentiated operationally (Burks et al., 2010). This includes operationally significant variations in freight ownership (i.e., for-hire vs. private), freight type (e.g., general vs. specialized), geographic area (regional, metropolitan vs. inter-city), predominant driving times (e.g., predominantly day vs. night), and average shipment size (truckload down to package pick-up and delivery). These factors affect the likelihood of driver drowsiness and fatigue.

Drivers vary markedly in their susceptibility to fatigue, and fatigue susceptibility appears to be a long-term, enduring personal trait (Van Dongen et al., 2004). Numerous studies cited here will note the wide individual differences seen among driver subjects. Other factors equal, studies will have larger sampling errors and greater potential for sampling bias in relation to a target population when subjects vary widely in underlying relevant characteristics.

Considering these multiple sampling challenges, a question in regard to virtually all driver fatigue studies is whether findings, even when valid for the sample tested, are robust enough to be generalizable to the entire CMV driver target population.

Of the CMV-specific studies reviewed here, only the Large Truck Crash Causation Study (LTCCS) was based on a population-based national sampling algorithm. The LTCCS, like the General Estimates System (GES) and some other U.S. DOT crash databases, was based on a stratified random sampling methodology in which the population was first divided into subgroups (strata) and there is then there was random sampling of specific crashes from those subgroups. Less rigorous but still to a large degree nationally representative was the 1990 NTSB study of fatal-to-the-driver truck crashes. NTSB sampled all qualifying crashes occurring in eight geographically dispersed U.S. states for one year. The fatigue-related estimations by Tefft (2012, 2014; see Section 3.6) were nationally representative, but for cars, not trucks. In their survey of CMV driver reactions to the 2003-2004 HOS rule changes, McCartt et al. (2005, 2008; see Section 4.6) collected large truck driver samples in two states over three calendar years. IIHS crash case-control studies have studies large crash and control samples in individual states (see Section 4.1). Otherwise, the fatigue studies described in this paper all involve limited samples with no realistic aspirations of national representativeness.

2.3 Research Designs

Scientific research may seek to answer questions which are exploratory, descriptive, or relational (Privitera, 2014). Research designs seek valid answers to such questions, particularly relational questions such as the relation between work schedules and driver alertness. This section reviews core concepts and issues relating to research design and how these concepts and issues are seen in fatigue studies.

2.3.1 Core Research Design Concepts

Privitera (2014) classifies research designs into three categories: nonexperimental, experimental, and quasi-experimental. Below are the definitions of these three types of designs:

> **Nonexperimental design** – Method in which behaviors/events are observed "as is" without researcher intervention. It may reveal correlations or other associations among variables, but does not demonstrate cause-and-effect.
>
> **Experimental design** – Method in which the experimenter fully controls specific conditions and subject experiences (i.e., independent variables or IVs) and measures their effects as dependent variables (DVs). To be a true experiment, there are three required elements of control: randomized assignments, manipulation, and a comparison/control group (see below). When properly conducted and analyzed statistically, experiments demonstrate cause-and-effect; i.e., a single, unambiguous explanation for an observed effect.
>
> **Quasi-experimental design** – A study structured like an experiment (e.g., for analysis) but where one or more element of control is lacking; e.g., non-random assignments; pre-existing, non-manipulated factor(s); or no comparison/control group. Quasi-experiments do not demonstrate cause-and-effect, but may *imply* cause-and-effect. Subtypes include:

- One-group designs (e.g., pre- and post-test)
- Time-series designs (e.g., series of tests)
- Developmental (e.g., longitudinal)
- Non-equivalent control groups.

Additional terminology central to describing and understanding fatigue research designs include the following:

Quasi-independent variable (quasi-IV) – A variable treated as an IV but which includes pre-existing, non-manipulated traits (e.g., gender, health status) or co-varying traits (e.g., time-of-day in relation to time-on-task) where assignment to conditions is not random.

Predictor – A general term describing any variable used as the basis for the prediction of some driver response or other outcome. This could include experimental independent variables, quasi-IVs, or a variable used in a non-experimental correlation. In this paper, the term *predictor* will be used to refer to quasi-IVs and correlational variables. The term *independent variable* will be reserved for variables manipulated in a true experiment.

Internal validity – The extent to which a design contains sufficient control to demonstrate cause-and-effect. True, well-conducted experiments have high internal validity while non-experiments have no internal validity. The internal validity of a quasi-experiment is intermediate and often uncertain.

External validity – The extent to which observations made in a study generalize beyond the specific manipulations and setting of the study. For example, the external validity of a driving simulator study is the degree to which its findings generalize to real-world driving. Subcategories include:

- Population validity; generalizability to the target population or to different subpopulations
- Ecological validity; generalizability across settings
- Temporal validity; generalizability over time
- Outcome validity; generalizability across different but related DVs (e.g., different measures of alertness or safety).

Privitera discusses five common threats to validity of research studies. These are confounding factors which vary systematically with the IV. Specific threats include:

- History/maturation; an unanticipated event co-occurs with the manipulation.

- Regression and testing effects; e.g., regression-to-the-mean or improvements due to experience taking the test.
- Instrumentation and measurement; e.g., errors in measurement occurring systematically with levels of the factor.
- Attrition or experimental mortality; e.g., rates of completion are different between study groups.
- Environmental factors; a condition of testing co-varies with an IV.

Knipling (2009) and other writers have used slightly different terminology to conceptualize experimental control. IVs and DVs are defined as above, but additional types of variables include *controlled* and *uncontrolled* variables. Controlled variables (CVs) are factors potentially affecting DVs and which are held constant, randomized, or otherwise counterbalanced. Uncontrolled variables (UCVs) are those which are not controlled and which are potential confounds. Figure 5 below illustrates this conception schematically. A strong experiment controls for its most threatening confounds. For example, a strong experiment testing either time-on-task or time-of-day effects would control for the other factor since each factor can confound the fatigue effects of the other.

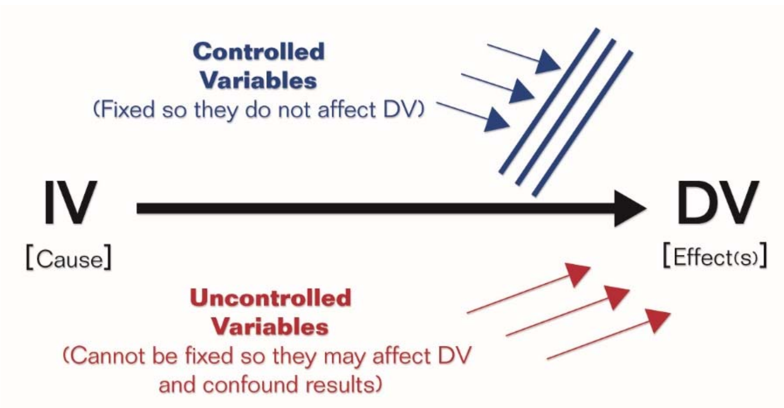

Figure 5. Schematic representation of experimental variables. **Source: Knipling, 2009.**

2.3.2 Research Design Issues Relevant to CMV Driver Fatigue Studies

Laboratory studies of fatigue are relatively easy to conduct as true experiments. For example, one can sleep deprive subjects (the IV or treatment) and measure multiple effects on alertness (DVs). Such studies by-and-large meet the required criteria for experiments: randomized assignments, manipulation, and comparison/control. Time-of-day of testing is a potential confound and threat to internal validity since it co-varies with sleep deprivation duration at any point-in-time. However, this confound can be addressed by multiple, counterbalanced testing sessions. The sleep dose-response study by Balkin et al. (2000; see Section 4.4) illustrates this experimental approach. Questions may be raised about the external validity of some laboratory studies, but concerns are less when effects are large and solidly based in human physiology.

Field studies of fatigue are much more problematic, however. Most studies of schedule parameters are quasi-experimental. That is, they do not manipulate schedules, but rather observe effects associated with pre-existing schedule conditions. Time-on-task (hours driving) is a prime predictor of interest, but it is not manipulated by the experimenter. Across multiple hours of driving there are concurrent, and potentially confounding, variations in time awake, time-of-day (circadian status), roadway types, and traffic conditions. Such studies have problematic internal validity due primarily to environmental threats as defined above per Privitera. Examples of such quasi-experiments reviewed include several large truck naturalistic driving studies (e.g., Hanowski et al., 2008; Blanco et al., 2011) and fleet studies relating HOS-related exposure to crash involvement (Jovanis et al., 2011). The use of crashes (Jovanis) and SCEs (Hanowski, Blanco) rather than fatigue-specific DVs further compromises internal validity.

Figure 6 illustrates the research design concern regarding such quasi-experiments. HOS parameters are quasi-IVs, presumed to affect crashes (or SCEs) by way of the construct "fatigue," perhaps itself due to some physiological factor such as sleep time. Validity threats include numerous non-HOS-related confounding variables with their own well-documented effects on CMV crash rates. Some of these confounds potentially create systematic bias, while others simply act randomly to add error to outcome measures.

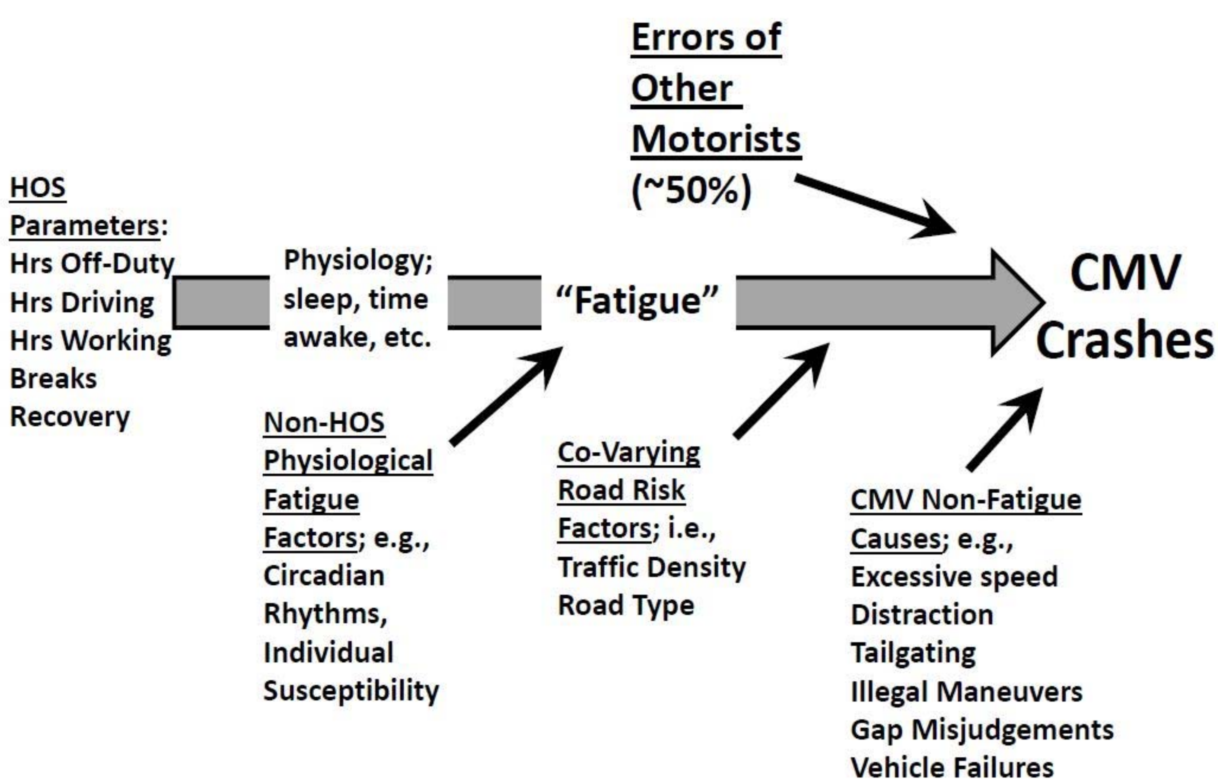

Figure 6. Potential confounds in studies relating HOS parameters to CMV crashes.

From left to right in Figure 6, the first set of confounds are non-HOS physiological fatigue factors like circadian rhythms and variations in individual susceptibility. Alertness varies greatly and systematically with circadian status, and largely independently of work *per se*. Circadian changes can be operating within work schedules if they are not controlled experimentally.

The next set of confounds are two pervasive road risk factors which may vary systematically across a work trip. Traffic density directly affects crash risk. Wiegand et al. (2008; see Section 4.3) found a truck SCE vs. baseline odds ratio of 7.2 for high traffic densities (Level of Service C-F) vs. low density (LOS A-B). Hanowski et al. (2008) found the correlation between truck SCE rate and average traffic density by TOD to be +0.83, and attributed the association of driving hours to SCE rate primarily to the traffic density confound (see Section 3.7). Kononov et al. (2011) found a 60% freeway rush hour traffic density increase to be associated with an 84% increase in crash rate per VMT (reflecting individual vehicle risk). Hickman et al. (2005) found that only 10% of tractor-semitrailer driving was on undivided roadways, but that 38% of SCEs occurred there. This yields an SCE odds ratio of 5.3 for driving on undivided roads. Fatal crash rates on such roads are about three times those on freeways (FHWA, 2000).

As already discussed, the errors of other motorists precipitate the majority of serious multi-vehicle truck crashes. Truck driver fatigue *could* contribute to these crashes, but not to a great extent. In the LTCCS, truck driver fatigue was present in 22% of truck at-fault involvements, but in only 3% of involvements where the other motorist was at-fault (Knipling and Bocanegra, 2008).

Finally, much driver error is not due to degraded performance, but rather simply due to voluntary misbehavior (Evans, 2004; Knipling, 2009). Misbehaviors like speeding, tailgating, and illegal maneuvers cannot be attributed to fatigue to any significant degree. Other human errors can occur without fatigue involvement. Section 3.5 will review a study by Barr et al. (2011) showing distraction and drowsiness to be largely opposites. Finally, not every crash is due to driver error. About 12% of LTCCS crashes were assigned non-human CRs, mostly vehicle-related failures.

In short, any quasi- or non-experimental study of HOS effects on crash rates must "survive" a gauntlet of potential confounds which threaten internal validity and weaken causal inference. More rigorous would be true experiments in which key confounds are controlled, and the use of DVs that are fatigue-specific rather than general and "contaminated" by non-fatigue causes.

There are large variations in individual susceptibility to drowsiness (Wylie et al., 1996; Van Dongen et al., 2004). Thus, experimental between-subjects comparisons may be perilous with small samples and/or non-random assignment to groups. Within-subjects experimental designs require fewer subjects and have relatively greater statistical power. When applicable, experimental studies reviewed below will be classified as between- or within-subjects.

3. STUDIES QUANTIFYING AND DESCRIBING FATIGUE AND OTHER CRASH FACTORS

This chapter and the next one describe and critique major CMV driver fatigue-related studies in regard to their methodologies and other features. Studies were selected for their prominence, relevance, and methodological distinctness. Most were major studies funded and published by FMCSA. The goal is to describe major studies which, in the aggregate, represent the most important research methodologies which have been applied to the subject. Methodologies may be instructive both in regard to their strengths and their weaknesses. The goal is *not* to comprehensively describe all relevant fatigue studies and findings, or to draw conclusions regarding specific HOS rules.

The study descriptions include the elements listed below. For brevity, only essential aspects of each study are delineated. The most essential aspects are those relating to methodology, including apparent flaws and potential improvements.

- **Overview and primary study purpose.** Major purposes include quantifying the fatigue crash problem, discerning schedule effects on driver fatigue, and discerning individual differences in fatigue susceptibility.
- **Study design.** The general design of each study (i.e., non-experimental, experimental, quasi-experimental) is stated, along with further classification and discussion. Much of the terminology used is consistent with Privitera (2014). In most cases this terminology was not used in the original study, but it is used here for consistency and to facilitate critical evaluation. This section often also includes a summary of statistical analysis methods.
- **Subjects and sample frame.** A brief description of study subjects and how they were sampled from their populations. "Subjects" may be humans, crashes, SCEs or other.
- **Predictors.** These include independent variables (IVs) in true experiments and quasi-IVs in quasi-experiments. In most studies, predictors are factors believed to affect alertness or safety.
- **Dependent variables (DVs).** These are *measures* of driver alertness, performance, safety outcome, or other presumed effect. The term DV is used equally here for experiments, quasi-experiments, and even non-experiments. Nevertheless, it should be understood that the validity of DVs as true *effects* depends on study design.
- **Notable controlled variables (CVs).** Factors which could affect the dependent variable(s), but which are held constant or counterbalanced (e.g., randomized) to nullify that effect.
- **Notable uncontrolled variables (UCVs).** Factors not manipulated or controlled, but which could affect DVs, and thus which constitute threats to internal validity.
- **Principal study findings.** These are stated to provide a full context for each study, but no general fatigue-related conclusions are drawn except those with implications for

methodology. Unless otherwise noted, stated findings are from original project reports, not from subsequent analyses.

- **Study limitations & potential improvements.** Limitations are typically threats to internal or external validity resulting from the study design or other aspects of its methodology. Anomalous or other questionable study findings may be noted. Potential improvements to address study limitations may be stated. In regard to external validity, note that almost all the studies have limited population validity since they involved relatively small numbers of subjects from particular fleet types. For brevity, this critique is not repeated for every applicable study.
- **Citation.** Full citation for study.

This chapter reviews six studies which primarily quantify and describe the role of fatigue in CMV crashes. Chapter 4, to follow, presents 14 studies with the general goal of quantifying and characterizing factors affecting fatigue. Typically these factors are HOS parameters (e.g., hours driving) or are otherwise closely related to HOS concerns. Some studies address both the fatigue crash problem size/characteristics and factors affecting fatigue. Thus there is some overlap between Chapters 3 and 4. Both chapters present studies in their approximate chronological order of publication.

The six studies presented in this chapter are:

1. Safety Study: Fatigue, Alcohol, Other Drugs, and Medical Factors in Fatal-to-the-Driver Heavy Truck Crashes (National Transportation Safety Board, 1990)
2. Large Truck Crash Causation Study (FMCSA, 2006; Starnes, 2006; other reports)
3. Fatigue Analyses from 16 Months of Naturalistic Commercial Motor Vehicle Driving Data (Wiegand et al., 2008)
4. Near-Crashes as Surrogate Safety Metric for Crashes (Guo et al., 2010)
5. An Assessment of Driver Drowsiness, Distraction, and Performance in a Naturalistic Setting (Barr et al., 2011; Hanowski et al., 2000)
6. Prevalence of Fatigue-Related Crashes Estimated from Multiple Imputation of Crashworthiness Data System (CDS) Unknowns (Tefft, 2012; Tefft, 2014).

3.1 Safety Study: Fatigue, Alcohol, Other Drugs, and Medical Factors in Fatal-to-the-Driver Heavy Truck Crashes (National Transportation Safety Board, 1990).

Overview and primary study purpose: This early, well-known NTSB crash investigation study identified the principal causal factors of 182 fatal-to-the-truck-driver heavy truck crashes in eight states. Nine of the crashes also involved fatalities in other vehicles, but most were single-vehicle crashes where only the truck driver died. Publicity from the study helped to make the truck driver fatigue problem more visible and also highlighted the fact that in-depth investigations find more driver fatigue than that seen in police accident reports (PARs).

Study design: Non-experimental study (in-depth investigations) of crashes meeting the fatal-to-the-truck driver criterion.

Subjects and sample frame: For a one-year period between Oct. 1, 1987 and Sept. 30, 1988, NTSB investigated (post-crash, on-site) every fatal-to-the-driver large truck crash occurring in CA, CO, GA, MD, NJ, NC, TN, and WI. This represented about one-fourth of such crashes in the U.S. for the same time period, making the study sample about 25% of the crash population. Standard NTSB investigative procedures included site and vehicle inspections, witness and police interviews, toxicology tests, and review of records including PARs, driver medical records, and driver logs.

Predictors: None as such.

Dependent variables (DVs): The probable cause matrix in the NTSB report listed 15 different causes for the crashes; most crashes had two or three causes indicated. Causal factors included: physical incapacity, fatigue, alcohol, drugs, driver inexperience, unsafe vehicle movement, disregarded signs/signals, failure to perceive dangerous situation or yield to other traffic, lack of occupant protection (safety belt), inadequate conspicuity, bad brakes, other mechanical deficiencies, signs/roadway, and load shift. The presence of fatigue was assessed by NTSB based on a combination of investigative information about the crash scenario (e.g., drift off road), driver sleep, time-of-day, and time-on-duty.

Notable controlled variables (CVs): The crash sample was defined and developed as described above. Standard NTSB investigative procedures were followed.

Principal study findings:

- Truck driver fatigue was the most frequent probable cause, reported for 57 of the 182 crashes (31%).
- This percentage is about three times higher than that found in PARs (10.6%) for truck fatal-to-the-driver crashes (Knipling and Shelton, 1999).
- Of the 57 fatigue-related crashes, a total of 40 other probable causes were indicated. Drugs and alcohol were among the most frequent factors cited together with fatigue. Of the 57 drivers judged to be fatigued, 19 were also impaired by alcohol or drugs.
- Overall, alcohol and/or drug use was cited for 53 drivers (29%) based on toxicological tests.
- Nineteen (19) of the crashes (10%) were attributed to driver medical conditions, principally cardiac arrest.
- For 65% of the involved trucks there was "some management deficiency in oversight of the driver or the proper condition of the vehicle . . . "

Study limitations & potential improvements:

- Although NTSB stated explicitly that their 31% fatigue estimate applied only to fatal-to-the-driver truck crashes and not to larger crash populations, many commentators have incorrectly generalized the finding to larger crash populations (Knipling, 2009). Fatal-to-the-truck driver crashes are significant in their own right, but they represent only about one in seven fatal truck crashes and one in 675 police-reported truck crashes overall. The *police-reported* fatigue rate in fatal-to-the-truck-driver crashes is nearly 30 times higher than that for all police-reported truck crashes (Knipling and Shelton, 1999).
- Population validity (to other crash types) is questionable for many study findings since these crashes reflect the worst crash causal scenarios. Also, temporal validity is questionable since alcohol and drug use by truck drivers were likely far greater in 1987-88 than currently. The combination of fatigue and alcohol/drugs is probably much less common today.
- Factors considered in the fatigue designation included amount of prior sleep, hours worked, and TOD. Thus analyses of fatigue in the study would be *circular* in relation to those same factors (see definition of circularity, Glossary).
- Only 13 (7%) of the crashes were coded as involving recognition failure (failure to "see or perceive a potentially dangerous situation and/or fail[ure] to yield to other traffic in such a situation"). Both ND and crash investigation studies in the decades since 1990 have consistently found far greater involvements of these driver errors in truck and other crashes.
- The study was conducted under the pre-2003 HOS rules which required only 8 hours off-duty daily and permitted only 10 hours of driving daily.
- There were no comparisons to other crash types or categories.
- There was no non-crash control group thus making estimates of relative crash risk impossible. There were also no comparisons to not-at-fault crashes since virtually all of the crashes were truck driver at-fault (see LTCCS discussion below).
- Crash investigation is an after-the-fact reconstruction rather than a "replay" of crash events. It is subject to various validity threats, including hindsight bias and circularity (Dilich et al., 2006; Knipling, 2009).

Principal Citation:

NTSB. *Safety Study: Fatigue, Alcohol, Other Drugs, and Medical Factors in Fatal-to-the-Driver Heavy Truck Crashes*. Report No. NTSB/SS-90/02. 1990.

3.2 Large Truck Crash Causation Study (FMCSA, 2006; Starnes, 2006; other reports)

Overview and primary study purpose: The congressionally mandated $20 Million LTCCS was one of the largest studies ever conducted by the U.S. DOT. FMCSA and NHTSA collaborated over six years to obtain and publish in-depth, on-scene crash investigations of 963 serious (injury or fatal) large truck crashes. The LTCCS provided important statistics on the fatigue crash problem size and also on many other crash causes and characteristics. Though it was non-experimental, its variables may be juxtaposed for parametric analyses.

Study design: Structured non-experimental crash investigations. The crash sample was obtained from 24 nationally representative areas (existing General Estimates System [GES] locations) during the years 2001-2003, before the major HOS rule changes published in late 2003. Quick-response investigation teams collected data on crash events, conditions of occurrence of the crash, and on the vehicles and drivers involved. Trained state inspectors also performed standardized Level I Commercial Vehicle Safety Alliance (CVSA) inspections on involved trucks and drivers. Most variables focused on pre-crash events; for each case there were more than 1,000 potential variables. Most LTCCS variables were lists of pre-defined, single-choice elements (choices).

Subjects and sample frame: Each of the 1,000+ variables was defined in relation to a crash (e.g., time of occurrence), vehicle (e.g., make/model, critical reason), or person (e.g., gender, age). Each case (or involved vehicle or person) was assigned a statistical weight, with the intention of matching the national profile of serious large truck crashes. As with GES, case weights were essentially the inverse of sampling percentages, which varied by crash severity and location (e.g., population density).

Predictors (quasi-IVs): No true IVs, but many variables have been treated as quasi-IVs or comparison groups in analyses. Notably, these include:

- Truck vs. car
- Type of truck; e.g., combination-unit vs. single-unit
- Crash severity (limited to K, A, and B in the KABCO crash severity scale)
- Critical Reason (CR) assignment (to truck/truck driver or to other involved vehicle/driver/person).
- Truck driver schedule, including reported sleep; e.g., hour-of-driving, hour-of-work, hours of prior sleep, hours since last main sleep period, time-of-day.
- Various environmental/roadway conditions of occurrence.

Dependent variables (DVs): Every variable and element within the variable could be considered a DV. Most notably in relation to fatigue, this includes:

- Critical Reason (CR) assignment (to truck/truck driver or to other involved vehicle/driver/person). The CR was the immediate reason for the physical events leading to the crash; most were specific driver errors but they also included vehicle failures and environmental factors affecting one vehicle. Only one CR was selected for each crash and was assigned to only one vehicle/driver; CR assignment could be considered tantamount to "fault."

- CR category, including physical (non-performance) failure, recognition failure, decision error, performance (response execution) error, vehicle failure, environmental/roadway factor.
- Specific CR, in particular "driver asleep." There were about 50 specific CRs, selected from a predefined, single-choice menu.
- Associated factors; notable factors present in the crash but explicitly *not* claimed to play a causal or even contributory role in the crash (FMCSA, 2006). Examples include fatigue, aggression, alcohol involvement, "emotion/experience," traffic, vehicle condition (e.g., brakes out-of-adjustment), weather factors, "speed/distance" factors. Each associated factor was a separate variable, and thus many could be coded for a particular crash.

Notable controlled variables (CVs): There were no controlled variables in the formal, experimental sense, but cases were selected, investigated, coded, and weighted per standardized protocols. Unlike Naturalistic Driving (ND) studies, the LTCCS had no accessible non-crash control sample to enable estimation of the relative risks associated with crash factors. FMCSA did advance the idea that relative risks for some factors (e.g., HOS violations) could be assessed by comparing truck-CR ("at-fault") crashes to nontruck-CR ("not at fault) crashes (FMCSA Analysis Division, 2007). This approach has at least two important limitations. First, it does not assess *crash* risk but rather crash *fault* risk. Second, to be valid, a compared factor would need to be determined independently of CR assignment. Otherwise, there would be a circular or biased comparison. For example, illegal maneuvers were associated with a 26-fold increase in "risk" (FMCSA Analysis Division, 2007) but "illegal maneuver" was a CR element and its coding was certainly not independent of CR assignment (Knipling, 2009a). A further elaboration of this approach (Knipling, 2009b, 2011c) especially relevant to driver impairment is to compare three categories: (1) Truck single-vehicle involvements (known to have the highest involvement of impairment); (2) Truck at-fault multi-vehicle involvements (much less impairment); and (3) Truck not-fault multi-vehicle involvements (minimal truck driver impairment).

Principal study findings: The LTCCS has generated hundreds of important research findings. Among those most relevant to HOS and fatigue are:

- The breakdown of CR categories for all 963 truck crashes assessed (including both single- and multi-vehicle) crashes was (FMCSA, 2006):
    - Truck driver physical failure/non-performance (includes asleep-at-the-wheel): 6.3%
    - Truck driver recognition failure: 15.5%
    - Truck driver decision error: 20.8%
    - Truck driver performance (response execution) error: 5.0%
    - Truck vehicle failure: 10.1%
    - Environmental/roadway failure affecting truck: 1.3%
    - CR assigned to other involved vehicle/driver: 45.4%.

- Truck driver asleep-at-the-wheel was the assigned CR in 3.8% of truck crash involvements (Starnes, 2006). Surprisingly, perhaps, this percentage was the same for both CUTs (usually long-haul vehicles) and SUTs (usually short-haul; Knipling and Bocanegra, 2008).
- The truck driver asleep-at-the-wheel percentage was starkly different for single-vehicle crash involvements (12.8%) versus multi-vehicle involvements (0.2%).
- In multi-vehicle crash involvements, the other driver was about nine times more likely to be asleep-at-the-wheel than the truck driver.
- Truck driver fatigue was an associated factor in 13% of truck involvements, corresponding to a "relative risk" of 8.0 per the comparison methodology (and its caveats) described above.
- More than half (62%) of truck driver asleep-at-the-wheel crash involvements occurred during the two-hour period between 4:01am and 6:00am (Knipling, 2009).
- Comparisons of truck single vehicle, at-fault multi-vehicle, and not-at-fault multi-vehicle involvements found significant differences (descending in that order) for fatigue as an associated factor, early morning (~dawn) driving, lack of recent sleep, and time since last sleep. No such relations were seen for hours driving, hours worked, or hours on-duty (Knipling, 2009b, 2011c).

Study limitations & potential improvements:

- The LTCCS was conducted in 2001-2003 under the pre-2003 HOS rules which required only 8 hours off-duty daily and permitted only 10 hours of driving daily.
- Although its 963 truck crashes are the most ever investigated in-depth, the sample size is still inadequate for many analyses, especially those involving crash sub-populations.
- As noted above, there was no non-crash control group thus making estimates of relative *crash* risk impossible. *Fault* risk estimates were possible as discussed above.
- Crash investigation is an after-the-fact reconstruction rather than a "replay" of crash events. It is subject to various validity threats, including hindsight bias and circularity (Dilich et al., 2006; Knipling, 2009).
- Although the LTCCS sampling and case weighing scheme was derived analytically from the national crash picture, the study probably over-weighted both truck single-vehicle crash involvements (Knipling, 2009) and those where three or more vehicles were involved. Thus, two-vehicle crashes were probably under-weighted.
- The "one-CR, one vehicle" scheme for the principal causal factor is a simplification of actual crash causation, though it may have the benefit of preventing over-attribution ("double-counting") of crash causes.
- Associated factors (e.g., Driver Fatigue) were coded for their presence, not for any presumed contributory role. There was no coding of contributory factors. This, combined with the lack of a non-crash control group, makes causal inferences speculative for many variables such as "fatigue." Also, the large number of different, independent associated factors leads easily to spurious over-attribution of crash causality to specific factors when they are considered individually (Knipling, 2009).

- The Driver Fatigue associated factor was coded "based on an evaluation of the driver's current and preceding sleep schedules, current and preceding work schedules, and a variety of other fatigue-related factors including recreational and non-work activities" (FMCSA & NHTSA, 2006). Thus, the variable is subject to *circularity* in analyses of the association of the variable with those factors (e.g., schedule).

Principal Citations:

FMCSA. *Report to Congress on the Large Truck Crash Causation Study*. MC-R/MC-RRA, March 2006.

Starnes, M. *LTCCS: An Initial Overview*. NHTSA National Center for Statistics & Analysis, DOTR HS 810 646, August 2006.

Extensive use of LTCCS data is also found in:

Knipling, R.R. *Safety for the Long Haul; Large Truck Crash Risk, Causation, & Prevention*. American Trucking Associations. ISBN 978-0-692-00073-1, 2009a.

Knipling, R.R. Three large truck crash categories: what they tell us about crash causation. Proceedings of the *Driving Assessment 2009* conference, Pp. 31-37, Big Sky, Montana, June, 2009b.

3.3 Fatigue Analyses from 16 Months of Naturalistic Commercial Motor Vehicle Driving Data (Wiegand et al., 2008)

Overview and primary study purpose: This study analyzed fatigue measures in 16 months of truck ND data from a previous VTTI study. It compared 1,217 Safety-Critical Events (SCEs) to 2,053 randomly selected baseline epochs from 34,230 total hours of driving. Two measures of driver fatigue for all events were Observer Rating of Drowsiness (ORD) and Percent Eye Closure (PERCLOS). These two fatigue measures were compared for SCEs and baseline epochs, and for other event conditions and characteristics. The counter-intuitive results reported in this study call into question the validity of ND SCEs as indicators of driver fatigue and their usefulness as sources of data on fatigue.

Study design: The study employed ND methods as described in this paper for Hanowski et al. (2008) and other ND studies. Wiegand et al. reanalyzed 1,217 SCEs, including 14 crashes (1%), 15 curb strikes (1%), 120 near-crashes (10%), and 1,068 crash-relevant conflicts (88%). Most SCEs were triggered by atypical driver responses and behaviors, including longitudinal decelerations (i.e., hard braking, 54%), short times-to-collision (14%), or swerves (20%). Baseline epochs were selected randomly to be proportional to driver exposure; i.e., one epoch per driver per work week. Using two measures of driver fatigue (ORD and PERCLOS), odds ratios derived to identify driving conditions and events associated increased driver drowsiness. The dataset was from the Drowsy Driver Warning System Field Operational Test (DDWS FOT) employing 46 DDWS-equipped CUTs. The DDWS system had no discernible beneficial effect in reducing drowsiness, however, and thus all of the data were aggregated for this and other fatigue- and causation-related analyses.

Subjects and sample frame: The data was from 46 CUTs and 103 drivers in normal truckload (one carrier) and less-than-truckload (two carriers) operations. The sample was "intended to be generally representative of the longhaul commercial vehicle driver population" (P. i). Drivers were 95% male, had an average age of 40, and an average 10 years of truck driving experience.

Predictors: SCEs were compared to baseline epochs in regard to drowsiness. In addition, various other event conditions and characteristics were compared. These included relation to junction [intersection], divided vs. undivided highway, roadway alignment, traffic density, and vehicle speed.

Dependent variables (DVs):

- ORD is a subjective but structured measure of drowsiness, developed and validated by Wierwille and Ellsworth (1994). Trained analysts observed video recordings of driver faces and behaviors for a 60-second period leading up to each SCE and for baseline epochs. ORD uses a 100-point scale; ORD scores ≥ 40 were classified as "drowsy." As a subjective measure, ORD was subject to inter-rater differences, although the three raters overall average ratings were not significantly different.
- PERCLOS is the proportion of time that the eyes are 80-100% closed. It is a measure of slow eyelid closure not inclusive of eye blinks. PERCLOS has been validated in past research against other fatigue measures including lane deviations and lapses of attention. A labor-intensive, manual method required analysts to view 3 minutes and 10 second recordings of each event and encode individual video frames (10 per second). The PERCLOS value for the event was the average of these measures. Scores ≥ 12 were designated drowsy.

Notable controlled variables (CVs):

- All 46 trucks were CUTs and were operated in the same general roadway environments.
- SCEs and baseline epochs were coded in a consistent manner based on the same data directory and other evaluation methods.

Notable uncontrolled variables (UCVs):

- As with other ND studies, drivers were in regular revenue-generating operations and did not adjust their schedules or driving for the study.

Principal study findings:

- Drivers were above the ORD drowsiness threshold in 26.4% of SCEs but 40.9% of baseline epochs.
- They were above the PERCLOS drowsiness threshold in 9.9% of SCEs but 15.8% of baseline epochs.
- Odds ratio calculations found the estimated relative risk of SCE involvement compared to baseline was:
  - 1.93 times greater (95% CI: 1.63 to 2.30) when the ORD rating was *below* the fatigue threshold (a rating of <40).
  - 1.70 times greater (95% CI: 1.30 to 2.23) when PERCLOS was *below* the fatigue threshold (a rating of <12%).
- Relative to exposure, SCEs were most likely to occur when ORD and PERCLOS values were lowest; i.e., when drivers were *most* alert.
- In a separate study identifying CRs for a large subset of the same SCEs, Hickman et al. (2005) found only *one* of 915 SCEs (0.1%) involved truck driver asleep-at-the-wheel as the CR. Eleven (11; 1.2% of 915) were attributed to high drowsiness but driver not asleep.
- Across all events, ORD and PERCLOS correlated weakly (r = +0.21).
- Data coders could choose up to four driver factors/behaviors contributing to the SCEs. "Drowsy, sleepy, fatigued, or other reduced alertness" was coded for 10.8% of the 1,217 SCEs.
- Drowsiness by both measures was greater in single-vehicle SCEs than in multi-vehicle SCEs. However, even single-vehicle SCE driver drowsiness did not exceed that of baseline epochs.
- No significant relations were seen between drowsiness and the presence of distractors (e.g., cell phones) in SCEs.
- High drowsiness odds ratios (3.9 for ORD, 2.1 for PERCLOS) were seen for SCEs occurring in the dark vs. light. Dark unlighted vs. dark lighted (i.e., by street lights) odds ratios were not significant.
- Both fatigue measures showed no overall differences between a.m. and p.m. times. The importance of this comparison is questionable, however. Both a.m. and p.m. 12-hour periods encompass high and low circadian periods as well as light and dark periods.
- Drowsiness was associated with non-junctions, divided highways, straight roads, sparse traffic, and vehicle speeds over 55mph.
- While SCE occurrence was inversely related to drowsiness, it was strongly and directly related to traffic density. An SCE vs. baseline odds ratio of 7.2 was found for high traffic densities (Level of Service C-F) vs. low density (LOS A-B).
- Compared to exposure, overall SCE rates were lowest during the overnight hours (midnight to 6am) and highest in the mid- to late-afternoon.
- Overall SCE rates were lowest in non-junction areas (i.e., away from intersections) whereas drowsiness in SCEs was highest in non-junctions.

- A supplemental analysis, published separately (Wiegand et al., 2009), found that obese (BMI $\geq$ 30) drivers in the sample were twice as likely as non-obese drivers to be fatigued in at-fault SCEs.

Study limitations & potential improvements:

- The *negative* association of both drowsiness measures with SCE involvement might suggest, superficially, that driver drowsiness *reduces* risk. Few would accept this conclusion, however. A better explanation is that SCEs are biased toward events where drivers are active and in traffic, while most drowsiness occurs under opposite conditions. This would make the negative association an artifact of the methodology. Relevant is the fact that SCEs are captured overwhelmingly by extreme driver responses, whereas fatigue results in reduced driver responsiveness. The methodological implication of this disparity is that unfiltered SCEs do not have construct validity in relation to fatigue. That is, SCEs in the aggregate do not measure or capture fatigue, which makes unfiltered SCE datasets inappropriate testbeds for studying driver fatigue. "Unfiltered" here refers to datasets taken in the aggregate and not analyzed for actual presence of fatigue or related causal factors.
- Use of SCEs would be more appropriate if they are individually evaluated for fatigue, with analysis focusing on those with high fatigue.
- As in other ND studies, only one of the two vehicles in any two-vehicle event was instrumented, and thus event detection and analysis was from the perspective of that vehicle. "This differential detection meant that the apportionment of events . . . as truck driver-initiated (truck "at fault") or other driver-initiated (truck "not at fault") did not represent the universe of such events that occurred." (P. v) Truck drivers were designated at-fault in 75% of all study SCEs, a percentage far greater than that seen in crashes (e.g., in the LTCCS).
- The study did not classify and disaggregate baseline epochs per the many variables used to classify SCEs. For example, drowsiness in baseline epochs was not classified for traffic density, relation to junction, TOD, etc. Such classifications would have permitted an array of separate, controlled comparisons. These controlled comparisons would likely have seen reduced negative SCE-drowsiness associations compared to those seen in the overall dataset.

Principal Citation: 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

3.4 Near-Crashes as Surrogate Safety Metric for Crashes (Guo et al., 2010)

Overview and primary study purpose: This study does not address driver fatigue *per se*, but it illustrates validations which have been performed of ND SCEs in relation to crashes. Employing data from the 100-Car Naturalistic Driving Study (Dingus et al., 2006), the study compared collected SCE near-crashes and crashes; i.e., an internal consistency check. A specific goal was to determine if near-crash samples could be added to crash samples in ND studies to greatly increase sample sizes (by ten-fold or more) for analyses without jeopardizing crash relevance.

Study design: Non-experimental *post hoc* reanalysis of SCEs and related ND data from the original "100-Car" study. Event triggers to detect both crashes and near-crashes included lateral accelerations, longitudinal accelerations, event button activations by drivers, short forward times-to-collision (TTCs), short rear TTCs, and yaw rates. Each had its own threshold, defined by the researchers. No data from outside the study were employed. In comparing crashes and near-crashes, three specific analyses were conducted:

- Sequential factor analysis: similarity of pre-crash scenarios including pre-incident maneuvers, precipitating factors, evasive maneuvers, and contributing factors.
- Crash/near-crash ratio analysis: correlations between circumstances resulting in a larger number of crashes and those resulting in a larger number of near-crashes. Poisson regression models were used. The following categories were considered: Driver, Level of Service (LOS, a measure of traffic density), Age, Gender, and Weather.
- Sensitivity analysis: For factors considered, estimation of risk for crashes only and for crashes/near-crashes combined, with comparison of estimated risks for these two cases.

Subjects and sample frame: In the original study (Dingus et al., 2006), 100 car drivers from Northern Virginia and Washington, D.C. were recruited as primary drivers to have their vehicles instrumented or to receive a leased vehicle for this one-year study. The study intentionally oversampled high-mileage and young adult drivers (e.g., college students) to increase the likely number of observed safety-related events.

Predictors: Not applicable, apart from crash/near-crash comparisons.

Dependent variables (DVs): The study compared crashes and near-crashes, defined as:

- Crash: any impact; 69 of 9,125 analyzed triggered events (0.8%). Severity categories included:
  - 5 police-reported airbag deployment or injury
  - 7 police-reported property damage only (PDO)
  - 21 non-police-reported PDO
  - 36 non-police-reported "low *g* physical contacts or tire curb strikes >10mph."
- Near-Crash: Any circumstance that requires a rapid, evasive maneuver by the participant vehicle or others involved. Evasive maneuvers included braking, steering, accelerating, or combinations thereof; 761 of 9,125 events (8.3%).
- Incidents (not analyzed in the current study): 8,295 of 9,125 events (90.9%).

Principal study findings: The Executive Summary (P. viii) stated the following: "The empirical study using 100-Car data indicates the following main conclusions: 1) there is no evidence suggesting that the causal mechanism[s] for crash and near-crash are different; 2) there is a strong frequency relationship between crash and near-crash; 3) using near-crashes will have biased results; however, the direction of the bias is consistent based on this empirical study, and 4) using near-crashes as surrogates can significantly improve the precision of the estimation. This result is analogous to the trade-off between bias and precision in many statistical estimation problems. For small-scale studies with limited numbers of crashes, using near-crashes as surrogate measures is informative for risk assessment and will help identify those factors that have a significant impact on traffic factors." Additional specific findings [including *post hoc* calculations performed here and indicated] included:

- Across 14 conflict types, the crash-near crash correlation of frequencies was +0.44 [calculated here from their Table 39]. Single-vehicle scenarios (conflict types single vehicle + object/obstacle + parked vehicle) were 37 of 69 crashes (54%) versus 59 of 761 near-crashes (7.8%).
- Drivers reacted to the crash threat in only 45 of 68 crashes (66%) versus 723 of 760 near-crashes (95%). This discrepancy was interpreted as follows (P. 23): "The significant difference in driver reaction for crashes and near-crashes implies that driver response is critical in distinguishing between these two types of events. However, this difference shall not be considered as evidence against the identical causal mechanism. The causal mechanism in this study is considered as the risk factors that trigger the safety events, not the driver's last response to avoid a crash. A crash and a near-crash can have exactly the same causal mechanism but a different safety outcome because of the evasive maneuver."
- A comparison of the *number* of contributing factors (e.g., distraction, surface conditions, traffic density, lighting, weather, visual obstruction) found similar numbers for crashes and near-crashes. For example, single-vehicle crashes had 1.58 factors identified, compared to 1.71 for single-vehicle near-crashes.
- The report presented crash and near-crash breakdowns for 54 precipitating factors. Across the 54 factors, the correlation between those for crashes and those for near-crashes was +0.18 [calculated here from their Table 48].
- "[T]here is a positive relationship between the frequency of crash and near-crash involvement" (P.29) by driver. The statistically significant crash-near-crash correlation coefficient was +0.21.
- Crash and near-crash distributions were similar for driver gender, driver age, lighting condition, road alignment, surface condition, and weather.
- Crash-to-near-crash ratios differed significantly by traffic density. A much higher percentage of crashes (41/69 = 59%) than near-crashes (244/761 = 32%) occurred under low-traffic (LOS A) conditions.
- Event and baseline videos were reviewed for driver drowsiness. The proportions were:
    - Crash: 14/69 = 20.3%
    - Near-Crash: 111/830 = 13.4%
    - Randomly selected baseline epochs: 599/17,344 = 3.5%.

- Regarding the relation of crashes and near-crashes, the report concludes: "There is no debate that crashes and near-crashes are two different types of events. This is not only true by operational definition but several results in this report demonstrate that the two cannot be completely identical. However, this does not eliminate using near-crashes as crash surrogates for a specific purpose." (P.48)

Study limitations & potential improvements:

- This was perhaps the easiest validation test imaginable for ND SCEs. It was an internal consistency test of events generated in the same study via the same sensors and methodologies. There were no external comparisons to existing crash datasets. SCE near-crashes and actual crashes were in adjacent categories differing only in whether an impact occurred. The study claimed that "there is no evidence suggesting that the causal mechanism[s] for crash and near-crash are different" (P. viii) but this statement is contradicted by the following:
    - The only moderate correlation between conflict types in crashes and near-crashes (+0.44) and the large difference in single-vehicle scenarios (54% of crashes, 7.8% of near-crashes).
    - The weak correlation (+0.18) between precipitating factors in crashes and near-crashes.
    - The much higher percentage of crashes (59%) than near-crashes (32%) in low-traffic conditions.
    - The much higher incidence of evasive maneuvers in near-crashes than in crashes (see below).
- The report found the presence of an evasive maneuver to be the primary distinguishing factor between crashes (often no) and near-crashes (yes), but did not consider this causally significant. Per the reports glossary, evasive maneuvers are performed in response to a precipitating event. Three of the four main categories of driver error CRs (i.e., non-performance, including fatigue; recognition failure [failure to respond to crash threats]; and response execution errors) constituting 65% of the truck-at-fault driver errors in the LTCCS (Starnes, 2006) involved absent or faulty evasive maneuvers. Extreme fatigue involves a driver relinquishing vehicle control and never executing evasive maneuvers. How could crash/non-crash differences in driver reactions "not be considered as evidence against the identical causal mechanism?"
- To this reviewer, it is hard to rectify the above findings and various statements in the report, such as those below:
    - ". . . there is no evidence suggesting that the causal mechanism[s] for crash and near-crash are different" (P. viii)
    - "In the context of naturalistic studies, the contributing factors for near-crashes and crashes should be similar or identical . . . and their differences should be merely of

severity. Only then can near-crashes be used to evaluate factors that affect traffic safety, instead of analyzing crash data directly." (p.16)

- "There is no debate that crashes and near-crashes are two different types of events." (P.48)

Principal Citations:

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

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.

3.5 An Assessment of Driver Drowsiness, Distraction, and Performance in a Naturalistic Setting (Barr et al., 2011; Hanowski et al., 2000)

Overview and primary study purpose: The Barr study intensively reanalyzed ND data collected in an early FMCSA-sponsored ND study of driver fatigue in local/short haul (L/SH) trucking operations (Hanowski et al., 2000). The study processed 871 hours of ND data from 42 truck drivers to identify and characterize episodes of drowsiness, relate them to driver and external factors, and relate driver drowsiness to distraction. Predictive models were developed to identify driver characteristics (e.g., age, years of commercial driving experience, sleep quality/quantity) and external factors (e.g., time of day, weather, traffic density) associated with the likelihood of driver drowsiness. The study is notable for its methodology (reviewing all driving to assess alertness and detect drowsiness and distraction), and for its finding that drowsiness and distraction were generally inversely related.

Study design: Non-experimental and quasi-experimental ND study with *post hoc* analysis of relationships among variables. There was no manipulation of IVs, but TOD and driving hours (time-on-task or TOT) were among those variables treated as quasi-IVs. The data used were collected as part of a VTTI ND study of driver drowsiness among L/SH truck operators (Hanowski et al., 2000). Cameras and other sensors were activated upon engine ignition; thus, data were recorded continuously while the trucks were in operation, rather than being recorded only when triggered by pre-defined critical events or near-crash situations as in more recent ND studies. Analysis of various fatigue-related variables included analysis of variance, linear discriminant analysis (e.g., to classify drivers as high- or low-fatigue), contingency table analysis (e.g., to compare drowsy to baseline epochs), stepwise linear regression, and logistic regression.

Subjects and sample frame: A total of 42 drivers from two L/SH trucking companies participated in the ND study. L/SH operations were defined as those primarily involving trips of 100 miles or less from the home base. L/SH drivers typically start and end their workdays at their home base. Each driver drove an instrumented truck for approximately two weeks. Drivers drove predominantly during daylight hours starting at around 6 a.m. Drowsy events were identified from video recordings by some initial driver behavior (e.g., yawning) and then further analyzed and classified.

Predictors (quasi-IVs):
- Time-of-Day (TOD)
- Hours driving/working (within workday and average across workdays)
- Driver characteristics
- Environmental/roadway conditions.

Dependent variables (DVs):
- A primary dependent measure was Observer Rating of Drowsiness (ORD), a 5-point scale (1 = not drowsy, 5 = extremely drowsy). Previous research (e.g., Wierwille and Ellsworth, 1994) asserted the reliability and predictability of this measure.
- PERCLOS (percent eye closure) was also used as a drowsiness measure.
- Other measures of visual attention included eye point-of-regard transitions (EYETRANS) and eyes off road (EYESOFF).
- A composite metric called the Fatigue Index quantified the overall drowsiness for individual drivers and encompassed frequency, duration, and severity of drowsiness.

Notable controlled variables (CVs): All drivers drove similar straight truck on L/SH runs.

Notable uncontrolled variables (UCVs): Since almost all runs were during the day, TOT and TOD generally co-varied in relation to each other (i.e., were cross-confounding). Traffic density and other factors also varied within work days and work weeks.

Principal study findings:
- A total of 2,745 drowsy events were identified in 871 hours of naturalistic driving video data. These were classified as: 1,636 ORD-2 events (slightly drowsy); 824 ORD-3 events (moderately drowsy); 160 ORD-4 events (very drowsy); and 125 ORD-5 events (extremely drowsy).
- Logistic regression analysis comparing high-fatigue and low-fatigue index drivers found strong associations of fatigue with younger (age 19-25) and less experienced (<1 year) drivers.
- A “strong and consistent relationship was found between drowsiness and time of day.” Drowsy driving events were twice as likely to occur between 6am and 9am, as compared to

baseline, or non-drowsy driving, and approximately 30 percent of all observed instances of drowsiness occurred within the first hour of the work shift." (Abstract)

- High drowsiness events were also more frequent between 2:00 and 4:00pm. Frequency was lowest in the early evening after 6:00pm.
- In regard to TOT, high drowsiness events were most frequent in the first hour of driving. There was a later rise between 7 and 10 hours driving/working, but then a fall between 10 and 13 hours.
- Overall, the TOD and TOT results suggest that drivers were not fully awake and refreshed when they began their work days.
- A "somewhat weak association between sleep quantity, quality, and drowsy driving was established in this study." (P. ix)
- Average daily driving time was associated with more drowsiness, but the relation fell short of statistical significance at $\boldsymbol{p} < 0.05$.
- Drowsy drivers experienced "tunnel vision," with fewer eye transitions and less time with eyes off the forward roadway. From these measures, it was inferred that drowsy drivers had lowered awareness of the overall driving environment.
- Drivers might engage in physical activity (shifting positions, singing) during drowsiness, but common distracting secondary activities like eating, reading, and cell phone use were associated with alertness, not drowsiness.
- Drowsiness and distraction were generally *inversely* related. EYETRANS and EYESOFF were greater when drivers were alert and lower when they were drowsy. Drowsy drivers narrowed their working visual fields, whereas distracted drivers widened them.
- The criterion for ORD-5 included a requirement that the observed drowsiness affected driving performance. Most ORD-5 events involved lateral lane breaks. There were 125 ORD-5 events (4.5% of all drowsy events) and they occurred at an overall rate of one per seven hours of driving.
- There were large individual differences among drivers in proportion of time drowsy, number of ORD-5 events, Fatigue Index, and other drowsiness metrics. The original Hanowski et al report stated that 4 of 41 drivers drove 7% of the study hours but had 39% of all the drowsy episodes.

Study limitations & potential improvements:

- The L/SH operations did not include nighttime trips and other driving conditions of greatest relevance to establishing HOS rules. Moreover, the risk of serious fatigue-related crashes is far less in L/SH than in long-haul operations. Thus the study's population and ecological validities are limited.
- Event sample sizes were small compared to more recent ND data collections.
- Study L/SH drivers were younger and less experienced than most current long-haul drivers.
- The original ND study (Hanowski et al., 2000) detected 249 SCEs, of which 77 were attributed primarily to truck driver errors (versus 137 to other drivers). Sixteen (16) of the 77

truck driver at-fault SCEs had high drowsiness. However, the Barr study did not explore these SCEs further and did not report any drowsiness-SCE relationships.

Principal Citations:

Barr, L.C., Yang, D., Hanowski, R. J., and Olson, R. *An Assessment of Driver Drowsiness, Distraction, and Performance in a Naturalistic Setting*. FMCSA-RRR-11-010, February, 2011.

Hanowski, R. J., Wierwille, W. W., Garness, S. A., and Dingus, T. A. *Impact of Local/Short Haul Operations on Driver Fatigue.* Final Report No. DOT-MC-00-203. Washington, DC: U.S. Department of Transportation, Federal Motor Carriers Safety Administration, September, 2000.

3.6 Prevalence of Fatigue-Related Car Crashes Estimated from Multiple Imputation of Crashworthiness Data System (CDS) Unknowns (Tefft, 2012; Tefft, 2014)

Overview and primary study purpose: This pair of studies estimated the prevalence of driver drowsiness in police-reported towaway passenger vehicle crashes in the NHTSA Crashworthiness Data System (CDS). They imputed driver drowsiness in cases when the driver's attentiveness was coded in the dataset as unknown, and analyzed these imputed values in conjunction with observed values from cases where drowsiness was known. This derived an estimate of the overall proportion of crashes in the entire sample that involved a drowsy driver. The procedure, known as multiple imputation, resulted in much larger driver drowsiness percentage estimates (usually by several fold) than would be the case if known drowsiness crashes were compared to the entire crash dataset. It is based on the key assumption that unknowns have the same statistical profiles as do knowns. Multiple imputation could potentially be applied to other crash datasets, including those for large trucks and buses. The method has been newly applied to drowsiness, but is well-established for other, similar purposes such as estimating alcohol involvement in crashes.

Study design: Structured non-experimental crash investigations. The crash sample was obtained from 24 nationally representative areas (existing National Automotive Sampling System [NASS] locations). The key variable for analysis was DRIVDIST, or driver attentiveness, which was coded for each driver based on driver interviews and review of crash documentation including police reports. Potential values of DRIVDIST included: driver attentive, driver looked but did not see, sleepy or fell asleep, distracted (with 13 separate codes for specific distractions), and unknown. Post-analysis adjusted the original CDS-reported drowsiness percentages by multiple imputation of unknowns on this variable. A crash was considered drowsiness-related if any driver in the crash was designated as drowsy. Multiple imputation is a Monte Carlo simulation technique in which the missing values are replaced by multiple simulated versions, each representing a different CDS variable known to be related to fatigue crashes. Examples include crash severity, number of involved vehicles, pre-crash maneuver, crash type, and time-of-day. Then each of the simulated complete datasets was analyzed by standard methods, and the results were combined to produce percentage point estimates and confidence intervals that incorporated missing-data uncertainty. A partial validation test involved comparing the predictions of unknowns to cross-predictions of knowns, as if they were unknowns. This procedure showed good concordance.

Subjects and sample frame: The 2014 study sample employed in the analysis was 14,268 U.S. crashes involving 25,528 drivers from the years 2009-2013 in which at least one passenger car, pickup truck, van, minivan, or sport utility vehicle was towed from the scene. These crashes were investigated as part of the NHTSA NASS CDS. The CDS towaway criteria results in a sample representing the most severe ~40% of police-reported passenger vehicle crashes. Trucks and buses were included only if they were involved in a crash with a passenger vehicle; thus there were no truck or bus single-vehicle crashes. The 24 sampling sites were selected to be nationally representative, and individual CDS cases were selected for investigation based on a stratified random sampling scheme. The CDS data collection regimen includes review of the Police Accident Report (PAR), vehicle and crash site investigation, reconstruction of crash trajectories, interviews with drivers and other persons, and review of medical records to determine the nature and severity of crash injuries.

Predictors (IVs/quasi-IVs): None as such, though comparisons were made across crashes of different severities.

Dependent variables (DVs): Drowsiness percentages for drivers and for crashes.

Notable controlled variables (CVs): There were no controlled variables in the formal, experimental sense, but cases were selected, investigated, coded, and weighted per standardized protocols. Case selection included vehicle type and crash severity criteria.

Principal study findings:

- In the 2014 study, prior to imputation, 35% of all CDS drivers were coded as attentive just prior to crashing, 5% were coded as "looked but did not see," 8% were coded as distracted, and 2% were coded as drowsy. Driver attention was coded as unknown for 51% of drivers.
- Following imputation, the estimated percentages of drivers who were drowsy were:
    - 3% of drivers involved in crashes that resulted in no injuries
    - 4% of drivers involved in crashes that resulted in injuries
    - 8% of drivers involved in crashes where a person was admitted to a hospital
    - 15% of drivers involved in fatal crashes.

- Prior to imputation, 3% of all CDS crashes involved at least one driver coded as drowsy, 33% of crashes had no drowsy drivers and no drivers of unknown attentiveness, and 64% had no drivers coded as drowsy but at least one driver whose attentiveness was unknown.
- Following imputation, estimates of the percent of motor vehicle crashes involving a passenger vehicle which involved a drowsy driver:
    - 6% of all towaway crashes, including those with injuries/fatalities
    - 5% of all non-injury towaway crashes
    - 7% of injury crashes (person received treatment for injuries)
    - 13% of crashes where a person was hospitalized
    - 21% of crashes where a person was killed.
- The initial 2012 study, involving a sample of 47,597 crashes from 1999-2008, found the following crash percentages after imputation:
    - 7% of all towaway crashes
    - 13% of crashes where a person was hospitalized
    - 16% of crashes where a person was killed.
- The above percentages are point estimates; 95% confidence intervals around these estimates were large. For example, the 95% confidence interval for the 6% all-towaway estimate above was 4-8% while that for the 21% fatal crash estimate was 13-28%. These somewhat large confidence intervals reflected both sampling variability and variability associated with multiple imputations of missing values. Estimation variance was especially high for fatal crashes as they had the highest percentage of unknowns for the attentiveness variable.
- As seen in the above statistics, both analyses showed drowsiness percentages varying directly with crash severity.
- The 2012 paper reported that crashes occurring between 11:00pm and 6:59am were nearly five times as likely to have involved a drowsy driver as were crashes that occurred between 7:00am and 10:59pm. Drowsiness was 2.4 times greater in Saturday and Sunday crashes than those occurring on weekdays.
- Driver age (young), gender (male), and number of passengers (driver alone) were among the other factors associated with higher drowsiness percentages.

Study limitations & potential improvements:

- Imputation involves generalizing percentages from the known part of a database to the entire database, including unknowns. Thus, there is an inherent but largely untested assumption that unknowns have the same causal characteristics as knowns.
- As Tefft noted in the 2014 paper, the "study did not investigate the cause [e.g., critical reason] of the crashes . . . [thus] it is possible that other factors besides drowsiness . . . may have contributed to some of the crashes." In other words, the study did not estimate drowsiness as the proximal crash cause but rather the involvement, to an unspecified degree, of drowsiness in the crash.

- The CDS encompasses only light passenger vehicle crashes. Large trucks and buses were included only if they crashed with a passenger vehicle. No dataset comparable to CDS exists for trucks and buses, though the same general method could be applied to other datasets.
- As its name implies, CDS investigations focus primarily on crash consequences and vehicle performance in protecting occupants, not on causation. A data collection system focusing on crash causation might provide different, or at least more detailed, results.
- The CDS inclusion of non-injury towaway crashes makes it broader, severity-wise, than the LTCCS (which included only injury and fatal crashes), but narrower than datasets containing all police-reported crashes such as the GES.
- Any coding limitations in the original data would also be carried forward in the imputations. For example, only one driver attention state could be coded. A driver could not be coded as both drowsy and distracted.

Principal Citations:

Tefft BC. Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999-2008. *Accident Analysis & Prevention*, 45(1): 180-186, 2012.

Tefft, B.C. *Prevalence of Motor Vehicle Crashes Involving Drowsy Drivers, United States, 2009-2013*, AAA Foundation for Traffic Safety, 2014.

4. STUDIES OF FACTORS AFFECTING FATIGUE

This chapter presents 14 studies with the general goal of quantifying and characterizing factors affecting fatigue. It employs the same presentation format as did Chapter 3. The 14 studies presented in this chapter are:

1. Case-Control Studies of Large Truck Crashes (Jones and Stein, 1987, 1989; Teoh et al., 2015)
2. Driver Fatigue & Alertness Study (DFAS; Wylie et al., 1996)
3. Effects of Operating Practices on Commercial Driver Alertness (O'Neill et al., 1999)
4. Effects of Sleep Schedules on CMV Driver Performance: (Balkin et al., 2000)
    a. (1) Actigraphic Assessment of Sleep of CMV Drivers Over 20 Days
    b. (2) Sleep Dose/Response Study
5. Stress and Fatigue Effects of Driving Longer Combination Vehicles (FMCSA, 2000)
6. HOS & Fatigue-Related Survey of Long-Distance Truck Drivers (McCartt et al., 2005, 2008)
7. Analysis of Risk as a Function of Driving-Hour: Assessment of Driving-Hours 1 Through 11 (Hanowski et al., 2008)
8. The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations (Blanco et al., 2011)
9. Hours of Service and Driver Fatigue: Driver Characteristics Research (Jovanis et al., 2011)
10. Motorcoach Driver Fatigue Study 2011 (Belenky et al., 2012)
11. Investigation of the Effects of Split Sleep Schedules on Commercial Vehicle Driver Safety and Health (Belenky et al., 2012)
12. Laboratory Study of the Efficacy of the 34-Hour Restart (Van Dongen & Belenky, 2010)
13. Field Study of the Efficacy of the New Restart Provision for Hours of Service (Van Dongen & Mollicone, 2013)
14. Effect of Circadian Rhythms and Driving Duration on Fatigue Level and Driving Performance of Professional Drivers (Zhang et al., 2014).

4.1 Case-Control Studies of Large Truck Crashes (Jones and Stein, 1987, 1989; Teoh et al., 2015)

Overview and primary study purpose: Crash case-control studies permit estimation of the relative crash risk associated with various factors. Crashes may be compared to corresponding "non-crashes" in regard to multiple characteristics. Quantitative and qualitative differences between the two may be used to derive odds ratios or other comparative statistics. Case-control studies do not determine causation perforce, but they can provide compelling evidence to suggest causation. The review here describes the general methodology used in two different studies by the Insurance Institute for Highway Safety (IIHS). Both an older (Jones and Stein, 1987; 1989) and new (Teoh et al., 2015) study have looked at various vehicle, carrier, and operational factors in crash involvement. At this writing the new study has no published results and thus is described only in terms of methodology. The case-control methodology is applicable to many HOS and other driver fatigue issues; one could compare cases to controls on virtually any HOS parameter such as hours of driving, hours of work, and day of work week. Other fatigue factors addressable include fatigue countermeasures (e.g., alertness monitors, electronic logs) and driver health and wellness.

Study design: Both studies were prospective case-control studies in which trucks involved in crashes were matched with non-crash involved trucks in regard to truck type, road location, day-of-week (or weekday vs. weekend), and time-of-day (TOD). The Jones and Stein 1987 report (and 1989 paper) assessed the association of crashes with driving hours, HOS violations, vehicle-related violations, and other factors. Other variables included truck weight, size, configuration, driver age and experience, type of trip, and carrier operations type. The crashes occurred on Interstate highways in Washington State. For each large truck involved in a crash, control trucks were randomly selected from the same traffic stream and at the same time, but one week later. All drivers and trucks were inspected using standard Federal roadside inspection procedures. Univariate case-control odds ratios and 95% CIs for them per Chi-Square tests were derived for various measures. A logistic regression model was used to analyze simultaneous effects of multiple factors.

In the new study, cases were large trucks (26,000+ lbs. GVWR) involved in serious crashes in North Carolina. For each case, one control truck was matched on the basis of truck type (single- vs. combination-unit), location (same roadway, same direction, within 5 miles of crash site), weekday vs. weekend, and time of day (4 hour blocks). Crash trucks were subjected to a CVSA level 1 post-crash inspection after the crash, whereas the control truck inspections were conducted within a couple months of the crash. For both cases and controls, both driver logs and supplemental schedule-related information (e.g., duration of last sleep period) were obtained from drivers. Data on driver/carrier safety records were obtained from FMCSA for trucks involved in the study. Matched pair odds ratios were derived for univariate comparisons.

Subjects and sample frame: Large trucks (plus their drivers and carriers) involved in crashes. The original study included data for approximately 330 CUT crashes occurring in Washington State in the mid-1980s. Sample sizes varied for different analyses. The new study involved serious CUT and SUT crashes occurring in North Carolina.

Predictors: Formally, the predictor was cases vs. controls. Prediction can be conceived in both directions, however. For example, driving hours may be considered predictive of involvement in cases (crashes).

Dependent variables (DVs):

- Hours of driving
- HOS log violations
- Equipment defects of various types and severities; e.g., out-of-service (OOS) brake violations.
- Driver age, experience, etc.
- Carrier characteristics.

Notable controlled variables (CVs):

- The Washington State analysis was limited to tractor-semitrailers (CUTs) because of their mechanical and operational differences from other truck types. The North Carolina study included both CUTs and SUTs, and controls were selected to match cases in truck configuration.
- Road location, day-of-week, time-of-day.

Notable uncontrolled variables (UCVs): Uncontrolled variables include any factor which might co-vary with crash vs. non-crash status, but not measured as a DV. For example, general driver and carrier safety conscientiousness likely affect both crash likelihood and safety practices such as vehicle maintenance.

Principal Washington State study findings:

- Driving >8 hours was associated with increased risk. The percentage of drivers who had driven more than 8 hours was 10% for crashes vs. 6% for controls. The 1.8 case-control odds ratio had a 95% confidence interval (CI) of 0.8 to 3.4.
- The elevated risk of driving >8 hours was greater for multi-vehicle crashes (2.6) but not significant for single-vehicle crashes.
- Adjusted case-control odds ratios and 95% CIs included:
    - HOS log violations: 3.0 (2.0 to 4.4)
    - OOS HOS log violations: 4.2 (2.0 to 8.7)
    - OOS steering defect: 2.6 (1.2 to 5.9)
    - Driver age <30: 1.3 (0.8 to 2.1)
    - Interstate carrier (vs. Intrastate): 2.1 ((1.2 to 3.6)
    - For to hire carrier (vs. private): 1.5 (1.0 to 2.4)
    - Empty trailer: 1.1 (0.7 to 1.8).
- The overall percentages of drivers with HOS log violations was 17% for crashes and 10% for controls.
- Crash-involved drivers under the age of 30 were more likely to be log violators.
- Out-of-service (OOS) violation rates for vehicle deficiencies were high: 41% for crash trucks and 31% for controls. Brake and tire defects were most common.

Study limitations & potential improvements:

- The principal study limitations are those inherent in case control studies. Cases and controls are "presented" to researchers, not created by them as would the case for true IVs. Thus any factor co-varying with crash involvement and the DVs measured could underlie statistical relationships. Most notably, unmeasured constructs like "safety conscientiousness" likely affect both crash likelihood and safety practices such as vehicle maintenance. Brake deficiencies, for example, suggest safety management deficiencies beyond brake maintenance *per se*.
- Another limitation inherent to case-control methods is the fact that any matched case-control parameter (e.g., TOD) is eliminated as a source of causal inference. A different source of control comparisons for those factors would be necessary. For example, one could use truck VMT by TOD for case-exposure comparisons, if such data were available.
- The 1987 report concluded, based on the multi- vs. single-vehicle odds ratios associated with driving >8 hours, that, "the effect of fatigue is more prevalent in multiple vehicle crashes." (P.11). This conclusion is contradicted by numerous other studies of fatigue-related crashes (e.g., the LTCCS).
- Findings relevant to driver fatigue and HOS were limited from these two specific studies, but the methodology could be applied more intensively to address these topics.

Principal Citations:

Jones, I.S. and Stein, H.S. *Effect of Driver Hours of Service on Tractor-Trailer Crash Involvement.* IIHS Report. September 1987.

Jones, I.S. and Stein, H.S. Defective equipment and tractor-trailer crash involvement. *Accident Analysis and Prevention*. Vol. 21, No. 5, Pp. 469-481, 1989.

Teoh, E.R., Carter, D.L., Smith, S., Lan, B., and McCartt, A.T. Risk factors for injury/fatal crashes of interstate large trucks in North Carolina. In preparation, 2015.

4.2 Driver Fatigue & Alertness Study (DFAS; Wylie et al., 1996)

Overview and primary study purpose. This large, early, on-road naturalistic driving study assessed fatigue under Canadian and U.S. operational truck driving schedules (the HOS regulations at the time, pre-2003) using a variety of DVs, many recorded during driving. Secondary objectives included developing and validating fatigue research methods, and gathering data in support of driver alertness monitoring. The study employed more than a dozen different fatigue measures, most or all of which were well-validated from prior research. The DFAS was likely the single most important study leading to the 2003 HOS rule changes.

Study design: Experiment (but with incomplete control) employing between-subjects comparisons of four truck driving schedules. Various statistical analyses were performed, most analyses of variance for independent groups. This captured both individual factor effects and interactions. The criterion for statistical significance was $p < 0.05$. Although classified here as an experiment, the study had significant uncontrolled variables. In addition, certain potential fatigue causes were treated as quasi-IVs in *post hoc* analyses.

Subjects and sample frame: 80 truck drivers driving real, revenue-producing long-haul less-than-truckload (LTL) operational runs in tractor-semitrailers. The 40 U.S drivers were from two different companies while the 40 Canadian drivers were all from a single company. Drivers drove for 16 weeks each. Drivers were age 25-65, had 1+ year of prior CMV driving, were "healthy," and alcohol-free.

Predictors (IVs, quasi-IVs): Driving schedule was the principal IV. Four conditions, all involving daily "turnaround" trips. The U.S. trips were between St. Louis and Kansas City while the Canadian trips were between Montreal and Toronto. Conditions were:

- C1: 10-hr daytime (5 consecutive days); 11 hours off-duty. U.S.
- C2: 10-hr rotating backward, starting 3 hours earlier each day; 8 hours off-duty. U.S.
- C3: 13-hr nighttime start (4 consecutive days); 8 hours off-duty. Canada.
- C4: 13-hr daytime start (4 consecutive days); 8 hours off-duty. Canada.

Within each of the four primary IV conditions, there were other fatigue factors which were treated as quasi-IVs in *post hoc* analyses. This included hours of sleep, hours working, hours driving, days driving, time-of-day, and schedule regularity. The truck cab ambient environment (e.g. heat, noise) was also recorded.

Dependent variables (DVs):

- Driving task performance:
  - Lane tracking (Standard Deviation of Lane Position or SDLP)
  - Steering wheel movement
- Surrogate non-driving tests:
  - Code Substitution test
  - Critical Tracking Test
  - Simple Response Vigilance Test (SRVT)
- Video recording of driver's face and road ahead
- Physiological measures:
  - Body temperature
  - Polysomnography (e.g., EEG) during sleep; enabled quantification of amount of sleep and sleep quality, including amount of time in each sleep stage.
  - Polysomnography (e.g., EEG) during driving

  - Vagal tone (electrocardiogram)
- Driver-supplied information:
  - Sleep history questionnaire
  - Daily HOS logs
  - Self-assessments of fatigue (Stanford Sleepiness Scale).

Units of measurement were specific to each DV. Different statistical tests were used for different types of variables.

Notable controlled variables (CVs):

- Within a condition, trips began at the same place and time, had the same mid-trip turnaround location, and covered the same roads.

Notable uncontrolled variables (UCVs):

- Each subject was exposed to only one of the four experimental conditions. This between-subjects design meant that subject variations could affect comparisons across major conditions.
- Trips were operational, revenue-producing runs with variations in traffic, roadway type, terrain, etc.
- Although time off-duty was controlled per HOS parameters, amount of sleep was not.
- Tractor make/model was partially uncontrolled but believed not to be a factor.

Principal study findings:

- Time-of-day (TOD) was the strongest and most consistent factor influencing driver fatigue and alertness.
- Drowsiness, especially in driver face video recordings, was much greater during night driving then during day driving.
- Time-on-task (hours driving) was not a strong or consistent predictor of fatigue.
- Number of days working/driving was not a strong or consistent predictor of fatigue.
- There were large individual differences in the incidence of fatigue; 11 of the 80 drivers (14%) had 54% of the drowsy episodes.
- Drivers obtained an average of 5.2 hours sleep per 24 hours, versus a self-reported ideal of 7.2 hours.
- Driver self-assessments of fatigue level were poor; there was little correlation between subjective and concurrent objective measures of fatigue (e.g., non-driving performance tests).
- Differences in video-observed drowsiness were primarily related to differences in exposure to night driving and other TOD differences.
- In a small percentage of driving (19 of 244,667 minutes or 0.008%), drivers were judged from polysomnographic data (e.g., EEG, EOG) to be in a loss-of-alertness state labeled “PSG-Drowsy Driving.”

- Video-judged drowsiness was generally the most robust measure of fatigue.
- Lane tracking and steering variability were subject to confounding from roadway conditions, but generally degraded in association with video drowsiness.
- Of the non-driving performance tests, the SRVT "may be the best . . . index of cumulative fatigue."
- Many drivers did not effectively manage their off-duty time to obtain the maximum possible time in bed within their off-duty hours.

Study limitations & potential improvements:
- See above notable UCVs.
- Conducted under old HOS rules which required only 8 hours off-duty daily, among other differences. Also, conducted entirely in Less Than Truckload (LTL) operations, whereas the majority of long-haul trucking is truckload (TL). Thus population, temporal, and ecological validity are all questionable.
- Numerous aspects of the methodology had never before been used under these conditions, and thus there was some trial-and-error and lost data. For example, lane tracker acquisition was only 33% (though there were no indications that this biased results).
- Some instrumentation was obtrusive (e.g., EEG) and data collection regimen was time-consuming and disruptive to normal operations.
- Some subject self-selection bias was possible since drivers had to agree to being subjected to obtrusive instrumentation, more obtrusive than in more recent studies.
- Study participation was limited to drivers with no documented history of Obstructive Sleep Apnea (OSA), though two participating drivers were diagnosed with OSA during the study.
- Video review for detecting drowsiness was not continuous but rather based on sampling. Observers made a simple judgment whether a driver "appeared drowsy." This included consideration of eyelid closure, but it was essentially a subjective judgment.
- The DFAS did not filter vehicle dynamics to identify extreme events; it did not capture "Safety-Critical Events" (SCEs).
- An improved study would address the above limitations and also employ more advanced instrumentation to increase its capabilities.

Principal Citation: Wylie, C.D., Shultz, T., Miller, J.C., Mitler, M.M., & Mackie, R.R., *Commercial Motor Vehicle Driver Fatigue and Alertness Study*, Federal Highway Administration, U.S. Department of Transportation, Washington, DC, 1996.

Additional analyses in: Mitler , M.M., Miller, J.C.,Lipsitz, J.J., Wash, J.K., Wylie, C.D. The sleep of long-haul truck drivers. *New England Journal of Medicine*, vol. 337: 755-761, 1997.

4.3 Effects of Operating Practices on Commercial Driver Alertness (O'Neill et al., 1999, other related reports)

Overview and primary study purpose. This was a truck driving simulator-based and simulated work study of working five consecutive 14-hour shifts which included 12 hours of mostly driving but with intermittent sessions of moving boxes. Its principal purposes were to assess the effects of the driving and work schedule and whether loading/unloading was followed by reduced driving performance.

Study design: Quasi-experiment with some experimental elements. All study subjects followed the same driving, working, rest, and recovery schedule over a 15-day period, with the exception of the loading/unloading schedule, which was counterbalanced across subjects between Week 1 and Week 2. Following two days of simulator and procedural familiarization, the schedule was:

- Days 1-5: Driving/working:
  - 10 hours off-duty
  - 14 hours on-duty during daytime/early evening (0700-2100) with 3 breaks totaling ~2 hours. Half of the drivers performed twice-per-day loading/unloading of 44 lb. book boxes for 90 minutes during these trips.
- Days 6-7: 58 hours off-duty (weekend). Multiple Sleep Latency Tests (MSLTs) and 10-minute PVTs were administered to subjects in the morning and early evening each day.
- Days 8-12:
  - 10 hours off-duty
  - 14 hours on-duty during daytime/early evening (0700-2100) with 3 breaks totaling ~2 hours. The other half of the drivers performed loading/unloading during these trips.
- Days 13-14: 58 hours off-duty (weekend), including the MSLTs and PVTs described above.
- Day 14: Final driving day to measure performance recovery.

Each subject performed the loading/unloading task twice daily on each of three work days. Half did this in Week 1 while the other half did it in Week 2. The 90-minute task included manual lifting and carrying of 44-lb boxes, and then moving a pallet of boxes with a pallet jack.

Subjects and sample frame: Ten (10) male CDL holders with long-haul experience. Drivers were non-smokers and completed a DOT physical and cardiac stress test to qualify.

Predictors (IVs, quasi-IVs):

- Hours-of-driving
- Time-of-day (daytime/early evening only)
- Physical work loading/unloading versus driving only (with rest breaks).
- Day-of-week
- "Weekend" recovery time.

Dependent variables (DVs): The First Ann Arbor Corporation (FAAC) DTS-2000 truck driving simulator with realistic truck cab and controls presented an 87-mile loop of varied driving, which included measures of driver performance such as:

- Vehicle speed control; e.g., adherence to speed limits
- Lane tracking (weaving)
- Gear-shifting performance; "grinds," engine stalls
- Brake usage
- Response to perturbation probes (crash threats or impending vehicle malfunctions); e.g., traffic stops ahead, oncoming vehicle in lane, merge/squeeze, oil pressure drop, air pressure drop, engine overheat, tire blowout, fog. Quality of driver response to each probe was rated on a 3-point scale by expert truck driver trainers.
- Video ratings of driver alertness on a 3-point scale by human factors researchers.

Other DVs included:

- Psychomotor Vigilance Test (PVT) administered 3 times daily on work days, twice daily on weekends.
- Recovery measures (based on EEGs and wrist-worm activity monitoring watches) included sleep patterns, sleep latency, and subjective sleepiness.

Units of measurement were specific to each DV. Different statistical tests were used for different types of variables. The criterion for statistical significance was $p < 0.05$.

Notable controlled variables (CVs):

- Physical loading/unloading task
- Timing and nature of perturbation probes (crash threats, vehicle problems) presented to drivers in the simulator.
- Other driving standardization possible through use of simulator; i.e., repeatable driving scenarios administered to all participants.

Notable uncontrolled variables (UCVs):

- Within the standardized duty tour, time-on-task, time awake, and TOD were all changing concurrently. Thus they were uncontrolled in relation to each other.
- On weekends, drivers could sleep, nap, and relax as they liked, except for the twice-daily testing.

Principal study findings:

- No major performance deteriorations over the duty tour; no statistically significant differences in responses to driving threats, lane keeping performance, or self-ratings of subjective sleepiness following 14 hours on duty versus driving following 10 hours on duty. Many driver response measures showed small, gradual declines over the duty tour, however.

- The only consistent and significant declines over the work schedule were in speed maintenance and gear-shifting performance.
- The authors attributed the above small/mixed effects more to time-of-day (e.g., mid-afternoon dips) than to time-on-task.
- Breaks (e.g., rest, eating) were almost always followed by performance improvements.
- Loading/unloading effects on subsequent alertness were not strong. Morning sessions were generally "invigorating," whereas afternoon sessions generally contributed to fatigue.
- No major declines in simulator driving over 5 days of work/driving; some small but statistically significant declines, however.
- Drivers averaged 6.3 hours nighttime sleep during work weeks. Weekend recovery sleep periods were longer (6.3-7.8 hours, including naps taken during the day).
- Driver weekend recovery of alertness (i.e., return to baseline performance as measured by EEG and MSLT) was generally complete within the first 24 hours of the 58-hour weekend.
- A preliminary study prior to the main study involved focus groups and driver surveys. Krueger and Van Hemel (2001) found that drivers' main fatigue concerns regarding loading/unloading related to the delays often involved, not to the physical labor required. Many long-haul drivers do not regularly load/unload their trailers, although unloading is more common than loading. The amount of loading/unloading by drivers varies by freight industry sector: grocery and household furniture carriers are among those most likely to require drivers to load/unload freight.
- Researchers suggested that subsequent studies might better focus on the beneficial effects of breaks rather than on deleterious effects of physical work. All three forms of breaks from driving (90-minute rest break, 90-minute loading/unloading, and 30-minute lunch break) generally enhanced driving performance, at least initially.

Study limitations & potential improvements:

- Questionable ecological validity:
    - Possible limited generalizability of simulated driving to real driving.
    - Only one daily shift (0700-2100) was tested, limiting generalizability to other shifts.
- Small number of subjects (10).
- The DV vehicle speed control (i.e., adherence to speed limits) has a questionable link to driver fatigue (construct validity); it is probably more related to driver impatience, frustration, or other underlying constructs.
- Although the FAAC DTS-2000 simulator was described as high-fidelity with accurate vehicle dynamics, one may question the fidelity of almost any simulator to real driving due to differences in the physical tasks and in risks, and also due to possible observation effects on subjects.
- An improved study would also include real driving measures and have more subjects.

Principal Citations:

O'Neill, T.R., Krueger, G.P., Van Hamel, S.B., & McGowan, A.L., *Effects of Operating Practices on Commercial Driver Alertness*, Office of Motor Carrier Safety Report No. FHWA-MC-99-140, Federal Highway Administration, U.S. Department of Transportation, Washington, DC, 1999.

O'Neill, T.R., Krueger, G.P., Van Hemel, S.B., McGowan, A.L. & Rogers, W.C. Effects of cargo loading and unloading on truck driver alertness. *Transportation Research Record,* Vol. 1686, Paper No. 99-0789; pp. 42-48, 1999.

Krueger, G.P. & Van Hemel, S.B. (2001). Effects of loading and unloading cargo on commercial truck driver alertness and performance. Federal Motor Carrier Safety Administration FMCSA Technical Report No. DOT-MC-01-107. US Department of Transportation, Washington, DC, 2001.

4.4 Effects of Sleep Schedules on CMV Driver Performance (Balkin et al., 2000).

Note: This project consisted of two separate studies. Study 2 was more extensive and important, but both are described here for completeness.

4.4.1 Study 1: Actigraphic Assessment of Sleep of CMV Drivers Over 20 Days

Overview and primary study purpose: Study 1 was a field study using wrist actigraphy to determine amounts and patterns of sleep in long- versus short-haul CMV drivers over 20 days.

Study design: Non-experiment; *in situ* observational study of 25 long-haul and 25 short-haul drivers.

Predictors: None *per se*; the principal factor of interest was driver work schedule.

Dependent variables (DVs):
- Manual subjective sleep and activity logs completed by drivers.
- Actigraph data identifying main sleep periods and naps.
- Combination of the above to best characterize sleep amounts and patterns.

Principal study findings:
- Both groups of drivers averaged 7.5 hours sleep per 24-hour day, including naps.
- Correlations between off-duty hours and main sleep hours were moderate to high: +0.42 for short-haul drivers ($p < 0.01$) and +0.82 for long-haul drivers ($p < .01$).
- Much of long-haul drivers' sleep was obtained in sleeper berths.
- "In both groups, however, there was no off-duty duration that guaranteed adequate sleep – for example, one driver obtained no sleep during a 20-hour off-duty period." (P. ES-5)

- For many drivers there were large day-to-day variations in total sleep. Some drivers showed chronic sleep restriction with intermittent bouts of extended recovery sleep.

Study limitations & potential improvements:
- Ecological validity; study was conducted under pre-2003 HOS rules which required only 8 daily off-duty hours for long-haul drivers.
- Both the reliability of actigraph readings in moving vehicles (for long-haul drivers in sleeper berths) and those of subject self-reports were questionable.

4.4.2 Study 2: Sleep Dose/Response Study

Overview and primary study purpose: Study 2 was a controlled laboratory between-subjects experimental study of the effects of various nightly times in bed (3, 5, 7, and 9 hours) on performance and alertness. A multiple-measure test regimen included driving on a desktop simulator. Results demonstrated the effects of sleep restriction (even minimal restriction) on alertness and were also used to optimize a Sleep-Performance Model (SPM) algorithm.

Study design: Experiment. Full 14.5-day regimen for each subject included 3 days of training/baseline with 8 hours in bed, 7 days with either 3, 5, 7, or 9 hours in bed (the four conditions) and 4 days recovery with 8 hours in bed. There was a variety of dependent alertness measures. Data were generally analyzed using a three-way mixed ANOVA for time-in-bed groups and across the days of the study. Main effects for sleep group (3, 5, 7, or 9 hours), day, and time-of-day were analyzed, as were their interactions (especially group X day).

Subjects and sample frame: Sixty-six (66) CMV drivers (CDL holders) aged 24-62, including 16 females (median age 43) and 40 males (median age 35). The sample included both truck and bus drivers. Their CMV driving experience varied widely.

Predictors (IVs, quasi-IVs): The principal IV was daily time-in-bed across 7-day test period (3, 5, 7, or 9 hours). Variables analyzed as quasi-IVs included TOD, time awake, and duration of the last sleep period.

Dependent variables (DVs): "The wide variety of performance and physiological measures . . . provide a comprehensive overview of the effects of sleep deprivation." (P. ES-7). This included:
- Psychomotor tasks; e.g.,
    - Walter Reed Performance Assessment Battery (PAB), which included serial addition and subtraction, choice reaction time measures, logical reasoning, "running" memory, code substitution, the Stroop color-word test, and delayed recall.
    - Performance on Systems Technology Inc. STISIM desktop driving simulator (medium fidelity)
    - Psychomotor Vigilance Test (PVT)

- Physiological measures included:
  - Polysomnographic measures, including electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG), and electrocardiogram (EKG). These were measured 24 hours per day and were used to identify sleep-alertness states, including microsleeps.
  - Oculomotor measures; e.g. pupil diameter, saccadic velocity.
  - Vital signs (e.g., heart rate).
  - Sleep latency.

Notable controlled variables (CVs):
- Fully controlled laboratory setting.
- Wake-up time was 7:00am for all four groups
- Standardized times for all performance tests and physiological measures for all groups across entire study.

Notable uncontrolled variables (UCVs):
- Subjects were heterogeneous with respect to age and CMV driving experience.

Principal study findings:
- There were statistically significant relationships between amounts of sleep the night before and subject performance (e.g., on the PVT) the following day.
- Sleep restriction affected simulator crash frequencies, with crashes increasing across days and with the 3-hour group experiencing the most sharply elevated risks.
- There was no strong relationship between lapses of alertness (as measured by EEG and EOG) and crashes while driving the simulator.
- The performance of the 7-hours-in-bed group was measurably poorer on some measures (e.g., PVT) than the 9-hour group, suggesting "that there was no compensatory or adaptive response to even this mild degree of sleep loss." (P. ES-8)
- Performance and physiological differences between the groups grew across the 7 days of differential sleep restriction. The 3-hour group, especially, experienced a large and cumulative alertness deterioration.
- Performance and physiological recovery from the severe sleep restriction (3-hour group) was not complete after 3 consecutive nights of recovery sleep (8 hours in bed).
- Daytime alertness and performance was a function of multiple factors, including circadian rhythm (TOD), time awake since last sleep period, duration of the last sleep period, and prior sleep extending back for at least several days. These factors can be incorporated into Sleep Performance Models.
- The 10-minute PVT was judged the most reliable and robust dependent measure for use in developing SPMs.
- There were large subject individual differences across almost all DVs.

Study limitations & potential improvements:

- The large number of DVs with limited administrations of each meant that some had low statistical power.
- Not every DV can be assumed to have ecological and outcome validity in relation to driver alertness and safety.
- Many tests could be administered only once or a few times daily, thus limiting TOD (circadian) comparisons and, for some, statistical power of comparisons.
- Between-subject design and subject individual differences contributed to error variances in group comparisons.

Principal Citation (for both studies):

Balkin, T.J., Thorne, D., Sing, H., Thomas, M., Redmond, D.P., Wesensten, N., Russo, M., Williams, J., Hall, S., & Belenky, G.L., *Effects of Sleep Schedules on Commercial Motor Vehicle Driver Performance*, FMCSA Technical Report No. DOT-MC-00-133, U.S. Department of Transportation, Washington, DC, 2000.

4.5 Stress and Fatigue Effects of Driving Longer Combination Vehicles (FMCSA, 2000)

Overview and primary study purpose: This is one of the few on-road driver fatigue tests employing a formal experimental design. Study drivers drove three different truck configurations, including two Longer Combination Vehicle (LCV) types, on standardized schedules and routes. Configurations included standard tractor semi-trailers (single trailer), tractors pulling triple trailers connected with conventional “A” dollies, and tractors pulling triple trailers connected with dual connection “Super-C” dollies (purported to increase vehicle stability). The purpose was to discern whether driving triples was significantly more stressful and fatiguing than driving a single and whether there was a difference between the two triple trailer dolly types. The study employed more than a dozen different fatigue measures, most or all of which were well-validated from prior research.

Study design: Experimental, within-subject comparisons. Counterbalanced sequence of subject exposure to three experimental conditions.

Subjects and sample frame: Twenty-four (24) experienced CMV drivers between the ages of 40 and 62. All had 9+ years of experience and had previously driven triples. Drivers were recruited from nearly a dozen companies, including large national and smaller regional carriers.

Predictor (IV): Truck-trailer configuration (three conditions).

Dependent variables (DVs):

- Self-reports:
  - Stanford Sleepiness Scale (SSS)
  - NASA Raw Task Load Index
  - Worksafe Australia Questionnaire
- Computer-based non-driving tests:
  - Critical Tracking Task (CTT)
  - Unprepared Simple Reaction Time
  - Two Finger Tapping Test (motor coordination)
  - Code Substitution Task (perception, short-term memory)
- Driving performance (lane tracking and steering):
  - Lane Deviation Squared
  - Maximum Lane Deviation
  - Standard Deviation of Lane Position (SDLP)
  - Large Steering Wheel Reversals
- Physiological:
  - Heart period/rate
  - Heart period/rate variability.

Units of measurement were specific to each DV. Most DVs were collected during driving or breaks from driving. Measures were also taken on recovery days following driving.

Notable controlled variables (CVs): Having 24 subjects enabled counterbalanced sequences of exposure to the three trailer configurations. Four subjects drove each of the six possible sequences (SAC, SCA, ASC, ACS, CSA, CAS). Trips were *non-revenue*, which allowed control of several key variables:

- Subject (within-subject design)
- Tractor
- Schedules and routes.
- Ancillary tasks (e.g., non-driving tasks as would normally occur in real operational trips).

Notable uncontrolled variables (UCVs): Although times and routes were controlled, weather and traffic could still vary.

Principal study findings:

- Across almost all measures, driving a standard single resulted in the least fatigue/stress, followed by the triple "C" dolly and then the triple "A" dolly. Results were cited as statistically significant but no further information was provided in the tech brief.
- The above effects were found during the trips and also on recovery days following the trips.

- In key respects, driver performance and status were superior when driving triple "C" dollies vs. "A" dollies; in particular, there were fewer lane exceedances.
- Driver individual differences were prominent in "all analyses." They represented 32-51% of mean squares for key lane-keeping and workload variables. Differences relating to truck configuration were small compared to driver individual differences.
- A rigorous experimental design and multiple DVs can be employed successfully in on-road driver fatigue studies.

Study limitations & potential improvements:

- Daytime trips only; no night driving (threat to ecological validity).
- Some instrumentation was obtrusive (e.g., heart rate monitors). Non-driving tests required stops.
- No capture of dynamic events; unknown validity in relation to the broader outcome of safety (versus alertness).
- Overall, this study was well-designed and executed for its purpose; i.e., to assess the causal relation between driving different LCV configurations and driver fatigue/stress.

Principal Citation: FMCSA. Stress and fatigue effects of driving long-combination vehicles. Tech Brief. No. FMCSA-MCRT-00-012, 2000. Earlier 1996 Report to Congress [citation not found].

4.6 HOS & Fatigue-Related Survey of Long-Distance Truck Drivers (McCartt et al., 2005, 2008).

Overview and primary study purpose: The Insurance Institute for Highway Safety (IIHS) conducted surveys of representative samples of long-distance truck drivers in Pennsylvania and Oregon immediately before (in 2003), one year after (2004), and two years after (2005) the major HOS rule change in late 2003. The survey attempted to capture rule-related changes in driver work schedules, sleep schedules, HOS compliance, and sleepiness while driving. It also quantified associations between reported rule violations and sleepiness. Major 2003-to-2004 HOS changes addressed included:

- Daily minimum off-duty requirement: 8 → 10 hours.
- Maximum hours of driving prior to going off-duty: 10 → 11 hours.
- Maximum tour-of-duty (beyond which driving is not permitted): 14 hours
- Initiation of 34-hour restart permitting reset of cumulative weekly hour limits (which themselves remained unchanged).

Study design: Anonymous "before and after" interviews were conducted with samples of drivers of large trucks passing through roadside weigh stations on Interstate highways in western PA and northwestern OR. Survey participation rates were high (88-98%), perhaps due to the weigh station setting where drivers were already stopped.

Statistical differences were tested using the Cochran-Mantel-Haenszel chi-square statistic ($p < 0.05$), stratified by state, cargo type (i.e., private carrier, for-hire carrier, owner-operator/other), and trailer type. This test was chosen because the distributions of sampled drivers varied significantly across these factors for at least part of the sample. The Cochran-Mantel-Haenszel chi-square statistic tested whether significant differences between the years existed for at least one of the strata. The study also computed odds ratios ($p<0.05$) for the strengths of associations between reported fatigued driving and reported frequency of rule violations, and that between reported rule violations and carrier, job, and driver characteristics. Analyses were conducted for each state and in some cases for the states combined. Results for each state were consistently in the same direction, which provided a rationale for combining them.

Subjects and sample frame: The sample included 1,921 drivers stopped at weigh stations on Interstates. Included were drivers who regularly made trips requiring them to spend at least one night away from home. Participants were paid $10 and answers were anonymous. Team drivers were 8-9% of the sample in PA and 19-20% in OR. Most drivers were middle-aged (age 40-59) and had 10+ years of experience driving large trucks. Across all samples, ~75% were in truckload operations and ~80% were for-hire. Fleet size varied widely.

Predictors:
- Before (2003) vs. one year after (2004) vs. two years after (2005) major HOS rule change.
- Driver-reported HOS violation types were used to predict driver-reported drowsiness.
- Adverse working conditions (e.g., excessive wait times) were used to predict HOS violations.

Dependent variables (DVs):
- Various measures of HOS compliance and non-compliance.
- Driver-reported drowsiness and dozing while driving.

Notable controlled variables (CVs):
- Surveys were taken at the same PA and OR weigh station locations using standard protocols.

Notable uncontrolled variables (UCVs):
- Numerous external factors could vary between the states, or across time. One specific state difference was that OR permitted ~25% of trucks to bypass the manual weigh stations based on pre-screening and prior clearances. Such exempted carriers tend to have better safety and compliance programs. This likely skewed the OR sample toward somewhat less compliant and safe drivers and carriers.

Principal study findings: Reported changes from before to after the rule change included:

- Following the rule change, more drivers increased their daily driving hours than decreased them. In the 2004 PA data, 22% reported more daily driving hours, 6% fewer, and 72% the same.
- More drivers increased their daily off-duty hours than decreased them. In 2004 PA, 31% reported more daily off-duty hours, 8% fewer, and 60% the same.
- However, the proportion of drivers who reported typically taking 10+ daily hours off-duty decreased between 2004 and 2005 (74–78% vs. 62%).
- In 2005, 38% of drivers in each state reported typically taking fewer than the required 10 hours off. The majority of these reported that they typically took fewer than 8 hours off.
- Following the rule change, more drivers reported increased daily sleep than reported decreased sleep. In 2004 PA, 35% reported more daily sleep, 7% less, and 58% the same.
- In 2004 PA, 39% of drivers reported getting <8 hours sleep daily after the rule change, versus 51% before the rule change.
- Across the two states and two post-rule years (2004 and 2005), ~80% of drivers reported using the 34-hour restart provision as part of their regular schedule.
- Percentages of drivers reporting that they drove while sleepy "at least once in the past week" were as high, or higher, after the rule change. The two states' progressions were:
    - PA: 2003-43% → 2004-48% → 2005-43%
    - OR: 2003-36% → 2004-36% → 2005-41%
- Percentages of drivers reporting that they "dozed at the wheel" in the past month were higher after the rule change. The two states' progressions were:
    - PA: 2003-13% → 2004-16% → 2005-19%
    - OR: 2003-12% → 2004-14% → 2005-21%
- Changes were mixed in percentages of drivers reporting that they had worked longer than permitted (i.e., violated the rules) during the past month. The two states' progressions were:
    - PA: 2003-25% → 2004-28% → 2005-29%
    - OR: 2003-30% → 2004-32% → 2005-24%
- In 2004, the percentages of drivers saying that they "often" violated specific rules were:
    - Drive >11 hours before taking required break: PA 7%, OR 13%
    - Drive past 14-hour tour-of-duty: PA 6%, OR 8%
    - Take fewer <10 hours off-duty: PA 12%, OR 11%
    - Drive more than weekly limit before taking required break: PA 4%, OR 7%
    - Omit hours worked in logbook: PA 17%, OR 21%
- All measures of the frequency of HOS rule violations were positive predictors of the two fatigue measures (driving sleepy in past week and dozing off in past month), with various 2004-2005 odds ratios ranging from 1.65 to 2.18. The study cited past research reporting the same association.

- Frequencies of reported HOS violations were most associated with having unrealistic delivery schedules, solo (vs. team) driving, longer detention (pick-up and drop-off) times, and frequent difficulty finding a legal place to stop or rest.

Study limitations & potential improvements:

- The biggest scientific question about the study relates to the accuracy of driver self-reports. In particular, driver self-assessments of drowsiness are known to be unreliable (e.g., Wylie et al., 1996, Section 4,2; and Van Dongen and Mollicone, 2013, Section 4.13). Further, the two drowsiness-related DVs involved memory over the past week or month in addition to self-assessment. Other questions required memory-based comparisons over one or two years.
- Two key findings appear contradictory. On the one hand, drivers reported taking more time off and getting more sleep following the rule change. On the other, they reported greater sleepiness. Almost all other fatigue studies show a strong positive relationship between amount of sleep and alertness.

Principal Citations:

McCartt, A.T., Hellinga, L.A., & Soloman, M.G. Work schedules of long-distance truck drivers before and after 2004 HOS rule change. *Traffic Injury Prevention,* 9:201-210, 2008.

McCartt, A.T., Hellinga, L.A., & Soloman, M.G. Work schedules before and after 2004 HOS rule change and predictors of reported rule violations in 2004: survey of long-distance truck drivers. *Proceedings of the 2005 Truck & Bus Safety & Security Symposium*, Alexandria, VA, November 14-16, 2005.

4.7 Analysis of Risk as a Function of Driving-Hour: Assessment of Driving-Hours 1 through 11 (Hanowski et al., 2008)

Overview and primary study purpose: This VTTI Naturalistic Driving (ND) study assessed the association of driving hours with Safety-Critical Event (SCE) incidence. Trucks were equipped with a full ND instrumentation suite to permit detection and analysis of SCEs (crashes, near-crashes, and crash-relevant conflicts). The study also looked at associations between TOD, traffic density, and driving shift (i.e., successive trips) with SCE incidence.

Study design: Quasi-experimental. There was no manipulation of IVs, but driving hours was treated as the nominal IV in analyses. The association of driving hour and SCE incidence was measured and assessed using logistic regression modeling. Supplemental analyses looked at odds ratios for SCE incidence by driving hour. Odds of occurrence were defined as the probability of an event (SCE) divided by the probability of non-occurrence. One approach used "generalized estimating equations (GEE) to account for correlations that might exist between drivers and within drivers (with respect to critical incident occurrence)." There was also a *post* *hoc* correlation of SCE incidence with national traffic density variations. The criterion for statistical significance was $p < 0.05$.

Subjects and sample frame: The dataset included 2.3 million VMT recorded from 98 drivers (97 male, 1 female, mean age of 40) and 45 instrumented trucks. Drivers drove the instrumented company trucks for an average of 12 weeks each. The participating companies were not described in the project report.

Predictors (quasi-IVs):
- Driving hours
- Time-of-day
- Traffic density.

Dependent variable (DVs):
- SCE incident rate was the principal DV and was "used as a surrogate for driver performance decrement" (FMCSA, 2008). The primary driving hour analysis included 819 SCEs (12 crashes, 12 tire-strike crashes, 85 near-crashes, and 710 crash-relevant-conflicts). SCE triggers included longitudinal accelerations/decelerations (e.g., hard braking), short times-to-collision (dynamic proximity to other vehicles), swerves, critical incident button activations by driver, and analyst-identified events. Individual trigger criteria were selected by researchers; for example, the hard braking criterion was decelerations of $\geq |-0.35g|$. No numeric breakdown of incident types was provided.
- SCEs were classified by fault (truck vs. other vehicle/person); thus a secondary DV as at-fault SCE incident rate.

Notable controlled variables (CVs): All drivers drove similar tractor semi-trailers on similar revenue-producing long-haul runs.

Notable uncontrolled variables (UCVs): Two uncontrolled variables in relation to time-on-task were time-of-day and traffic density. These were addressed in *post hoc* analyses.

Principal study findings:
- Generally no statistical difference in in SCE rates between the 2^{nd} and 11^{th} driving hours.
- Results were "consistent with Wylie et al. (1996) [DFAS] with regard to time-on-task being a poor predictor of crashes and safety-related traffic events."
- Highest SCE rate was in 1^{st} hour of driving; significantly greater than hours 2 through 11.
- Results for at-fault SCEs were very similar to those for all SCEs. This was not surprising since 618 of the 819 SCEs (75.5%) were truck at-fault.

- Supplemental analysis "found a strong time-of-day effect which . . . appeared to have resulted from hour-by-hour traffic density variations" (FMCSA, 2008) and not to physiological circadian effects.
- SCE rates were lowest between 2:00 and 6:00am, highest between noon and 5:00pm.
- Supplemental analysis reported a +0.83 correlation (r) between SCE rate and average U.S. traffic density across the 24-hour day.
- 1st hour spike was attributed to exposure to dense early traffic and possibly to sleep inertia.
- No consistent relationships between successive driving shifts and SCE incidence.

Study limitations & potential improvements:

- No demonstrated outcome validity in relation to crashes. The ND SCE dataset consisted of multiple, disparate dynamic events, each with a threshold criterion chosen by researchers. The ND SCE dataset composition was not analytically-derived from, or statistically related to, any crash dataset.
- SCE relative frequency was "used as a surrogate for driver performance decrement" (FMCSA, 2008). It is not clear, however, whether the intended construct was driver fatigue *per se*, driver risk in general, or something in between.
- ND does not capture at-fault and not-at-fault events equally; only one vehicle is instrumented and most events captured are at-fault. Thus, SCE rate cannot be taken as indicative of overall crash risk, even if the requirements noted above had been met.
- Some subject self-selection bias was possible since drivers had to agree to the ND surveillance protocol (though it was not physically obtrusive).
- Because the study was conducted under real operational conditions, there was no control or counterbalancing of driving times or other conditions.
- Much of the data was collected when a Drowsy Driver Warning System (DDWS) was operative in the vehicle; however, this was not judged to be a factor in the results because the system had been ineffective.
- No records of non-driving work activities.
- No disaggregation of environmental factors (roadway conditions) likely to be associated with daily trips and traffic density, such as trafficway flow (divided vs. undivided highway). Such analyses would have tested and perhaps demonstrated internal validity of findings.
- SCEs were counted and there was a separate analysis of at-fault events. However, there was no analysis of their characteristics or specific causal scenarios; e.g.,
    - Single-vehicle vs. multi-vehicle
    - Event scenario ("accident" type; e.g., rear-end, lane change, etc.)
    - Specific Critical Reason (CR); e.g., driver asleep, distraction,
    - Associated factors.
- No video evaluation of driver alertness (e.g., PERCLOS, Observer Ratings of Drowsiness). No DDWS data was reported as part of this study.

- There were no measures of continuous driver performance, driver physiology, or non-driving performance (e.g., the PVT).
- The use of SCEs "as a surrogate for driver performance decrement" is questionable given the lack of analytic derivation from crashes, the lack of SCE analysis to demonstrate performance mechanisms, and the high correlation of SCE incidence with traffic density. Prior ND studies employing the much of the same data (and/or the same paradigm) found a very low causal role of fatigue in SCEs compared to serious crashes (Hickman et al., 2005; Knipling 2009) and an *inverse* relation between SCE involvement and observed drowsiness (Wiegand et al., 2008). In the latter, both Observer Ratings of Drowsiness (ORDs) and Percent Eye Closure (PERCLOS) were *lower* for SCEs than for random "baseline" periods.
- The overall SCE rate was more than three times higher during daytime than during the circadian danger period of 2:00 to 6:00am. If SCEs were reflective of driver fatigue, one would expect the opposite.
- The TOD confound could have been removed by selecting control events to match SCE "cases" in TOD. One could also match other known confounds such as roadway type and weekday vs. weekend. This analysis, which still could be performed today using existing data, would better isolate HOS parameters.
- An improved study would validate the composition of its SCE dataset against crashes, analyze them in greater detail to determine their nature, and control for potential confounds such as roadway type.

Citations:

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.

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.

4.8 The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations (Blanco et al., 2011)

Overview and primary study purpose: This was a VTTI ND investigation of the 14-hour truck driver workday employing the same essential methodology as Hanowski et al. (2008) but with modifications to the mix of SCEs in the dataset. Its goals were to characterize the long-haul (truckload) and line-haul (less-than-truckload) driver workday and to examine the relationship between SCEs and three potential fatigue factors: driving hours, work hours, and breaks. Unlike the 2008 ND study, this study incorporated more information outside of driving *per se*, such as driver breaks and activities during breaks. It did not, however, separately consider at-fault SCEs, as had the 2008 study. The study contains extensive analyses of various criteria for counting

SCEs and various subsets of the data, such as separate analyses of trips including the full legal 11 hours of driving. It was a principal basis for the 2011 HOS rulemaking.

Study design: Quasi-experimental. There was no manipulation of IVs, but driving hours, work hours, and breaks were nominal IVs. Associations of these with SCE incidence were quantified and assessed using mixed-effect Negative Binomial regression models. This included models considering driving hour as a continuous variable and models where it was considered discrete. Supplemental analyses looked at odds ratios for SCE incidence by driving hour. Driver activities were captured using "hybrid" activity registers recording both driving and non-driving activities.

Subjects and sample frame: The study included 97 truck drivers (75 TL, 21 LTL), nine instrumented trucks, and about 735,000 miles of continuous driving data. Of 96 drivers providing demographic information, 91 were male while 5 were female; the average age was 44. Each driver drove for approximately 4 weeks. Four trucking companies participated; two were very large companies while two were of medium size (approximately 180 employees each).

Predictors (quasi-IVs):
- Driving hours
- Work hours (total hours working, driving + non-driving)
- Breaks (30+ minutes), including several subtypes.

Dependent variables (DVs):
- SCE incidence was the only DV. There were 2,197 SCEs, which included 1,118 Unintended Lane Deviations (ULDs) and unstated numbers of longitudinal accelerations/decelerations (e.g., hard braking), short times-to-collision (dynamic proximity to other vehicles), swerves, critical incident button activations by driver, and analyst-identified events.
- SCE rates were calculated both based on the full SCE number (i.e., SCEs/Hours) and "SCE-collapsed," which counted only one SCE per hour (i.e., Hours with SCEs/Hours).
- To increase statistical power for work hour comparisons over the 14 hour work tour, work hours were aggregated into three groups: beginning (work hours 1-5), middle (6-9), and end (10-14).
- "Unintentional Lane Deviations" (ULDs) were added to the mix of SCEs and were 51% of the dataset (1,118 of 2,197). A ULD occurred when the truck edge crossed over the lane edge without prior signaling and where no hazard (e.g., guardrail, ditch, vehicle, etc.) was present.

Notable controlled variables (CVs): All drivers drove similar tractor semi-trailers on similar long-haul runs.

Notable uncontrolled variables (UCVs): Two uncontrolled variables in relation to time-on-task were TOD and traffic density. Unlike Hanowski et al. (2008), these were not addressed in separate *post hoc* analyses.

Principal study findings:

- Drivers' workdays consisted of 66% driving, 23% non-driving work, and 11% resting/eating.
- Analyses on driving hours (i.e., driving only) and SCE risk found a significant time-on-task effect across all hours. Specific paired driving hour comparisons (e.g., Hour 11 vs. Hour 10) were generally not significant, though some were (e.g., Hour 11 vs. Hour 1). There was no 1st hour spike in SCE rates as previously seen by Hanowski et al. (2008).
- Analyses on work hours (i.e., driving plus other work) found that the risk of being involved in an SCE increased as work hours increased. That is, overall, shift end (Hours 10-14) SCE rates were highest, followed by shift middle (6-9), followed by shift beginning (1-5). This suggests that time-on-task effects may not be related to driving hours alone. When drivers began the day with several hours of non-driving work, followed by driving that went deep into the 14-hour workday, SCE risk was found to increase.
- Breaks from driving were followed by reduced SCEs (during 1-hour window after a break) and were effective to counteract the negative effects of time-on-task. SCE rates fell by 30-50% following breaks.
- Study data permitted analysis of individual driver SCE rates. Driver individual differences in SCE rates were great, with four extreme outliers. These four (of 97) drivers had more than 100 SCEs each and 36% of all the SCEs. Some hourly SCE rate comparisons became non-significant when these extreme drivers were removed from the dataset.

Study limitations & potential improvements:

- Both the internal and external validities of the study may be questioned. The threats to internal validity include multiple potential confounds (as illustrated earlier in Figure 6). The external validity question relates to the outcome validity of SCEs in relation to actual crashes.
- The limitations of the ND SCE methodology, as cited above for Hanowski et al. (2008), apply to this study also. As with that study, there was no demonstrated outcome validity in relation to crashes; i.e., no analytical derivation of SCE dataset parameters in relation to crashes.
- SCEs were not analyzed to determine their nature. The content and construct validity of SCEs as indicators of driver fatigue or driver performance deterioration are questionable for the same reasons as cited above for Hanowski et al., 2008.
- The addition of a large number of ULDs (51%) to the dataset raises questions about their meaning in relation to crashes and driver fatigue. ULDs are sensitive to driver fatigue, but they are not specific to fatigue. A prior ND study (Olson et al., 2009) found 77% of ULDs to be were *distraction*-related (involving a non-driving task), and distraction and drowsiness

appear to be negatively associated in SCEs (Barr et al., 2011). Also, there are questions regarding ULD criteria in light of varying road lane widths and ensuring that the lane breaks were indeed unintentional (Knipling, 2011a).

- No rationale was given for the number of ULDs added to the dataset, and their resulting proportion, in relation to the composition of crash types. This might be likened to "cooking without a recipe."
- The study did not analyze SCEs in relation to traffic density or consider traffic density as a confound, in spite of previous ND findings that SCEs were strongly driven by traffic density (i.e., Wiegand et al., 2008; Hanowski et al., 2008).
- The study also did not consider TOD, known as a dominant factor in fatigue (circadian rhythm) and overall crash rates (mainly reflecting traffic density) from virtually every past study. The TOD confound could have been eliminated by supplemental comparisons of SCEs to control points matched for TOD and DOW. Case and control points could be compared in relation to hours of driving and work. This analysis could still be performed *post hoc* today using existing data.
- One step toward more validated SCE datasets would be to decide on, and explicitly state, which crash population SCEs are intended to represent. They might be posited to represent all crashes, all serious injury crashes, all truck driver at-fault crashes, all fatigue-related crashes, or some other population. Then efforts could begin to index SCE datasets to defined, important crash populations. Such indexing would provide outcome validity.
- As the study authors pointed out, future studies would benefit from capturing actual driver sleep in addition to off-duty hours and activities. For example, naps during breaks were not captured in the current study. Future studies of breaks would benefit from being able to determine whether the break included a nap.
- The authors also noted that study participation was voluntary and not all drivers who were asked agreed to participate. Thus there could have been some self-selection bias.

Principal Citation: 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.

4.9 Hours of Service and Driver Fatigue: Driver Characteristics Research (Jovanis et al., 2011)

Overview and primary study purpose: This FMCSA-funded study by Paul Jovanis and his colleagues at Penn State University was a comparison of HOS logs for drivers on truck trips where a crash occurred to those where no crash occurred. Control logs included those from different non-crash-involved drivers and also those from the same crash-involved drivers in the hours before their crash. It was a principal basis for the 2011 HOS rulemaking. Predictors (quasi-IVs) of principal interest included driving hours (i.e., time-on-task), multiday work patterns, and duty breaks. The DV was relative crash likelihood; there were no measures of driver fatigue or driver performance *per se*. The study employed data from 2004-05 and 2010; these two time periods had the same pre-2011 HOS rules. Notably, these rules allowed 11 hours of driving over a 14-hour tour-of-duty and permitted a 34-hour restart without any time-of-day restrictions.

Study design: The quasi-experimental case-control design compared driver HOS logs for periods of 1-2 weeks prior to the crash to a random sample of non-crash logs (two for each crash log) for drivers from the same company, terminal, and month using a case-control time-dependent logistic regression formulation. Control data also included hours logged by crash-involved drivers before their crashes. A survival model was used to capture the fact that a crash in a particular hour (e.g., hour 7) implied that the driver "survived" the first 6 hours but crashed in the 7th. Control hours included those from non-crash-involved drivers and also those from other crash-involved drivers prior to their crash. Multiday patterns were analyzed using cluster analysis of different shift schedules. The criterion for statistical significance was $\boldsymbol{p} < 0.05$, though the study also cited "barely significant" findings of $\boldsymbol{p} < 0.20$.

Subjects and sample frame: Log data were collected from 1,564 drivers. For truckload (TL) operations (two companies) there were 318 crash-involved logs and 560 controls. For LTL operations (three companies) there were 224 crash-involved logs and 462 controls.

Predictors (quasi-IVs):
- Driving hours
- Multi-day work patterns
- Duty breaks in which a driver went from driving status to either sleeper berth or off-duty status. Note that these were often two hours or longer, in contrast to the current HOS requirement of a 30-minute break after 8 hours of driving.
- "Pseudo-violations," defined as a driving pattern which was legal but which would have violated previous HOS rules. This was examined in the context of the 34-hour restart provision.

Dependent variables (DVs):
- Crashes vs. non-crashes. Crashes met DOT-reportable criteria (i.e., involved a fatality, treated injury, and/or a vehicle towed).
- There were no descriptions or analyses of crash fault, causal factors, number of involved vehicles, or roadway conditions of occurrence.

Notable controlled variables (CVs):
- Control logs were selected randomly from the same company, terminal, and month as the case (crash) logs.

Notable uncontrolled variables (UCVs):

- Non-crash control logs were from different drivers than the crash-involved drivers (i.e., between-subjects rather than within-subjects). Crash- and non-crash drivers could have differed in their typical schedules, susceptibilities to fatigue, and overall driving risk.
- Time-of-day (TOD) is well-known as a confounding factor in crash risk due to predictable daily traffic density variations, likely road changes (e.g., from Interstates to local roads) during work shifts, and also to the high-risk early morning circadian physiological low period.

Principal study findings:

- "Considering all the data, there was a consistent increase in crash odds as driving time increases." (Abstract) This included a gradual increase across hours 1-10 and then a marked increase from hour 10 to hour 11.
- "Breaks from driving reduced crash odds." (Abstract)
- "There was, however, an increase in crash odds associated with the return to work after a recovery period of 34 hours or more." (Abstract)
- The "pseudo-violation" analysis (of one carrier) found that "schedules compressed over 2 days are associated with increased crash odds on subsequent days. Compressed schedules over 1 day . . . are not associated with increased crash odds on subsequent days." (P. 58)

Study limitations & potential improvements:

- The internal validity of the study is questionable because of numerous threats to validity, as illustrated earlier in Figure 6 and the associated discussion.
- The word "fatigue" is in the report title, key words, and introduction, but is found only once thereafter in the report. "Fatigue" is not defined in the report glossary, nor elsewhere. Though fatigue was the central construct, the study contained no measures of fatigue. Further, it did not provide any evidence of fatigue involvement, nor evidence of any other specific driver performance deterioration or error. Fatigue was implicitly assumed as the intervening variable (construct) between schedule features and crash risk, but alternative explanations are possible and were not addressed. For example, the marked increase in crash risk between hours 10 and 11 could have been due to drivers hurrying to complete their trips rather than due to fatigue. Or it could have been due to conditions external to the driver; e.g., variations in traffic density.
- Similarly, "driver characteristics" was in the title of the report, but the study did not report or address any characteristics of its drivers other than their recent HOS schedules.
- There was no description of study crashes (other than that they were DOT-reportable) and no disaggregations of the dataset by crash characteristics. Disaggregations could have demonstrated internal validity by providing evidence consistent with actual fatigue involvement. This would include disaggregation by crash severity (fatigue incidence is

greater in severe crashes), single- vs. multi-vehicle (fatigue is much greater in single-vehicle crashes) and roadway type (fatigue is greater on divided highways).

- There was also no disaggregation of crashes by fault or "preventability" (a standard surrogate for fault used by commercial fleets). If fatigue were operative, one would expect much stronger schedule effects for at-fault ("preventable") crashes.
- Time-of-day (TOD) is an environmental factor threatening the internal validity of virtually all quasi-experimental studies of driving schedules. TOD was an embedded factor in schedule clusters analyzed, but no clear descriptive statistics on crash TOD were presented. For example, the study presents a summary of crashes, non-crash exposure, crash exposure, total exposure, and crashes/exposure by driving hour, but there is no similar table for TOD. A time-on-task by TOD matrix of crash risk (crashes/exposure) would have clearly shown the relative roles of these factors and interactions, but none was provided (Knipling, 2011b).
- Newly published data from the same Jovanis dataset (Chen and Xie, 2015) show a 5-fold range in hourly crash frequencies across the 24 hours and a 3-fold jump beginning at 5:00am, the initial hour of a very high fatigue period as seen in the LTCCS (see Section 3.2). Jovanis considered only time-on-task, but his overnight drivers could have been experiencing greater end-of-shift risks due to other factors, i.e., circadian-based fatigue, early morning rises in traffic, and/or end-of-shift roadway shifts from freeways to local roads. Only disaggregation of these disparate but co-varying factors would provide an answer.
- TOD as a confound could have been eliminated by supplemental comparisons of crashes to control log points matched for TOD and day-of-week. Cases and these TOD-matched controls could be compared in relation to hours of driving and work. This analysis could still be performed *post hoc* today using existing data.
- The 11th hour rise in risk was much stronger for LTL than TL drivers. Most LTL trips were overnight whereas most TL trips were during the day. The LTL larger rise could have been due to the circadian low period of 3:00 to 7:00am (perhaps interacting with morning rush hour traffic) when most overnight trips were ending or near their end.
- The LTL analysis clustered similar driving shift patterns, but did not consider TOD as a separate variable. The TL analysis considered TOD but added it as a final step after driving hours, multiday patterns, and driving hour-pattern interactions, thus potentially masking independent TOD effects (Knipling, 2011b).
- The critical 11th hour statistics were based on 16 of 542 total crashes (3.0%) and 50 of 9,862 total exposure hours (0.5%).

Principal Citation: 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.

Additional information in: Chen, C. & Xie, Y. Effects of driving hours and time of day on large truck safety based on multilevel discrete-time survival analysis. *TRB 2015 Annual Meeting,* Paper 15-5331, 2015.

4.10 Motorcoach Driver Fatigue Study 2011 (Belenky et al., 2012)

Overview and primary study purpose: This study gathered data on the daily work and sleep cycles over a one-month period for 84 motorcoach drivers, totaling 1,710 duty days. Its aim was to determine motorcoach driver duty hours, sleep time, fatigue, and performance while operating within the limits of U.S. motorcoach HOS rules (which are different than truck rules). The study reflected concerns about features of motorcoach HOS rules which might not support optimal sleep and rest. Their rules require only 8 consecutive hours off-duty daily, a duration which would not likely permit sufficient restorative sleep. Daily driving hours (maximum = 10) and work hours (maximum = 15) might result in drivers having "short" daily work-rest cycles of less than 24 hours with resulting backward schedule rotation and cumulative fatigue.

Study design: Non-experimental, with correlations calculated between some measures (e.g., daily shift duration and sleep time). Time during shift (i.e., beginning vs. end) was treated as a quasi-IV in relation to measures of driver alertness. Descriptive statistics (e.g., mean, standard deviation) were calculated for various measures. "Further, where appropriate, the data were analyzed using mixed-effects ANOVA models with Statistical Analysis Software (SAS) v9.2 Mixed Procedure (MP)." (P. 6)

Subjects and sample frame: The 84 motorcoach drivers (64 males, 20 females) recruited for the study constituted a convenience sample, not statistically representative of any driver population. Drivers self-identified as driving for Charter, Tour, Regular Route, or Commuter Express operations. Most drivers were middle-aged and were overweight or obese.

Predictors (quasi-IVs): Much of the study was purely descriptive. However, time during shift (i.e., beginning vs. end) was analyzed as a quasi-IV.

Dependent variables (DVs):

- Drivers kept a duty/sleep diary to provide data on duty start times, duty break times, and duty end times.
- Data from wrist-worn actigraphs were analyzed by computer scoring to measure sleep/wake history.
- The 3-minute PVT was administered via a smart phone when drivers were going on and off duty, as well as before and after any mid-duty breaks.
- Karolinska Sleepiness Scale (KSS), intended to capture the construct drowsiness. The KSS is a 9-point semantic differential scale which reads:
    - 1: extremely alert
    - 2: very alert
    - 3: alert
    - 4: rather alert

  - 5: neither alert nor sleepy (the neutral point)
  - 6: some signs of sleepiness
  - 7: sleepy, but no difficulty remaining awake
  - 8: sleepy, some effort to keep alert
  - 9: extremely sleepy, fighting sleep.
- Samn-Perelli Fatigue Scale (SPFS), a self-rating ordinal scale of fatigue, as distinguished from sleepiness. The scale ranges from 1 (fully alert) to 7 (completely exhausted). The conceptualized construct was general fatigue.

Notable controlled variables (CVs): None other than consistencies of test administration. There was also *post hoc* disaggregation of data by the four motorcoach operations types to assess whether measures varied by operations type.

Notable uncontrolled variables (UCVs): Multiple UCVs since no operational variables related to driving were controlled. Most notable confounds include:

- Comparisons between fatigue measures (e.g., PVT, self-ratings) at shift start vs. shift end were characterized as reflecting the effects of work, but the testing also occurred at different circadian times of the day (i.e., mean 8:43am vs. 5:51pm). Light conditions could also differ, as more end-of-shift testing would be at twilight or in darkness. The small 2.5% mean change in PVT speed could easily have reflected one or both of these factors.
- Comparisons across motorcoach operations could be confounded by inter-subject differences, especially the presence of subject outliers within one or more groups.

Principal study findings: "From the data: 1) duty start times clustered in the morning; 2) average total duty time for duty days was slightly more than 9 hours [9.14 hours]; 3) average total sleep time per 24 hours was in the range of 7 to 9 hours, with less sleep during on-duty days and more sleep during off-duty days. During on-duty days, longer total duty times were associated with shorter sleep. Drivers performed worse on the PVT and reported increased sleepiness and fatigue at the end of a duty period relative to the beginning. These findings were in the context of an estimated average of 43 hours on duty per week. Thus, drivers in the sample on average started work in the morning, worked approximately 9-hour days, and slightly more than a 40-hour week, and obtained satisfactory amounts of sleep. On average, drivers did not push the limits of the hours-of-service regulations." (Abstract) Additional findings included:

- The mean driver BMI was 31.9, which is in the obese range.
- Drivers worked 13 hours or more on 11% of duty days.
- Twelve (12) of the 84 drivers (14%) worked an average of 60 hours or more per week.
- The clustering of shift start times in the morning (mean start = 8:43am) "suggests a normal diurnal schedule with normal diurnal circadian phase synchronization" (P. 12) or, in other words, reasonable schedule regularity.

- Mean total daily sleep was 7.3 hours when drivers were working and 9.1 hours when drivers were not working.
- The overall correlation (Pearson **r**) between shift start time and total daily sleep was +0.34, suggesting that later start times enable more sleep.
- The overall correlation between daily shift duration and daily sleep by duty day was -0.20, suggesting that longer work shifts result in less overall driver sleep. However, this correlation was reduced to near zero (and was non-significant) when the two measures were first aggregated by driver. Thus the negative correlation could largely reflect between-driver differences.
- PVT speed decreased (i.e., reaction times increased) between shift start and end, but only slightly: 2.5%.
- The two subjective, ordinal scale measures of self-rated fatigue and sleepiness changed much more, relatively. The mean KSS value rose from 2.6 to 4.6, while the mean SPFS rose from 2.2 to 3.9. This is consistent with other studies finding that work results in greater changes in subjective self-ratings than in objective performance (e.g., Wylie et al., 1996). Note, however, that both early- and late-shift KSS values were on the alert side of the scale mid-point of 5.0. Thus, drivers overall were not rating themselves as sleepy, even at the end of shifts.

Study limitations & potential improvements:

- Computer-scored actigraphy appears to overestimate actual sleep, since wakeful but sedentary activity can be misread as sleep. Therefore, the driver sleep picture may be less sanguine than suggested by the study.
- Supporting the above view is the fact that mean non-workday sleep was nearly two hours greater than workday sleep. The difference is consistent with a driver sleep debt developed during the work week and reduced by increased sleep during weekends.
- Wide recruiting was necessary to obtain 84 subjects, suggesting the possibility of considerable self-selection bias in the subject pool. Frequent HOS violators would probably be far less likely to volunteer for the recording regimen than would those compliant with HOS rules. Drivers who knew themselves to be fatigue-susceptible might also be reluctant to volunteer. Self-selection bias would reduce or otherwise change the range of driver characteristics, which could in turn affect correlations with other driver characteristics factors such as daily sleep, self-ratings, and measured alertness.
- An extension of the study, and also an internal validity check, would have been to enter driver sleep data, time-of-day, and perhaps time awake into a sleep-performance model to generate driver alertness predictions. These could have been correlated to PVT scores, a measure often used to validate and refine sleep-performance models.
- PVT performance and subjective fatigue differences between the start and end of the work shift were characterized as representing the effects of work; e.g., "Participants demonstrated decreased PVT performance at the end of the duty period relative to the beginning. Similarly,

participants reported an increased level of fatigue and sleepiness at the end of a duty period relative to the beginning." (P. x). The report never noted the testing confound between work (start vs. end of shift) and circadian phase (morning vs. late afternoon), even though the report Introduction presented circadian phase as one of the two factors most predictive of alertness (along with prior sleep).

- As noted, the increases in mean self-rated sleepiness on the KSS across shifts did not cross the scale mid-point of 5.0. Drivers were not generally on the sleepy side of the scale, even at shift end.

Principal Citation: Belenky, G. Wu, L. J., Zaslona, J.L., Hodges, J. Motorcoach *Driver Fatigue Study, 2011*, FMCSA Report No. FMCSA-RRR-12-042, December 2012.

4.11 Investigation of the Effects of Split Sleep Schedules on Commercial Vehicle Driver Safety and Health (Belenky et al., 2012).

Overview and primary study purpose: This experiment addressed the question of whether split sleep is as beneficial as consolidated sleep in sustaining driver performance and, over the long term, in sustaining driver health. The study compared daily sleep split into two periods per day to two consolidated sleep conditions, nighttime sleep and daytime sleep. Dependent measures included total sleep time, subject objective performance (simulated driving, PVT), subjective state (sleepiness, mood, and effort), and biomedical parameters (e.g., blood chemistries) associated with long-term health. A general conclusion of the study was that nighttime sleep was superior to split sleep and to daytime sleep. Split sleep was in many ways superior to daytime sleep. These differences were not found on all measures, however. Most notably, objective performance measures generally did not differ. Currently, truck drivers are permitted to take a split-sleep period of only two hours apart from their main off-duty period, though prior to 2003 the HOS rules were more flexible. Study results may lead to consideration of a return to more flexible split-sleep rules.

Study design: To maximize experimental control, the study was conducted in an in-residence laboratory and using all healthy young male subjects. Three sleep conditions were examined: consolidated nighttime sleep, split sleep, and consolidated daytime sleep. A between-subjects experimental design was used whereby 53 subjects were each assigned to one of the three conditions. The testing regimen covered 10 days, which included 2 baseline days, a 5-day simulated workweek, and a 2-day recovery period. Ten hours was chosen as the daily sleep opportunity time to be consistent with a 14 hours on, 10 hours off CMV driver schedule. Statistical analyses included between-group comparisons and three-way ANOVAs (e.g., Condition by Workday by Testing Session).

Subjects and sample frame: The 53 participants were young men (age 22-40) and were healthy and non-obese (BMI < 30). Subjects had to meet more than 25 health and related criteria to participate. A healthy, homogeneous sample was used to reduce random factors affecting the data. The sample was not intended to be representative of CMV drivers, nor was the study setting intended to be representative of CMV driving.

Predictor (IV): The IV was sleep schedule with 3 conditions, each with the same 10-hour total sleep opportunity: consolidated nighttime sleep (10pm to 8am), split sleep (3am to 8a.m. and 3pm to 8pm), and consolidated daytime sleep (10am to 8pm).

Dependent variables (DVs):
- Total daily sleep time, measured polysomnographically (EEG, EOG, EMG, etc.) twice during baseline, twice during the work week, and one during recovery.
- Driving performance on a simulator (40-minute drive):
  - Average speed on straightaways
  - Speed variability on straightaways
  - Standard deviation of lane position (SDLP)
  - Emergency braking reaction time (to pedestrian threat).
- Non-driving performance measures:
  - Psychomotor Vigilance Test (PVT) lapses
  - Digit-Symbol Substitution Test (DSST)
- "Neurobehavioral test battery" to assess subjective state:
  - Measures of sleepiness; e.g., Karolinska Sleepiness Scale (KSS)
  - Measures of mood and emotion (positive and negative)
  - Measures of performance and effort.
- Biomedical parameters:
  - Blood glucose, interleukin-6 (IL-6), leptin, and testosterone
  - Blood pressure.

Notable controlled variables (CVs):
- Total opportunity for sleep (10 hours)
- Test and evaluation protocols
- External stimulation: subjects had no contact with the outside world during the 9-day testing.

Notable uncontrolled variables (UCVs):
- Although the laboratory protocol permitted excellent experimental control, there was one notable exception. Subjects in the daytime condition were also participating in a separate, concurrent study requiring 16 days in the laboratory rather than 10 as for the other subjects.
- Because of limited equipment and staff, some testing sequences were counterbalanced rather than fully standardized.

Principal study findings:

- During the 5-day simulated workweek, measured sleep was:
  - Nighttime sleepers: 8.4 hours ± 13.4 minutes standard error of the mean
  - Split sleepers: 7.2 hours ± 14.2 minutes
  - Daytime sleepers: 6.4 ± 15.3 minutes.
- REM sleep was greatest for night sleepers and least for day sleepers.
- Split sleepers slept more during the overnight/morning period than during the afternoon/evening period.
- During the 5-day simulated workweek no significant differences were found in performance on the PVT, driving simulator, or the DSST.
- During the 5-day simulated workweek, KSS subjective sleepiness scores (1 = very alert, 9 = very sleepy) were:
  - Nighttime sleepers: 3.5 ± 0.2
  - Split sleepers: 3.5 ± 0.2
  - Daytime sleepers: 4.3 ± 0.2.
- Subjective sleepiness was greater for the daytime sleep condition than for others, but the mean values for all three conditions were on the alert side of the scale midpoint of 5.0.
- Other subjective measures (e.g., mood) did not differ by condition.
- Blood glucose and testosterone were increased in the daytime group compared to the other two groups. Otherwise there were no significant differences.
- There were no significant changes or differences in blood pressure.
- Across the testing days, day sleepers showed some fatigue increases; e.g., greater lane deviations, KSS increases.
- In conclusion, the "Results of this study suggest that when consolidated nighttime sleep is not possible, split sleep is preferable to consolidated daytime sleep" (FMCSA, 2012).

Study limitations & potential improvements:

- By design, the study did not use CMV drivers as subjects and did not attempt to simulate CMV driving conditions. The authors envisioned that their laboratory work would be followed by a field study with greater fidelity to CMV driving. In other words, external validity in relation to CMV driving was sacrificed in order to ensure high internal validity.
- The study used young, healthy subjects to reduce subject variability and possible confounding effects of sleep disorders. Yet younger drivers are generally more susceptible to drowsiness during driving than are older drivers (e.g., Barr et al., 2008; Knipling, 2009a). Thus, subject population validity is problematic.

Principal Citations:

Belenky, G. Jackson, M. L., Tompkins, L., Satterfield, B., and Bender. A. *Investigation of the Effects of Split Sleep Schedules on Commercial Vehicle Driver Safety and Health.* FMCSA Report No. FMCSA-RRR-12-003, December 2012.

FMCSA. *Investigation of the Effects of Split Sleep Schedules on Commercial Vehicle Driver Safety and Health.* Research Brief, December 2012.

4.12 Laboratory Study of the Efficacy of the 34-Hour Restart (Van Dongen & Belenky, 2010)

Overview and primary study purpose: This laboratory study compared the effectiveness of 34-hour restarts involving primarily night sleep (the "best case"condition) and those involving primarily daytime sleep (the "worst case" condition). Activities and tests simulated two 5-day work periods separated by 34-hour restart periods. The main IV was the "circadian placement" of the sleep/wakefulness cycle, both across the work weeks and during the restart period. DVs included PVT lapses and simulated driving performance. This study was the laboratory basis for a later field test addressing the same questions (and described in Section 4.13).

Study design: The report's Abstract stated, "A sample of 27 healthy subjects was studied in an in-residence laboratory study with frequent testing of cognitive performance and driving performance on a high-fidelity driving simulator. A "worst-case"–"best-case" between-groups comparison was made of two 5-day (14-hour/day) work periods separated by a 34-hour restart period. Half the sample was randomized to the "best-case" condition, which entailed daytime wakefulness and work (and nighttime sleep) throughout the study. The other half was randomized to the "worst-case" condition, which entailed nighttime wakefulness and work (and daytime sleep) during the two 5-day work periods, while transitioning back to a daytime schedule during the 34-hour restart period." The primary statistical design involved within-subject comparisons of performance during the first 5-day work period with performance during the second 5-day work period (i.e., repeated measures), and between-groups comparisons between the two conditions. The researchers employed two-way mixed-effects ANOVAs and focused on the interaction of group by work period (session). Additional analyses involved three-way mixed-effects ANOVAs of condition by session by day, and three-way mixed-effects ANOVAs of condition by session by TOD.

Subjects and sample frame: The 27 participants were healthy young men and woman (age 22-40) without known sleep disorders or other medical issues. The healthy, homogeneous sample was not intended to be representative of CMV drivers.

Predictors (IV and quasi-IV): The IV was the "circadian placement" of the sleep/wakefulness cycle, both across the work weeks and during the restart period. Within each group and work period (before and after the restart), factors such as work day and TOD were employed as quasi-IVs for supplemental analyses.

Dependent variables (DVs): Multiple DVs included:

- Driving performance during a rural drive scenario on a PatrolSim IV driving simulator (MPRI, Salt Lake City, UT). This is a high-fidelity driving simulator widely used to train professional drivers. The simulated drives (40-minute drive, 4 times daily) measured:
  - Braking response to pedestrian or dog-on-road threat
  - Lane deviations during uneventful straightaway driving
  - Speed variability during straightaway driving
  - Fuel use during straightaway driving.
- Non-driving performance measures:
  - Psychomotor Vigilance Test (PVT) lapses. PVT sessions were 10 minutes each and the lapse criterion was 500 msec.
  - Digit-Symbol Substitution Test (DSST)
- "Neurobehavioral test battery" to assess subjective state:
  - Subjective measures of sleepiness; e.g., Karolinska Sleepiness Scale (KSS)
  - Subjective measures of mood and emotion (positive and negative)
  - Measures of performance and ratings of perceived effort.

Notable controlled variables (CVs):

- Amount of scheduled work (simulated driving and testing)
- Durations of scheduled wakefulness and sleep periods
- Test and evaluation protocols
- External stimulation: subjects had no contact with the outside world.

Notable uncontrolled variables (UCVs):

- Driving scenarios were standardized, but this meant that they did not create realistic day-night variations; e.g., more traffic during the day and darkness at night.
- In most respects, the two groups' schedules were day-night mirror images. This meant, however, that the "worst case" group was exposed to circadian-related disadvantages throughout the testing, not just during the restart period.

Principal study findings:

- In the "best case" condition, average PVT performance in the 5-day work period was the same before and after the 34-hour restart period, indicating that the restart was effective in maintaining performance.

- In the “worst case” condition, average PVT performance after the 34-hour restart was impaired relative to before, indicating that the restart period was not effective.
- Total sleep time was consistently reduced in the “worst case” condition relative to the “best case” condition during the two 5-day work periods.
- The level of PVT performance impairment reached in the “worst case” condition was modest when compared to the documented effects of one night of acute total sleep deprivation or a week of sustained sleep restriction to 6 hours per day. “Worst case” subjects were allowed to nap prior to transitioning to night work, which perhaps accounted for their relatively good performance compared to those seen in more extreme conditions.
- “Worst case” subjects were subjectively sleepier than “best case” subjects, but did not report increased sleepiness after the 34-hour restart period. Subjective sleepiness did not accurately track objective performance impairment.
- “Worst case” subjects displayed an increase in lane deviation over the hours of the night. This was accompanied by an increase of up to 1% in fuel use. Indices of driving impairment (speed variability, lane deviation, emergency braking reaction time, and fuel use) were significantly predicted by lapses on the PVT administered before driving. “PVT performance was . . . predictive of simulator driving performance, suggesting that mathematical models of fatigue and performance predictive of PVT performance may be predictive of driving performance as well.” (P.56)
- A follow-on supplemental laboratory study, labeled “Phase II” (FMCSA, 2010), used a within-subjects design to show that, for nighttime drivers, 58-hour restart periods with opportunities for two nighttime sleep periods were more effective in restoring alertness than were 34-hour, one biological night restarts.

Study limitations & potential improvements:

- Although the study was ostensibly a comparison of two restart regimens, the two groups were treated differently in other important ways. Most notably, the “worst case” group was on a day-sleeping, night-working schedule throughout the work weeks. They were getting less sleep and performing under more adverse circadian conditions. During the 34-hour restart periods, they were forced to transition back to night sleeping for their one nightly sleep period. Throughout the testing, the “worst case” group had to “work” through the overnight period from 3am to 6am where time awake and circadian low periods combine to maximize fatigue. Thus “worst case” conditions were adverse throughout the protocol, not just during the restart period.
- Recall from Section 2.3 that *external validity* is the extent to which observations made in a study generalize beyond the specific manipulations and setting of the study. Three types of threats to external validity in the current study are:
    - Population validity; generalizability from the current subjects to CMV drivers. The study used young, healthy subjects to reduce subject variability and possible confounding effects of sleep disorders. Yet younger drivers are

generally more susceptible to drowsiness during driving than are older drivers (e.g., Barr et al., 2008). These young-old discrepancies heighten concerns about the study's population validity.
- Ecological validity; generalizability across settings.
- Outcome validity; generalizability across different but related DVs (e.g., different measures of alertness or safety). Specifically, this study used mostly fatigue-related DVs, but these cannot necessarily be generalized to overall driving safety because of other, non-fatigue factors, most notably TOD-related variations in traffic density. Determination of fatigue effects is appropriate as a first step, however.

Principal Citations:

Van Dongen, H. and Belenky, G. *Investigation into Motor Carrier Practices to Achieve Optimal Commercial Motor Vehicle Driver Performance, Phase I*, FMCSA Report No. RRR-10-005; December 2010.

FMCSA. *Investigation into Motor Carrier Practices to Achieve Optimal Commercial Motor Vehicle Driver Performance, Phase I*, Tech Brief No. RRR-10-006; December 2010.

FMCSA. *Duration Restart Period Needed to Recycle with Optimal Performance, Phase II*, Tech Brief No. RRR-10-062-TB; December 2010.

4.13 ND Field Study on the Efficacy of the New Restart Provision for Hours of Service Report to Congress (Van Dongen & Mollicone, 2013)

Overview and primary study purpose: As a follow-up to its laboratory study of HOS restarts, FMCSA sponsored an ND study to assess fatigue in CMV drivers working their normal schedules and performing their normal duties. Researchers compared driver fatigue between weekly duty cycles preceded by a restart break with *one* nighttime rest period to weekly duty cycles preceded by a restart break with *two or more* nighttime rest periods. The study was conducted between January and July of 2013, just prior to the effective date of the HOS restart rule requiring two overnight (1am to 5am) periods in a 34-hour restart. Thus it compared future "illegal" to "legal" restart schedules, though all were legal at the time. Behavioral measures included sleep time, reaction time performance, sleepiness/alertness ratings, and driving performance across two duty cycles and the intervening restart breaks for participating CMV drivers.

Study design: Comparisons were made among the following four groups:
- Group A: 20 drivers with restarts with one nighttime period preceding both of two week-long duty cycles.

- Group B: 5 drivers with restarts with one nighttime period preceding their first duty cycle and with 2+ nighttime periods preceding their second duty cycles.
- Group C: 26 drivers with restarts with 2+ nighttime periods preceding their first duty cycle and with one nighttime period preceding their second duty cycles.
- Group D: 55 drivers with restarts with 2+ nighttime periods preceding both duty cycles.

Primary comparisons were between restart periods containing one overnight (1am to 5am) rest period and those containing 2+ overnight periods. The study's primary statistical approach involved mixed-effects analysis of variance (ANOVA) of 24-hour patterns of outcome measures between subjects as well as within subjects between week-long duty cycles.

Subjects and sample frame: The CMV drivers included 100 men and 6 women, ages 24 to 69. Their commercial driving experience ranged widely around a mean of 12.4 years. Drivers represented several types of trucking operations: 44 were local, 26 were regional, and 36 were over-the-road. The field study encompassed 1,260 total days of data and 414,937 miles of driving. Each driver in the study sample contributed data from two weekly duty cycles.

Predictors (quasi-IVs): The principal quasi-IV was number of overnight (1am to 5am) periods in the restart period; i.e. 1 vs. 2+. TOD was also employed as a quasi-IV in study analyses.

Dependent variables (DVs): Drivers were issued smartphones which they used to enter sleep/wake logs, record caffeine use, take PVTs, and report subjective sleepiness ratings (KSS). Smartphone features such as calling and texting were disabled, and PVT application was blocked during vehicle motion. The PVT test used was a 3-minute version presented via a smartphone app; 355 msec was the criterion for attentional lapses. Drivers' sleep-wakefulness activity was measured with wrist activity monitors. Vehicle-based measurements included SDLP.

Notable controlled variables (CVs): Test administration protocols were standardized. Study testing was estimated to require only about 30 minutes per subject per day.

Notable uncontrolled variables (UCVs): As a quasi-experimental ND study in real operations, the study did not control (manipulate) its principal predictor, number and timing of rest periods. It also did not assign drivers to conditions randomly. Specific relevant UCVs included total restart rest time (see ATRI criticism, below), work and driving times, specific routes, and other operational factors.

Principal study findings: In its report to Congress, FMCSA wrote: "The results of this naturalistic field study indicate that having at least two nighttime periods from 1am until 5am in the restart break helps to mitigate fatigue as measured both objectively and subjectively. This constitutes evidence in support of the efficacy of the new restart rule." In the weekly duty cycle after the restart, drivers whose restarts had only one overnight period:

- Drove and worked more at night, while sleeping more during the day.
- Spent more time driving, with a larger proportion of driving at night.
- Obtained less average daily sleep (6.0 hours vs. 6.2 hours).
- Exhibited more PVT lapses, especially at night. The group averages were 2.0 ± 0.3 lapses (>355 msec) per 3-minute test bout following restarts with one nighttime period, compared to 1.7 ± 0.3 lapses following restarts with 2+ overnights. Note, however, that these averages are both at the alert end of the KSS scale, where 1 is "extremely alert," 2 is "very alert," and 3 is "alert."
- Showed greater SDLPs at night, in the morning, and during the afternoon, but not during the evening. The overall SDLP difference was small (0.1cm) and not statistically significant.
- Greater reported subjective sleepiness per the KSS, especially near the end of their duty cycles (which was early morning for night drivers). However, average end-of-cycle self-ratings never exceeded the scale mid-point of 5.0 for sleepiness for any group or time period.

During the restart period, both groups slept primarily at night and obtained nearly equal daily amounts of sleep: 8.8 hours for one overnight drivers and 8.9 hours for 2+ overnight drivers.

Study limitations & potential improvements:

- The PVT and SDLP data might have supported supplemental case-control analyses. Cases might be defined as PVT high-lapse and/or high SDLP readings. These could have been compared to normal readings matched by TOD and other factors. This might have isolated restart period differences and provided a supplemental validation test of study findings.
- PVT bouts were 3 minutes and the lapse criterion was 355 msec, and tests were administered via Smartphones. A more robust regimen might have been 10-minute bouts and/or use of a 500 msec lapse criterion, and administered using standard, dedicated PVT instrumentation.
- The American Transportation Research Institute (ATRI), a research organization supporting the trucking industry, published a critique of the research (Brewster and Short, 2014). ATRI's principal criticisms, conveyed here to represent their views, included the following:
    - The two groups could have differed significantly in total restart time. By definition, the one overnight group was limited to 52 maximum hours off-duty, whereas there was no upper limit to the 2+ overnight group off-duty hours.
    - A relatively "small sample size and short study duration."
    - The study did not address a separate feature of the new rule; i.e., the restriction of restart use to once per week.
    - Concerns have been raised by other PVT researchers about the "veracity and reproducibility" of the shortened, 3-minute PVT administered via Smartphone.
    - The difference in average number of PVT lapses between the two groups (2.0 vs. 1.7 per session) was statistically but not practically significant.

  - PVTs taken during the restart period had significant effects for TOD but not for group (condition); thus, comparisons were confounded by TOD.
  - The practical significance of key group differences are questionable:
    - The two groups' average post-restart SDLPs differed by just 1mm (1/10 of a cm) and lane position variations were mostly within lanes. Moreover, the overall lane tracking methodology was problematic.
    - The difference in average 24-hour sleep time during the restart period was just 6 minutes (8.8 vs. 8.9 hours).
    - Post-restart average KSS differed by only 0.2 points on the 9-point scale, and both averages (3.1 and 3.3) were between "alert" and "rather alert."
  - Adverse productivity effects of the rule are greater than estimated by FMCSA.
- Another ATRI criticism, supported by this reviewer, is that the study's DVs simply reflect true fatigue differences that would be expected between day drivers and night drivers. Restarts containing only one overnight period may well be associated with greater night driving, which is in turn associated with greater driver fatigue. But fatigue is not itself a safety outcome. Overall, night driving is likely associated with *lower* CMV crash rates than day driving due to reduced traffic conflicts at night. Traffic density affects the likelihood of many more types of crashes than does driver fatigue. ATRI cites FMCSA statistics from the 2011 Motor Carrier Management Information System (MCMIS) showing large truck fatal/injury crash rates to be approximately 60% higher between 6am and 6pm than during the 12 nighttime hours. Thus a rule resulting in shifts toward more day driving likely increases overall CMV crash risks. The relative risk increase is even greater for the public, since daytime crashes are more likely to involve other motorists. Knipling (2009) reached a similar conclusion, though cautioning that available statistics are not definitive due to uncertainties about the representativeness of mileage exposure data.

Principal Citations:

Van Dongen, H. and Mollicone, D. J. *Field Study on the Efficacy of the New Restart Provision for Hours of Service*, FMCSA Report No. RRR-13-058; September 2013.

FMCSA. *Field Study on the Efficacy of the New Restart Provision for Hours of Service Report to Congress*, January 2014.

FMCSA. *Field Study on the Efficacy of the New Restart Provision for Hours of Service: Final Report*, Research Brief, January 2014.

4.14 Effect of Circadian Rhythms and Driving Duration on Fatigue Level and Driving Performance of Professional Drivers (Zhang et al., 2014)

Overview and primary study purpose: This small on-road study examined independent and interacting effects of TOD and hours of driving on several indicators of fatigue. These included a subjective self-assessment measure, the Karolinska Sleepiness Scale (KSS) and two driving performance measures. In spite of several deficiencies, the study illustrates an on-road experimental approach which could be improved and applied more widely.

Study design: Between-subjects experimental design with TOD as the manipulated IV. Five drivers drove an instrumented car for six hours each starting at either 9:00 (morning), 13:00 (afternoon), or 21:00 (night). Driving duration was a quasi-IV for each group.

Subjects and sample frame: Fifteen (15) middle-aged, experienced taxi drivers. Subjects were randomly assigned to one of the three groups. No driver had a known sleep disorder.

Predictors: (IVs, quasi-IVs): The IV was TOD group representing circadian periods (i.e., morning, afternoon, night). Driving duration was analyzed as a quasi-IV.

Dependent variables (DVs):
- Karolinska Sleepiness Scale (KSS) self-ratings based on a 9-point semantic differential scale from 1 (extremely alert) to 9 (extremely sleepy). An observer in the vehicle requested the driver's self-rating every 5 minutes.
- Standard deviation of lane position (SDLP); more details of instrument and measurement criteria provided in paper.
- Steering wheel reversal rate; more details provided in paper.
- Attempts to include eye measures in the study were unsuccessful.

Notable controlled variables (CVs):
- Standardized out-and-back 600-km round trip on the China G70 highway.
- In-vehicle temperature and noise were controlled.

Notable uncontrolled variables (UCVs):
- Time awake (hours since 7:00 awakening) co-varied with both TOD and hours of driving.
- Though the trips were standardized, traffic and weather could vary.

Principal study findings:
- On the KSS, the night group reported the greatest subjective sleepiness, followed by the afternoon group and then the morning group.

- For lane tracking (SDLP), the night group was worst but the other two groups were not significantly different.
- Self-assessed fatigue was greatest during circadian lows (14:00-16:00 and 02:00-0400). Lane tracking was also poor during these periods.
- Self-reported fatigue (KSS) correlated with both TOD and hours of driving, although changes with the latter were not always linear.
- Significant associations were seen between TOD (circadian lows) and SDLP, and between self-rated fatigue (KSS) and SDLP.
- Steering reversals had weaker associations with other fatigue measures than did lane tracking (SDLP). Unexpectedly, the morning group had higher reversals than the other two groups.

Study limitations & potential improvements:

- Small sample (N = 15), between-subjects design. Subjects were taxi drivers, not CMV drivers.
- Limited driving times and lack of full coverage of 24-hour day. Specifically, no driving between 4:00 and 9:00am
- Non-control of time awake as a potential confound to both TOD and driving hours.
- Though the driving was on-road, the overall setting was not naturalistic. An observer in the vehicle requested driver ratings every 5 minutes.
- Subjective driver self-assessments of alertness/sleepiness have weak correlations with objective alertness measures, as shown by several studies cited in this paper. The Zhang study did, however, did find a significant association between KSS scores and SDLP.
- Furthermore, the "demand characteristics" of the testing may have adulterated self-ratings. In the DFAS, drivers were seen as basing their fatigue self-assessments more on their self-expectations ("if I've been driving a long time, I must be tired") than on objective changes.
- KSS scores were averaged across the five taxi drivers in each group, even though the KSS is based on a Likert-like 1-9 ordinal scale, not on an interval or ratio scale.

Principal Citation:

Zhang, H., Yan, X., Wu, C, & Qiu, T.Z. Effect of circadian rhythms and driving duration on fatigue level and driving performance of professional drivers. *Transportation Research Record*, No. 2402, *Truck and Bus Safety; Roundabouts*, 2014, Pp. 19-27.

5. CONCLUSIONS

This paper has presented background facts and behavioral science concepts relevant to research on commercial driver fatigue and HOS rules. It has reviewed and critiqued 20 research studies, encompassing numerous different research designs and methodologies. These include crash investigations naturalistic driving studies of various types, case-control studies, on-road experiments, simulator driving studies, laboratory sleep restriction studies, and surveys. Conclusions are drawn here in two areas: suggested best practices in HOS- and other driver fatigue-related research, and research needs in these same areas. There is some overlap in content among the items, but each articulates a specific idea.

5.1 Suggested Best Practices

The following 16 suggested best practices are based both on innovative ideas from the studies reviewed, and on identified shortcomings. Not all would apply, or be feasible, for every study.

(1) Link dependent variables to defined target crash populations. HOS- and other driver fatigue-related research findings must be extrapolated to the CMV crash population as part of any countermeasure implementation. A pervasive implicit assumption seems to be that decreases in fatigue resulting from a countermeasure will result in proportional decreases in crashes. This is clearly not the case, since the majority of CMV crashes are not discernibly fatigue-related. Both the rigor of research and extrapolation of findings would be improved if the linkages between DVs and the CMV crash population were examined and stated prior to the research. The defined target population of crashes would thus function like a sampling frame or accessible population. The principal examples, going from smaller to larger, are asleep-at-the-wheel (AATW) crashes, crashes where fatigue contributes, and crashes where fatigue is present (as associated factors were defined in the LTCCS). Other target crash populations might be single-vehicle crashes, at-fault crashes, and all crashes. Linkages might not be precise, but they would put research findings in better context and perhaps help in interpretation.

(2) Disaggregate crashes and SCEs by key fatigue-relevant categories. A number of crash dimensions strongly correlate with fatigue incidence. Causal inference about fatigue factors and correlates is strengthened when events are classified by these dimensions to reveal comparisons. Three principal dimensions are:

- *CMV single-vehicle vs. multi-vehicle at-fault vs. multi-vehicle not-at-fault.* These three salient categories differ sharply in fatigue incidence and many other causal factors (Knipling, 2009b, 2011c). Respective LTCCS percentages for fatigue presence were 30%, 14%, and 3%. For AATW as the CR, the percentages were 13%, 1%, and 0%. For many studies, the focus should be on single-vehicle events, perhaps with *post hoc* extrapolation to all crashes.

- *Severity.* Fatigue-related crashes are generally more severe than non-fatigue crashes, and the role of fatigue varies directly with crash severity (e.g., KABCO levels). Thus, factors purported to affect fatigue should ordinarily have greater effects on severe crashes.
- *Roadway type.* Similarly, the role of fatigue varies by roadway type, with freeways and other highways having higher incidence rates than do local roads.

(3) Focus on severe crashes. The above suggestion related to the richness of samples and therefore the strength of causal inference. Another reason for focusing on severe crashes is that that is where the preponderance of human harm resides. Zaloshnja and Miller (2007) estimated that serious crashes in the top levels of the KABCO severity scale (specifically K, A, and B) represented 11% of police-reported large truck crashes but 78% of crash costs, 91% of reduced quality-of-life years, and 92% of lost productivity. Relevance to KAB crashes seems required for any study claiming safety significance. The paramount importance of severe crashes even more true for fatigue-related crashes, which on average are much more severe than non-fatigue-related crashes. The genesis of severe crashes differs in many ways from that of minor crashes (Knipling, 2009; Evans, 2004; FMCSA Analysis Division, 2014). Thus, for both scientific and safety-effectiveness reasons, most research should focus on, or be validated against, the most severe fatigue-related crashes.

(4) Demonstrate construct validity for fatigue measures. Fatigue is a construct, a conceptual variable known (or assumed) to exist but which cannot be directly observed nor measured (Privitera, 2014). Use of constructs (also termed intervening variables; e.g., by Shinar, 2007) is routine in behavioral science. Other examples are anxiety, motivation, resilience, cognition, intelligence, personality, and love. Constructs are operationally defined to permit observation and measurement, but construct validity cannot be assumed or ignored. It must be demonstrated by showing that measurements behave in ways that would be expected if they indeed measure the construct. Demonstrating content validity (i.e., that the contents of the measure correspond to elements of the construct) is closely related and also important. The term "construct" is almost never seen in major driver fatigue studies and most do not explicitly address the construct validity of their measurements. Construct validity should be addressed explicitly in project planning and interpretation.

(5) Employ the best-validated fatigue measures. It follows from the above that the best measurements are likely to be those with the greatest evidence of construct validity. The PVT, SDLP, and PERCLOS are among the most highly-validated fatigue measures. Unfiltered SCEs and crashes are among the least-validated as measures of fatigue. In fact, based on Wiegand et al. (2008; see Section 3.3), the only known applicable study, SCE rate "measures" alertness better than it "measures" fatigue, given that its only positive association is with alertness. Subjective self-assessments of alertness/sleepiness appear to have partial validity, but several studies reviewed have pointed out discrepancies between subjective and objective measures.

Van Dongen and Belenky (2010, P. 56) stated, "This discrepancy between actual impairment and introspective awareness is a common finding in sleep and performance research . . . individuals cannot be relied upon to accurately self-identify fatigue-induced impairment." Even when drivers know they are sleepy, they can't accurately quantify their sleepiness or accurately predict how imminent loss of consciousness might be (Itoi et al., 1993).

(6) Standardize key fatigue measures in major studies. Key fatigue measures should also be standardized in regard to their specific definitions and protocols. PVTs reported in this paper have differed in bout duration (e.g., 10-minute vs. 3-minute), lapse criterion (500 msec vs. 355 msec), and test apparatus (desktop counter display vs. Smartphone). These differences likely affect study results and inferences. Measurement protocols may also vary across studies for lane tracking, steering, eye closures, and observer drowsiness ratings. Measurement standardization across all studies is not warranted or feasible, but measures should be standardized across major studies influencing HOS or other policy decisions.

(7) In naturalistic driving, focus on steady driving periods rather than on SCEs. This paper has questioned the validity of SCEs in relation to serious crashes, and in particular to serious fatigue-related crashes. A better ND-based paradigm would be analysis of driver performance during steady driving periods. A "steady driving period" might be defined as a period (e.g., 5 minutes) in which vehicle speed is constant (with or without use of cruise control) and there is no interaction with other traffic; i.e., "lonely" highway driving. This approach would provide a high level of control to enhance the validity of continuous physiological (e.g., PERCLOS) and performance (e.g., lane deviation) measures. It recalls seminal driving simulator studies from decades ago (e.g., Dingus et al. 1987; Wierwille, 1999) which elucidated driver state and performance changes during sleep-deprived driving on monotonous desert highways. Figure 7 shows fatigue-related deteriorations in driver state and driving performance during steady driving. The data shown is from a simulator study (reported in Knipling and Wierwille, 2004), but the same kinds of data could be obtained in ND.

Performance degradation under such narrow, standardized conditions is not generalizable to crash risk under all conditions, but it does provide a "pure" measure of driver fatigue for associations with HOS parameters. It would also lend itself to within-subjects designs and have far greater external validity than the use of SCEs. ND research on steady driving periods might also reveal new experimental measures of fatigue effects on driving. For example, a declining frequency of mirror glances might be a robust fatigue indicator, as implied by findings of Barr et al. (2011).

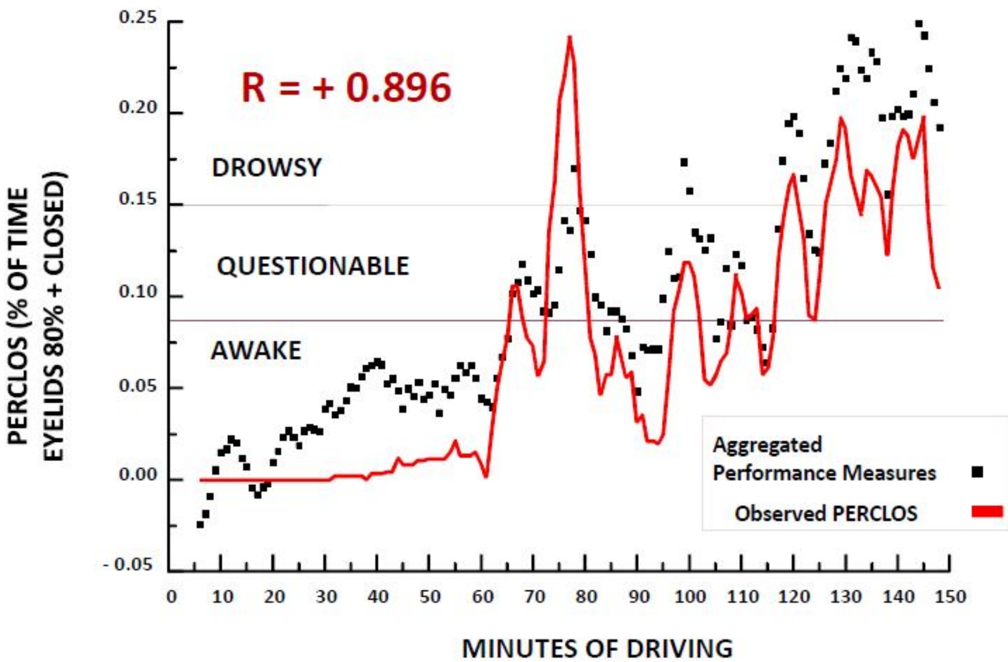

Figure 7. Concurrent, correlated changes in driving performance (mostly lane tracking measures) and eyelid closure (PERCLOS) for a sleep-deprived driver during "steady driving" on a simulator. **Knipling and Wierwille, 1994.**

(8) When possible, perform true experiments. Only true experiments can determine cause-effect relationships unequivocally and to the limits of statistical inference. True experiments require three elements of control: manipulation of an independent variable (e.g., hours of driving) by the experimenter, randomized assignments (e.g., drivers to conditions), and a comparison/control group (Privatera, 2014). Most FMCSA-sponsored driver fatigue research is intended to demonstrate cause-effect relationships between HOS parameters and driver alertness (and/or safety). Quasi-experimental studies are subject to contamination by confounding variables co-varying with nominal IVs, such as when TOD co-varies with an HOS parameter of interest. The validity of causal inference from quasi-experimental studies (e.g., Blanco et al., 2011; Jovanis et al., 2011) is questionable unless there is supplemental evidence of causation. True experiments can be performed in laboratories (e.g., Balkin et al., 2000) but also in real driving (e.g., the FMCSA LCV fatigue study [Section 4.5] and the Zhang et al. 2014 study [Section 4.14]). To reduce costs and maintain external validity, such studies might be performed in large LTL or private fleets with the flexibility to schedule trips in accordance with experimental design requirements. This would provide both a real-world setting and experimental control.

(9) Reduce confounding in quasi-experiments and analysis of their findings. Causal inference from quasi-experiments (e.g., where HOS parameters are quasi-IVs) could be enhanced by applying prior controls on events and exposure. For example, studies could be limited to travel on highways where most fatigue-related crashes occur and confounds are reduced. Post-analysis of events could address confounds and validate findings. For example, events (crashes or SCEs) could be examined to see if they are discernibly fatigue-related. ND permits practically comprehensive data mining. Crash and corresponding exposure data are less detailed, but could still be used to support these methods.

(10) Perform laboratory studies prior to field studies. Several controlled laboratory fatigue studies have been reviewed in this paper. As Van Dongen and Belenky (2010, P. 52) stated in regard to their restart study (see Section 4.12), "Running the study in the laboratory (as opposed to in the field) helped to eliminate environmental confounds, allowed for the use of sensitive laboratory performance measures, simplified the logistics, and moderated the sample size requirement as corroborated by a power calculation performance in advance of the study." In many cases, these advantages outweigh the principal disadvantage of reduced external validity. External validity threats can be reduced by subsequent field studies based on laboratory findings.

(11) When possible, use within-subjects rather than between-subjects designs. There are extreme individual differences in fatigue susceptibility. For example, the single worst of 80 drivers in the DFAS (Wylie et al., 1996) had more drowsy incidents than the 49 least-drowsy drivers in the study combined (Knipling 2009a). Obstructive Sleep Apnea is one strong factor causing these differences, but strong differences are also found among healthy people (Dinges et al., 1998; Van Dongen et al., 2004). Fatigue susceptibility appears to be an enduring individual trait, with wide differences between individuals (Van Dongen et al., 2004). Such subject variability dictates the use of within-subjects designs whenever feasible.

(12) To the extent possible, control or account for time awake as a confounding co-variate of time working or driving. HOS rules regulate continuous working and driving hours, but, probably, the more critical temporal factor affecting alertness is time awake or time since the last main sleep period (Dijk et al., 1992; Rosekind, 2005). The biological "sleep-wake homeostat" contributes to declines in alertness and cognitive functioning with increasing hours awake regardless of ongoing activities. Alertness declines particularly after 14-16 hours of wakefulness (Dawson and Reid, 1997). Most of the 14 studies reviewed in Chapter 4 employed hours working or driving as predictors, but only a few considered time awake, the more likely underlying factor affecting alertness. While HOS rules cannot regulate time awake, rules like the 14-hour tour-of-duty limit are based on human limitations in daily time awake. Associations of alertness or performance with hours driving or working may largely reflect the influence of time awake as a hidden co-variate.

(13) Stratify HOS associations by TOD, and publish the statistics. TOD is a strong confounding variable in almost any attempt to relate HOS parameters to fatigue, or to larger safety outcomes. "A confounding variable is a variable that is not manipulated or controlled by the researcher . . [but which] . . . behaves in a way that is similar to the independent variable and thus, in retrospect, makes it impossible to determine whether the effect [is] . . . due to the independent variable . . ." (Shinar, 2007; P. 26). At least two powerful causes are embedded in TOD. The biological circadian rhythm is among the strongest factors affecting alertness and also the ability to sleep. Thus any measure of fatigue or sleep is likely to vary by TOD. Circadian rhythms also affect crash propensity, but across the 24-hour day their effects are small compared to the effects of traffic density on crash risk. Traffic density affects the likelihood of many more types of crashes than does driver fatigue (Hanowski et al., 2008; Wiegand et al., 2008; Knipling, 2009; Brewster and Short, 2014). It affects the likelihood of most crash types and across the 24-hour day. Another TOD-related factor confound is shifts from driving on local roads to freeways (often occurring early in work shifts) and from freeways to local roads (late in shifts). All of these embedded factors can interact. Accordingly, almost any measures of schedule-related fatigue or safety should be stratified by TOD and presented accordingly in study reports.

TOD statistics have been conspicuously absent from major HOS research in recent years. The 2011 fleet crash case-control by Paul Jovanis et al. analyzed time-on-task associations exhaustively, but published no statistics on crash frequencies or relative rates by TOD. Yet newly published statistics from the same dataset (Chen and Xie, 2015) show a 5-fold range in hourly crash frequencies across the 24-hour day and a 3-fold spike in crashes beginning at 5:00am and extending through the morning rush. Similarly, the Blanco et al. (2011) VTTI ND HOS study presented no TOD statistics on SCE frequencies or rates, even though the previous major truck ND study at VTTI had attributed its principal effects to TOD-related variations in traffic density. Figure 8 shows an example of tabular statistics which could have been derived and presented in both studies. Figure 8's horizontal axis (labeled at the top) shows driving hours (time-on-task), but it could as easily be other HOS parameters. The fraction ***c/e*** is crashes/exposure. For ND studies, ***c/e*** would be SCEs/exposure. Full disclosure of TOD-stratified statistics would elucidate study findings and also allow other researchers to independently analyze and apply study findings.

Hours Driving: Time-of-Day:	1	2	3	4	5	6	7	8	9	10	11	Sum
12:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
1:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
2:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
3:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
4:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
5:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
6:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
7:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
8:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
9:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
10:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
11:00AM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
12:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
1:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
2:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
3:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
4:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
5:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
6:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
7:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
8:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
9:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
10:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
11:00PM:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e
Sum:	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e	c/e

Figure 7. Sample blank time-on-task (hours driving) by time-of-day (TOD) matrix which should be derived and presented to address TOD confounding. c = crashes (or safety-critical events), e = exposure.

(14) When possible, perform case-control comparisons. Almost any study capturing crashes, SCEs, or discrete high-fatigue events could compare those to controls (e.g., non-crashes, non-SCEs) matched by TOD and/or other fatigue-relevant confounds. This would help to isolate HOS parameters of interest such as hours of driving. If subject (driver, carrier, vehicle) characteristics are measured, comparisons of cases and controls would provide estimates of their associated risk.

(15) Set stringent standards for statistical significance, and seek to document practical significance. Achieving statistical significance in a study does not mean that study findings are important or practically applicable to CMV operations. This is especially true for traffic safety studies like those reviewed in this paper. Many of the studies reviewed have had multiple confounding variables and threats to external validity. One way to reduce Type I errors (falsely rejecting a null hypothesis and thus falsely accepting an “effect”) is to set very stringent criterion levels of significance. An even higher standard would be to require a quantification of the practical implications of a finding. For example, Dinges (2014) has modeled PVT lapse durations in terms of vehicle distances traveled while driving.

(16) Model traffic exposure and other non-fatigue safety effects of HOS rule changes. There is currently a debate regarding the safety benefits and possible disbenefits of the new 34-hour restart rule requiring two overnight (1am to 5am) off-duty periods. The summary of Van Dongen and Mollicone (2013; Section 4.13) discusses the issues. The sharpest debate seems to be on whether the fatigue-reduction benefits of the new rule, as suggested by their study, are outweighed by exposure of trucks to greater traffic during daytime driving. An HOS rule truly reducing fatigue could still increase crashes due to such unintended effects. Most CMV crashes are *not* discernibly fatigue-related, but most *are* discernibly traffic-related. No conclusions are drawn here regarding net benefits/disbenefits of the new restart rule, but the fact of the debate demonstrates that traffic-related and other non-fatigue effects of HOS rule changes should be assessed and modeled as part of the decision-making process. Another example is possible overflow parking impacts of rules requiring breaks from driving. Parking shortages create roadway hazards due to congestion in rest areas and the many trucks parked on shoulders, often illegally (Hamilton, 1999). Increased breaks likely reduce fatigue, but it could come at the cost of more rest parking-related collisions. Both potential effects should be assessed.

5.2 Research Needs

Following are 13 fatigue-related research/development needs identified from this review. These research needs could be addressed by future FMCSA-funded research or by any of the many other organizations concerned with driver fatigue. They derive from the same sources as did the suggested best practices above. They include a number of needs which might be considered basic research on the problem, as opposed to specific applied research on HOS parameters.

(1) Perform a video-based crash causation study. Post-crash investigations like the LTCCS have the inherent deficiency of being after-the-fact reconstructions rather than direct observations. ND SCEs are lacking as a fatigue testbed; they have not been validated against serious crashes and do not adequately capture fatigue as a crash cause. Both of these shortcomings could be met by a large study capturing in-vehicle videos of serious crashes and accompanied by LTCCS-like post-crash analysis. The study could also include non-crash case-controls to greatly strengthen causal inference. One would want to further ensure that the gathered crash dataset is representative of a target national crash population (e.g., serious crashes as profiled in GES or similar datasets). Obtaining a large video-based crash sample would probably not be feasible using current ND methods (recall Blanco's 4 crashes in 2,197 SCEs) but might be possible using very large samples already equipped with commercial in-cab video event recorders. The data-capture capabilities of systems such as DriveCam® for crash causation research has been shown (Marburg et al., 2015). Such a "crash video study" would address the weaknesses inherent in both conventional crash investigation (e.g., the LTCCS) and in ND. Post-crash investigations have the inherent deficiency of being after-the-fact reconstructions rather than direct observations. Non-crash SCEs are deficient because they are not validated against serious crashes and they do not adequately capture fatigue. Both of these shortcomings could be met by a large study capturing videos of serious crashes. Such a study could also include LTCCS-like post-crash investigation and comparisons to non-crash case-controls.

(2) Validate and elucidate crash causation model(s) in relation to fatigue. Scientific models are heuristic; they generate testable hypotheses which, when tested, lead to further refinements and elaborations to the models. Two possible crash causation models were presented in Section 1.2. These models both have intuitive appeal but both lack scientific validation. For the Risk-Cause model, research could better describe how driver fatigue works as a risk factor vs. as a proximal cause. If it is a risk factor, how does it interact with other operative risk factors? For the Swiss Cheese model, does fatigue behave quantitatively as one would expect; i.e., increase fatigue → increase crash risk? Does it interact with other risk factor "cheese slices" as one would expect? If so, then fatigue could be said to affect the risks of many kinds of crashes, not just those currently known as fatigue-related.

(3) Develop a multi-component model of fatigue's role in CMV crash risk. Driver fatigue is an element of crash risk, but many other strong elements of crash risk are not known to be fatigue-related. Fatigue can be a principal cause (i.e., AATW as the CR) or it can contribute in different ways. A multi-component model with differentiated fatigue influences might better capture and quantify the overall role of fatigue in crash causation. In such a model, "fatigue" could encompass more than drowsiness *per se*. Fatigue could also encompass attentional lapses and misjudgments. Care should be taken, however, to avoid over-attribution. Numerous factors interact to cause crashes; separate consideration of individual causes like fatigue leads almost inexorably to over-counting. Driver errors occur readily without any known fatigue.

(4) Quantify the role of fatigue-related attentional lapses in CMV crashes. Known fatigue-related crashes are mostly drift-out-of-lane road departures, but another mechanism is fatigue-related lapses resulting in recognition failures. Driver recognition failures were 30% of truck driver at-fault crash involvements in the LTCCS (Starnes, 2006). Rear-end crashes in particular have long been associated with driver inattention, distraction, and other recognition failures, but most have not been considered related to fatigue. Sleep deprivation is known to result in attentional lapses (e.g., on the PVT; Dinges, 2014) but this does not mean that a large percentage of attentional lapses in real driving is fatigue-related. Recall the findings of Barr et al. (2011) that drowsiness and distraction are more opposite than alike. The envisioned study might examine ND SCEs (if analytically linked to real crashes) and compare them to concurrent indicators of fatigue to estimate the percent of recognition failures related to fatigue. This research need overlaps with #3 above but is more focused on attentional lapses.

(5) Determine causal mechanisms underlying reported associations between HOS parameters (e.g., driving hours) and safety outcomes. Blanco et al. (2011) and Jovanis et al. (2011) were two principal studies published by FMCSA and forming the scientific rationale for its HOS rulemaking. Both were quasi-experiments, not true experiments. Both asserted relationships between its events and HOS parameters (hours of driving, cumulative work, and breaks), yet neither study described, classified, nor analyzed its events. Driver fatigue was presumed, but never demonstrated. The search for driver performance and other causal mechanisms within these and similar studies should include control of potential confounds and analysis of events (SCEs or crashes). Control for confounds could be achieved by stratifying data by "competing" factors or by employing targeted case-controls. The strongest HOS-relevant confound is probably time-of-day (TOD). Roadway type, traffic density, and light condition (light vs. dark) are other potential confounds which could be controlled in ND studies. Event (SCE or crash) analysis could include almost every descriptive variable found in crash databases; e.g., conditions of occurrence, event scenarios, number of involved vehicles, associated factors (including fatigue), and CRs including AATW. SCEs could be assessed for CRs, driver avoidance maneuvers, driver drowsiness (e.g., Observer Rating of Drowsiness, PERCLOS), or and other driver behavior visible in videos. Analysis of the events would test the fatigue content and construct validity of study findings. Event analysis would also reveal a wealth of information about fatigue incidents and crashes relevant to other fatigue countermeasures, including technologies, driver monitoring, enforcement, and education. All of the source data for these studies presumably still exists, so these analyses could still be performed today. The TRB Committee on Truck and Bus Safety (ANB70) has recognized this research need; it is articulated more fully on the TRB research need website.

(6) Delineate crash harm resulting from fatigue-related crashes. The preponderance of driver fatigue-related crash harm likely resides in the most severe crashes; e.g., KAB crashes per the KABCO scale. Analysis of known, police-reported fatigue-related crashes in datasets like GES and FARS could delineate their distribution in relation to crash severity and numerous other factors. Reported injuries and fatalities could be used to generate harm distribution estimates. There is widespread agreement that police reports understate the role of fatigue but that does not mean that parametric data from them are unreliable. In fact, police-reported fatigue statistics show impressive consistency with fatigue-related expectations across various parameters. Examples were seen in Massie et al. (1997), Knipling and Wang (1995), and Knipling and Shelton (1999). Delineation of fatigue crash harm could provide justification for its use as a primary fatigue crash problem size metric. In other words, the fatigue crash problem size could be defined based on its percentage of crash harm rather than its percentage of crashes.

(7) Delineate national CMV crash rates by TOD, roadway types, and vehicle type. HOS rule changes are likely to affect CMV exposure (i.e., VMT) by TOD. Industry concerns about the current restart rule (i.e., the requirement for two overnight periods in the 34-hour restart) center on traffic exposure shifts. Weighing the possibly-opposite effects of driver fatigue reductions versus traffic exposure increases requires knowledge of CMV crash rates per VMT by TOD. Statistics on crash numbers and severities by TOD are readily available in major databases such as GES and FARS. These are numerator statistics for the calculation of crash rates. Improved traffic monitoring data on mileage exposure (the denominator) has in recent years been made available by the Federal Highway Administration (FHWA). Telematic data from instrumented trucks is another ready source which was not available in years past. Both numerator and denominator statistics should be stratified by CMV type (e.g., CUT vs. SUT vs. motorcoach) and by roadway type. These statistics would improve the accuracy of predictions of HOS rule effects on "bottom line" safety. The information would also help some fleets to shift their trip times and routes toward safer choices. This research need has been articulated and endorsed by the TRB Truck and Bus Safety Committee (ANB70) and can be found on the TRB research needs database at *http://rns.trb.org/dproject.asp?n=25339*.

(8) Develop methods for improved ND SCE crash and harm representativeness. This paper has questioned the validity of mixed ND SCE datasets in relation to crashes, especially serious crashes. ND SCE datasets contain almost no serious crashes and, worse, no CMV-relevant efforts have linked SCEs to serious crashes analytically. SCEs are qualitatively different from crashes; they are mostly abrupt avoidance responses, while many crashes occur due to the lack of an avoidance response. SCE and crash statistical profiles differ sharply on some key descriptive variables. The gap could be bridged, however, by improving SCE sampling in relation to serious crashes and, especially, by differentially weighting SCEs to match the profiles of serious crashes. The matching could be based on objective crash and SCE characteristics; i.e., established crash descriptors of when, where, and how crashes occur. These are already standard variables in datasets like GES and FARS. So, for example, if 2% of serious CMV crashes were roadway departures at curves on rural highways between midnight and 6am, then SCEs would be sampled and weighted to match this. Further, if 4% of CMV crash *harm* met these same criteria, then corresponding SCEs could be weighted at 4% for a separate harm-linked profile. The TRB Committee on Truck and Bus Safety (ANB70) has recognized this research need; it is articulated more fully on the TRB research need website.

(9) Differentiate the driving effects of time awake from those of time driving and time working. Time awake is well established as a physiological factor in alertness (Krueger, 2004). In almost any CMV driver schedule, driving hours and work hours co-vary with time awake to a high degree. Time awake is probably more critical as a temporal factor affecting alertness (Rosekind, 2005). Few if any studies have clearly distinguished time awake driving effects from time-on-task effects. The two may have different HOS and other fatigue management implications, however. The envisioned research would seek to differentiate the two types of temporal effects and implications for fatigue management. The same study could address unresolved questions about time-on-task fatigue effects; recall, for example, that the DFAS (Wylie et al., 1996) found no significant association between driving hours and driver alertness, and that the LTCCS saw no association between driving or work hours and truck driver fault in crashes (Knipling, 2009b, 2011c).

(10) Develop methods and guidelines for CMV driver sampling. Most driver fatigue studies have involved 100 or fewer CMV driver subjects recruited from a few companies with similar operations at a few geographic locations. The CMV driver population, however, is huge and diverse. There are wide variations in operations types, vehicles, traffic environment, physical job requirements, and other characteristics. CMV driver fatigue susceptibility also varies widely, in part because of the high incidence of obesity and other medical conditions, but also due to "natural" individual differences. Several studies described have used younger, healthy subjects to reduce subject variability, but young drivers are thought to be *more* susceptible to drowsiness, and they likely differ in other ways from CMV drivers. The envisioned study would delineate key CMV driver characteristics which should be the basis for sample development and validation, and suggest other methods to improve sample representativeness. Its applications would extent to other safety topics beyond fatigue. It could also be expanded to encompass sampling of motor carriers and their drivers, a research need already articulated by the TRB Truck and Bus Safety Committee (see *http://rns.trb.org/dproject.asp?n=28338*).

(11) Validate driving simulators as a testbed for driver fatigue studies. Driving simulators offer numerous advantages over real driving as research testbeds. These include subject safety, scenario and test event standardization, repeatability, and sophisticated measurement. On the negative side, simulator sickness (attributed to computer-generated imagery) and the overall fidelity of simulated driving to real driving usually prompt questions of ecological validity. Ecological validity might be especially problematic when driving sessions are of long duration, as is the case in many fatigue studies. Thus, research is needed to validate and improve fatigue-related research using driving simulators.

(12) Assess the health associations with CMV driving-related fatigue. The scope of this paper has not included fatigue effects on health or medical factors affecting fatigue. When drivers do not feel well due to headache, back pain, or whatever, their conditions undoubtedly affect their levels of alertness and fatigue. When drivers take medications (whether prescription of over-the-counter), there are often negative alertness and performance effects (Krueger, 2010; Krueger, Leaman & Bergoffen, 2011). The importance of these issues is acknowledged here, even though the paper has not focused on them. One question within this discussion is the extent to which CMV driving results in long-term health problems (e.g., obesity) versus the extent to which CMV drivers self-select for the often sedentary work and thus bring their health problems to the job. Longitudinal driver studies with non-driver controls (including family members) might differentiate CMV driving job-related health factors (e.g., long hours, excessive sitting, lack of exercise) from non-job-related factors (e.g., genetic/biological predispositions, health-relevant demographics, and social/family norms).

(13) Perform foundational R&D for complementary fatigue management paradigms. This paper has emphasized the difficulties in establishing valid causal links between HOS parameters and CMV crash rates. There are simply too many strong non-fatigue and/or non-HOS-relevant forces affecting CMV crash rates and operating as confounds in research. This is not just a research dilemma – it is a fundamental limitation of HOS rules. HOS are necessary and must be enforced, but the effects of specific rules on overall crash rates are uncertain and perhaps very limited. HOS rules and enforcement are the most obvious and visible countermeasures to fatigue, but there are other approaches. The following approaches already exist but would benefit from additional research and development to make them more effective, more standardized, more acceptable, and more universal in CMV transport:

- Motor carrier fatigue management training and formal certification.
- Alertness-optimizing carrier management of driver work and rest schedules within HOS parameters.
- Driver performance monitoring, including development of standards and tamper-proof devices; e.g.,
  - Continuous in-vehicle; e.g. PERCLOS, SDLP
  - Personal; e.g., activity monitoring watches and associated algorithms.
- Assessments of driver fatigue susceptibility
  - Medical qualifications; e.g., OSA
  - Functional testing of drowsiness susceptibility, perhaps based on physiological indicators.
- Driver history surveillance with exclusions or remediation for critical events; e.g., single-vehicle or other crashes suggesting driver impairment.
- Changes to laws, policies, or regulations to reduce driver and carrier incentives for HOS rule violations or other unsafe practices; e.g., reduction of driver detention (waiting times), driver employment status (e.g., employee vs. independent contractor), and pay policies (e.g., pay method, overtime).

GLOSSARY

Below are selected terms used in this report which might be unfamiliar to some readers. They are defined in the context of CMV safety and consistent with common usage in the field. Although specific reference citations are given in many cases, most of the terms below are widely used and in multiple scientific contexts.

Associated Factors (LTCCS) -- Human, vehicle, or environmental conditions present at the time of the crash. A causal or contributory role was not required. Comparison of associated factors for samples of different types of crashes could lead to causal inferences, however (FMCSA, 2006).

Circadian rhythm – A 24-hour physiological activity and rest cycle that is inherent in almost all animals. Circadian peaks tend to occur in the morning and early evening. There is a dip in the early- to mid-afternoon and a deeper trough during the overnight (very early morning) hours. The timing of the two daily circadian lulls in body physiology, mood, and performance differs slightly from individual to individual, but within a person is resistant to daily alteration.

Circularity – A subtle problem in crash investigation and data analysis, most notably in relation to driver fatigue and schedule factors. For example, when a crash occurs late in a driving shift or driver's work week, the police investigator may attribute the crash to driver fatigue, based in part on the driver's long work hours. Later, crash data statistical analysts note the correlation of fatigue with long work hours, and conclude a causal relationship. Circularity can be avoided by not basing crash data *analysis* on the same factors used to classify them; for example, classifying fatigue crashes using only scenario information (i.e., interviews and the nature of the crash) if the analysis goal is to understand schedule or other temporal factors in fatigue (Knipling, 2009).

Confound (or confound variable) – An unanticipated (or otherwise unaccounted for) variable which could be causing observed changes in measured variables (Privitera, 2014).

Construct (aka hypothetical construct) – A conceptual variable known (or assumed) to exist but which cannot be directly observed or measured. Fatigue, however defined, is a prime example. "Safety" might also be considered a construct since there may be multiple measures of it (Privitera, 2014).

Controlled variable – Factor held constant in a study to reduce confounding of the independent variable. For example, intra-subject comparisons (rather than inter-subject) across conditions in fatigue studies reduce confounding effects of individual differences (Knipling, 2009).

Convenience sampling – Sampling in which subjects are selected because they are easy or convenient to reach and recruit (Privitera, 2014).

Critical Event (CE) – In the LTCCS, the vehicle action or event that put the vehicle or vehicles on a course that made the crash unavoidable (FMCSA, 2006)

Critical Reason (CR) – In the LTCCS, the human, vehicle, or environmental failure leading to the Critical Event and thus to the crash (FMCSA, 2006). Simplistically, it is the immediate or proximal cause of a crash (Knipling, 2009).

Dependent variable (DV) – The variable believed to change in the presence of the IV or other predictor. It is the response shown by humans or other subjects, and the presumed effect in a cause-effect relationship (Privitera, 2014).

Disaggregation – Crash data analysis may be more valid and meaningful when there is separation by major crash subcategories. Important disaggregations for better understanding crash causation include crash severity, truck type, single-vehicle vs. multi-vehicle crash, type of crash (rear-end, lane change, etc.) and divided vs. undivided highway.

Experiment – Scientific method in which an experimenter fully controls specific conditions and subject experiences (i.e., independent variables or IVs) and measures their effects on dependent variables (DVs). To be a true experiment, there are three required elements of control: randomized assignments, manipulation, and a comparison/control group (see below). When properly conducted, experiments demonstrate cause-and-effect; i.e., a single, unambiguous explanation for an observed effect (Privitera, 2014).

Exposure – Vehicle miles traveled (VMT), hours driving, or other denominator to determine crash rates. A pervasive deficiency in national crash databases is lack of exposure data (Knipling, 2009).

External validity – The extent to which observations made in a study generalize beyond the specific manipulations and setting of the study. For example, the external validity of a driving simulator study is the degree to which its findings generalize to real-world driving. Subcategories include:

- Population validity; generalizability to the target population or to different subpopulations
- Ecological validity; generalizability across settings
- Temporal validity; generalizability over time
- Outcome validity; generalizability across different but related DVs (e.g., different measures of alertness or safety) (Privitera, 2014).

Fault/At-Fault – In this paper, the words *fault* and *at-fault* have been used to designate the vehicle/driver assigned the CR (e.g., LTCCS), or whose driver made the critical error. Overwhelmingly, this would also be the vehicle/driver with legal fault in the crash, but the term as used here does not refer to legal fault.

"Harm" – A quantitative measure of the combined human and material loss from traffic crashes based on economic valuation (Zaloshnja and Miller, 2007). Using crash "harm" as a metric permits objective comparisons across different vehicle types, crash types, crash severity levels, and ways of assessing risk (Knipling, 2009).

Hindsight Bias – In crash investigation and naturalistic driving event analysis, this is the tendency to seek an expected or "logical" causal explanation for the crash/event rather than judging it totally objectively (Dilich et al., 2006). Hindsight bias has also been called the *knew-it-all-along effect*. For example, crash reconstructionists investigating run-off-road crashes may tend to look for one of the better known and *expected* causes of such crashes (e.g., speed, slippery roads, fatigue) rather than truly weighing all possible causes and contributing factors objectively. In naturalistic driving data reduction, an observer may tend to rate pre-event driver drowsiness or errors greater knowing that a traffic incident occurred than if there had been no incident (Knipling, 2009).

Hypothetical Construct – An inferred intervening factor or state thought to mediate associations between IVs (or factors conceptualized as IVs) and measured DVs (Shinar, 2007). Fatigue/drowsiness is the hypothetical construct assumed to mediate the relationship between HOS and safety outcomes. A critical question to ask, however, is whether fatigue or some different intervening variable is operating.

Independent variable (IV) – The variable manipulated in an experiment. IVs are often called "treatments" and are seen as the cause in any cause-effect relationship identified through experimentation. In this report, the term IV is used only for variables actually manipulated in an experiment, not for other predictor variables such as "quasi-IVs" in quasi-experiments (to be discussed below) (Privitera, 2014).

Internal consistency – The extent to which different types of measures of a variable are similar. One might consider the internal consistency of different fatigue measures in a study, for example (Privitera, 2014).

Internal validity – The extent to which a design contains sufficient control to demonstrate cause-and-effect. True, well-conducted experiments have high internal validity while non-experiments have no internal validity. The internal validity of a quasi-experiment is intermediate and often uncertain (Privitera, 2014).

Motorcoach – Intercity or charter bus (not a transit or school bus).

Naturalistic driving (ND) – Vehicle research method where vehicles are instrumented with unobtrusive video cameras and various dynamic sensors.

Nonexperimental design – Method in which behaviors/events are observed "as is" without researcher intervention. It may reveal correlations or other associations among variables, but does not demonstrate cause-and-effect (Privitera, 2014).

Nonresponse bias – Sampling bias due to some individuals choosing not to participate. From a different perspective, the same phenomenon is often called *self-selection bias* (Privitera, 2014).

Operational definition – The external manifestation of a construct that is observed and measured (Privitera, 2014).

PERCLOS (Percent Eye Closure) – A well-validated measure of driver drowsiness defined as the percent of time that the eyelids are 80% or more closed (Wierwille, 1999).

Probability sampling – Sampling in which the probability of selecting each individual in a population is known. In most studies, each individual has an equal probability of selection (Privitera, 2014). This is virtually unattainable in CMV driver studies.

Quasi-experimental design – A study structured like an experiment (e.g., for analysis) but where one or more element of control is lacking; e.g., non-random assignments; pre-existing, non-manipulated factor(s); or no comparison/control group. Quasi-experiments do not demonstrate cause-and-effect, but may imply cause-and-effect. Subtypes include:

- One-group designs (e.g., pre- and post-test)
- Time-series designs (e.g., series of tests carried out over days)
- Developmental (e.g., longitudinal)
- Non-equivalent control groups (Privitera, 2014).

Quasi-independent variable (quasi-IV) – A variable treated as an IV but which includes pre-existing, non-manipulated traits (e.g., gender, health status) and where assignment to conditions is not random (Privitera, 2014).

Reliability – Consistency, stability, or repeatability of one or more measures or observations. Reliability may be defined and/or measured differently in different studies; e.g., inter- versus intra-rater reliability (Privitera, 2014).

Representative sample – One in which the key characteristics of the sample correspond to those of the target population (Privitera, 2014).

Risk factor – Any factor – driver, vehicle, environmental, carrier – operative prior to a crash and affecting crash probability.

Sampling (selection) bias – Sampling where certain individuals are favored over others, thus threatening study validity (Privitera, 2014).

Sampling error – Random variations in sample characteristics which may threaten study validity (Privitera, 2014).

Sampling frame (accessible population) – The portion of the target population that can be clearly identified or sampled from (Privitera, 2014).

Sleep hygiene – The collection of behavioral health habits that drivers and others can adopt to maintain or improve their personal alertness, safety, health, and happiness (Knipling, 2009).

Stratified random sampling – Sampling in which the population is first divided into subgroups (strata) and there is then random sampling from those subgroups. The LTCCS and other DOT crash data systems (e.g., General Estimates System or GES) have employed stratified random sampling (Privitera, 2014).

Target population – All members of a group of interest; e.g., all CMV driver, all CMV drivers covered by a specific HOS rule (Privitera, 2014).

Traits vs. states – Traits are long-term personal characteristics (e.g., medical conditions, personality), whereas states are short-term characteristics (e.g., alertness level due to recent sleep, moods). (Knipling, 2009)

Truck – Unless otherwise stated, “trucks” refers to large trucks; i.e., heavy vehicles with a Gross Vehicle Weight Rating (GVWR) of 10,000 lbs. or greater. The two major configurations of large trucks are combination-unit trucks (CUTs, generally tractor-semitrailers) and single-unit trucks (SUTs, also called straight trucks). The distinction between these two subtypes is important because they have different physical characteristics and operational uses, and thus have different crash profiles. Light trucks (e.g., pickup trucks, vans) are not included as “trucks” per this definition nor in most statistics on truck crashes.

Uncontrolled variable – Factor not held constant which could potentially confound the effects of an IV. For example, time-of-day, if uncontrolled, is a potential confound to time-on-task effects, and vice versa (Knipling, 2009).

Validity – The extent to which a measurement of a variable or construct actually measures what is purports to measure. Four types are important and relevant:

- *Face validity.* Does the measure *appear* to measure the construct?
- *Construct validity.* Does the measure actually measure the construct?
- *Criterion*-related validity. Does the measure predict or correlate with an expected outcome?
- *Content validity.* Do the contents of the measure represent the features of the construct? (Privitera, 2014)

Variable – Any value or characteristic that can change from one person to another or one situation to another (Privitera, 2014).

Workload – Mental and physical effort required to perform a task such as driving. “Work” refers primarily to the *mental* tasks of driving – perceiving, identifying crash threats, deciding, and performing. Activities that increase workload (e.g., operating controls, talking on a cell phone) reduce available resources for attention to the road and traffic (Knipling, 2009).

CITED REFERENCES

Balkin, T.J., Thorne, D., Sing, H., Thomas, M., Redmond, D.P., Wesensten, N., Russo, M., Williams, J., Hall, S., & Belenky, G.L., *Effects of Sleep Schedules on Commercial Motor Vehicle Driver Performance*, FMCSA Technical Report No. DOT-MC-00-133, U.S. Department of Transportation, Washington, DC, 2000.

Barr, L.C., Yang, D., Hanowski, R. J., and Olson, R.. *An Assessment of Driver Drowsiness, Distraction, and Performance in a Naturalistic Setting*. FMCSA-RRR-11-010, February, 2011.

Belenky, G. Jackson, M. L., Tompkins, L., Satterfield, B., and Bender. A. *Investigation of the Effects of Split Sleep Schedules on Commercial Vehicle Driver Safety and Health*. FMCSA Report No. FMCSA-RRR-12-003, December 2012.

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.

Blower, D.F. The relative contribution of truck drivers and passenger-vehicle drivers to truck/passenger-vehicle traffic crashes. UMTRI Research Review, UMTRI, Ann Arbor, MI. Vol. 30, No. 2, pp. 1-15, Apr-June 1999.

Brewster, R. and Short, J. *Technical Memorandum: Assessment of the FMCSA Naturalistic Field Study on Hours-of-Service Restart Provisions*. ATRI. April 2014

Burks, S.V., M. Belzer, Q. Kwan, S. Pratt, and S. Shackelford. *Trucking 101; an Industry Primer*, TRB Circular E-C146, Trucking Industry Research Committee (AT060), December 2010.

Chen, C. & Xie, Y. Effects of driving hours and time of day on large truck safety based on multilevel discrete-time survival analysis. TRB 2015 Annual Meeting, Paper 15-5331, 2015.

Council, F.M., Harkey, D.L., Khattak, A.J., and Mohamedshah, Y.M. Examination of “fault,” “unsafe driving acts,” and “total harm” in car-truck collisions. *Transportation Research Record 1830*, Pp. 63-71, TRB, 2003. (Also see FHWA Summary Report FHWA-HRT-04-085, HRDS-06/07-04.

Craft, R. 2008 National Truck and Bus Crash Picture. FMCSA Webinar, Feb. 17, 2010.

Dawson, D., Noy, Y.I., Harma, M., Akerstedt, T., and Belenky, G. Modelling fatigue and the use of fatigue models in work settings. Accident Analysis and Prevention 43, Pp. 549-564, 2011.

Dawson, D., Reid, K., Fatigue, alcohol and performance impairment. *Nature,* 388, 235, 1997.

Dijk D.J., Duffy J.F., and Czeisler C.A. Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *J Sleep Res.* 1:112-117, 1992.

Dilich, M., Kopernik, D., & Goebelbecker, J. Hindsight judgment of driver fault in traffic accident analysis; misusing the science of accident reconstruction. *Trans Res Record* No 1980, TRB, Pp. 1-7, 2006.

Dinges, D.F. What is drowsy driving and what causes it? Presentation at the NTSB Forum *Overcoming the Dangers of Drowsy Driving*. Washington DC, October 2014.

Dinges, D. F., Mallis, M.M., Maislin, G.M., and Powell, J.W. *Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and the Basis for Alertness Management*. NHTSA Report No. DOT HS 808 762, April, 1998.

Dingus, T.A., Hardee, H.L., & Wierwille, W.W. Development of models for on-board detection of driver impairment. *Accident Analysis and Prevention*. 19, No. 4, Pp. 271-283, 1987.

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.

Evans, L. *Traffic Safety*. Science Serving Society, Bloomfield Hills, MI. ISBN 0-9754871-0-8, 2004.

FHWA. Converting two-lane highways to four-lane can reduce crashes. *Research & Technology Transporter*. FHWA-RD-00-015, July 2000.

FHWA OMC (Office of Motor Carriers). Driver-related factors in crashes between large trucks and passenger vehicles, Analysis Brief, MCRT-99-011, April 1999.

FMCSA. Stress and fatigue effects of driving long-combination vehicles. Tech Brief. No. FMCSA-MCRT-00-012. 2000.

FMCSA. *Report to Congress on the Large Truck Crash Causation Study*. MC-R/MC-RRA, March 2006.

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.

FMCSA. Pocket Guide to Large Truck and Bus Statistics. FMCSA Analysis Division. *The Large Truck Crash Causation Study*. FMCSA-RRA-07-017, July 2007.

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.

FMCSA. *An Evaluation of Emerging Driver Fatigue Detection Measures and Technologies*. Tech Brief FMCSA-RRR-09-066, June 2009.

FMCSA. *Investigation into Motor Carrier Practices to Achieve Optimal Commercial Motor Vehicle Driver Performance, Phase I*, Tech Brief No. RRR-10-006; December 2010.

FMCSA. *Duration Restart Period Needed to Recycle with Optimal Performance, Phase II*, Tech Brief No. RRR-10-062-TB; December 2010.

FMCSA. *Investigation of the Effects of Split Sleep Schedules on Commercial Vehicle Driver Safety and Health*. Research Brief, December 2012.

FMCSA Analysis Division. *The Large Truck Crash Causation Study*. FMCSA-RRA-07-017, July 2007.

FMCSA Analysis Division. *Large Truck Crash Facts 2011*. FMCSA-RRA-13-049, October 2013.

FMCSA Analysis Division. *Large Truck Crash Facts 2012*. FMCSA-RRA-14-004, June 2014.

FMCSA & NHTSA. *Large Truck Crash Causation Study: Analytic User's Manual*. Washington, DC: U.S. DOT. 2006.

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

Hamilton, P. *Rest Area Forum: Summary of Proceedings*, Federal Highway Administration Report No. FHWA-RD-00-034, December, 1999.

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

Hanowski, R. J., Wierwille, W. W., Garness, S. A., and Dingus, T. A. *Impact of Local/Short Haul Operations on Driver Fatigue.* Final Report No. DOT-MC-00-203. Washington, DC: U.S. Department of Transportation, Federal Motor Carriers Safety Administration, September, 2000.

Hernán, M. S. & Robins, J. M. *Causal Inference*, draft book, May 14, 2014.

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.

Itoi, A., Cilveti, R., Voth, M., Bezalel, D., Hyde, P., Gupta, A., and Dement, W.C. Can Drivers Avoid Falling Asleep at the Wheel? Relationship Between Awareness of Sleepiness and Ability to Predict Sleep Onset. AAA Foundation for Traffic Safety, February 8, 1993.

Jones, I.S. and Stein, H.S. *Effect of Driver Hours of Service on Tractor-Trailer Crash Involvement.* IIHS Report. September 1987.

Jones, I.S. and Stein, H.S. Defective equipment and tractor-trailer crash involvement. *Accident Analysis and Prevention*. Vol. 21, No. 5, Pp. 469-481, 1989.

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.

Knipling, R.R. *Safety for the Long Haul; Large Truck Crash Risk, Causation, & Prevention.* American Trucking Associations. ISBN 978-0-692-00073-1, 2009a.

Knipling, R.R. Three large truck crash categories: what they tell us about crash causation. Proceedings of the *Driving Assessment 2009* conference, Pp. 31-37, Big Sky, Montana, June, 2009b.

Knipling, R.R. Naturalistic driving events: no harm, no foul, no validity. *Driving Assessment 2015*, Salt Lake City UT, June 22-25, 2015.

Knipling, R.R. Peer Review Critique of VTTI Study: *The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operation.* Critique placed on the FMCSA Hours-of-Service (HOS) rulemaking docket (FMCSA-2004-19608), May 2011a.

Knipling, R.R. Peer Review Critique of Penn State Study: *Hours of Service and Driver Fatigue: Driver Characteristics Research.* Critique placed on the FMCSA Hours-of-Service (HOS) rulemaking docket (FMCSA-2004-19608), May 2011b.

Knipling, R.R. The Good, the Bad, and the Ugly: Three Large Truck Crash Categories and What They Tell Us About Driver Fatigue. Paper placed on the FMCSA Hours-of-Service (HOS) rulemaking docket (FMCSA-2004-19608), May 2011c.

Knipling, R.R. & Bocanegra, J. *Comparison of Combination-Unit Truck and Single-Unit Truck Statistics from the LTCCS*. FMCSA & Volpe Center Project report. Contract No. DTRS57-04-D-30043. 2008.

Knipling, R.R. and Nelson, K.C. *Safety Management in Small Motor Carriers*. CTBSSP Synthesis 22, TRB, ISBN 978-0-309-22340-9, http://www.trb.org/Publications/PubsCTBSSPSynthesisReports.aspx, 2011.

Knipling, R.R. and Shelton, T.T. (1999). Problem size assessment: large truck crashes related primarily to driver fatigue. *Proceedings of the Second International Large Truck Safety Symposium,* E01-2510-002-00, University of Tennessee Transportation Center, Knoxville, Pp. 3-12, October 6-8, 1999.

Knipling, R.R., and Wang, J.S. *Crashes and Fatalities Related to Driver Drowsiness/Fatigue.* NHTSA Research Note, 1994.

Knipling, R.R. and Wang, J.S. Revised estimates of the U.S. drowsy driver crash problem size based on General Estimates System case reviews. *39th Annual Proceedings, Association for the Advancement of Automotive Medicine*, Chicago, October, 1995.

Knipling, R.R. and Wierwille, W.W., Vehicle-based drowsy driver detection: current status and future prospects. *Proceedings of the IVHS America 1994 Annual Meeting,* Pp. 245-256, Atlanta, April 17-20, 1994.

Kononov, J., Lyon, C., and Allery, B.K. Relating flow, speed and density of urban freeways to functional form of an SPF [Safety Performance Function], Paper 11-2070, *Transportation Research Board Annual Meeting*, 2011.

Krueger, G.P. *Technologies and Methods for Monitoring Driver Alertness and Detecting Driver Fatigue: A Review Applicable to Long-Haul Truck Driving*. Unpublished report for ATRI and FMCSA. June 2004.

Krueger, G.P., Psychoactive medications, stimulants, hypnotics, and nutritional aids: Effects on driving alertness and performance. *Journal of the Washington Academy of Sciences*, Vol. 96, No. 4, 2010, pp. 51–85, 2010.

Krueger, G.P., Leaman, H.M. & Bergoffen, G. Effects of psychoactive chemicals on commercial driver health and performance: Stimulants, hypnotics, nutritional, and other supplements. TRB Commercial Truck and Bus Safety Synthesis Program (CTBSSP) Report No. 19. Washington, DC: National Academies' Transportation Research Board. June 2011.

Marburg, T.L., Hickman, J.S., and Hanowski, R.J. Common data elements in the large truck causation study and commercially available onboard monitoring systems. Presentation 15-3245 at 2015 TRB Annual Meeting, Washington DC, January 2015.

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.

McCartt, A.T., Hellinga, L.A., & Soloman, M.G. Work schedules before and after 2004 HOS rule change and predictors of reported rule violations in 2004: survey of long-distance truck drivers. *Proceedings of the 2005 Truck & Bus Safety & Security Symposium*, Alexandria, VA, November 14-16, 2005.

McCartt, A.T., Hellinga, L.A., & Soloman, M.G. Work schedules of long-distance truck drivers before and after 2004 HOS rule change. *Traffic Injury Prevention,* 9:201-210, 2008.

McCartt, A.T, Rohrbaugh, J.W., Hammer, M.C., & Fuller, S.Z. Factors associated with falling asleep at the wheel among long-distance truck drivers. *Accident Analysis and Prevention*, 32, Pp. 493-504, 2000.

Miller, J.C. Detecting fatigue: lessons learned. Presentation to *International Congress of Aviation and Space Medicine*, 62nd Annual Meeting, Mexico City, Oct 12-16, 2014.

Mitler , M.M., Miller, J.C.,Lipsitz, J.J., Wash, J.K., Wylie, C.D. The sleep of long-haul truck drivers. *New England Journal of Medicine*, vol. 337: 755-761, 1997.

Moore-Ede, M. *The Twenty-Four Hour Society*. Addison-Wesley Publishing Co., ISBN 0-201-57711-9, 1993.

NTSB. *Safety Study: Fatigue, Alcohol, Other Drugs, and Medical Factors in Fatal-to-the-Driver Heavy Truck Crashes*. Report No. NTSB/SS-90/02. 1990.

Olson, R.L., Hanowski, R.J., Hickman, J.S., & Bocanegra, J. (2009). *Driver Distraction in Commercial Vehicle Operations* (Report No. FMCSA-RRR-09-042). Washington, DC: USDOT, FMCSA. September, 2009.

Olson, R.L., Hickman, J.S., Knipling, R.R., Hanowski, R.J., and Carroll, R.J. Factors and driving errors associated with fatigue in a naturalistic study of commercial drivers. Paper and presentation in preparation for the *Fatigue Management in Transportation Operations International Conference*, Seattle, September 11-15, 2005.

Orris, P., Buchanan, S., Smiley, A., Davis, D., Dinges, D. and Bergoffen, G. *Synthesis Report #9: Literature Review on Health and Fatigue Issues Associated with CMV Driver Hours of Work.* TRB CTBSSP. ISSN 1544-6808, ISBN 0-309-08826-7, 2005.

Privitera, G. J. *Research Methods for the Behavioral Sciences,* Sage Publications, Inc., ISBN 978-1-4129-7511-7, 2014.

Reason, J. *Human Error*. Cambridge Univ. Press, ISBN 0-521-30669-8, 1990.

Rosekind, M.R. *Managing Safety, Alertness and Performance through Federal Hours-of-Service Regulations: Opportunities and Challenges*. Alertness Solutions. FMCSA rulemaking docket #FMCSA-2004-19608. 2005.

Shinar, D. *Traffic Safety and Human Behavior*. Elsevier. Amsterdam. ISBN 978-0-08-0450029-2, 2007.

Short, J., Boyle, L., Shackelford, S., Inderbitzen, R.E., and Bergoffen, G. *Synthesis 14: The Role of Safety Culture in Preventing Commercial Motor Vehicle Crashes*. TRB Commercial Truck & Bus Synthesis Program, ISSN 1544-6808, ISBN 978-0-309-09891-5, 2007.

Starnes, M. *LTCCS: An Initial Overview*. NHTSA National Center for Statistics & Analysis, DOTR HS 810 646, August 2006.

Tefft BC. Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999-2008. *Accident Analysis & Prevention*, 45(1): 180-186, 2012.

Tefft, B.C. *Prevalence of Motor Vehicle Crashes Involving Drowsy Drivers, United States, 2009-2013*, AAA Foundation for Traffic Safety, 2014.

Thiffault, P. *Addressing Human Factors in the Motor Carrier Industry in Canada,* Canadian Council of Motor Transport Administration, May 2011.

Transportation Research Board Committee on Truck & Bus Safety (ANB70). Research Needs Statements available on the TRB website (*http://rns.trb.org*):

- Toward Naturalistic Driving Crash Representativeness (23-2015)
- Driver Performance and Other Causal Mechanisms in Quasi-Experimental Hours-of-Service (HOS) Studies (24-2015).

Van Dongen, H., Baynard, M.D., Maislin, G., & Dinges, D.F. Systematic inter-individual differences in neurobehavioral impairment from sleep loss: Evidence of trait-like differential vulnerability. *Sleep*, 27(3), Pp. 423-433, 2004.

Van Dongen, H. and Belenky, G. *Investigation into Motor Carrier Practices to Achieve Optimal Commercial Motor Vehicle Driver Performance, Phase I*, FMCSA Report No. RRR-10-005; December 2010.

Van Dongen, H. and Mollicone, D. J. *Field Study on the Efficacy of the New Restart Provision for Hours of Service Report to Congress*, FMCSA Report No. RRR-13-058; September 2013.

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

Wiegand, D.M., Hanowski, R.J., McDonald, S.E. "Commercial drivers' health: A naturalistic study of body mass index, fatigue, and involvement in safety-critical events. *Traffic Injury Prevention* 10: 573-579, 2009.

Wierwille, W.W. Historical perspective on slow eyelid closure: whence PERCLOS? Washington, DC: Federal Highway Administration (FHWA) Report No. FHWA-MC-99-136, *Ocular Measures of Driver Alertness Technical Conference Proceedings*, Pp. 31-53, 1999.

Wierwille, W.W. and Ellsworth, L.A. Evaluation of driver drowsiness by trained observers. *Accident Analysis and Prevention*. Vol. 26, No. 5, Pp. 571-581, 1994.

Wylie, C.D., Shultz, T., Miller, J.C., Mitler, M.M., & Mackie, R.R., *Commercial Motor Vehicle Driver Fatigue and Alertness Study*, Federal Highway Administration, U.S. Department of Transportation, Washington, DC, 1996.

Zaloshnja, E. and Miller, T. *Revised Costs of Large Truck- and Bus-Involved Crashes*. Pacific Institute for Research & Evaluation, Final Report, FMCSA Contract # DTMC75-01-P-00046, November, 2002.

Zaloshnja, E. and Miller, T. *Unit Costs of Medium & Heavy Truck Crashes*. Final Report, Pacific Institute for Research & Evaluation for FMCSA, available at http://ai.volpe.dot.gov/carrierresearchresults/pdfs/crash%20costs%202006.pdf, March 2007.

Zhang., H., Yan, X., Wu, C, & Qiu, T.Z. Effect of circadian rhythms and driving duration on fatigue level and driving performance of professional drivers. *Transportation Research Record*, No. 2402, *Truck and Bus Safety; Roundabouts*, 2014, Pp. 19-27.

Author Contact:
Ronald R. Knipling
President, *Safety for the Long Haul Inc.*
5059 North 36th Street
Arlington, VA 22207-2946
(703) 533-2895
rknipling@verizon.net
www.safetyforthelonghaul.com