ELSEVIER

Accident Analysis and Prevention 38 (2006) 1127–1136

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

The development of a naturalistic data collection system to perform critical incident analysis: An investigation of safety and fatigue issues in long-haul trucking

Thomas A. Dingus[a,1], Vicki L. Neale[a,2], Sheila G. Klauer[a,*], Andrew D. Petersen[a,3], Robert J. Carroll[b]

[a] Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg, VA 24061, USA

[b] Federal Motor Carrier Safety Administration, 400 Virginia Avenue SW#600, Washington, DC 20024, USA

Received 4 March 2005; received in revised form 26 April 2006; accepted 2 May 2006

Abstract

Traditionally, both epidemiological and empirical methods have been used to assess driving safety. This paper describes an alternative, hybrid, naturalistic approach to data collection that shares advantages with each traditional approach. Though this naturalistic approach draws on elements of several safety techniques that have been developed in the past, including the Hazard Analysis Technique, instrumented vehicle studies, and fleet studies of driving safety interventions, it has a number of unique elements. Sophisticated instrumented vehicles collected over 400,000 km of commercial vehicle data to address the long-haul trucking application described in this paper. The development of this data collection and analysis method and data collection instrumentation has resulted in a set of valuable tools to advance the current state-of-the-practice in driving safety assessment. An application of this unique approach to a study of long-haul truck driver performance, behavior, and fatigue is described herein.

Keywords: Driver fatigue; Naturalistic truck driving study; Critical incident analysis

1. Introduction

This paper describes the application of an innovative approach for assessing the problem of fatigue in the truck driving industry. The technique used is a critical-incident-based, naturalistic data collection approach using sophisticated instrumented vehicles to accurately and efficiently collect and analyze over 400,000 km of commercial vehicle data. A brief description of the unique aspects of this approach, as applied to the context of long-haul fatigue, is provided below.

Although crash statistics and case studies reported in the literature indicate that driver fatigue is a major cause of long-haul truck crashes (e.g., NTSB, 1990, 1995; Haworth et al., 1988), there exists no accurate relationship of fatigue causal factors (e.g., time-on-task, circadian effects, and low quality of sleep) to other crash contributing factors (e.g., night visibility or interaction with other traffic). Possible reasons for this are the lack of a specific check box for fatigue on some state police report forms, lack of firm evidence to support a finding of fatigue, lack of awareness on the part of the drivers of their own fatigue, and the significant number of crashes involving drift out of lane, which is not cited as drowsiness related (Knipling and Wang, 1994). Nonetheless, survey and focus group research indicates a severe problem. For example, in a survey of drivers at truck inspection stations, Braver et al. (1992) found that 19% of tractor–trailer drivers reported falling asleep at the wheel one or more times during the previous month. In addition, focus group results reported by Neale et al. (1998) specify several contributors to fatigue on the road, including issues with single and team driving arrangements, equipment, and rest stop facilities.

The significant shortcoming in the quality and type of precrash data available from traditional epidemiological crash databases, in conjunction with massive data requirements for analysis (i.e., many crashes and/or tens of millions of miles

* Corresponding author. Tel.: +1 540 231 1564; fax: +1 540 231 1555.
E-mail addresses: tdingus@vtti.vt.edu (T.A. Dingus), vneale@vtti.vt.edu (V.L. Neale), charlie@vtti.vt.edu (S.G. Klauer), andy@vtti.vt.edu (A.D. Petersen), Robert.carroll@fmcsa.dot.gov (R.J. Carroll).
[1] Tel.: +1 540 231 1501; fax: +1 540 231 1555.
[2] Tel.: +1 540 231 1514; fax: +1 540 231 1555.
[3] Tel.: +1 540 231 1516; fax: +1 540 231 1555.

0001-4575/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2006.05.001

of data), have spawned a variety of empirical approaches for assessing driving safety. These empirical methods utilize driving simulators, test tracks, or on-road instrumented vehicles to collect driving performance data and other safety surrogate measures, so called because they do not directly measure safety in terms of crash frequency or severity (Dingus, 1995). The measures that are often collected as part of experimental research, such as vehicle speed, eye movement, and lane position data, provide valuable insights for safety assessment. However, they are largely unproven as safety surrogates because they have not been directly tied to crash causal factors. In addition, the ability of experiments to accurately simulate complex driving environments is often difficult to assess especially because the empirical approach can lead to modified driver behavior. Therefore, the results of such research can provide insight into relative risk but cannot be directly translated into risk of injury or fatality.

An alternative to these traditional methods is something of a hybrid approach. This approach utilizes empirical naturalistic data collection on a large scale. The scale is typically much smaller than epidemiological requirements, but it still requires substantial data to be collected and analyzed. This has several advantages. First, the data can be analyzed via a critical incident and near crash approach because these occur more frequently than crashes. Second, there is growing evidence that critical incidents and true near crashes provide a valid safety surrogate relative to many traditional empirical measures (Dingus et al., 1999). Third, the collection of natural driver behavior provides important information about the complex circumstances and scenarios that lead to crashes; this information could be useful for mitigating crashes.

This incident-based, naturalistic approach draws on elements of several safety techniques that have been developed in the past, including the Hazard Analysis Technique, instrumented vehicle studies, and fleet studies of driving safety interventions. The Hazard Analysis Technique is a method successfully used in other areas of safety research to determine the relative safety of a system using a smaller, more detailed data set of unsafe acts or near misses (Heinrich et al., 1980). Using this technique, data are collected to identify incidents where potentially unsafe actions, behaviors, or equipment malfunctions have occurred. These incidents are then classified as either safety-related errors or as near crashes (in surface transportation applications). Hanowski et al. (2003) operationally defined safety-related errors for driving research as "precautionary braking, swerving, or lane change with minimal risk of a near crash". Dingus et al. (1995) also operationally defined a near miss as "a rapid controlled or uncontrolled acceleration, deceleration, swerve, lane change, or stopping to avoid a crash".

The Hazard Analysis Technique has been successfully applied to driving safety by several researchers (Dingus et al., 1995, 1999; Koziol et al., 2000). In the Dingus et al. (1995) study, an in-vehicle map display was tested and evaluated in a naturalistic driving study in which subjects drove a camera-equipped, instrumented vehicle on a given route through both urban and rural areas. The study identified the map display that produced the fewest critical errors and best overall driving safety performance.

Hanowski et al. (2000) used a similar Hazard Analysis Technique in a local short-haul truck driving study. Short-haul trucks were instrumented for collection of continuous driving performance and video data. Video data reductionists reviewed the continuous video and identified critical incidents, near crashes, or crashes when triggers appeared on the video. This study found that fatigue does indeed play a significant role in safe driving ability.

1.1. Assessing safety-related fatigue issues for long-haul truck drivers

Assessing driver performance, behavior, and fatigue in the long-haul truck environment required a significant advance in the critical incident data collection methods described above due to several factors. First, the numerous data collection system requirements (75 electronic data measures and 4 video channels) in conjunction with the long trip lengths (up to 10 days) precluded continuous data collection due to storage constraints. Second, the requirements for reduction and analysis of continuous video data were beyond the resources available for the project. Third, the opportunity to analyze and classify a large number of critical incidents, true near crashes, and precrash factors for even a few crashes was greater than ever before. This provided the opportunity to create a new method to structure, analyze, and store the data, with this new method having elements of both an empirical study and a crash database. Fourth, the data were purposely structured with the ability to answer many additional questions beyond the immediate objectives of the current application.

As will be shown via the long-haul truck application described in this paper, the development of this data collection and analysis method and data collection instrumentation have resulted in a valuable set of tools to advance the current state-of-the-practice in driving safety assessment.

2. Method

2.1. Drivers

The long-haul truck study included 47 male and 9 female participants, constituting 13 teams (2 drivers) and 30 single drivers. All participants were recruited from one of four for-hire commercial trucking companies and were paid up to $630 for full and cooperative participation. The average age of the drivers was 43 years (range: 28–63 years old) with an average of 13 years of driving experience (range: 1–42 years of driving experience).

The drivers used two VTTI-owned tractors to perform the data collection while hauling their normal cargo on regularly scheduled revenue-producing runs. These tractors included a 1997 Volvo L4 VN-series tractor and a 1995 Peterbilt 379. Functionally identical instrumentation packages and data collection systems were installed in each truck.

Table 1
Trigger types and descriptions

Trigger type	Description
Steering	Driver turned steering wheel faster than 3.64 rad/s
Lateral acceleration	Lateral motion equal or greater than 0.3 g
Longitudinal acceleration	Acceleration or deceleration equal to or greater than 0.25 g
Critical incident button	Activated by the driver upon pressing a button located on the dashboard when an incident occurred that he or she deemed critical
Lane deviation	Driver crossed the solid lane border (Boolean occurrence)
Time-to-collision	Truck came within a time-to-collision (range/range-rate) of 4 s from a lead vehicle
Perclos	Perclos monitor detected that eyes were closed 8.0% of any given 1-min period
Karolinska sleepiness rating	Driver subjectively assessed own drowsiness as extremely fatigued/difficult to stay awake (rating of 7 or above on sleepiness scale)
Karolinska sleepiness rating, no response	Driver did not respond to the Karolinska rating query
Timed trigger	Baseline data triggered pseudorandomly every 45–75 min
Lane departure and steering	Lane departure (tractor crossed a lane line) immediately followed by a steering event. Disabled if turn signal was activated

2.2. Critical incident method

A critical incident was operationally defined as a measured variable that exhibits a predetermined signature or exceeds a trigger criterion that may be indicative of fatigue, a lapse in performance, a safety-related external event, or potentially hazardous driving behavior. The system relied upon a set of a priori criteria to determine the presence of a critical incident. The types of triggered events for which the data acquisition systems (DASs) had been programmed are shown in Table 1. Values listed were determined experimentally. Whenever the system detected one of these triggers, the computer collected video and driving performance data for a period of 1.5 min before and 0.5 min after the triggering event.

2.3. The data acquisition system

The DAS was created by the Virginia Tech Transportation Institute (VTTI) Technical Operations staff. Any commercially available components integrated into the instrumentation package are specifically noted in the following system description.

The core of the DAS was a Pentium-based laptop computer that ran custom software and communicated with a distributed data acquisition network. A power inverter supplied the power from the truck batteries. The total system required less than 150 W of power and did not draw sufficient current to damage or drain the batteries. While the truck was in motion, the video, vehicle and driver performance, environmental, and sleep monitoring subsystems were all operational. During extended periods of inactivity, selected subsystems were deactivated.

The small size of the individual components used in the systems allowed them to be concealed in locked compartments under the bunks in the sleeper berth of each truck. The only devices visible to the drivers were a Perclos alertness monitor, a panel with subjective sleep scale pushbuttons, and a critical-incident pushbutton installed on the trucks' instrument panels as described below.

Driver interaction with the DAS was minimal. The drivers' only required activities were to: (1) indicate when an unusual critical incident occurred by pressing a button located on the dashboard, (2) respond to pseudorandom inquiries as to their subjective feelings of sleepiness by pressing one of nine buttons mounted on a panel on the dashboard, (3) don the sleep monitoring device before going to sleep, and (4) rate their subjective feelings of sleepiness when they woke up by pressing one of nine buttons on a panel mounted in the sleeper berth. No other operation of data collection hardware was required of the drivers. For the interested reader, details of the configuration of the data collection system are explained in Dingus et al. (2001). A few highlights of the system are noted below.

The video cameras were strategically located to observe the driver's face, road scene ahead of the truck, and view down either side of the tractor. The multiplexed video signal (Fig. 1) and the audio signals from the two microphones were then passed to a bank of VCRs that made up the real-time triggering-event recording system.

The DAS recorded various vehicle- and driver-related performance measures, including vehicle velocity and pedal position

Fig. 1. Multiplexed image showing the scenes observed by the four strategically located video cameras (clockwise from upper left): the forward view, the driver's face (obscured here to preserve the driver's anonymity), and the views down the left and right sides of the tractor. The letters on the left are critical incident codes (S: steering, T: time-to-collision), and the numbers are video frame numbers to synchronize the video and computer data.

as well as steering wheel velocity and position. Abrupt steering wheel movements, measured by an absolute optical encoder, were used to trigger critical incidents. Lateral, longitudinal, and vertical accelerations experienced by the tractor were also used as critical incident triggers.

A radar-based front-to-rear crash-avoidance sensor was installed for the purpose of using time-to-collision with a forward vehicle as a trigger criterion. Eaton-Vorad radar sensors were integrated into the data collection systems on the tractors. Similar sensors were also used as side clearance detectors.

Lane position was recorded with a lane-tracking device (SafeTRAC) based on forward-looking machine vision technology. Unintended lane deviation was used as a trigger criterion.

To determine the effects of variations in sleep quality on driving performance and potentially unsafe driving behavior, it was important to assess driver alertness while driving. The Perclos system (Grace et al., 1999; Wierwille et al., 1994) monitors eyelid position through the use of infrared light and machine vision technology to locate and assess the relative size of the driver's corneal reflection. To measure subjective sleepiness, a bank of nine switches was placed within easy reach of the driver. Each switch corresponded to one of the nine ratings used in the Karolinska Sleepiness Scale (Gilberg et al., 1994). A second set of identical switches was placed in the sleeper berth. The driver was prompted to rate subjective feelings of sleepiness by pressing the appropriate switch at pseudorandom intervals ranging from 45 to 75 min.

A pushbutton was mounted on the dashboard of each truck to be activated by the drivers if they experienced a critical incident. This system acted as a redundant cue to trigger the data collection system in the event that the driver detected some kind of critical or unusual incident that did not have a signature detectable by the system.

2.4. Training the drivers

An experimenter met with each driver prior to participation to present detailed instructions on tasks, procedures, and an orientation to the heavy vehicle. The experimenter initially gave a brief overview of the study, informed the driver of the nature of the study, and inquired if the driver was agreeable to participating (all drivers who met with the experimenter agreed to participate). After the driver verbally agreed to participate, the experimenter asked the driver to read and sign an informed consent form.

After explaining the questionnaires and nightcap operations, the experimenter gave the drivers an orientation to the truck cab. The entire training procedure lasted approximately 60–90 min.

2.5. Procedure for data reduction and data preparation for analysis

The video data were recorded on as many as seven 8-mm videotapes. A performance file contained all of the driving performance data. Finally, drivers completed two types of surveys during the course of their trips.

The complex nature of the data required that data go through an extensive reduction process as described below. Five video reduction analysts were trained to reduce the video data for this project. All of the analysts had prior experience reducing video for other on-road studies. All of the analysts were trained in three sessions: first the trainer demonstrated the reduction software, then the trainee worked with another experienced video analyst, and finally the analyst practiced reducing video with the trainer present.

The trainer performed spot checks of all the analysts' work throughout the entire data reduction process. The trainer held weekly meetings with all the analysts to discuss any questions or issues that developed from either the spot checks or incidents that the reduction analysts encountered. The weekly meetings and spot checks were critical in maintaining a high level of consistency between data reductionists.

2.5.1. Event description

After viewing the video for an entire event, the analyst determined the criticality of each trigger, which could be categorized as either "invalid", "noncritical", or "critical". A trigger was considered invalid if it was activated when it was clearly not appropriate, for example, a time-to-collision activated by a vehicle in an adjacent lane. A trigger was deemed noncritical if the corresponding behavior (i.e., the cause of the trigger) was a safe and legal driving maneuver that nevertheless met the trigger criteria. The timed (i.e., baseline driving) triggers were always considered noncritical.

The trigger severity category described the apparent danger of the event, ranging from no danger (driver error without hazard) to physical contact with another object or person (collision). The four trigger severity types are listed and operationally defined in Table 2, where the trigger description category reflects the

Table 2
Trigger severity definitions

Trigger severity type	Description
Collision	Any contact between the truck and any other fixed or moving object, animal, or pedestrian
Near collision	Any conflict between moving vehicles or situation of very close speed/distance proximity between the truck and any other fixed or moving object, animal, or pedestrian that required a rapid, evasive maneuver to avoid a crash
Driver error with hazard present	The commitment of a driving error such as an unplanned lane deviation, improper lane change with a vehicle present, judgment error related to tailgating, etc. in close proximity to another vehicle or fixed or moving object, animal, or pedestrian that *did not* require a rapid evasive maneuver to avoid a crash
Driver error without hazard present	The commitment of a driver error, as described above, where there was not close proximity to another vehicle or fixed or moving object, animal, or pedestrian

Table 3
Trigger description definitions

Trigger description	Definition
Undetermined	The cause of the event cannot be determined from the video. May be due to the event being out of camera view, etc.
Normal driving	The driver is exhibiting safe driving behavior and is following all rules of the road. Must only be used to describe a noncritical or invalid trigger
Obstacle present	There is an unexpected obstacle in the driver's path, *excluding other vehicles*. May be used when the driver reacts to a pedestrian, debris, or an animal in his or her path
Other vehicle present	*Another vehicle* obstructs the driver's path and the driver is not at fault
Impediment present	The driver must react to an unexpected but deliberate traffic obstruction such as construction zone traffic cones, a police officer directing traffic, or a speed bump
Driver distraction	The event is the result of the driver's inattention to the primary driving task
Judgment error	The driver exhibits poor judgment in driving in an otherwise safe situation, for example, cutting off another driver or following another vehicle too closely
Other	Any event that cannot be categorized by the above descriptions

apparent cause of the event. The eight trigger description types are listed and operationally defined in Table 3.

2.5.2. Road conditions

Nine road condition categories described the general environment and traffic conditions at the time of the event. The analyst indicated the number of lanes in a single direction on the roadway in the "number of lanes" box. The traffic density was rated by the analyst using the level of service definitions (Transportation Research Board, 1997). The environment and traffic were further described by the following six categories: road geometry, visibility, road type, road condition, weather, and exterior illumination. Table 4 outlines the six categories and their respective options. The analyst could also enter any relevant information not covered by the above categories in the "other" category. Finally, the analyst wrote a short narrative of the event in the detailed trigger description box. The narrative included any relevant information concerning the situation or environment.

2.5.3. Observer rating of drowsiness (ORD)

In the interest of accurately quantifying each driver's level of fatigue, analysts made an observer rating of drowsiness (ORD) at 1-min intervals during each event epoch. The ORD scale was based upon that developed by Wierwille and Ellsworth (1994), which is a form of the Likert (descriptive graphics) scale. The continuous scale contains five descriptors: not drowsy, slightly drowsy, moderately drowsy, very drowsy, and extremely drowsy. Judgments of drowsiness were made from the driver's facial expression, body movements, and eyelid droop based on the technique described by Wierwille and Ellsworth.

3. Results and discussion

The data collection system and reduction process previously described resulted in a data set capable of addressing many hypotheses related to long-haul truck driving. The data set has many classification variables, similar to a crash database, and detailed measurement data, like a traditional instrumented vehicle study. Hypotheses related to driver fatigue, distraction, truck-car interaction, hours-of-service regulation compliance, driving and sleep patterns, time-on-task effects, and driver experience and age effects, among others, could all be studied using these data. This is the strength and perhaps the primary contribution of this method.

These analyses evaluated measures of driving performance described in the previous sections including the number of, severity of, and causal factors associated with critical incidents detected during the data collection runs. The number and type of critical incidents and driver errors were analyzed across the major independent variables associated with this study. Those discussed in the following sections include driver type (single versus team) and time of day.

For selected analyses, given that each driver or team of drivers operated the instrumented vehicle for varying lengths of time each day and for varying numbers of days, rate values were

Table 4
Road condition categories

Road geometry	Visibility	Road type	Road condition	Weather	Exterior illumination
Straightaway	Unlimited	Parking lot/loading area	Dry	Clear/dry	Dawn
Curve left	Rain	Alleyway	Wet	Cloudy	Daylight
Curve right	Snow	One-way road	Icy/snow	Drizzle	Dusk
S-curve	Fog	Rural undivided	Gravel/sand on road	Hard rain	Night
Intersection on a straightaway	Darkness	Rural divided (median)	Gravel road	Light snow	Other
Intersection on a curve	Glare from sun	Rural divided (lane)	Other	Hard snow	
Loading area/parking lot	Glare from headlights	Urban undivided		Sleet	
Merge lane from right	Twilight (dusk/dawn)	Urban divided (median)		Other	
Merge lane form left	Other	Urban divided (lane)			
Other		Other			

calculated to account for exposure. Calculating the frequency of incidents that occurred for each hour of driving allowed the data to be compared while controlling for these exposure differences. For example, many more drivers drove at 3 p.m. than at 3 a.m. Thus, any comparison in the raw frequencies of incidents would be somewhat misleading. The number of critical incidents per hour of shift and the number of critical incidents per hour of day were also calculated for each driver to determine whether drivers were involved in more or fewer critical incidents over the course of a day as fatigue potentially became more of an issue.

For the analyses described below, 13 single drivers and 7 sets of team drivers (27 drivers total) were chosen from the data set based upon the number of days that data were available. When there were fewer than 3 days of data collected (due to truck failure or other extenuating circumstances), the driver was not included in the participant pool.

3.1. *Number of critical incidents and driver errors for single and team drivers*

A chi-square test determined if there was a difference in the frequency of critical incident triggers and timed triggered events for single and team drivers. The result from this chi-square test was significant, χ^2 (1, $N = 4841$) = 167.34; $p < 0.01$. The frequencies for the triggered events (i.e., critical incidents) and time-triggered (i.e., baseline) events for single and team drivers are shown in Fig. 2. As can be seen, single drivers had substantially more critical incidents than did team drivers. The total number of critical incidents for single and team drivers was 1898 and 564, respectively. The mean number of critical incidents for the two groups of drivers was 146.0 for the single drivers and 40.3 for the team drivers. A *t*-test of these means indicated that the difference in the number of critical incidents was also significantly different between the two groups, $t_{24} = 13.16$; $p < 0.01$.

As a check of the amount of data collected (number of hours driven) per driver type, Fig. 2 also shows the number of time-triggered events for each driver. Recall that time-triggered events were baseline events recorded every 45–75 min for each driver. The total number of time-triggered events was 1024 for single drivers ($M = 78.77$) and 849 for team drivers ($M = 60.64$). A *t*-test of these mean values indicated that these values were not significantly different, $t_{25} = 1.61$; $p > 0.05$. Therefore, the number of hours driven by single drivers versus team drivers cannot account for the large discrepancy in the number of critical incidents for single drivers versus for team drivers.

The findings regarding age ($t_{39} = 0.04$; $p > 0.05$) and experience ($t_{36} = 2.75$; $p > 0.05$) also suggest that these factors did not play a role in the increased number of critical incidents for single drivers. The point of this discussion is to highlight the surprisingly large discrepancy between the number of critical incidents involving each group of drivers; this discrepancy cannot be accounted for by the number of hours driven, age of drivers, or driver experience. This non-result suggests that the reason for fewer critical incidents for team drivers is that they drove more conservatively and cautiously than single drivers, perhaps to allow their partners a smooth ride and quality rest.

This apparent consideration factor clearly also has substantial safety implications. The drivers who participated in the focus groups (Neale et al., 1998) commented that it was very important to team with drivers who were conscientious of their partners sleeping behind them. A team driver must take certain precautions that will promote a smoother ride out of consideration for the partner, such as allowing a greater following distance to minimize the need for quick braking maneuvers and, in general, driving less aggressively. The focus group participants explained that it was necessary to have a driving partner that one could "trust with your life". In order to investigate this second hypothesis, the data were examined to look at whether many of the teams in this study were regular teams or teams who drove together just for the purpose of participating in this study. Two teams had driven together for only a few months, and only one team had never driven together. Aside from this one team, it could be assumed that the team drivers were relatively comfortable with their teammates' driving ability prior to this study.

Therefore, single drivers were involved in over three times the number of critical incidents that team drivers were. It is also important to note that 8 of the 13 single drivers used in this analysis had a greater number of critical incidents than *any* of the team drivers.

3.2. *Number of critical incidents by hour of day*

The ANOVA for critical incident rate (dependent variable) for this analysis was a two-way, mixed-factor design with driver

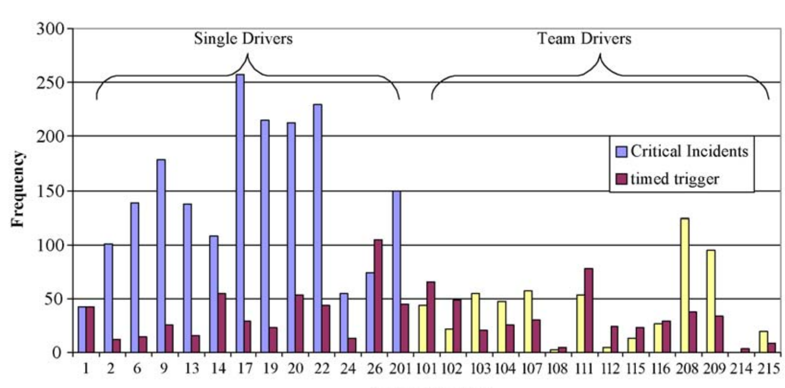

Fig. 2. Frequency of critical incidents and timed triggers for single and team drivers.

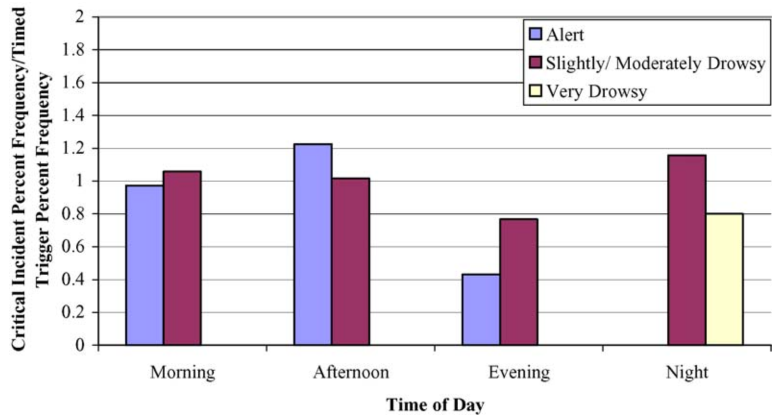

Fig. 3. Ratio of critical incident percent frequency to timed trigger percent frequency for team drivers.

type (single, team) and hour of day as independent variables. The results showed that the driver type, $F_{1,25} = 8.49$ and $p = 0.007$, and hour of day, $F_{23,480} = 1.93$ and $p = 0.006$, were significant. The highest critical incident rates occurred during daylight hours between 11 a.m. and noon and between 3 p.m. and 6 p.m. In contrast, the lowest rates occurred during the hours starting at 8 p.m., 9 p.m., and 11 p.m. The rates at night, even very late at night, were low. The highest rate of critical incidents occurred during the afternoon hours, and the lowest rates generally occurred at night. This finding is contrary to the hypothesis that drivers exhibit poorer driving performance late at night due to circadian-rhythm-induced fatigue. It appears likely that the higher traffic volumes during daylight hours could have had a greater impact on the number of driver errors than did fatigue.

To get a clearer picture of the general effects of factors such as meals, circadian rhythms, and traffic, a second analysis was conducted by grouping hour of day into 4 day segments: morning (4:00–11:59), afternoon (12:00–17:59), evening (18:00–21:59), and night (22:00–3:59) (Pinel, 1997). Rates of the occurrence of critical incidents were calculated by dividing the number of critical incidents occurring per time segment by the number of hours driven in that same time category. This procedure was conducted to control for the amount of driving that occurred for each driver during all four time segments. The resulting mean rates for each day segment for team and single drivers are shown in Figs. 3 and 4. An ANOVA conducted on the rates proved to be significant, with a difference across time segments, $F_{3,71} = 4.23$; $p = 0.008$. The results are broken down in each figure based upon the ORD rating of generally "alert", "somewhat drowsy", or "very drowsy". Note that the team drivers (Fig. 3) only have cases of "very drowsy" ratings for incidents that occurred at night. In contrast, the single drivers (Fig. 4) have "very drowsy" ratings throughout the day. As also shown in Fig. 4, not only was the presence of critical incidents high for single drivers for the afternoon segment, but the cases of an ORD rating of "very drowsy" was also very high during this time. This finding indicates that the truck driver fatigue problem for single drivers is a serious problem during most periods of the day. In addition, the single-driver fatigue problem may be most serious not in the late night hours as would be hypothesized due to circadian rhythm effects but in the afternoon and evening hours.

On a related note, analyses investigating the amount of sleep that single versus team drivers received while on the road indicated that team drivers obtained significantly more hours of sleep than the single drivers. These findings remained consistent across all days of the long-haul trip. These related findings lend more evidence that fatigue poses more of a safety problem for the single drivers than for the team drivers.

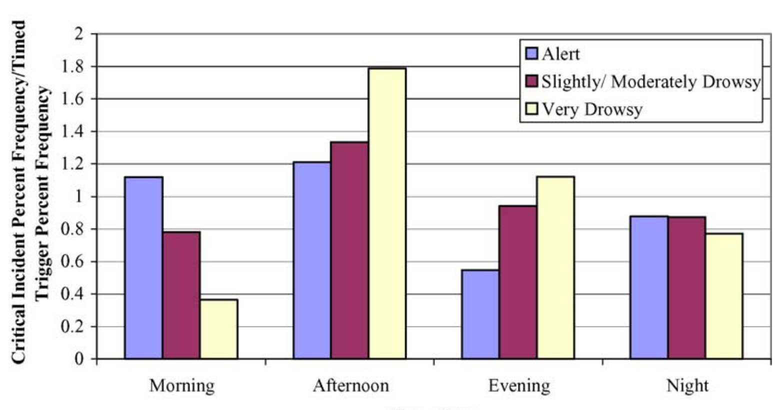

Fig. 4. Ratio of critical incident percent frequency to timed trigger percent frequency for single drivers.

3.3. Critical incident severity analysis

As part of the data reduction and classification process for this project, the severity of each critical incident was determined. Severity levels included collision, near collision, driver error with hazard present, and driver error with no hazard present.

The most informative of these data are often contained in the cases of the most severe classifications. Specifically, the collisions and near collisions are the cases where the driver was at the greatest risk, and was often out of control of the vehicle for at least an instant. There were two collisions captured as part of this study. Both were non-injury, non-police-reported crashes. One was a lane deviation collision with a road sign where the driver was extremely fatigued. The second was a collision with an animal where fatigue was not an apparent factor.

In addition to the collisions, there were 22 near collisions recorded for a total of 24 severe incidents. There were a variety of causes and circumstances associated with these incidents. Specifically, of the 24 incidents:

- 7 were clearly caused by another driver;
- 19 involved single drivers and 5 involved team drivers; 15 separate drivers were involved in total;
- 7 were caused by one single driver; 13 (over half) were caused by 4 single drivers;
- in contrast, no team driver was involved in more than one incident;
- 4 involved cases of extreme fatigue;
- 4 occurred in work zones.

Even though there were relatively few cases, the most severe incidents could almost all be classified in several categories, specifically:

- seven were caused by close headway with a rapid deceleration of the forward vehicle;
- two were near forward collisions with a stationary obstacle or vehicle in the roadway;
- three were caused by an unplanned lane deviation;
- two were caused by an attempted lane change with a vehicle present in the adjacent lane;
- of the seven “other” vehicle cases, four were lane change cut-offs.

Several of the findings summarized above are of particular interest to this study. As was described in a previous section that analyzed all critical incidents, single drivers had a much higher rate of involvement in these most severe incidents than did team drivers. This difference was statistically significant, χ^2 $(1, N=22)=8.16$; $p<0.01$. This finding may indicate that the single drivers were not just driving more aggressively because they did not have to be concerned about waking or disturbing a partner but that they were driving less safely in general. An important related finding is that only four of the single drivers (about 15% of the drivers present in this analysis) accounted for over one-half of the most severe incidents. This finding is similar to one found by Hanowski et al. (2000) where a small number of local short-haul operators accounted for the majority of the unsafe acts that occurred during a naturalistic observation study.

Furthermore, 4 of the 17 cases initiated by the truck driver involved extreme fatigue, 3 by single drivers and 1 by a team driver. Although these are low numbers in general, they serve to confirm that there is some degree of extreme fatigue present in long-haul trucking that affects driving safety.

4. Conclusions

The current study utilized a new approach to studying driving safety applied in a long-haul trucking environment. As described in detail, the approach involved a critical-incident-based, naturalistic driver assessment using sophisticated instrumented vehicles. This approach can use either specially instrumented vehicles or portable instrumentation installed in a commercial or personal vehicle to collect a relatively large data set unobtrusively and in situ. This increases the validity of the results obtained because drivers are in a real driving context. Unlike fleet or instrumented vehicle studies that have been run in the past, this technique utilizes sophisticated, critical-incident-triggered instrumentation and data reduction techniques to efficiently collect more detailed performance, precrash, and near crash data in addition to crash frequency and severity data. These techniques use analysis of vehicle-based critical events and near crashes that allow the assessment of precursor factors that are similar to crash causes. However, since these critical events occur much more frequently than crashes, they can be measured and assessed in an a priori empirical setting.

The use of this method for the long-haul truck study provided many important findings; however, the analyses presented here represent a small segment of the data that actually exist within this extensive data set. A primary finding was that single drivers had many more critical incidents, at all levels of severity, as compared to team drivers. The difference was very large, with ratios of between 4:1 and 2.5:1 depending on severity level. The difference in total number of critical incidents was not limited to a few drivers, as has been the case with other similar studies (e.g., Hanowski et al., 2000), as eight of the single drivers had more incidents than *any* of the team drivers. Analyses of possible alternative explanations for these differences (including miles driven, age, experience, and company) showed that no systematic differences were present. Therefore, the conclusion regarding this finding is that single drivers drive significantly more aggressively than do team drivers. Based on the focus group results for this project (Neale et al., 1998), a plausible explanation is that team drivers drive more carefully so that their partners get higher quality sleep and are not alarmed by their driving.

Another important finding was that the frequency of critical incidents and fatigue-related critical incidents varied significantly by the hour of the day. However, contrary to what might be expected, the largest number of incidents (even corrected for exposure), as well as the largest number of cases of very drowsy single drivers, occurred in the late afternoon and early evening hours. It was apparent from this analysis that interaction with heavier traffic had a greater impact upon the occurrence of critical incidents, and most likely the greatest impact on crash risk overall, than did fatigue due to circadian rhythm effects. Thus, there is a trade-off that must be considered when attempting to address the truck driver fatigue issue. For example, proposals have been made that would limit truck drivers' late-night driving to minimize fatigue. In considering such a proposal, however, these results show that one must also consider the relative risk reduction due to the presence of far less traffic during those time periods.

In looking at only the most severe of the critical incidents, including 2 collisions and 22 near collisions, the pattern of a few drivers involved in many incidents did emerge. One single driver caused 7 of the 24 most severe incidents, and 4 single drivers caused 13 of the 24 incidents. No team driver had more than one severe critical incident. These results show that while the singles drove more aggressively in general compared to team drivers, a smaller number exhibited dangerous driving behavior on multiple occasions. These results show that traveling with on-board monitoring systems or a using an additional driver screening process prior to permanent licensing could potentially help reduce risk in long-haul trucking.

The severe critical incident analysis also found that 4 of the 24 severe incidents were caused by extreme (i.e., head-bobbing) fatigue. This finding confirms the presence of these levels of fatigue in the long-haul trucking industry. The only interesting pattern regarding these four events was that three of them occurred with single drivers. A primary issue regarding this study was whether or not team drivers were more fatigued because they might not have gotten a high enough quality of sleep in a moving sleeper-berth truck. It is important to note that there were very few occurrences of this extreme fatigue recorded. Therefore, these findings suggest that team drivers may not be at the greatest risk for extreme fatigue.

4.1. *A final word on the potential of the method and the data set*

The above study, independent of its own contribution to the understanding of long-haul truck driving, demonstrates the potential of this method to help bridge the gap in safety assessment knowledge that exists between results obtainable through traditional empirical and epidemiological methods. The rapid development of data processing and storage technology, video compression techniques, camera and sensor capabilities, and wireless data transmission options promise to make these methods even more useful, thus allowing for larger-scale studies in the near future.

It is also important to note that the data sets created via this approach have significant value beyond their original purpose. For example, in addition to the findings above for which the reported study was designed, the method and resulting data set are also solving additional important safety problems. A strength of the incident-based approach is that the data are archived much like a crash database but with much more detailed information in many respects. This allows the data to be used for many purposes over a period of years. For our data, a second, post hoc study of truck-car interaction errors has already been accomplished, and several more studies investigating driver distraction issues will be conducted in the near future. This aspect makes the approach described in this paper a powerful addition to traditional driving safety assessment techniques. The application of this method has great potential to improve our understanding of the complex relationships of causal and contributing factors for both commercial and private vehicle crashes.

References

Braver, E.R., Preusser, C.W., Preusser, D.F., Baum, H.M., Beilock, R., Ulmer, R., 1992. Who Violates Work Hour Rules? A Survey of Tractor–trailer Drivers. Insurance Institute for Highway Safety, Arlington, VA.

Dingus, T.A., 1995. Moving from measures of performance (MOPs) to measures of effectivness (MOEs). In: Safety Evaluation of Intelligent Transportation Systems Workshop Proceedings. ITS America, Washington DC.

Dingus, T.A., Hetrick, S., Mollenhauer, M., 1999. Empirical methods in support of crash avoidance model building and benefits estimation. ITS J. 5, 93–124.

Dingus, T., McGehee, D., Hulse, M., Jahns, S., Manakkal, N., Mollenhauer, M., Fleishman, R., 1995. Travtek Evaluation Task C_3-Camera Car Study, Report No. FHWA-RD-94-076. U.S. Department of Transportation, Federal Highway Administration, Washington, DC.

Dingus, T.A., Neale, V.L., Garness, S.A., Hanowski, R.J., Keisler, A.S., Lee, S.E., Perez, M.A., Robinson, G.S., Belz, S.M., Casali, J.G., Pace-Schott, E.F., Stickgold, R.A., Hobson, J.A., 2001. Impact of Sleeper Berth Usage on Driver Fatigue. Federal Motor Carriers Safety Administration, Washington, DC.

Gilberg, M., Kecklund, G., Akerstedt, T., 1994. Relations between performance and subjective ratings of sleepiness during a night wake. Sleep 17 (3), 236–241.

Grace, R., Byrne, V.E., Bierman, D.M., Legrand, J.M., Gricourt, D.J., Davis, R.K., Staszewski, J.J., Carnahan, B., 1999. A drowsy driver detection system for heavy vehicles. In: Proceedings of the Ocular Measures of Driver Alertness Conference, Herndon, VA.

Hanowski, R.J., Keisler, A.S., Wierwille, W.W., 2003. Light Vehicle-Heavy Vehicle Interactions: A Preliminary Assessment Using Critical Incident Analysis, Final Report, Report no. FMCSA-RT-03-013. U.S. Department of Transportation, Federal Motor Carrier Safety Administration, Washington, DC.

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

Haworth, N.L., Triggs, T.J., Grey, E.M., 1988. Driver Fatigue: Concepts, Measurement and Crash Countermeasures, Report No. CR 72. Department of Transport and Communications, Federal Office of Road Safety, Document Retrieval Information, Canberra, Australia.

Heinrich, H.W., Petersen, D., Roos, N., 1980. Industrial Accident Prevention. McGraw-Hill Book Company, New York.

Knipling, P., Wang, J., 1994. Crashes and Fatalities Related to Driver Drowsiness/Fatigue, Research Note. U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC.

Koziol, J., Inman, V., Carter, M., Hitz, J., Najm, W., Chen, S., Lam, A., Penic, M., Jensen, M., Baker, M., Robinson, M., Goodspeed, C., 2000. Evaluation of the Intelligent Cruise Control System: vol. I—Study Results, Report No. NHTSA-98-3. National Highway Traffic Safety Administration, Washington, DC.

National Transportation Safety Board, 1990. Safety Study: Fatigue, Alcohol, Other Drugs, and Medical Factors in Fatal-to-the-Driver Heavy Truck Crashes (vol. 1), NTSB No. PB90-917002. Washington, DC.

National Transportation Safety Board, 1995. Safety Study Factors that Affect Highway Fatigue in Heavy Truck Accidents (vol. 1: Analysis), NTSB No. PB95-917001. National Technical Information Service, Springfield, VA.

Neale, V.L., Robinson, G.S., Dingus, T.A., Davis, R.E.L., 1998. Long-haul drivers' perspective on sleeper berth usage and fatigue in the trucking industry. SAE Technical Paper Series, 982784. SAE International, Warrendale, PA.

Pinel, J.P.J., 1997. Biopsychology. Allyn and Bacon, Washington, DC.

Transportation Research Board, 1997. Highway Capacity Manual, Special Report 209, 3rd ed. National Research Council, Washington, DC.

Wierwille, W.W., Ellsworth, L.A., 1994. Evaluation of driver drowsiness by trained raters. Accid. Anal. Prevent. 26 (5), 571–581.

Wierwille, W.W., Ellsworth, L.A., Wreggit, S.S., Fairbanks, R.J., Kirn, C.L., 1994. Research on Vehicle Based Driver Status/Performance Monitoring: Development, Validation, and Refinement of Algorithms for Detection of Driver Drowsiness, Contract Final Report DOT HS 808 247. National Highway Traffic Safety Administration: Washington, DC.