# Naturalistic Driving Safety-Critical Event Datasets: What Exactly Do They Represent?

ANB70 Research Needs Statement 10/27/18

Available online at: https://rns.trb.org/details/dproject.aspx?n=42638

## Background/Description

This research needs statement describes an analytic project to identify and specify, as precisely as possible, the large truck target crash population represented by Naturalistic Diving (ND) Mixed Safety-Critical Event (M-SCE) datasets.

The goal of truck safety research is to understand causes and to identify interventions relevant to crash harm; i.e., fatalities, injuries, and damage resulting from crashes. Any research method used to study truck safety should be closely and plausibly linked to harm from real world crashes. More specifically, the representativeness of research dependent variables (i.e., measures of effects) in relation to definable harmful crashes should be verifiable. In other words, it should be possible to quantitatively extrapolate research dependent measures to real-world measures of harm. This might include benefits from safety enhancements (e.g., technologies) or from degradations due to adverse conditions (e.g., excessive work hours). Such extrapolations are most crucial in regulatory or other cost-benefit analyses.

The ND M-SCE methodology has been dominant over the past decade in truck safety research funded and managed by the Federal Motor Carrier Safety Administration (FMCSA). This has been particularly true in research relating to driver fatigue and HOS research (e.g., Hanowski et al., 2008; Blanco et al., 2011; FMCSA, 2015; Blanco et al., 2016). FMCSA's website indicates that it spends nearly one million dollars annually to fund this type of data collection. M-SCE datasets consist mostly of driver evasive maneuvers or other aberrant dynamic events captured using videos and other instrumentation. ND researchers decide on the types of SCEs (e.g., hard-brakings, swerves, lane breaks) to be included in M-SCE datasets. They also control event type proportions by adjusting dynamic thresholds for each type. This method raises external validity questions, however, because there are no empirical, analytical, or *post hoc* verified links between M-SCE dataset composition and actual harmful event populations. Two such populations are serious injury crashes and asleep-at-the-wheel crashes (Knipling, 2017a & 2017b). No one questions the advantages or the validity of *individual* ND event recordings featuring multiple videos and an array of dynamic event recordings. A video with accompanying dynamic data enables a more accurate causal assessment than that possible through post-crash interviews and reconstructions. Questions are raised, however, about whether *aggregated* M-SCE datasets relate to a discernible population of harmful crashes. In other words, can M-SCE statistics be generalized to defined real-world outcomes?

The suggested research would thoroughly analyze M-SCE dataset composition and that of various known target crash populations, seeking the best matches based on important objective event characteristics. An "important" event characteristic is one known to be correlated with crash causes. "Important event characteristics" are also known as *auxiliary variables* in survey research (TRB ANB20, 2016). Gender, age, and race are three common auxiliary variables in survey research. Every survey has to reasonably match its target population in gender, age, and race. Number of involved vehicles is perhaps the simplest crash example; single- and multi-vehicle crashes have known significant differences in crash causal profiles (Knipling, 2009, 2017b) just as men and women have known differences in political/social opinions. Therefore, a significant difference between the surrogate dataset and the target crash population in proportion of single- versus multi-vehicle events would likely invalidate the surrogate dataset's causal profiles in relation to crashes, just like an opinion survey sample with a lopsided gender distribution.

Concerns about ND M-SCE external validity are based on at least three troubling aspects of the methodology and resulting data. These include the following:

1. **No sampling from population**. Unlike most crash datasets (and indeed most empirical safety and health research), M-SCE datasets are not sampled from target problem populations. The Large Truck Crash Causation Study (LTCCS), for example, was a stratified random sample of U.S. serious large truck crashes, specifically K, A, and B crashes on the KABCO scale (FMCSA, 2006). Therefore LTCCS data represents this harmful crash populations within its sampling error and the reliability of its sampling methodology. M-SCE datasets are not drawn from a target population or constructed using a sampling algorithm. Rather they are constructed judgmentally without an analytic link to a target population.

2. **Tenuous theoretical foundation**. The historic roots of the M-SCE paradigm are in the writings of H. W. Heinrich, a mid-20th century industrial engineer. Based on studies of industrial accidents (not traffic crashes), Heinrich formulated the premise that "the predominant causes of no-injury accidents are, in average cases, identical with the predominant causes of major injuries, and incidentally of minor injuries as well" (Heinrich, 1941). The truth of this assumption has been acknowledged by ND practitioners as being essential to the validity of ND near-crashes and other non-crash SCEs in relation to crashes (Guo et al., 2010a & 2010b). It is still espoused in ND reports; e.g., "The non-crash events [SCEs] were operationally defined for this study as having elements identical to a crash scenario, with the exception that a successful evasive maneuver was also present" (Blanco et al., 2016, P. 184). Within this statement is the implicit assumption that crashes *avoided* (i.e., SCEs) are causally representative of crashes *which actually occur*.

The Heinrich assumption of homogeneous causal mechanisms has been disavowed by the National Safety Council in relation to industrial accidents, Heinrich's principal domain (Manuele, 2014; see also Dunlop, 2013). In traffic safety, Knipling (2017b) has published numerous examples of the heterogeneity of causal factors in both heavy truck and light vehicle crashes. Crash causal profiles vary markedly as a function of when, where, and how they occur (i.e., auxiliary variables). "Horizontal" heterogeneity refers to the variety of scenarios seen within any crash severity level. For example, the two major categories of at-fault crash involvements, single- and multi-vehicle, have sharply disparate causal profiles. Vertical heterogeneity is seen in the fact that fatigue and asleep-at-the-wheel are roughly five times greater in fatal crashes as in minor ones (Tefft, 2014). Heterogeneity applies to many other crash factors besides fatigue, and is the general rule. These consistent findings seem to refute the Heinrich assumption (Knipling, 2017b).

3. **Objective characteristics discrepant from ostensible target populations**. Although FMCSA has specified no target crash population(s) for its M-SCE datasets, it seems implicit that top candidates would include actual truck crashes or truck driver fatigue-related crashes in the case of studies relevant to fatigue management or HOS rules. Yet there are sharp discrepancies between SCE datasets and those of these potential target crash populations in regard to when, where, and how they occur. Some of these discrepancies are presented below under the discussion of possible candidate target populations.

Specific candidate target crash populations include the following ten. None of these can be considered *a priori* choices, since none of them were targeted through sampling or other analytic links prior to M-SCE dataset construction.

- **"Safety Impacts"**. In its 2015 HOS "Restart" study plan, FMCSA stated that its primary measure of "safety impacts" would be ". . . the number of SCEs captured via the OBMSs [Onboard Safety Monitoring Systems]. These include electronically-recorded hard brakes, hard accelerations, swerves, contact with other objects, and driving in excess of posted speed limits." It did not define this safety metric further in relation to real-world outcomes. In delineating the extent and pervasiveness of crash heterogeneity, Knipling (2017b) concluded that abstract phrases like "safety impact" have no definite meaning without an accompanying crash population referent. That's because specific designated crash target populations vary widely in their causal profiles. The same might be said of "risk." "Risk" has no specific or quantitative meaning without answering, "Risk of what outcome?"
- **Crash Risk**. A 2016 report of ND M-SCE results (Blanco et al., 2016) stated that, "The main objective of this on-road study was to collect ND data that could be used to investigate issues related to CMV crash risk." And, as cited above, the study conceptualized non-crash SCEs as equivalent to crash scenarios without actual impacts. However, the study presented no objective comparisons of truck M-SCE data to truck crashes. Only five of the 2,889 SCEs

reported were actual impacts, so no reliable comparisons were possible within the study. A top-level comparison to officially reported crashes (i.e., those meeting state reporting thresholds and reported nationally) revealed lopsided differences. Eighty-one percent (81%) of Blanco's SCEs involved only the truck; for police-reported truck crash involvements nationally, this percentage is 21% (FMCSA, 2018). Large truck single- and multi-vehicle crash involvements differ starkly in their causal profiles (Starnes, 2006; Knipling, 2017b). A dataset this different from police-reported crashes on the dimension of number of vehicles has doubtful representativeness in relation to these crashes. Numerous other such differences could be cited (e.g., Knipling, 2015; TRB ANB70, 2015).

- **Serious/Fatal Crash Risk**. If SCEs are unrepresentative in relation to all police-reported crashes, one would expect them to be even more unrepresentative in relation to the most severe crashes. Numerous differences are seen between police-reported property damage only (PDO) crashes and more serious crashes (Blincoe et al., 2015; TRB ANB70 2015). One would expect SCEs and serious crashes to be even more different in their causation. Nonetheless, an assumption made in HOS regulatory impact analyses is that research findings from ND studies can be extrapolated directly to fatal crashes (e.g., FMCSA 2011, P.6-6) for the purpose of benefits estimation. Thus it would seem particularly important to assess the SCE-serious crash relationship even though it is likely to be weak.
- **Crash Harm Risk**. "Harm" is an analytic metric intended to capture and measure all negative consequences of crashes (Blinco et al., 2015). Economic cost elements include damage, traffic congestion, medical, legal, insurance, and lost income. Total societal harm includes economic loss but adds monetary valuations of lost quality-of-life. Harm risk is conceptually the ideal dependent measures of traffic safety studies since it captures human impacts. Crash harm risk from any factor will likely correlate highly with serious/fatal crash risk, since 80-90% of large truck crash harm resides at these levels (Zaloshnja and Miller, 2007). A conceptual and analytic complication is that harm is not tabulated based on crash counts but rather on counts combined with data on the consequences of crashes of various severities.
- **Driver Fatigue**. HOS rule development has been a principal application of ND M-SCE studies. Therefore one assumes that the government's intention is to represent driver fatigue or drowsiness. However, a 2008 Virginia Tech Transportation Institute (VTTI) comparison of driver drowsiness in SCEs in comparison to that of baseline controls found markedly *less* drowsiness in SCEs (Wiegand et al., 2008). In 2011, FMCSA changed its recipe for M-SCE datasets to ostensibly increase their relevance to driver fatigue. Specifically, 1,118 unintentional lane deviations were added to an unspecified mix of 1,075 other types of events because, "unintentional lane deviations provide a reliable indicator of fatigue" (Blanco et al., 2011; P.28). No validation of the 2011 M-SCE dataset was provided. A prior VTTI study (Olson et al., 2009) found 78% of lane deviations to be distraction-related; i.e., drivers were performing tertiary tasks. Later publication in 2016 of causal data from the same Blanco et al data collection indicated that only 8.9% of the SCEs had an assigned Critical Reason of

"reduced alertness" and only 0.5% were due to asleep-at-the-wheel. Knipling (2018) compared the time-of-day distribution of the 2016 Blanco SCEs to that of actual truck driver fatal fatigue crashes. The former was lowest in the early morning and peaked in late afternoon. The latter peaked in the early morning consistent with the well-known circadian trough. The two time-of-day distributions correlated negatively (R = -0.50) across the 24-hour day (Knipling, 2018). SCEs seem to mainly reflect active driver maneuvers in traffic whereas fatigued drivers are inactive and often driving in rural areas with little traffic (Knipling, 2016).

- **Driver "Performance Deterioration"**. ND studies have implicitly or explicitly conceptualized SCE rate as a measure of driver performance where rate increases indicate performance deterioration (e.g., Blanco at al., 2011). The title of that report states "driving performance" to be the study's dependent variable. In describing its major 2008 ND study (Hanowski et al., 2008), FMCSA (2008) stated that the principal dependent variable, SCE incident rate, was "used as a surrogate for driver performance decrement." "Performance deterioration" is inferred here to be a broader construct than fatigue/drowsiness *per se*. It might, for example, include other types of driver errors such as attentional lapses. Perhaps research could define and quantify a target population of crashes consistent with this causal construct. A difficulty might be found, however, in distinguishing crashes due to driver performance declines from those of high-performance drivers who engage in risky behaviors by habit or by conscious decision. Defining the construct "performance deterioration" in relation to a crash population would be problematic, and there is little indication of this in either crash (e.g., Starnes, 2006) or ND (Blanco et al., 2016) causal analyses. One does see a lot of driver distraction (e.g., 68.5% of Blanco's truck SCE Critical Reasons), but distraction primarily reflects an over-active driver state, not a state of activity decline (Barr et al., 2011).
- **At-Fault Crash Involvements.** M-SCE datasets are counts of *involvements* in events. Thus the most apt comparisons would be to driver- or vehicle-level data as opposed to crash-level data. A simplifying assumption made in crash causation studies is to assign a Critical Reason (proximal cause) to one involved vehicle in each crash. This is tantamount to assigning fault. At-fault and not-at-fault involvements have many differences (Starnes, 2006; Blower and Campbell, 2006; Knipling, 2017b). One could argue that inclusion of not-at-fault involvements in any crash or SCE dataset simply dilutes the dataset from the standpoint of causal analysis. Both truck safety researchers and industry are most interested in crashes caused principally by trucks and their drivers. Forty percent (40%) of the multi-vehicle event involvements in Blanco et al. (2016) were not-at-fault; i.e., the truck driver made no error contributing directly to causation. A target crash population for consideration in validation studies would be at-fault truck crashes, with modification of SCE datasets to include only at-fault SCEs. Overall, truck drivers were at-fault in 90% of Blanco's SCEs versus 55% of truck crash involvements in the LTCCS (Starnes, 2006).
- **Single-Vehicle Crashes**. M-SCE datasets could be reconceptualized as representing single-vehicle crash involvements. As noted, 81% of SCEs in Blanco et al. (2016) involved only

the truck. M-SCE datasets could be parsed to include only single-vehicle events, and then the representativeness of SCEs in relation to single-vehicle crashes could be examined. Driver fatigue is much greater in single-vehicle than multi-vehicle events. In the LTCCS, asleep-at-the-wheel was the assigned Critical Reason for 13% of single-vehicle involvements, versus just 1% of at-fault multi-vehicle involvements (Starnes, 2006; Knipling 2015, 2017b). The corresponding percentages for fatigue presence were 30% and 14%. Thus, for the purposes of HOS research, single-vehicle involvements might be a richer analytic target than all crash involvements. Some pertinent single-vehicle SCE-crash differences are known, however. For example, SCE rates are strongly associated with traffic density (see below), whereas single-vehicle crashes most often occur in low-traffic rural areas (Knipling, 2009; Blincoe et al., 2015).

- **Crashes with Avoidance Maneuvers**. As already noted, the doctrine underlying the M-SCE methodology theory assumes that crashes *avoided* are representative of crashes that *occur*. In Blanco et al. (2016), 99% of the 2,894 non-crash SCEs had an avoidance maneuver (this is how they were detected), versus only one of five actual crashes in the study (20%). In the LTCCS, only 46% of truck drivers were known to have attempted avoidance maneuvers in the seconds before their crash (Knipling and Bocanegra, 2008). In National Motor Vehicle Crash Causation Survey (NMVCCS) car-truck crashes, just 29% of car drivers and 32% of truck drivers (both weighted percentages) were coded as having attempted one or more avoidance maneuvers in response to critical precrash events. Thus crash involvements without avoidance maneuvers constitute a large part of crash involvement populations. It seems highly likely that ND M-SCE representativeness is far higher for crash involvements with avoidance maneuvers than for those without. Research could confirm and quantify this assertion, and then assess whether M-SCE datasets should be reconceptualized as representing only crash involvements with avoidance maneuvers.
- **Traffic-Related Crash Involvements**. Traffic density has a strong influence on SCE rates. Hanowski et al. (2008) reported that their, ". . . results found a strong positive correlation [of SCE rate] to national traffic density data" with a calculated Pearson's R of +0.83. A separate analysis performed at VTTI found an SCE-to-baseline odds ratio of 7.2 for high-traffic conditions (Wiegand et al., 2008). Blanco et al. (2016) also reported significant associations between SCE rates and traffic, even though 81% of their events were single-vehicle. Blanco's SCE rates peaked in late afternoon traffic. Ambient traffic is known to play a far smaller role in single-vehicle crashes than in multi-vehicle crashes, yet SCE rates were affected by traffic density even when the SCE dataset consisted predominantly of single-vehicle events.

Possible methodologies for identifying and validating M-SCE target crash populations are discussed here under "Tasks."

## Objective

This analytic project would identify and specify, as precisely as possible, the large truck target crash population represented by Naturalistic Diving Mixed Safety-Critical Event datasets. An enabling objective associated with one possible methodology for the project would be to analyze truck crash causes as a function of key objective crash characteristics, thereby increasing basic knowledge on crash causation.

## Potential Benefits

Identification and articulation of the target crash population would greatly improve the accuracy of estimates of real world benefits and disbenefits based on M-SCE data. This would increase the practical value of the method for studies of the effects of safety interventions such as Hours-of-Service rule changes.

## Related Work

**Lessons to be learned from medical surrogates**. Surrogate physiological measures are used in medical research, but there are analogous concerns about their validity in relation to health outcomes (British Journal of Anaesthesia, 2008). In the medical context, a surrogate dependent variable is a laboratory measurement used as a substitute for a clinical outcome. Medical interventions (i.e., treatments or any external influence) are tested for their effects on surrogate measures. To be fully valid, however, a surrogate measure must (a) predict the clinical outcome of importance; (b) capture all effects of the intervention; and (c) contain or closely emulate the same biological mechanisms or pathways leading to the outcome (DeMets, 2015). There have been many research cases where the use of surrogates as the primary or only outcome measure has been spurious. They include:

- Lowered cholesterol without survival benefit
- Increased bone density without decreased fractures in osteoporosis
- Increased cardiac function in congestive heart failure without improved survival
- Decreased arrhythmias without improved survival
- Lowered blood sugar without reduced diabetic complications or improved survival.

The requirements for traffic safety surrogate validity are directly analogous to those for medical surrogate validity. Borrowing from the above delineation, a traffic safety surrogate measure should (a) predict a crash harm outcome of importance; (b) capture all of the effects of the safety intervention (e.g., HOS rule changes); and (c) closely emulate the same psychological or psychobiological mechanisms leading to the outcome (e.g., an increase in driver drowsiness or attentional lapses associated with HOS-relevant work schedules).

**Prediction of individual driver risk**. This research needs statement has not addressed nor questioned the validity or usefulness of onboard safety monitoring of dynamic events for the prediction and reduction of individual crash risk. Onboard monitoring is well-established as a behavior-based technology intervention to reduce crashes (Hickman et al., 2007; Toledo et al., 2008). Dynamic measures like driver speeds, hard-braking, accelerations (longitudinal and lateral), and unsignalled lane breaks may be similar or identical to SCEs used in crash causation (including HOS) research. Af Wåhlberg (2008) has defined "celerations" as a synthetic measure of driver speed change behaviors (i.e., speed, close following, braking and steering control measures) which is predictive of driver at-fault crash involvements. Pradhan et al. (2017) have shown that feedback to truck drivers following recorded speeding and other kinematic exceedances can result in reductions in such events and in generalized positive driving behavior change.

Individual risk assessment and crash causation analysis are two different applications, however. Prediction of individual risk requires satisfactory SCE-crash correlations but does not require understanding or emulation of causal mechanisms. In contrast, the use of surrogate dynamic measures for understanding crash causes (e.g., HOS schedule effects) does require emulation of causal mechanisms, as discussed above under medical surrogates. In particular, dependent measures in surrogate-based research on HOS parameters must represent the safety target of HOS regulations; i.e., driver fatigue and asleep-at-the-wheel crashes. Further, factors affecting the surrogate (e.g., prior sleep, time-of-day, hours-of-driving) should operate in the same ways as they operate on the real-world referent.

**External & construct validity**. The types of scientific validity concerns discussed above are not limited to naturalistic driving. They arise widely in science as questions of external and construct validity (Privatera, 2014). *External* validity is the truthfulness of generalizations from studies to real-world phenomena of importance. If you are studying white mice, do your results apply to people? *Construct* validity is another type. A psychological construct is an underlying factor known to exist but which cannot be directly observed. It is inferred from data. "Fatigue" is a classic example. We may say and think we are measuring fatigue in a study, but can we prove that? Without clear evidence of external and construct validity, the meaning of many kinds of scientific findings cannot be understood with any degree of confidence.

Trochim (2006) described the difference as follows: ". . . where external validity involves generalizing from your study context to other people, places or times, construct validity involves generalizing from your program or measures to the *concept* of your program or measures. You might think of construct validity as a "labeling" issue. . . . [for example] When you measure what you term "self-esteem" is that what you were really measuring?"

Trochim describes validation as a process of pattern matching. In the current context, do the objective characteristics (or patterns) of SCEs match the objective characteristics of possible target populations? Two kinds of evidence are sought: convergent and discriminant validity. One demonstrates convergent validity when measures that are theoretically supposed to be highly interrelated are in fact highly interrelated. For example, if we believe that M-SCE datasets represent fatigue, do their circadian patterns match those already well-established for fatigue? One demonstrates discriminant validity when one demonstrates that measures that shouldn't be related are in fact not. For example, if we believe that M-SCE datasets do not simply reflect traffic interactions, we attempt to show that their incidence patterns are discrepant from simple variations in traffic.

**Validity Evidence Cited by VTTI**. In committee review discussions of this research needs statement, leading ND practitioner VTTI cited six publications as providing evidence for Mixed-SCE validity. These publications (Guo et al., 2010a & 2010b; Kusano et al., 2015; Perez et al., 2017; Simons-Morton et al., 2012; Wuk and Jovanis, 2011) are listed in the reference section below.

## Tasks

One possible methodology would first use existing crash causation data, including both crash investigations and ND data, to identify those objective crash characteristics (auxiliary variables) having the greatest influences on crash causality. Crash causality profiles vary widely, but often consistently, across different objective categories (Knipling, 2017b). Objective crash characteristics are those describing the “who,” “when,” “where,” and “how” of crashes. The following are established variables which might be top candidates: Driver Age, Hour of Crash, Relation to Junction, Interstate Highway (or not), Trafficway Description (e.g., undivided, divided with median, etc.), Work Zone (or not), Lane Use (urban vs. rural), Number of Vehicles in Crash, Crash Involvement Type (vehicle’s role; e.g., rear-end striking, lead vehicle stopped). Such variables have been used commonly in crash causation studies such as the LTCCS, the National Motor Vehicle Crash Causation Survey (NMVCCS), and ND studies. Key variables describing causation include Critical Reason and Associated Factors. To cite simple example, pairing Crash Involvement Type (objective description) with Critical Reason (causal inference) for bivariate analysis across studies would likely reveal distinctive causal topographies. Some of this work for crash investigations has already been reported (e.g., Starnes, 2006; Knipling and Bocanegra, 2008; Knipling, 2017b). Following identification of the most salient objective descriptors in relation to causation, the objective composition of ND M-SCE datasets in relation to those of candidate target crash populations could be assessed. Logically, the best match would identify the “best fit” target crash population. The same method could likely generate quantitative measures of goodness-of-fit.

Representativeness might be then improved artificially by indexing M-SCE objective characteristics to known target population characteristics. A research needs statement by the TRB Committee for Safety Data, Analysis, and Evaluation (TRB ANB20, 2016) suggests that M-SCE dataset crash-related validity and representativeness might be improved by differentially weighting SCEs to index objective SCE profiles to corresponding crash profiles. This would treat SCE datasets in the same statistical manner as one would treat an unrepresentative sample taken from a population. The size of weights required to bring M-SCE datasets into reasonable conformity with candidate target crash populations in regard to objective characteristics would be another measure of discrepancies between datasets. The most concordant target crash population for M-SCE datasets would be that with the fewest and smallest differences.

Another opportunity for SCE-versus-crash analysis lies in the mix of dynamic triggers chosen for M-SCE datasets. Unlike most crash samples (e.g., LTCCS), M-SCE dataset composition is not guided by a sampling algorithm. Researchers can mold datasets toward external validity objectives by adjusting thresholds of different triggers and assessing external validity effects. Here, different M-SCE mixes can be compared to the various possible target crash populations listed above. This combined with SCE indexing might result in acceptable concordance with one or more target populations of harmful crashes.

Note that these approaches respect and utilize the superior “why” answers (e.g., Critical Reason) provided by ND videos for individual events. But they attempt to reconcile or at least assess large dataset differences in “who,” “when,” “where,” and “how.”

Other analytic approaches and quantitative methods might be applied to this problem. The above suggestions are not intended to preclude other analytic options.

## Implementation

Project findings would be implemented in improved mixes of trigger types and proportions in ND datasets and, most notably, plausible extrapolations of M-SCE statistics to real-world outcome measures.

## Funding

The initial analysis of causation data (both crash and ND) to identify the strongest objective characteristics (auxiliary variables) might require $100-200K in funding. A study to then assess and seek to improve M-SCE vs. hypothesized target crash population concordance might cost an additional $200-300K. As noted, parts of this overall study could be performed separately with smaller levels-of-effort.

**Research Period**

This analytic program and specific projects of the type described would use existing ND and crash data and would not require special equipment or facilities. The work could be performed over a period of one year or less.

**Relevance**

This project is relevant to all past, present, and future studies employing the ND Mixed-SCE methodology. Algorithms for characterizing and quantifying corresponding outcome effects from SCE statistics could be applied to any study employing this methodology and with the recorded data variables required by the algorithm. Individual studies could address any possible specific target crash population. Thus the project would be appropriate for independent (e.g., university) researchers as well as for contracted research.

**Source Information**

Some of the information and issues discussed here are derived from another ANB70 RNS entitled *Toward Naturalistic Driving Crash Representativeness*, and from discussions within the ANB70 Data Needs Subcommittee. Another information source has been an ANB20 RNS entitled *Indexing Naturalistic Driving Events to Crashes*.

**RNS Developer**

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

**Cited References**

Af Wåhlberg, A.E., Driver celeration behaviour and accidents - an analysis. *Theoretical Issues in Ergonomics Science*, 9:5,383-403, September 2008. Available at: http://dx.doi.org/10.1080/14639220701596722.

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, Washington DC: Department of Transportation (DoT). 2011.

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.

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

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

Blower, D.F. and Campbell, K.L. *Methodology of the Large Truck Crash Causation Study*. Large Truck Crash Causation Study Analysis Series. Report No. FMCSA-RI-05-035. February, 2005.

British Journal of Anaesthesia. Editorial: Validity, credibility, and applicability: the rise and rise of the surrogate. 101 (5): 595–6, doi:10.1093/bja/aen292, 2008

DeMets, D.L. Role and potential of surrogate outcomes. Univ. of Wisconsin, presentation downloaded February 8, 2015. Available at http://depts.washington.edu/ssbiost/PRESENTATIONS/DeMets.pdf

Dunlop, L. Beyond the safety triangle. Safety & Health Practitioner, 31, (10), Pp. 45-46, October 2013. Available at: https://www.shponline.co.uk/beyond-the-safety-triangle/.

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, *2010-2011 Hours of Service Rule, Regulatory Impact Analysis*, RIN 2126-AB26. 2011

FMCSA. *Commercial Motor Vehicle Driver Restart Study Plan*. Submitted to the U.S. Department of Transportation Office of Inspector General in accordance with Sec. 133 (d) of the Consolidated and Further Continuing Appropriations Act of 2015. February 12, 2015.

FMCSA. Large Truck Crash Facts 2016. FMCSA-RRA-17-016. U.S. DOT, May 2018.

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.

Guo, F., S. G. Klauer, J. Hankey, and T. A. Dingus. Near-Crashes as crash surrogate for naturalistic driving studies. *Tranportation Research Record*, No. 2147, Pp. 66-74, December 2010a.

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

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.

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

Hickman, J.S., Knipling, R.R., Hanowski, R.J., Wiegand, D.M., Inderbitzen, R.E., & Bergoffen, G. CTBSSP Synthesis Report #11: *Impact of Behavior-Based Safety Techniques on Commercial Motor Vehicle Drivers*. TRB Commercial Truck & Bus Synthesis Program. ISSN 1544-6808, ISBN 978-0-309-09876-2, 2007.

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

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. Critical Review of Driver Fatigue & HOS-Related Research Methodologies. Commissioned paper for National Academy of Sciences Commercial Driver Fatigue Panel, Available at http://sites.nationalacademies.org/DBASSE/CNSTAT/CMV_Driver_Fatigue_Long-Term_Health_and_Highway_Safety/index.htm, 2016.

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

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

Knipling, R.R. Toward naturalistic driving crash representativeness. 10th SHRP2 Safety Data Symposium: From Analysis to Results. October 6, 2017. Transportation Research Circular E-229, Pp. 60-68, February 2018.

Knipling, R.R., Bocanegra, J., Hickman, J.S., and Hanowski, R.J. *Comparative Analysis of LTCCS and Naturalistic Driving Data*. Project work plan for FMCSA. November, 2008.

Kusano, K.D., R. Chen, A. Tsoi, and H.C Gabler. Comparison of event data recorder and naturalistic driving data for the study of lane departure events. *TRB 2015 Annual Meeting* Paper No. 15-0147. January 2015.

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

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

Olson, R.L., Hanowski, R.J., Hickman, J.S., & Bocanegra, J. (2009). *Driver Distraction in Commercial Vehicle Operations*. FMCSA-RRR-09-042. Wash. DC: DoT. 2009.

Perez, M.A., J.D. Sudweeks, E. Sears, J. Antin, S. Lee, J.M.Hankey, and T.A. Dingus. Performance of basic kinematic thresholds in the identification ofcrash and near-crash events within naturalistic driving data. *Acc Anal and Prev*. 103 (2017) 10-19, 2017.

Pradhan, A.K., B.T. Lin, C. Wege, and F. Babel. Effects of behavior-based driver feedback systems on commercial long haul operator safety. *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Manchester Village VT, June 26-29, 2017.

Privitera, G. J. *Research Methods for the Behavioral Sciences,* Sage Publications, Inc. 2014.

Simons-Morton, B., Z. Zhang, J. C. Jackson, and P.S. Albert Do elevated gravitational-force events while driving predict crashes and near crashes? *Am J Epidemiol*. 2012;175(10):1075–1079. 2012.

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

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

Toledo, T., Musicant, O., & Lotan, T. In-vehicle data recorders for monitoring and feedback on drivers' behavior. Tranportation Research, Part C. 2008.

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

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

Trochim, W.M.K. Construct Validity. Research Methods Knowledge Base; Web Center for Social Research Methods. 2006, accessed 8/11/16. Available at: http://www.socialresearchmethods.net/kb/constval.php

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

Wu, K-F and P.P. Jovanis. Defining, screening, and validating crash surrogate events using naturalistic driving data. *3rd International Conference on Road Safety and Simulation*, Indianapolis, September 14-16, 2011.

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.