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

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

Final Revision Submitted: March 5, 2017

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

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

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

1. INTRODUCTION

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

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

Selected Acronyms

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

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

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

2. METHOD

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

- The National Motor Vehicle Crash Causation Survey (NMVCCS) was conducted by NHTSA between 2005 and 2007. Trained researchers investigated 5,471 crashes, each of which involved a light passenger vehicle (*1*). NMVCCS included all five police-reported "KABCO" severity levels: K = Killed; A = Incapacitating injury; B = Non-incapacitating

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

injury; C = Possible injury; O = No injury (also known as Property Damage Only). Thousands of specific characteristics could be recorded for each crash and each involved person or vehicle. Weighed NMVCCS statistics represented 2,189,166 crashes with 4,031,226 involved vehicles.

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

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

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

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

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

3. HORIZONTAL CRASH HETEROGENEITY

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

3.1 Heterogeneity of Involvement Types

Numerous examples may be cited. In the LTCCS, truck driver asleep-at-the-wheel was the CR in 19% of road departures but 1% or less of rear-end, sideswipe, and intersection/crossing

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

path involvements. Excessive truck speed (for curve/turn or for responding to other vehicles) as the CR was 21% of road departures, 32% of head-on strikes where the truck was the encroaching vehicle, 25% of rear-end strikes into moving vehicles, but just 6% of rear-end strikes into stopped vehicles.

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

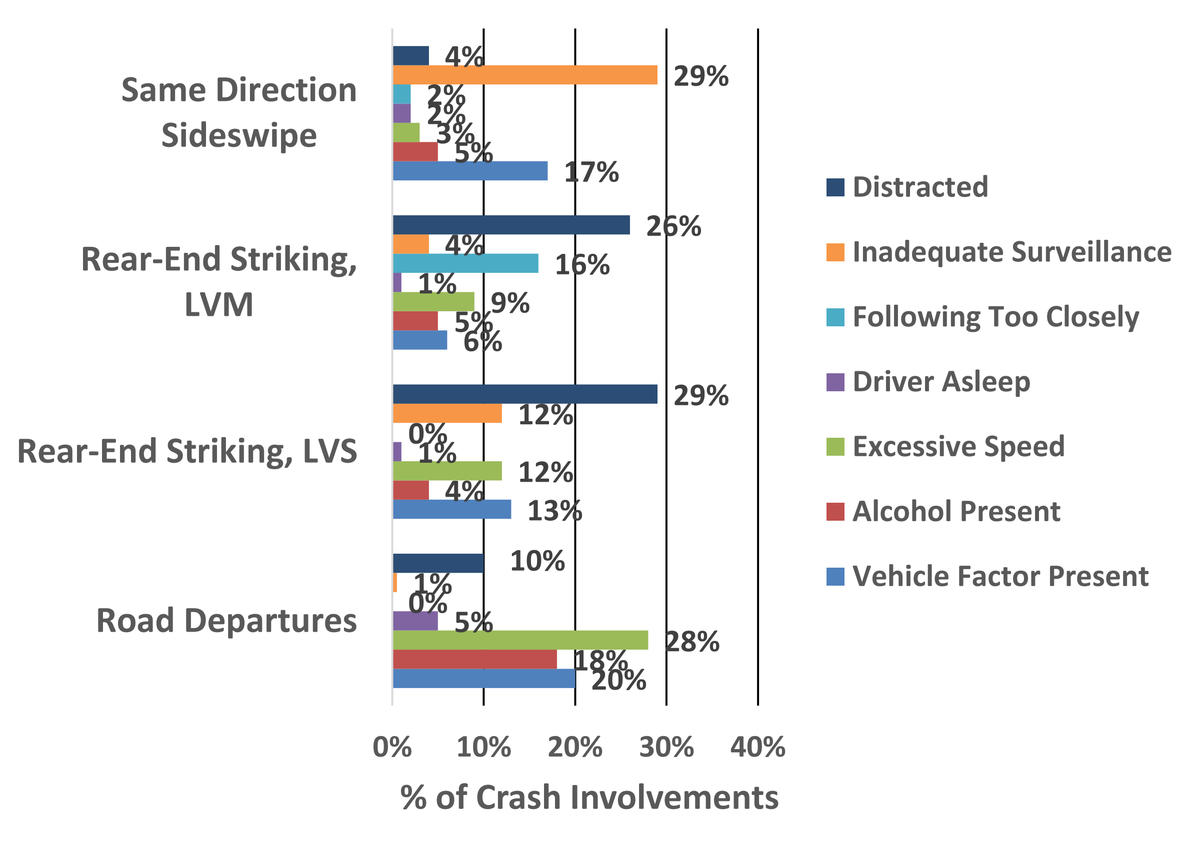


FIGURE 1 Weighted Distributions of Selected CRs for Four Crash Involvement Types in NMVCCS. LVM = Lead Vehicle Moving, LVS = Lead Vehicle Stopped. Distraction includes two types: internal and external. Excessive speed includes three: excessive speed for conditions, to be able to respond, and for a curve/turn.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

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

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

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

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

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

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

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

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

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

Table 2 Three Involvement Categories: Weighted Critical Reason Percentages

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

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

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

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

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

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

4. VERTICAL CRASH HETEROGENEITY

4.1 Systematic Differences Associated with Severity

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

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

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

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

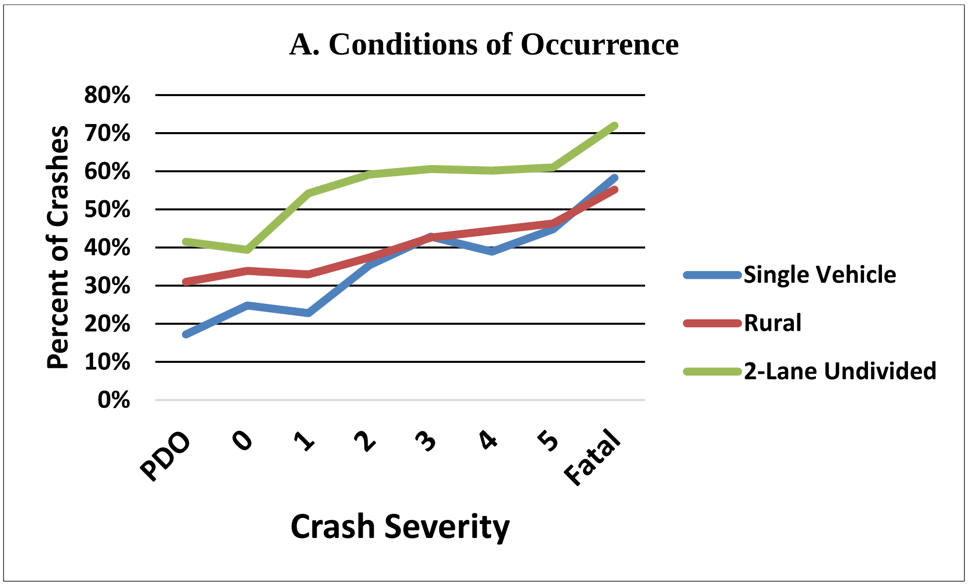
A 2015 NHTSA report by Blincoe et al. (*7*) shows the varying presence of six factors in crashes (including both police-reported and non-police-reported) of different severities. NHTSA estimated factor presence for the years 2008-2010 based on its National Automotive Sampling System (NASS) and related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

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

These differences demonstrate that crashes cannot be accurately characterized without reference to specific severity levels. Quantitative statements about crash risks and the

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

prevalence of causal factors (e.g., distraction, alcohol) require a stated target crash population to be meaningful.



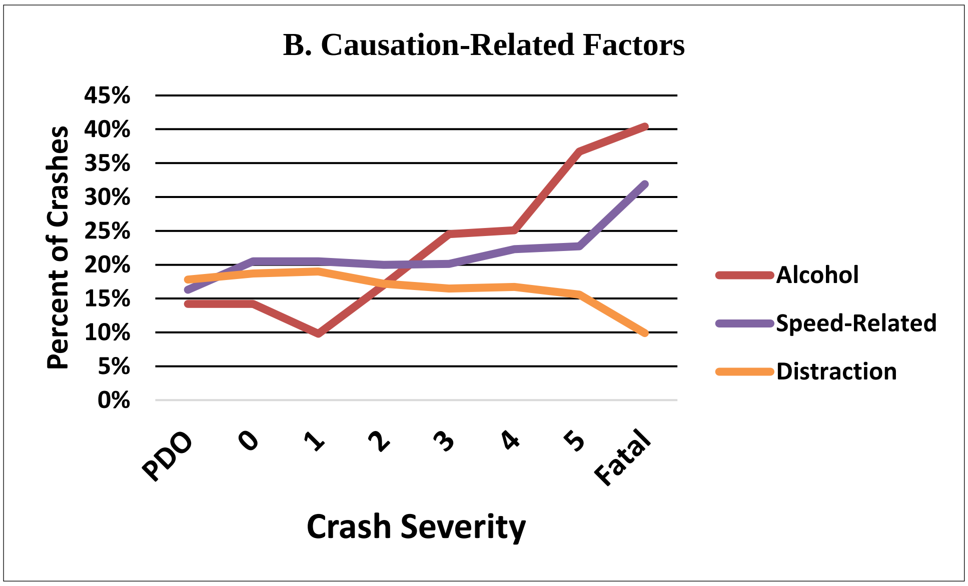


FIGURE 2 Trends in Crash (A) Conditions of Occurrence and (B) Causation-Related Factors with Increasing Severity (*7*). Note: Data for 2-lane undivided includes rural crashes only.

4.2 Is "Heinrich's Law" Invalidated?

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

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

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

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

4.3 Crash Severity and Total Crash Harm

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

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

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

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

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

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

5. IMPLICATIONS FOR NATURALISTIC DRIVING STUDIES (NDSs)

5.1 Creation of Mixed-SCE Naturalistic Driving Datasets

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

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

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

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

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

5.2 Mixed-SCE Datasets and Heinrich’s Triangle

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

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

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

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

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

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

5.3 All-Crash SCE Datasets

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

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

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

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

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

IIHS reported the two predominant SHRP2 crash types to be rear-end and road departure, with the proportions of these two types varying sharply by event severity. For

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

example, rear-end scenarios were 54% of the 98 most severe events, but only 12% of the 597 minor contacts. They are about one-third of police-reported crashes nationally (*19*).

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

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

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

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

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

6. IMPROVING EXTERNAL VALIDITY

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

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

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

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

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

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

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

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

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

7. CONCLUSIONS

This paper asserts the following:

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

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

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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