ATA
Driving Trucking's Success

American Trucking Associations
950 N. Glebe Road, Suite 210, Arlington, VA 22203

June 8, 2011

Via email

The Honorable Anne Ferro
Administrator
Federal Motor Carrier Safety Administration
1200 New Jersey Ave., S.E.
Washington, D.C. 20590

RE: Hours of Service of Drivers - Docket Number FMCSA-2004-19608

Dear Administrator Ferro:

Thank you for the opportunity to review and comment on several new research reports identified and/or funded by FMCSA and placed in the Hours-of-Service docket on May 9, 2011. The purpose of this letter is to forward to you a peer review critique of the study performed by The Center for Truck and Bus Safety at the Virginia Tech Transportation Institute entitled *The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations.* We have also placed a copy of this peer review, conducted by Dr. Ron Knipling, in the hours of service docket.

As you know, Dr. Knipling is exceptionally well-qualified to assess these studies and the relevance of their findings to the proposed hours of service rule. He has more than 30 years of experience in traffic safety with emphasis on driver human factors and motor carrier safety. For six years he served as the Chief of FMCSA's Research Division before accepting a position as a Senior Research Scientist and Senior Transportation Fellow with the Virginia Tech Transportation Institute (VTTI). He has authored nearly 250 technical reports, papers, and conference presentations, as well as the first and only comprehensive textbook on large truck safety, *Safety for the Long Haul; Large Truck Crash Risk, Causation, & Prevention.* In recognition of the book, he received the International Road Transport Union's (IRU) Order of Merit award, the first given to an American scientist.

At ATA's request, Dr. Knipling reviewed the VTTI study and prepared the attached report. In his report, Dr. Knipling finds that the study has significant limitations such as:

- The study is based on safety critical events (SCE) such as unintentional lane deviations, but not "real harm." Only 4 of the 2,197 SCEs in the study were actual crashes.
- In the subject NPRM, FMCSA acknowledged that "A definitive link between critical incidents and crash risk has not yet been established."
- The methodology raises questions related to representativeness. For instance, one company's 75 drivers constituted 78% of the study sample.
- The researchers failed to disaggregate the data.
- The study failed to address how previous research reflected a strong *positive* association between unintended lane deviation and driver distractions and the *negative* association between driver distractions and fatigue.

Based on these and other concerns, Dr. Knipling concluded that "…more probing and self-challenging analyses must be performed before [the] study['s] findings can be accepted as sound science."

We trust that Dr. Knipling's report will be read and reviewed by Agency staff and very carefully considered in this process. Again, thank you for the opportunity to comment on these studies.

Sincerely,

David J. Osiecki
ATA Sr. Vice President for Policy & Regulatory Affairs

Enclosure

Peer Review Critique of VTTI Study:

The Impact of Driving, Non-Driving Work, and Rest Breaks on Driving Performance in Commercial Motor Vehicle Operations

Prepared for the American Trucking Associations
June 7, 2011

Ronald R. Knipling, Ph.D.
Safety for the Long Haul Consultancy
5059 36th N. Street; Arlington, VA 22207-2946
(703) 533-2895; rknipling@verizon.net

This is a review and critique of a Virginia Tech Transportation Institute (VTTI) research report placed on the FMCSA website and in the commercial driver Hours-of-Service (HOS) rulemaking docket (FMCSA-2004-19608) on May 6, 2011. The citation follows:

> 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.

The VTTI study was an analysis of naturalistic driving data collected in the Naturalistic Truck Driving Study. Data were collected from 97 commercial drivers as they drove instrumented company trucks in normal revenue-producing runs. Each driver drove for about four weeks, and the study accumulated about 735,000 miles of driving data. Driver performance was assessed through the occurrence of "safety-critical events" (SCEs), which included crashes, near-crashes, crash-relevant conflicts, and unintentional lane deviations. In addition, each driver maintained a daily 24-hour activity register in which he or she recorded 14 different activity categories for 15-minute intervals. The study related driver schedules as classified by the registers to their SCEs captured through naturalistic driving. Four major issues relating to HOS regulations were addressed:

- Workday characterization; i.e., activity composition
- Driving hours or time-on-task (TOT)
- Work hours; TOT effects based on work hours as opposed to driving hours
- Breaks; including different kinds of breaks and their interaction with TOT.

This critique first addresses the advantages and disadvantages of naturalistic driving for crash-related research, with emphasis on its suitability as a testbed for driver fatigue. It highlights potential confounding co-variables, and additional analyses which could have been done within the naturalistic driving paradigm to address these. It also addresses specific findings relating to the four main study issues, as listed above.

Naturalistic Driving as a Testbed for Driver Fatigue

The immense capabilities of the naturalistic driving methodology are now well-known to traffic safety researchers. These include the following (Dingus et al., 2006, Hanowski et al., 2008, Knipling, 2008; Knipling, 2009a):

- **Real world**. Its data collection is in real, everyday driving and, in trucking, usually in operational, revenue-generating runs.
- **Video observation of drivers and surrounding environment**. Video cameras (typically five for truck studies) look directly at the driver and the forward and rearward roadway scenes. These can be replayed and analyzed in detail. For example, Barr et al. (2011) have just published a report based on detailed observations of drivers who were fatigued and those who were distracted. This and previous studies (Olson et al., 2005; Hanowski et al., 2008) provided Observer Rating of Drowsiness (ORD) scores for each SCE so that the association of drowsiness with SCEs could be assessed.
- **Other direct sensor measurements**. Vehicles are equipped with multiple dynamic sensors relevant to vehicle control and risk. For example, lane tracking devices can continuously measure both the vehicle's position in the lane as well as measuring the roadway lane width itself.
- **Incident capturing**. Sensor data in the overall data stream can be filtered to identify potential SCEs based on *a priori* thresholds for each sensor type.
- **Incident review, verification, classification, and analysis**. Once incidents are captured, their video records and other sensor data are reviewed to verify their safety relevance. SCEs may be analyzed in detail using a data directory of safety-relevant variables, much like an in-depth crash investigation. For many driver-related factors, the analysis can be more revealing than crash investigations because SCEs are directly observed. In addition, environmental conditions like weather and traffic density can be seen directly, rather than reconstructed *post hoc*. SCE conditions may be compared to those of randomly selected baseline epochs to assess their association with SCE occurrence.

  The addition of a driver activity register, as in the current study, offers a new level of analysis and insight into driver factors. For example, nightly sleep could have been compared to SCE involvements the next day. Key alertness-related activity variables (e.g., amount of sleep, time awake, and time-of-day) could have been entered into a Sleep-Performance Model (Balkin et al., 2000; Rosekind, 2005) to determine the correspondence of SCE rates to model predictions.
- **Huge data haul**. Most vehicle instrumentation suites collect data on dozens of dynamic variables concurrently and continuously. The current VTTI study, for example, was based on 735,000 miles of data recordings. Another recent study (Hanowski et al., 2008) collected 2.3 million miles of data. SCEs are far more numerous than crashes and thus can be studied

quantitatively with far more precision and statistical power.

- **Extensive data on each driver**. The current study collected about four weeks of driving for each of its 97 subjects, and other studies (e.g., Dingus et al., 2006) have collected a year or more of data per driver. This permits elegant studies of individual differences in driver behavior and risk. In the current study, just four of the 97 drivers accounted for 36% of all observed SCEs. In a previous study (Hickman et al., 2005; Hickman et al., in press), 13% of the drivers were associated with 66% of all high-drowsiness at-fault events.
- **Excellent testbed for driver distraction**. Naturalistic driving SCEs often occur during vehicle-vehicle traffic interactions or in other situations where recognition failures (failure to perceive crash threats) play a major role. In a study of driver distraction representing 203 drivers and 3 million miles of exposure, Olson et al. (2009) found that 63.9% of driver at-fault events involved a prior tertiary task; that is, a task extraneous to driving such as eating or cell phone use. This included 40.0% of crashes, 50.0% of near-crashes, 57.4% of crash-relevant conflicts, and 77.5% of unintentional lane deviations.

  Driver distraction is both frequent in SCEs and *causally linked* to their occurrence. Most notably, tertiary tasks like text messaging and interacting with a dispatching device are associated with sharp increases in SCE risk. The SCE-baseline odds ratio of 23 for text messaging reported by Olson et al. (2009) has been widely publicized and featured prominently in DOT's campaigns against driver distraction. The high SCE-baseline risk for various tertiary tasks, and their associated Population Attributable Risk percentages have left no doubt about the role of distraction in SCE genesis. Researchers can be confident that tasks, devices, or other factors increasing distraction will also increase SCE rates.
- **Excellent testbed for experiments**. Though not relevant to the current study, naturalistic driving may provide extensive and exquisite dependent variable data not obtainable otherwise.
- **Derivation of relative risk estimates**. In crash investigation there are no easily accessible non-crash controls, but that's not the case with naturalistic driving. Control data is readily available based on total exposure (as in the current study) and/or based on randomly selected baseline epochs (as in Olson et al., 2009; Hickman et al., 2005; Knipling et al., 2005). This permits derivation of risk odds ratios for ambient conditions like roadway type and traffic density. The "Phase I" large truck naturalistic driving study, for example, derived a 5.3 incident risk odds ratio for undivided highway vs. divided highways (Hickman et al., 2005; Knipling et al., 2005). In crash research, relative risk determination is far more problematic and generally limited to normally recorded variables such as those in driver logs (e.g., hour-of-day, hour-of-driving). In naturalistic driving, exposure data comes from the same source as the SCE data; i.e., the body of collected driving data. In most crash research, exposure data must come from a different source, such as roadside traffic counts.

Juxtaposed to the capabilities of naturalistic driving are its limitations and disadvantages, in particular in regard to the study of driver fatigue. Naturalistic driving disadvantages include the following:

- **Tenuous relation to crash harm**. Naturalistic driving may capture the "real world," but it does not capture "real harm." All 963 crashes in the Large Truck Crash Causation Study (LTCCS) resulted in a serious injury or death (Starnes, 2006). Based on statistics in Zaloshnja and Miller (2007), LTCCS crashes were the most severe ~11% of police-reported large truck crashes, but they represented 80-90% of all truck crash harm, including both human and material consequences. In the current VTTI dataset, only four (4) of the 2,197 SCEs (0.2%) were crashes, the criterion for which was "any contact" (Page 16). The four crashes were not described, but it is likely that most were non-police-reported, non-DOT-reportable, and/or non-injury. Because of their small crash percentage and their minor nature, truck SCE datasets probably represent less than 1/1,000 of crash harm.

  The concern here is not simply regarding "face validity." Rather, it is regarding the causal mechanisms underlying SCEs vs. those in crashes. In the "Phase I" truck naturalistic driving study (Olson et al., 2005), only *one* of 887 SCEs (0.1%) involved truck driver asleep-at-the-wheel as the Critical Reason (CR). In the LTCCS, 3.8% of truck serious crash involvements were asleep-at-the-wheel. That's a 38-fold difference.

- **Synthetic, *a priori* dataset construction**. Crash databases employ *outcome* criteria for crash inclusion. The Fatality Analysis Reporting System (FARS) requires a fatal crash, while the LTCCS required a serious injury (i.e., K, A, or B in the police-report KABCO crash scale). In contrast, naturalistic driving develops its dataset by aggregating different triggering sensor events, including longitudinal acceleration (e.g., hard-braking), time-to-crash, swerves, and, in this study, lane deviations. A small number of events are added via a critical-incident button or by analyst judgment. Each dynamic trigger type has its own *a priori* quantitative criterion, but these are not indexed externally to crash probability or harm. Thus, prior judgments made about each trigger criterion determine its proportion of the overall dataset and thus the overall profile of SCEs. In the current study, 1,118 unintentional lane deviations (ULDs) were added to the other 1,079 SCEs generated from other types of dynamic triggers and classified by severity (i.e., 4 crashes, 7 curb strikes, 46 near-crashes, 1,022 crash-relevant conflicts). The overall dataset may be considered "synthetic" in that it was *devised* by researchers rather than *derived* from observations of the phenomena of interest; i.e., crashes.

  In its NPRM discussion of its prior HOS-related naturalistic driving study (Hanowski et al., 2008) on Page 82179, FMCSA stated the concern that, ". . . the VTTI study does not appear to be definitive. . . .the study looks at the risk of critical incidents, which include near-crashes and crash-avoidance responses, as well as actual crashes. A definitive link between critical incidents and crash risk has not been established."

- **Questionable testbed for driver fatigue studies.** Unlike driver distraction, driver fatigue is not easily seen in naturalistic driving events. As already noted, only one of 887 SCEs (661 of which were at-fault) in the "Phase I" truck study (Olson et al., 2005) had a CR of asleep-at-the-wheel. In 10 more, high-fatigue was designated as the CR. That's 11 of 661 at-fault events, or 1.7%. A total of 13% had fatigue coded as an associated factor, though in most of these some active, voluntary behavior like speeding or tailgating was the CR. One might be tempted to attribute a significant portion of driver distraction to underlying fatigue, since, experimentally, sleep deprivation generates attentional lapses (Balkin et al., 2000). However, that experimental relationship does not prove that distraction in real driving is fatigue-related. In a newly released in-depth analysis of truck naturalistic driving addressing drowsiness and distraction, Barr et al. (2011) found the two to be inversely related. Their two measures of distraction (termed EYETRANS and EYESOFF) tended to be greater when drivers were alert and lower when they were drowsy. Fatigued drivers tended to narrow their working visual fields, whereas distracted drivers tended to widen them. EYETRANS and EYESOFF were lower during fatigue events than during baseline (control) events. When there was observable drowsiness, these distraction measures varied *inversely* with the level of drowsiness.

  In a separate fatigue-focused re-analysis of 3,270 truck SCEs, Wiegand et al. (2008) found inverse SCE-baseline odds ratios associated with fatigue. That is, the relative risk of an SCE compared to baseline 1.93 times greater when ORDs were *below* the fatigue threshold (i.e., 39 or less on a 100-point scale). Similarly, the SCE-baseline odds ratio was 1.69 when percent eye closure (PERCLOS) was below the fatigue threshold (i.e., 12.49% or less).

  Of course, this meant that SCEs in general could not be attributed to fatigue in the study. In the distraction study cited earlier, Olson et al. (2009) found the total Population Attributable Risk of tertiary tasks (i.e., distraction) to be 57%. That is, mathematically speaking, 57% of at-fault SCEs would be eliminated if distraction were eliminated. Applying the same algorithm to Wiegand's SCE ORD statistics yields a 36% Population Attributable Risk associated with *being alert*. No one would accept that as meaning that alertness actually increases crash risk, but it does lead one to wonder why drowsiness seems to function in a reverse manner in naturalistic driving data.

  Wiegand also found fundamental differences between high ORDs and SCEs. ORDs were higher on divided roadways while SCE risk was greater on undivided roads. ORDs were higher in low traffic density, while SCE were more frequent in high traffic density. The odds of a driver experiencing an SCE, when compared to baseline epochs, were 7.16 times greater when traffic density was high (coded LOS C-F) as when it was low

(LOS A-B).

The current study added UDLs to its mix of SCE types (as 51% of the total) because, "Decades of previous research have suggested that unintentional lane deviations provide a reliable indicator of fatigue" (Page 30). Yet, as noted above, 77.5% of the ULDs in Olson et al. (2009) were *distraction*-related, and distraction and drowsiness appear to be negatively associated in SCEs (Barr et al., 2011).

One might conceptualize the question of SCE validity vis-à-vis fatigue as one conceptualizes the validity of assessment tests and measurements (e.g., as in Anastasi, 1968). The small percentage of fatigue-related SCEs suggests low content validity. The apparent inverse relationship between fatigue and SCEs suggests not only low construct validity but even "reverse" construct validity. Contrast this with distraction, for which SCE content and construct validity are both very high as demonstrated in the Olson et al. (2009) study.

The current study makes the implicit assumption that driving and work hours could increase driver fatigue and that that fatigue would be reflected in increased SCE rates. For example, the SCE-reduction effect of work breaks in the study is presumed to be associated with fatigue reduction. Yet, based on the three VTTI studies cited above, the opposite could be true! No one really believes that being drowsy makes one a better driver or that work breaks increase fatigue. Rather, the take-away is that the number of naturalistic driving SCEs is not a sensitive or reliable indicator of fatigue.

- **Sensitivity to roadway factors**. SCE occurrence is highly sensitive to roadway conditions like roadway type and traffic density. This relationship has been validated by crash research demonstrating that crash rates are strongly affected by roadway factors (Harwood, 2006; FMCSA Analysis Division, 2010; FHWA, 2000; Kononov et al., 2011). Yet this validated safety relationship can be a source of confounding if road and traffic conditions are not held constant in comparisons. In the "Phase I" truck naturalistic driving study, the SCE-baseline odds ratio for undivided roads vs. divided ones was 5.3. Suppose that two comparison samples were equivalent except that one had 20% driving on undivided roads while the other had 10%. There would then be a 30% difference in overall SCE rate between these conditions due entirely to the difference in road type mix. Traffic density effects on naturalistic driving comparisons are similar.
- **Expensive and labor-intensive**. Naturalistic driving studies are expensive, requiring elaborate instrumentation suites and frequent maintenance. More relevant to the current study is the labor required to fully utilize the research capabilities described above. Full causal analysis requires that every SCE be viewed repeatedly and coded using a data directory of safety-relevant variables, much like an in-depth crash investigation. This has the great benefit of characterizing

each event in terms of its key causes (e.g., CRs) and conditions of occurrence such as the roadway factors discussed above. If individual SCEs are not analyzed fully, SCE datasets cannot be disaggregated to unmask underlying factors and control for known confounds.

- **Representativeness issues**. In its December 2010 NPRM, the agency noted several specific concerns about the prior HOS-related naturalistic driving study by Hanowski et al. (2008). On Page 82179, the agency noted that the study "involved a small sample size of 102 drivers that was not representative of the trucking industry." It also "involved drivers who were, with their knowledge, observed by video cameras and other electronic equipment. It is possible that this may have led drivers to behave more carefully than drivers would have in the absence of observation . . ." In addition, the agency expressed the concern that, ". . . drivers and carriers who participated in the video-surveyed study did so voluntarily, which could skew the study towards participation from more safety-conscious drivers and carriers." The same issues apply to all naturalistic driving studies to date. For example, the current study had just 97 drivers. Its Executive Summary (Page xii) states, "As with any study that uses volunteers, the drivers may not be representative of the entire population of commercial drivers."

  Representative concerns encompass carriers, drivers, and exposures to risk. Risk exposure varies widely depending on truck type (combination- vs. single-unit), operations type (e.g., private vs. for hire, long-haul vs. line-haul), type of cargo hauled, and region. The amount and patterns of exposure to urban traffic strongly affect both crash and SCE risk. The current study sample consisted of 75 drivers from a single long-haul carrier, and a total of 21 drivers from three line-haul carriers. While the 75:21 (3.6:1) ratio conforms well to the industry, the fact that one company's 75 drivers constituted 78% of the study sample obviously raises concerns about its national representativeness.

- **Paucity of data on other drivers and on vehicle factors**. Though they are probably not directly relevant to the current study, two other limitations may be noted. First, naturalistic driving provides vastly different perspectives of the two involved vehicles and drivers in two-vehicle events. The subject vehicle and driver are seen from the inside, while the other vehicle and driver are seen from the outside. Second, while there are elegant measurements of vehicle motion and dynamics, to date there have been few concurrent measurements of vehicle mechanical condition, including brake condition. Thus, a driver's "inadequate evasive maneuver" may be due principally to weak brakes.

Needed Disaggregations of the Dataset

As discussed above, a powerful application in naturalistic driving is that of event classification and analysis. When event videos are reviewed and verified as safety-critical, they can also be classified. In past studies such as the 100-Car study (Dingus et al., 2006), Local/Short-Haul driver fatigue study (Hanowski et al., 2000), and the "Phase I" truck naturalistic driving study (Hickman et al., 2005; Olson et al., 2005; Knipling et al., 2005), events were coded for multiple variables using a data directory. This included causal variables (e.g., CR, "fault"), alertness measures (e.g., ORD, PERCLOS), and conditions of occurrence (e.g., weather, road type, traffic density, relation to junction). In the current VTTI study, event videos were apparently reviewed to verify them as SCEs, but there was no further classification, at least in the published report. Below are discussions of three different disaggregations based on SCE coding which could have validated (or invalidated) study conclusions and would have provided more insight into driver fatigue and other crash risk factors:

- **At-fault vs. not-at-fault (NAF)**. Previous truck naturalistic driving studies (e.g., Hanowski et al., 2000; Hickman et al., 2005; Hanowski et al., 2008) disaggregated SCEs based on fault; i.e., commercial driver/truck at-fault vs. not-at-fault (NAF). In the "Phase I" study (Hickman et al., 2005; Hickman et al., in press), Critical Reasons (CRs) were identified for either the commercial driver/vehicle or for the other vehicle. CR assignment determined event fault for the purpose of analysis.

  Truck naturalistic driving studies assign a higher fault percentage to the truck/truck driver than is found in crash investigations. That's because SCEs are captured largely via evasive maneuvers, which are more frequent among at-fault than NAF vehicles (Knipling and Bocanegra, 2008). In addition, almost all the errors of truck drivers are seen, while many errors of other drivers are unseen. In the 2008 VTTI driving hour study (Hanowski et al., 2008), 75% of the SCEs were at-fault, 25% NAF. In contrast, LTCCS trucks were at-fault in just 55% of their involvements, which included both single-vehicle and multi-vehicle events (Knipling and Bocanegra, 2008).

  Although the percentage of NAF events in naturalistic driving studies is smaller than in crash investigations, their inclusion dilutes the dataset and weakens the links between study findings and commercial driver performance. All crashes and SCEs, whether at-fault or NAF, are influenced by a multitude of ambient factors, including weather, lighting conditions, road surface, road type (i.e., divided vs. undivided), and traffic density. These factors confound all driver-related analyses, but especially those where the role of subject driver error is minimal or non-existent. Truck driver errors should play a much smaller role in NAF than in at-fault SCEs. Truck driver fatigue (and, another step removed, factors reputed to cause fatigue) *could* play a small role in NAF SCEs, but isolating and documenting that role amidst all the other interacting factors is highly problematic and easily subject to confounding. A more incisive analysis would have disaggregated the dataset by SCE fault, and used only at-fault SCEs for primary study. The residual NAF SCE dataset could have been used as a secondary, "control" subsample. If TOT, rest break, or other schedule related changes in SCE frequencies were observed and were due to driver performance changes, one would expect to see changes in the at-fault vs. NAF distribution. For example, one would expect the proportion of at-fault SCEs to increase near the end of 14-hour work shifts, and to decrease following rest breaks.

As cited above, previous VTTI studies assigned a CR to each SCE and thereby disaggregated SCEs by fault, truck vs. other vehicle. CR coding also identified SCEs triggered directly by fatigue as well as the full array of other possible causes. Among the four studies recently released by FMCSA, the two FMCSA-issued studies of Florida transit bus drivers (Sando et al., 2010a, 2010b) used a dataset consisting entirely of crashes reviewed and deemed preventable (nearly synonymous with "at-fault") by its participating transit agencies. This was the scientifically correct procedure.

- **Single-vehicle (SV) vs. multi-vehicle (MV)**. A particular disaggregation of importance in fatigue studies, and studies of driver error in general, is that of single-vehicle (SV) vs. multi-vehicle (MV) events. Disaggregating SCEs by SV vs. MV would have provided a strong validation test of any driver performance-related hypothesis. SV crashes typically indicate a failure of driver *vehicle control*, whereas MV crashes reflect primarily a failure of *response to traffic events*. Of all LTCCS crashes where truck driver fatigue was coded as an associated factor (not necessarily a contributing factor), 65% were SV crashes. For truck driver asleep-at-the-wheel as the CR, 91% were SV crashes. Corresponding percentages for other physical impairment and performance/response execution error were 76% and 74%, respectively. On the other hand, 93% of LTCCS crashes with the associated factor "traffic" were MV crashes.

  Most SCEs are not crashes, but they can and should be coded as whether a crash, had it occurred, would have been SV or MV. The "Phase I" VTTI study (Hanowski et al., 2005; Olson et al., 2005) coded number of vehicles involved (or potentially involved had the event progressed). There were 267 SV SCEs (30%) vs. 620 MV SCEs (70%). A similar disaggregation in the current study would have permitted validation of findings by assessing relative TOT and other effects on the SV and MV subsamples. To the extent that work and/or driving TOT affect SCE rates, one would expect SV effects to be greater than MV effects. If the MV effects were equal or greater, it would suggest other factors were at play.

- **Divided vs. undivided road**. As discussed earlier, *trafficway flow* (i.e., divided vs. undivided roads) exerts a strong effect on crash and SCE risk. It varies systematically within many truck trips, which often end at locations away from Interstates and other divided highways. Many truck terminals and customer locations are miles away from divided highways, and trucks must negotiate these miles to reach their end-of-trip locations. This late trip risk increase has been called "landing risk."

  In the "Phase I" truck study (Hickman et al., 2005; Knipling et al., 2005), analysts classified trafficway flow in each SCE video and in randomly selected "baseline" driving periods. Only 10% of tractor-semitrailer driving was on undivided roadways, but 38% of SCEs occurred there. This yielded an SCE odds ratio of 5.3; in other words, driving on undivided roads had 5.3 times the SCE odds of driving on a freeway or other divided road. This finding is not surprising when one considers the design differences between divided and undivided roads. On divided roads, vehicles are all traveling in the same direction at relatively uniform speeds. On undivided roads, there are traffic signals, stops and starts, crossing vehicles, turning vehicles, pedestrians, many opportunities for distraction, and little margin-of-error. The current study should also have coded trafficway flow for SCEs and random baseline epochs. This would have revealed whether road type varied systematically across work days, in particular for the last few hours of driving.

Plus, undivided roads are narrower. Lanes of local roads may be 10' or even 9' in width, whereas the Interstate standard is 12', especially on rural Interstates. Urban arterials are usually in-between at 11' or 10' (Harwood, 2003; AASHTO, 2004). The narrow the road, the closer the traffic interactions and the greater the risk. There is little discernible effect of lane width on crashes on low traffic volume roads, but effects become dramatic with increasing traffic volumes. AASHTO's (2004) estimated Accident Modification Factors for roads with traffic volumes of 2,000 vehicles per day or greater (comparisons are with 12' lanes) are: (a) 11' lanes: 1.05; (b) 10' lanes: 1.30; (c) 9' lanes: 1.50.

Variations in lane width create a risk confound, and they also create an SCE *criterion* confound related to the functioning of the in-vehicle lane tracker. This will be discussed separately, below.

Time-of-Day (TOD) and Traffic Density as Potential Confounds

TOD operates via two different mechanisms to strongly affect crash risk. The first association is with traffic density. In the "Phase I" VTTI large truck naturalistic driving study (Hickman et al., 2005; Knipling et al., 2005), trucks drove in medium-to-heavy (Level-of-Service B-F) traffic 27% of the time (based on random control observations) but had 41% of their 915 SCEs under these conditions. That's an odds ratio of 1.9 for SCE likelihood in heavy/medium vs. light traffic. In a separate re-analysis which included the "Phase I" SCEs plus more than 2,000 more (3,270 total), Wiegand et al. (2008) reported an SCE-baseline odds ratio of 7.2 between Level-of-Service C-F and Level-of-Service A-B (i.e., heavy traffic compared to light-to-medium traffic). In the previous VVTI HOS-related study, Hanowski et al. (2008) found the correlation between truck SCE rate and the national average traffic density by TOD to be 0.83, meaning that 69% of the variance in SCE rate was associated with traffic density. The 2008 report attributed its TOT-related findings primarily to the influence of traffic density. A new analysis of freeway safety (Kononov et al., 2011) finds that crash rates per vehicle mile traveled rise slowly with increased traffic density *until a critical density is met*. Then they rise sharply, much like the sigmoid (S-shaped) dose-response curve often seen with medications. In the study, a 60% increase in freeway traffic beyond the critical density caused an 84% increase in crash *rate* per vehicle miles traveled. In other words, each vehicle's crash risk almost doubled. If trucks end their trips during rush hours and especially on undivided roads during rush hours, they will have higher crash rates during these times independently of driver performance factors. The current VTTI study did not analyze SCEs in relation to traffic density nor did it discuss traffic density as a potential confound.

The second major influence of TOD is via driver alertness. Circadian rhythms strongly affect human alertness every day. The deepest circadian trough is between 4:00am and 7:00am. Another, shallower trough occurs mid-afternoon. In the DFAS (Wylie, 1996), TOD was the “strongest and most consistent factor influencing driver fatigue and alertness.” In the LTCCS, 62% of truck driver asleep-at-the-wheel crashes occurred in the *two-hour period* between 4:01am and 6:00am. Only 20% occurred in the 14-hour period between 6:01am and 10:00pm (Knipling, 2009a).

Previous VTTI naturalistic driving studies (e.g., Hanowski et al., 2008; Hickman et al, 2005; Knipling et al., 2005) have confirmed strong TOD effects on SCE involvement rates. Indeed, TOD associations with SCE rate have been among the strongest of any factor, with traffic density inescapably apparent as the underlying determinant. Figure 1 below, plotted from data published in Hanowski et al. (2008), shows that rates rise sharply with the morning rush hour, and the gradually rise further in the afternoon.

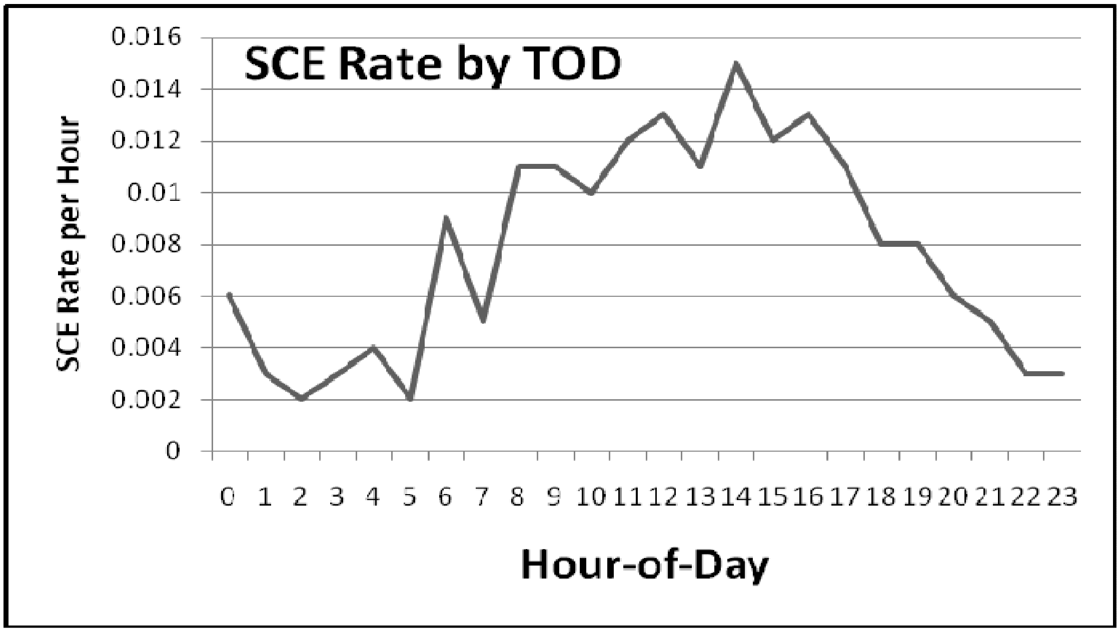

***Figure 1. SCE rate by time-of-day (TOD) for 98 drivers in Hanowski et al. (2008).* Data represent 819 SCEs over 2.3 million miles of driving.**

A TOD plot of SCE rates with SV and MV events disaggregated from the “Phase I” truck study (Hickman et al., 2005; Knipling et al., 2005) corroborates the above finding and also provides separate insights on SV and MV SCEs. Figure 2 shows relative event rate across the 24-hour day for 290 SV (i.e., truck only) and for 625 MV SCEs. These are compared to 1,072 baseline epochs to derive relative rates. Three-hour rolling averages are used here to smooth the curves. For simplicity, the rates were indexed to 1.0. Values less than 1.0 indicate reduced risk; values higher than 1.0 indicate elevated risk.

Figure 2 shows two different patterns for SV and MV SCEs. The SV relative rate line shows some fluctuations around 1.0, but no sharp deviations or consistent trends. One cannot conclude that any period of the day was consistently more or less risky for SV SCEs, not even the early morning hours with known, strong circadian influences on alertness. For MV events, however, the late afternoon hours were the riskiest. MV events were 68% of this dataset and 83% of all injury-producing truck crashes in 2008 (FMCSA Analysis Division, 2010).

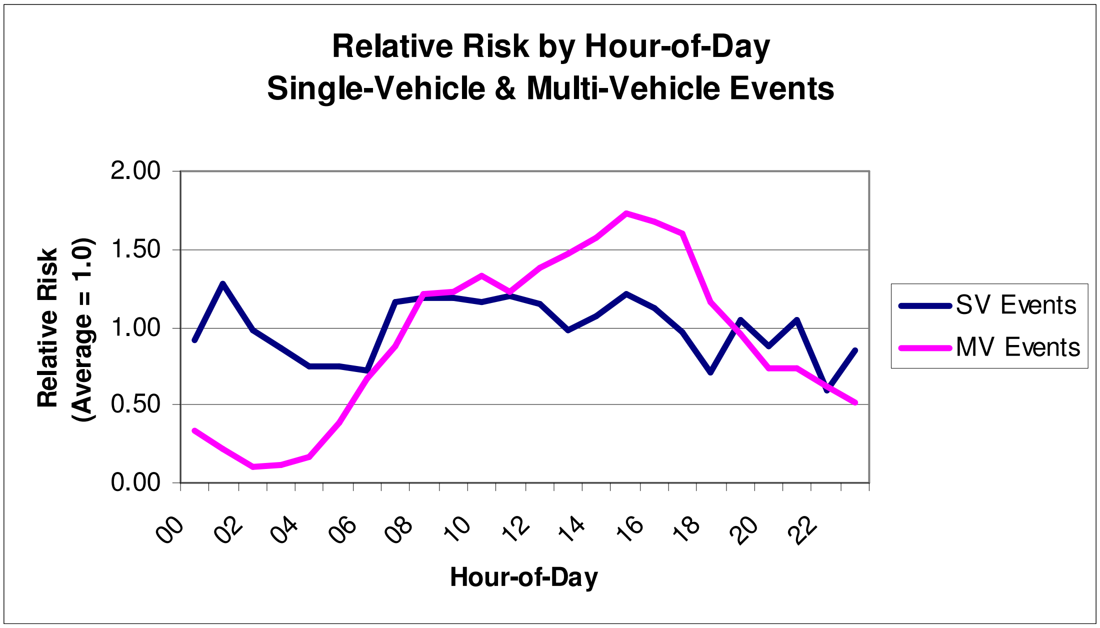

Figure 2. Relative event risk (events/exposure) by hour-of-day for long-haul CTs in the VTTI CMV naturalistic driving study (Hickman et al., 2005).

These statistics confirm TOD as a major driver of SCE risk. If TOD is not controlled, it is a confound. Further, SCE variations as a function of TOD are overwhelmingly associated with traffic density. Circadian effects were not apparent in this dataset, casting further doubt on SCE rate as a valid indicator of driver fatigue.

How could TOD have been addressed in the current study? Simple descriptive statistics and graphic presentations of SCEs as a function of TOD would have provided insight. One could use a logistic regression sequence similar to that used by Penn State in its report (Jovanis et al., 2011) released concurrently with the VTTI report. The regression sequence however, would need to have entered TOT and TOD as the initial two variables in the model, followed by their interaction terms.

A conceptually simpler approach, suggested in Knipling and Engler (2007), would be to populate the blank TOT × TOD matrix below in Figure 3. Each cell could contain an SCE count (s), exposure measure (e), and derived incidence proportion or risk (s/e). Risks could have been compared quantitatively both in vertical (time-of-day) and horizontal (TOT) directions using simple ANOVA methods. This analysis method would partially address the "landing" issue as well as the overall issue of confounding by TOD.

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

Figure 3. Blank time-on-task (hours driving) by time-of-day (TOD) matrix that could have been used to address TOD confounding. s = safety-critical events, e = exposure.

SCE Criterion Confound Related to Lane Width

Relative crash risks as a function of lane width were discussed earlier. For example, with moderate-to-heavy traffic volumes, crash risks on 10' lanes are 30% greater than those on 12' lanes (AASHTO, 2004). On low-volume roads, risk changes with lane width are not significant.

The use of a lane tracker to generate unintentional lane deviation (ULD) SCEs may have introduced a measurement/criterion confound as well. Section 2.2.5 of the report (Page 15) describes the lane tracking device. On Page 16, a ULD is defined as, "Any circumstance where the subject vehicle crosses over a solid lane line (e.g., onto the shoulder) where no hazard (e.g., guardrail, ditch, vehicle, etc.) is present. Table 7 on the next page defines the ULD trigger as, "Lane tracker status = abort. Distance from center of lane to outside of lane line less than 44 inches." The presumption here is that the authors meant to say, "Distance from center of *truck* to outside of lane line less than 44 inches."

If VTTI's criterion was <44" from the center of the truck to the lane. A typical tractor is 102" wide, or 51" from its center line to its side. Thus, for all lane widths the truck would be 7" or more over the lane edge. No one would deny that that is an undesirable lateral position for the vehicle. The question, however, is whether the criterion is reliable in relation to driver performance. Table 1 below compares four lane widths (12', 11', 10', 9') in relation to the available lane width on each side of the truck, the lane deviation criterion used, and the lateral deviation from lane center required to reach that criteria. While the lane tracker 44" criterion was consistent, the vehicle deviation from lane center required to meet that criterion varied directly and markedly with lane width. Which would have been the better criterion to assess driver performance?

Table 1. Lane Widths and Lane Deviation Criteria

Lane Width: Metric:	12' 144"	11' 132"	10' 120"	9' 108"
Distance between lane center (and centerline of perfectly centered truck) and lane edge.	**72"**	**66"**	**60"**	**54"**
ULD criterion distance between truck centerline and lane edge (i.e., vehicle encroachment toward lane edge.	**44"**	**44"**	**44"**	**44"**
ULD criterion distance between lane center and truck centerline (i.e., vehicle deviation from center).	**28"**	**22"**	**16"**	**10"**

VTTI could have addressed this conundrum in two ways. Its lane tracker was continuously measuring the distances to both lane edges, and thus it was capturing lane width. SCEs could have been disaggregated by metric to address risk confounding. Secondly, since the system could have easily calculated the vehicle deviation from center metric (last row in Table 1), an alternative ULD SCE criterion based on that could have been used for analysis. The point here is not that lane breaks on narrow roads create less risk, but rather that the algorithm used in the study to identify them introduced a criterion confound.

Differential Risk Among Driver Subjects

Page 39 of the VTTI report briefly addresses the issue of differential risk among the driver subjects. Just four of the 97 drivers had more than 100 SCEs each and 36% of all the SCEs. Their removal reduced the overall 11th driving hour SCE rate by 47% and the 11th/10th hour ratio from 1.53 to

1.06, rendering the difference non-significant. Nevertheless, "It was decided these four drivers represented a type of driver in the CMV driver population and should be included in the [overall] analysis" (Page 40).

If the scientific goals of the VTTI study were narrow – determining associations between work schedules parameters and SCEs – then this decision was correct. The four outliers were indeed part of the study sample and are probably representative of high-risk drivers within the greater CMV driver population. However, if the study had broader goals such as *understanding* CMV crash risk and effective countermeasures, then the decision was incorrect. From this larger perspective, differential risk within the driver sample was an essential topic for study.

This reviewer has never encountered a study of driver risk that did not contain compelling evidence of extreme individual differences among drivers. Instrumented vehicle studies of both commercial and non-commercial drivers reliably show that some drivers have markedly higher rates of involvement in at-fault incidents, while the remaining majority of drivers have relatively low involvement risk (Hickman et al., 2005; Barr et al., 2011; Knipling et al., 2004; Knipling, 2009a). Among 95 drivers in the "Phase I" study (Hickman et al., 2005; Knipling et al., 2005), the worst 19% of drivers had 53% of all at-fault safety-critical events (SCEs). Hour-for-hour, these high-risk drivers were 4.9 times riskier than were other drivers. Individual differences in fatigue susceptibility were even more extreme; 13% of the drivers had 66% of the high-drowsiness SCEs. Their SCE odds were 13 times those of the other 87% of the drivers. The Driver Fatigue and Alertness Study (DFAS; Wylie et al., 1996) also employed drivers with national fleets and found that 14% of the drivers had 54% of observed drowsiness (odds ratio = 7.2).

Differential fatigue susceptibility can also be demonstrated experimentally, even among healthy subjects with no sleep disorders or other notable medical conditions. In a laboratory study, Dinges et al. (1998) sleep-deprived 14 healthy adult males for 42 hours. Alertness was measured every two hours using the Psychomotor Vigilance Test (PVT) and other established tests. Six of the 14 subjects (43%) had nearly 70% of all lapses. These six subjects were almost as vulnerable to fatigue during the first 21 hours of sleep deprivation as the other eight subjects were during the second 21 hours. Some of the subjects were put through the same regimen months later; they duplicated their initial amounts and patterns of deterioration almost exactly.

Martin Moore-Ede (2007) uses the term *chronotype* to refer to each individual's set of sleep- and alertness-related characteristics. This is includes fatigue susceptibility, "morningness" vs. "eveningness," desire to nap, optimal length of sleep and naps, and sleep schedule flexibility/rigidity. Both "normal" genetic variation and medical conditions contribute strongly to each person's chronotype.

The pervasiveness of differential fatigue risk and general crash risk suggests that safety research, whenever possible, should seek to understand the human mechanisms underlying risk, not just top-level associations with risk. Based on all this evidence, would one say that further restricting all drivers' hours is the solution to driver fatigue? To this reviewer, the solution lies in identifying and restricting (or removing) high-risk drivers. This would be a targeted solution to a major underlying cause of the problem. Limiting research to top-level associations leads one to "blanket" restrictions which would be unnecessary for most drivers and inadequate for those at most risk.

Additional Comments on Specific Work Schedule Topics

By-and-large, this critique has addressed the overall study methodology rather than specific findings. Below are discussions of some specific issues relating to study findings on specific work schedule areas: workday characterization, driving hours, work hours, and breaks.

- **Workday characterization**. This reviewer appreciates the difficulties in obtaining accurate activity data from drivers, especially when the data was "24/7" for a 4-week period. The post-data collection corrections to the activity registers, as described in Section 2.5.1, appeared to be necessary and systematic. Nevertheless, the frequency and size of the driving time corrections raise possible concerns. More than half of the shifts required additions of 30+ minutes of driving time, and the average time adjusted per shift was 106 minutes (Page 20). These adjustments of nearly two hours, if not fully warranted and accurate, could skew study findings regarding TOT or TOD. This would include any comparisons across a one-hour period and even those across a two-hour period.

  In Figures 13 and 14 of the report, long-haul drivers spent 12.8% of their time resting while line-haul drivers rested for only 2.6% of their time. This large difference, combined with the many other operational differences between the two, suggests that there should have been more long-haul/line-haul comparisons for other variables in the study. The concurrent Penn State study (Jovanis et al., 2011) disaggregated all its data by long-haul (truckload) vs. line-haul (less-than-truckload).

- **Driving hours**. Almost all of this review's earlier comments have been framed in the context of the study's findings on driving hours and work hours. Although the study's parametric comparisons were between driving hour (TOT) and SCE rates, numerous hidden factors had to be operating. In spite of these potential influences, there were no overall significant effects of driving hour on SCEs and only a few specific hour-to-hour significant differences. The lack of primary effects associated with late driving hours is consistent with the previous VTTI HOS-related studies (Hanowski et al., 2005; 2008) and the bulk of the literature on TOT. This includes the DFAS instrumented vehicle study (Wylie et al., 1996), an experimental simulator study (O'Neill et al., 1999), and previous VTTI truck naturalistic driving studies (Hanowski et al., 2005; 2008). LTCCS crash data, though coving only 10 hours of legal driving and lacking a non-crash control group, show no overall differences in crash involvement patterns associated with TOT (Knipling, 2009b, 2011). The LTCCS comparisons were among three categories of crashes (truck SV, truck-at-fault MV, and other vehicle at-fault MV) with marked differences in truck driver contribution.

- **Work hours.** There were positive study findings in regard to work hour-SCE associations, specifically late hours of driving within 14-hour shifts. These effects could be due largely or entirely to hidden confounds, as already discussed. Verification of the study's work hour findings will require in-depth analyses of TOD, traffic density, road type, driver ORD/PERCLOS, driver fault, etc. Since the previous VTTI studies (Hanowski et al, 2005; 2008) found no such effects, perhaps corroboration of this study's findings via an entirely new data collection is indicated.

  The aggregation of work periods into 4-5 hour blocks (beginning: hours 1-5; middle: 6-9; end: 10-14) was justified as a way of raising sample sizes and statistical power. Yet this aggregation probably had the effect of increasing traffic density confounding, as today's urban and suburban rush hours span 4-5 hours also. Even small city rush hours typically span 3-hour periods (Shrank and Lomax, 2007). Full 14-hour tours of duty most often begin in the early morning and end just after the evening rush. Many other tours begin in the early evening and end during the morning rush hour. Most morning deliveries and pick-ups must be scheduled when receiver or shipper dock crews are on-duty; e.g., after 7:00am or 8:00am. The middle of driver work shifts usually occurs during lower-traffic mid-day hours or during night hours. Consistent with that were Table 32 statistics showing that none of the middle work hour categories had higher rates than either the beginning or end hours. Three beginning work hours had greater risk than middle hours, contrary to a simple TOT hypothesis.

  The report appears to have a computational error relating to the interpretation of work hour effects. Figure 3 (Page xvii) and Figure 42 (Page 65) present identical data but report different correlation coefficients between the variables. Figure 3 reports a Pearson R of +0.41 between work hour category and SCE occurrence. Figure 42 (Page 65) presents the same data but reports an **R** of +0.16. This difference is more than six-fold when one considers that R^2 represents SCE variance accounted for by work hour category (i.e., 16.8% of SCE variance vs. 2.5%). Visual features of the data (i.e., the wide vertical spreads and shallow line slope) suggest that the lower R and R^2 values are correct, which would suggest further that the association is very weak.

  Even if an association between late work shift driving and crash risk is validated by future research, the underlying causal mechanism is likely via *time awake.* Time awake above 16 hours is well established as an underlying physiological factor in alertness (Rosekind, 2005). It is a principal element in most Sleep Performance Models (Krueger, 2004). One of these, the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model, has been used by FMCSA to assess its HOS rules. In almost any driver schedule, work hours and driving hours co-vary with time awake. FMCSA and VTTI could isolate effects in the current study data by using the driver activity registers to determine time awake and using video review (i.e., CR, ORD, PERCLOS) to determine alertness effects. This might validate work hours as a significant schedule-related covariate.

VTTI tested the effect of removing the four highest risk drivers from the driving hour dataset, but not from the work hour dataset. Removal of high-risk drivers would likely have similar or greater effects for work hours. That's because higher risk drivers would likely be the ones pushing themselves harder (as was concluded in the VTTI Sleeper Berth study; Dingus et al., 2001). Or, if physiological factors were indeed operating, encroaching the limits of time awake would likely differentiate drivers more than encroaching maximum driving hours *per se*.

Collapsing multiple SCEs within an hour erased or reduced many 11th hour SCE effects. Page 54 states, "It is important to point out that there were no statistically significant differences when the 10th and 11th driving hours were compared. The 11th driving hour only showed a significantly higher odds ratio when compared to the 5th driving hour at the Beginning of the shift. No statistically significant results were obtained when the 10th and 11th hours were compared to other hours at the End of the shift (work hours 10 –14)."

The present study also corroborated ATRI's (2010) finding that 11th-hour safety events are a small absolute percentage of overall risk. In the ATRI data (shown as Figure 5 in the VTTI report), 11th hour crashes were just 2.0% of "legal" (hour 1-11) crashes. In the VTTI dataset, the 11th hour accounted for just 31 of 1,262 driving hours with SCEs (2.4%). Just 321.48 of 14,250.52 total exposure "opportunities" (2.3%) were in the 11th hour. Even a genuine risk elevation from late shift driving would affect only a small percentage of CMV driving.

- **Breaks**. VTTI's findings regarding the SCE effects of driver breaks were of high magnitude (30-50%) and, as noted on Page 73, "somewhat controlled" compared to other schedule-related findings spanning longer work and driving periods. Nevertheless, one may ask the same basic questions about this effect as have been asked about study effects in general. Most notably, are SCE reductions associated with actual alertness increases? This question could be answered, perhaps affirmatively, by comparing SCE CR profiles, driver ORDs, and/or driver PERCLOS before and after the break. One would also want to control for TOD, traffic density, and road type. Many commercial drivers stop for rest when they encounter heavy traffic (e.g., late afternoon rush) and then start driving again when traffic clears (early evening). Recall that SCE odds ratios can be as high as 7.2 for different levels of traffic density (Wiegand et al., 2008). Even small percentage pre- vs. post-break differences in these factors could reduce the real residual break effects.

  The study did not specifically separate naps from non-nap breaks in its analysis. This is unfortunate, since the consensus of sleep scientists (e.g., Moore-Ede, 1993; Rosekind, 2005) is that napping has far greater alertness benefits than non-napping rest. A policy requiring mandatory rest breaks may be less effective than policies and other practices (e.g., education) encouraging napping *per se*. A separate nap vs. no-nap analysis of the study data would be easy since napping was recorded on driver activity registers.

VTTI's explanation of negative binominal (NB) modeling results for the safety effects of breaks (Table 35) bears closer scrutiny. On Page 68, the authors assert that a potential explanation for the non-significance of 'breaks' and 'working hours' in the model presented in Table 35 is the high correlation (or multicollinearity) between 'breaks' and 'working hours'. They then present the Pearson correlation coefficient between the two variables as $r = 0.35$. The value $r = 0.35$ is far below the threshold of concern for multicollinearity, which typically involves variables that are correlated so highly they essentially duplicate one another (i.e., $r = 0.90+$). Furthermore, if this were a legitimate concern of the authors, it could have been readily addressed. Most statistical software packages make the option readily available to test for multicollinearity among variables in a model. Resultant statistics include the Variance Inflation Factor, or VIF, and its inverse, Tolerance. Multicollinearity is present if VIF is greater than 10 or Tolerance is less than 0.10, which should then be reported. Conversely, a lack of multicollinearity could easily have been ruled out (and would have been had the respective statistics been generated and reported).

Assuming that the study's rest break findings are robust enough to survive closer scrutiny as suggested above, an analytical modeling question is whether or not the time utilized for mandatory breaks would cause risk increases later in the work tour due to greater traffic exposure or to validated work hour effects on safety.

Conclusions

Naturalistic driving has immense capabilities but also critical limitations in regard to crash research. Most troubling for the current study is the validity question; are unfiltered SCEs valid indicators of driver fatigue? Given the small percentage of SCEs that can be shown to be fatigue-related and the apparent *negative* relationship between driver drowsiness measures (i.e., ORD and PERCLOS) and SCE rates, SCE validity for fatigue studies is dubious.

In addition to the core validity question are several related concerns about naturalistic driving. This includes the relevance of SCEs to serious crashes, the synthetic composition of SCE datasets (constructed via multiple trigger types, each with its own *a priori* criterion), national representativeness, and the high sensitivity of SCEs to road and traffic conditions.

Naturalistic driving studies are expensive, and perhaps it was cost that prevented FMCSA from funding additional, proven analysis checks of the VTTI data. Most notably, SCE TOD was not analyzed as a companion and "competing" variable to TOT – this in spite of the fact that VTTI's previous large HOS-related naturalistic driving study (Hanowski et al., 2008) attributed its *primary findings* to TOD-related traffic density.

Close behind the lack of TOD analysis was the lack of driver alertness-related analysis. This includes determining SCE critical reasons (CRs), driver ORD, and/or driver PERCLOS. Lane deviations (ULDs) were added to the SCE mix to improve its face validity in relation to fatigue, even though another previous VTTI study (Olson et al., 2009) had determined that ULDs were dominated (77.5%) by driver distraction. Moreover, distraction and drowsiness appeared to be negatively related. In addition, the ULD criterion used may have resulted in lane-width-related criterion confounding.

Multiple other disaggregations of the data (e.g., at-fault vs. NAF; SV vs. MV; trafficway flow; traffic density) might have addressed the issues of SCE validity and TOD confounding. Driver performance deterioration as a function of work schedule was the implicit, underlying explanation for positive study findings. Disaggregating the dataset by at-fault vs. NAF and by SV vs. MV would have provided easy and compelling validation checks. Comparing SCE event distributions to random baseline epochs from the same driving hours (and other schedule conditions) would have determined whether these confounds were operating within the TOT and other work schedule comparisons. Data mining of inter-subject differences would have shown whether observed effects were best understood as schedule effects or differential driver risk effects. The two would have different implications for countermeasures and for policy.

All of the specific findings on the study's four schedule topics (work characterization, driving hours, work hours, breaks) should be viewed through the lens of the above-stated methodological concerns. Nevertheless, several observations about them are warranted. Overall driving hour effects on SCEs were not significant. Late driving hour effects were seen for work hours at the end of 14-hour shifts, though there was a notable discrepancy in the reported correlation coefficient quantifying that effect. The SCE-reduction effects of breaks were considerable but require several easily performed validation checks before they can be accepted as true fatigue-reduction effects.

In some ways, the naturalistic driving methodology is a Pandora's Box. It opens a treasure trove of data for in-depth analysis and scientific understanding. Yet its full contents cannot be restrained or ignored. This study was based on several narrow HOS-based hypotheses and a seemingly restricted analysis plan. The data collected, however, contains much more than that. It contains almost everything needed to validate (or invalidate) study HOS-related conclusions, as well as to provide deeper insights into driver fatigue and safety. These more probing and self-challenging analyses must be performed before study findings can be accepted as sound science.

Cited References

American Association of State Highway Transportation Officials (AASHTO). *A Policy on Geometric Design of Highways and Streets. Washington, DC., 2004*

American Transportation Research Institute. (2010). *Hours-of-service rules safety impacts: 2010 Analysis* (Technical Report). Arlington, VA: American Transportation Research Institute.

Anastasi, A. *Psychological Testing, Third Edition*. MacMillan Publishing Co., 1968.

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.

Blanco, M., Hickman, J., Olson, R., Bocanegra, J., Hanowski, R. J., Nakata, A., et al. (in press). *Investigating critical incidents, driver restart period, sleep quantity, and crash countermeasures in commercial vehicle operations using naturalistic data collection.* Washington DC: Federal Motor Carrier Safety Administration, USDOT.

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., 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*. Interim Project Report for NHTSA Contract No. DTNH22-00-C-07007, Task Order 6; 2006.

Dingus, T., Neale, V., Garness, S., Hanowski, R., Keisler, A., Lee, S., Perez, M., Robinson, G., Belz, S., Casali, J., Pace-Schott, E., Stickgold, R., and Hobson, J. (2001). Impact of sleeper berth usage on driver fatigue, final project report (Report No. FMCSA-RT-02-050). Washington, DC: Federal Motor Carrier Safety Administration, USDOT.

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

FMCSA Analysis Division. *Large Truck Crash Facts 2008*. FMCSA-RRA-10-043, March 2010.

Hanowski, R.J., Dingus, T.A., Sudweeks, J.D., Olson, R.L., and Fumero, M.C. *Assessment of the Revised Hours-of-Service Regulations: Comparison of the 10th And 11th Hour of Driving Using Critical Incident Data and Measuring Sleep Quantity Using Actigraphy Data*, Virginia Tech Transportation Institute, report prepared for the FMCSA, June 2, 2005.

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 LocalShort 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.

Hanowski, R.J., Hickman, J., Fumero, M.C., Olson, R.L., and Dingus, T.A. (2007). The sleep of commercial vehicle drivers under the 2003 hours-of-service regulations. *Accident Analysis and Prevention, 39*(6), 1140–1145. 2007

Hanowski, R.J., Hickman, J.S., Olson, R.L., and Bocanegra, J. (2009). Evaluating the 2003 revised hours-of-service regulations for truck drivers: The impact of time-on-task on safety-critical event risk. *Accident Analysis and Prevention*, 41, 268–275.

Hanowski, R.J., Olson, R., Bocanegra, J., Hickman, J., Dingus, T.A., and Sudweeks, J.D. (in press). *Safety-critical events that occur in the 10th and 11th hour of driving in commercial vehicle operations: Does risk increase in the 11th hour?* (Technical Report). Washington DC: Federal Motor Carrier Safety Administration, USDOT.

Harwood, D. Roadway design comments in, *Future Truck and Bus Research Opportunities*. TRB Conference Proceedings 38. ISSN 1073-1652, ISBN 0-309-09422-4, 2006, Pp. 85-86.

Harwood, D.W., Torbic, D.J., Richard, K.R., Glauz, W.D., and Elefteriadou, L. *Review of Truck Characteristics as Factors in Roadway Design*. NCHRP Report 505, ISSN 0077-5614, ISBN 0-309-08779-1, 2003.

Hickman, J.S., Knipling, R.R., Olson, R.L., and Hanowski, R.J. (in press). *Commercial vehicle data collection & countermeasure assessment project. Phase I: Preliminary analysis of data collected in the drowsy driver warning system field operational test—Task 3: Preliminary analysis of partial countermeasure data.* Contract No. DTNH22-00-C-07007, Task Order 21. Washington, DC: Federal Motor Carrier Safety Administration, USDOT.

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.

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. What does instrumented vehicle research tell us about crash risk & causation? *Journal of the Washington Academy of Sciences*, 94, No. 3, Pp. 61-78, Fall, 2008.

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. The good, the bad, and the ugly: three large truck crash categories and what they tell us about driver fatigue.; submitted to Docket FMCSA-2004-19608, June 7, 2011.

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., Boyle, L.N., Hickman, J.S., York, J.S., Daecher, C., Olsen, E. C. B., and Prailey, T.D. *Synthesis Report #4: Individual Differences and the High- Risk Commercial Driver*. TRB Commercial Truck & Bus Synthesis Program. ISSN 1544-6808, ISBN 0-309-08810-0, 2004a.

Knipling, R.R. & Engler, D. Peer review critique of Penn State Jovanis et al. report: Crash Risk and Hours Driving: Interim Report II. Prepared for the ATA and submitted to FMCSA HOS Docket, September, 2007.

Knipling, R.R. Hanowski, R.J.; Hickman, J.S., Olson, R.L., Dingus, T.A. and Carroll, R.J. Exposure-risk analysis of large truck naturalistic driving data. *Proceedings of the 2005 Truck & Bus Safety & Security Symposium*, Alexandria, VA, November 14-16, 2005.

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.

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

Moore-Ede, M. *Consensus Between Circadian/Sleep Experts and Truck Drivers on the Impact of 2005 U.S. Federal HOS Split-Sleep and 14-Hour Clock Regulations on Truck Driver Sleep and Alertness*. Circadian International, Inc. 2007.

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.

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.

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.

Sando, T., Angel, M., Mtoi, E., & Moses, R. Analysis of the relationship between operator cumulative driving hours and involvement in preventable collisions. Paper 11-4165 submitted to the TRB Annual Meeting, Session 665, Submitted November, 2010a.

Sando, T., Mtoi, E., & Moses, R. Potential causes of driver fatigue: a study on transit bus operators in Florida. Paper 11-3398 submitted to the TRB Annual Meeting, Session 665, Submitted November, 2010b.

Schrank, D. & Lomax, T. *The 2007 Urban Mobility Report*. TTI. DOT Grant No. DTRT06-G-0044, September 2007.

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

Wiegand, D.M., Hanowski, R.J., Olson, R., & Melvin, W. *Fatigue Analyses from 16 Months of Naturalistic CMV Driving Data*, Draft Final Report, FHWA Report No. TBD, in press, 2008a.

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. *Unit Costs of Medium & Heavy Truck Crashes*. Final Report, Pacific Institute for Research & Evaluation for FMCSA, March 2007.

Author Bio:

Dr. Ronald R. Knipling is the author of the first and only comprehensive textbook on large truck safety, *Safety for the Long Haul; Large Truck Crash Risk, Causation, & Prevention*. In recognition of the book, he received the International Road Transport Union (IRU) Order of Merit award, the first given to an American scientist. Dr. Knipling has more than 30 years experience in traffic safety with emphasis on driver human factors and motor carrier safety. Specialty areas include crash data analysis, driver risk, crash causation, driver fatigue, hours-of-service, naturalistic driving, technology assessment, and carrier safety management. Accomplishments include nearly 250 technical reports, papers, and conference presentations.