Large Truck Crash Causation Study Analysis
DTRS57-04-D30043, TRACX

Michael H. Belzer, Ph.D.
Sound Science, Inc.
2281 Traver Road
Ann Arbor, MI 48105

Report of Analysis
Truck Crashes and Work-Related Factors Associated with Drivers and Motor Carriers

April 28, 2009

The Large Truck Crash Causation Study is an extensive data collection effort undertaken by the Federal Motor Carrier Safety Administration (FMCSA) in an effort to understand the causes of truck crashes. FMCSA contracted with Sound Science, Inc. through the Volpe Center and its contractor, Chenega Advanced Solutions & Engineering, LLC to study the economic causes of large truck crashes.

Literature Review: Empirical Evidence

Since large trucks engage in commercial activities, one of the most important areas of interest involves the economic forces at play in the commercial delivery of freight. Research has demonstrated a strong relationship between truck driver pay and safety. These findings hold true both for driver pay rates and pay raises as well as for unpaid non-driving (labor) time, and they hold true across the many components of compensation. Looking at all compensation variables, a cross-sectional carrier-level study showed that at the mean, for every 10% in truck driver compensation, motor carriers have a 9.2% lower crash rate. This study demonstrated that every 10% in driver pay rate is associated with a 5% lower carrier crash rate and every 10% in unpaid driver labor time further is associated with an additional 1% lower crash rate (Belzer, Rodriguez, and Sedo 2002). A study of individual drivers at a single firm showed, using an event-history analysis, that at the mean, every 10% in driver pay rate is associated with a 34% lower probably of crash, and every 10% in driver pay raise is associated with an additional 6% lower crash probability. This means that higher pay rates alone predict as much as a 40% lower crash probability (Belzer, Rodriguez, and Sedo 2002; Rodriguez et al. 2003; Rodriguez, Targa, and Belzer 2006).

The foregoing study used conventional labor economic theory to construct a labor supply curve for truck drivers, based on a survey of drivers conducted by the University of Michigan Trucking Industry Program. This analysis showed that at the margin, truck drivers and their employers (when they have an influence on this choice, which they frequently do) will tend to prefer to work more hours if a higher pay rate is offered for the work, up to the current average mileage pay rate in the labor market. As the driver's pay rate rises to and exceeds the mean, the driver (and his employer) will choose to work fewer hours, trading more leisure for labor as anticipated in conventional economic theory. At the time the data were collected, in 1997 and 1998, road drivers worked on average approximately 65 hours per seven-day week (almost 10% more than the legal limit), but as driver pay increased above the mean, drivers' growing preference for leisure tends to overcome this preference for more work hours and they reduce their working time to the legal limit and below as wage rates increased (Belzer, Rodriguez, and Sedo 2002). We do not know whether this reduction in working time is voluntary on the part of the drivers or is preferred by the carriers paying the higher rates, but we know it happens. While economists know that workers make tradeoffs between wages and working time, this direct relationship between truck driver pay – for both driving and non-driving time – and hours of work seems to have received less emphasis in motor carrier safety research than one might expect (Belzer et al. 1999).

Research has not confirmed that there is a significant difference in hourly v. contingent pay, and measurement flaws in the LTCCS data set hamper this analysis again, so no firm conclusions can be drawn from LTCCS. Since most local drivers work on an hourly basis and most over-the-road drivers work on a mileage basis (or on a percentage-of-revenue basis that conflates mileage and revenue factors), while "MileagePayThisTrip" is a significant predictor, as discussed below, measurement errors in these data make it risky to draw too many conclusions. Missing data in this variable cause multivariate analysis to lose 109 observations, which is a very significant fraction of valid cases and may bias the results. "HourlyPayThisTrip" only causes the loss of 26 cases, but is a strongly insignificant predictor. Again, this may not be due to the pay method itself but rather be caused by differences in work organization and pressures related to the differences in operations.

Substantial research has shown both theoretically and empirically that we should expect that truck driver compensation is an important predictor of safety. Efficiency wage theory, along with empirical research testing it, supports the hypothesis that carriers that pay drivers more than they would otherwise receive in a comparable non-trucking labor market (as well as better total compensation than they would otherwise expect working for a carrier in a lower-rate segment of the trucking service market) will reap superior performance and greater workforce stability. Reciprocity, or "fair wage theory", further suggests that drivers who earn better compensation will reciprocate because they believe their employer (or freight broker or motor carrier to which they are contracted) is treating them fairly, and this reciprocity will include both greater productivity and greater safety. Drivers who anticipate deferred compensation in the form of pension or other retirement benefits also will protect those benefits by driving in a responsible manner (Lazear 1990; Belzer, Rodriguez, and Sedo 2002).

Other researchers have paid insufficient attention to the influence of market pressures on occupational health and safety in trucking (notable exceptions include (Mayhew and Quinlan 1997, 2000, 2006; Quinlan 2001; Quinlan, Mayhew, and Johnstone 2006; Williamson and Feyer 1992; Quinlan, Wright, and National Transport Commission 2008). A great deal of the research on trucking safety seems to focus on the effectiveness of various engineering interventions, such as information technology, mechanical design, and materials technology, on trucking safety. Additional research focuses on interventions that we might characterize as "human engineering"; among those are regulations designed to limit driver behavior, such as hours-of-service regulations. While such efforts are important, they do not address the organizational and market problems directly. As long as economic competition in trucking provides incentives for drivers and motor carriers to seek economic advantage by undercutting these standards, pressure on these standards will remain strong and truckers will obey them only insofar as regulators and enforcers can maintain the pressure. In other words, markets and regulations will continue to have opposite internal logics that remain in tension with each other and do not necessarily result in safe operations. Safety benchmarking programs have developed in an effort to align market and safety incentives (for example, see Belzer 2004), but a full scale practical test of its effectiveness has never taken place.

LTCCS Data

Hedlund and Blower argue that the only way to conduct statistical analysis of the LTCCS is by using "induced exposure." This method requires analysts to separate out cases in which the truck is assigned the "critical reason for the critical event" from cases in which another vehicle is assigned the critical reason. This method is far from perfect, but at least it allows researchers to separate out the crashes in which the truck driver is considered to have been the last operator capable of avoiding the crash from those in which the final action that made the crash inevitable belonged to another person or vehicle operator (Hedlund and Blower 2006; Blower and Campbell 2005). This Sound Science study uses this induced exposure approach and attempts to determine statistical relationships between economic factors and the critical reason for the critical event.

Certain data quality problems have been identified, however, that might make it difficult to determine whether such relationships exist. Donaldson, for example, has made a litany of criticisms of the LTCCS (Donaldson 2005). The Transportation Research Board, on behalf of the National Research Council, created a Committee that provided oversight for the LTCCS and issued a letter report in 2003 critical of the methodology used by the developers of the data collection effort. This letter report articulated a particular concern with data quality. It appeared from the Committee's review of the developing study that key variables – especially the economic ones that the present analysis evaluates – probably would not be accurate (Council et al. 2003). Unfortunately, this concern seems to have been well founded and this limits analysis of these data.[1]

Appendix 1 (30 pages) provides extensive and detailed evaluation of scores of variables. In general, compensation data are fundamentally flawed across the board. The contractors working for the National Automotive Sampling System (NASS), which performed the data collection for FMCSA, proposed to collect far too much data, creating a kind of grab-bag of every kind of data that FMCSA or NASS or the National Highway Traffic Safety Administration (NHTSA) would ever want to know, much of it by using duplicative and unedited surveys created for other purposes. The sheer volume of information collected ensured that both data collectors and respondents would experience severe survey fatigue. Survey fatigue systematically creates problems with data that are more difficult to obtain, the last to obtain, or deemed least essential by the data collectors. In addition, the NASS contractors had little familiarity with the economic data and little

[1] Dr. Michael H. Belzer, President of Sound Science, Inc. and currently Associate Professor of Economics at Wayne State University, was a member of this oversight committee.

understanding of its significance, and this lack of comprehension shows up in the data in the form of missing or misinterpreted information.

Appendix 1 provides extensive detail on these data problems so the report will not reiterate them here, but a key question regarding basic information on compensation provides a good example of this problem. “VariableCompensation”, in the DriverAssessment Data Set, is supposed to document “whether or not the driver is compensated in accordance with a variable compensation package such that the driver is not paid on a consistent basis” (“LTCCS Analytical User's Manual”, 30 June 2006, page 122 of 634). All available experience and research tells us that although local drivers frequently are paid by the hour, almost all truck drivers in over-the-road operations are paid for their driving on a “piecework” basis (Belzer 2000); that is, they are paid by the mile or a percentage of revenue (which itself is based on miles, value of freight, and other contingencies such as weight, volume, and handling characteristics that contribute to the freight rate), or by the load (which is the same as the foregoing only more arbitrary). Some drivers are paid by the hour for some or all non-driving labor time (such as loading, unloading, waiting for freight or repairs), yet the LTCCS Codebook indicates that all drivers are paid on a fixed basis (that is, not variable). An examination of the non-public interview data shows that the data were recorded entirely differently – drivers are paid according to the framework described here – with the additional error of ascribing to some drivers pay both by the mile and by the hour for driving time, which cannot be true, and inaccurately categorizing some drivers’ compensation as “pay by the load”, which simply obscures the true compensation method. This is just one example of the scores of errors found in the data and analyzed in Appendix 1.

Appendix 2 (219 pages) provides descriptive statistics on each of the variables considered eligible for analysis for this report. In some cases multiple variations are provided on each variable, in many cases providing descriptive statistics controlling for whether the vehicle is a truck by the definition used in the study or is a truck by the definition I used to attempt to narrow the analysis to what would generally be considered a large commercial truck. This appendix also includes descriptive statistics for derived variables created by Sound Science in an effort to gain better analytical precision. The descriptive statistics appendix can be used as a resource when analyzing the variables used in the logistic regression model that provides the final output of the report. The main section of the report will not recapitulate the findings in the descriptive statistics appendix except to note certain highlights.

1. In almost exactly 50% of all crashes, the truck was assigned the critical reason for the critical event. This means that this variable, used as a dependent variable in the logistic regression, is evenly distributed between truck and non-truck attribution.
2. For this variable, as for most, controls for “GVE Truck” and “GVE CDL Truck” did not change the outcome materially. In fact, there is little difference in most variables between the latter two categories. This shows up in the logistic regression, also, as the truck designation is not statistically significant in any model.
3. Problems with missing data plague all of the data analysis. More than half of the cases are excluded because data collectors did not collect data symmetrically on automobiles and their drivers as well as trucks and truck

drivers, and because 107 cases had to be thrown out because they came from the pilot phase of the LTCCS. Approximately half the actual trucking cases, however, were excluded from analysis because of missing data (again, detailed in Appendix 1; some of the data were missing in the LTCCS final data set and many more had to be excluded by Sound Science because of inconsistent and inappropriate answers).

4. Some variables that might have been useful could not be used for the regression analysis simply because data had been collected on so few cases that analysis was impossible. In addition, this amount of missing data leads to biased results and as indicated elsewhere, it is difficult or impossible to determine the direction of the bias in any individual analysis.
5. Fatigue seems to be attributed stably as a factor in about 15% of all crashes and almost only in the case of the truck driver. Data collectors did not systematically and insistently collect fatigue data from automobile drivers so there is no real way to compare the groups. Finally, "fatigue" is not a discrete variable but rather a complex derived concept that only crudely approximates the true condition of the driver, so readers should interpret any analysis using this variable with caution. Researchers call this kind of variable a "dummy" because it is a "dumb" placeholder for a concept, and when the concept actually is variable and not fixed, using such an indicator variable (1/0) may be misleading.
6. WorkPressureCount, a variable that exists in the data, provides a sketchy measure of the pressure factor. Confirmatory factor analysis of a wider range of variables, used in "WorkPressureTotal", an index created by Sound Science, suggests that better results would be obtained by using a more scientifically constructed factor. Factor analyses shows that this variable can be articulated into eighteen factors, only one of which stands out as distinctive with an eigenvalue of 3.554, explaining 20.9% of the explained variance, which is quite high (see Table 1). Since the first eigenvalue is almost double the size of the second one, and since the second and subsequent eigenvalues decline gradually in magnitude, it is safe to conclude that the factor construct based on theory and experience in industrial relations is reasonably sound. While some variables are weak, all have the same sign, which is not the case in subsequent factors. Moreover, the use of a derived variable constructed from a much larger set of variables for the WorkPressureTotal variable allows it to be treated as a continuous variable, since drivers have from zero to seven of these variables. Although 1,314 of 1,456 responses are zero, the added information from the positive responses provides sufficient information to confirm the relationship between these work pressures and the likelihood of being assigned the critical reason for the critical event.
7. Sound Science tested WorkPressureTotal[2] using Ordinary Least Squares

[2] The WorkPressureTotal variable used actually was WorkPressureTotal+1, to avoid any inadvertent effects of the "0" in the model; this conversion produced no differences.

Analysis of Covariance (OLS ANCOVA). We can use OLS ANCOVA because the variable is continuous, though it does not have a normal distribution. The ANCOVA confirms the validity of the work pressure factor. All variables are highly significant and have the same signs (see Table 3). The model fit is extraordinarily high, with $R^2 = 0.996$. The size of this correspondence suggests that industrial relations determinants inherent in the economic relationships have great internal consistency. Subsequent analysis of the effects of this factor on AssignedCriticalReason shows that it is the strongest industrial relations factor, only somewhat less than AggressionCount and Fatigue – both constructed variables – in its contribution to prediction of the ACR dependent variable.

8. While sleep variables appear in Appendix 2, they do not appear in the regression analysis since sleep factors are not a primary concern of this analysis. Instead, this analysis uses "Fatigue" rather than to try to construct a sleep index to plug into the regression. If another analysis of the LTCCS data has determined a way to introduce an accurate factor for sleep, that could be integrated into this analysis but only insofar as it pertains to economic issues. For the most part, the latter already are included in the work pressure variable.

This study seeks to determine the effects of these work factors on truck crashes. To do so, we use the following model, reflective of the factors discussed above.

$$\text{AssignedCriticalReason} = \alpha + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7 + \beta_8 + \varepsilon$$

where β_1 = WorkPressureTotal, β_2 = AggressionCount, β_3 = Fatigue, β_4 = ClassYears, β_5 = ClassYearsSq, β_6 = SafetyBonus, β_7 = HoursDriving, β_8 = MileagePayThisTrip(Driver), and ε = error. These variables are used because theory and previous research support the hypothesis that these concepts related to the organization of work and the competitiveness of the trucking industrial process predict motor carrier and truck driver safety, but analysis is limited because of the poor execution of the data collection and database development process, as discussed elsewhere.

Table 1: Principal Component Analysis of WorkPressureTotalD
2284 total cases of which 1336 are missing

Table 1(a): EigenValues

	Values	Variance Proportion
e1	3.554	20.9
e2	2.037	12.0
e3	1.416	8.3
e4	1.308	7.7
e5	1.183	7.0
e6	1.072	6.3
e7	1.007	5.9
e8	0.941	5.5
e9	0.879	5.2
e10	0.797	4.7
e11	0.740	4.4
e12	0.676	4.0
e13	0.451	2.7
e14	0.394	2.3
e15	0.324	1.9
e16	0.182	1.1
e17	0.038	0.2

Table 1(b): EigenVectors

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17
NewPosition_m	0.081	-0.042	0.314	0.552	-0.205	-0.073	-0.163	0.051	0.023	0.029	-0.457	0.514	-0.174	-0.086	0.004	0.011	0.004
ShippingDeadline_m	0.132	0.571	-0.007	-0.025	0.039	-0.136	0.060	-0.092	-0.055	-0.064	-0.172	-0.044	-0.224	0.569	-0.429	0.063	0.153
EXPWorkSchedule_m	0.199	0.090	-0.033	-0.235	-0.280	0.213	0.352	0.253	0.411	0.516	-0.215	0.121	0.259	0.119	0.076	0.011	0.027
Quotas_m	0.045	0.553	-0.023	0.053	0.073	-0.111	0.069	-0.185	-0.290	-0.071	0.048	0.264	0.497	-0.064	0.458	-0.059	-0.082
ExtraLoads_m	0.284	-0.116	-0.469	0.283	-0.029	-0.162	-0.083	0.116	-0.113	-0.005	-0.073	-0.069	0.493	-0.186	-0.473	0.178	0.044
Demoted_m	0.092	-0.052	0.456	0.344	-0.049	-0.204	0.231	-0.051	-0.293	0.462	0.192	-0.440	0.082	0.042	-0.049	-0.147	-0.033
SelfInducedOther_m	0.114	-0.006	0.180	-0.295	0.235	0.109	-0.589	0.364	-0.318	0.342	0.126	0.231	0.080	0.122	-0.106	-0.041	-0.009
OtherPressure_m	0.055	0.034	0.073	0.050	-0.587	0.243	-0.439	-0.438	0.186	-0.022	0.314	-0.040	0.175	0.157	-0.081	-0.051	-0.020
LoadPressure_m	0.399	-0.077	0.337	0.027	0.151	0.187	0.128	0.101	0.109	-0.352	0.206	0.052	0.147	-0.035	0.019	-0.049	0.655
LoadPressureIndicator_m	0.448	-0.032	0.268	-0.024	0.120	0.085	0.095	0.109	0.144	-0.332	0.069	-0.020	0.057	0.093	-0.080	0.085	-0.722
WorkFatigueCount	0.406	-0.120	-0.022	-0.207	-0.072	-0.265	-0.022	-0.286	-0.113	0.178	0.050	0.046	-0.244	-0.068	0.242	0.663	0.088
ScheduledExtensions_m	0.319	-0.042	0.049	-0.395	-0.101	-0.378	-0.157	-0.189	0.026	-0.062	-0.382	-0.156	-0.012	-0.238	-0.009	-0.541	0.028
UnscheduledExtensions_m	0.298	-0.135	-0.365	0.117	0.098	0.050	0.233	-0.201	-0.087	0.187	0.373	0.426	-0.292	0.071	-0.067	-0.421	-0.062
ShortNoticeTrips_m	0.293	0.039	-0.331	0.330	0.042	0.091	-0.288	0.254	0.080	-0.006	-0.099	-0.383	-0.159	0.268	0.513	-0.113	0.035
FillInTrips_m	0.118	0.543	-0.006	0.068	-0.026	0.167	-0.084	0.145	0.146	0.118	0.192	-0.130	-0.320	-0.647	-0.138	0.028	-0.022
UnpaidLoading_m	0.075	-0.041	0.016	0.057	0.445	0.561	-0.040	-0.510	-0.041	0.190	-0.380	-0.129	0.068	-0.077	-0.063	0.037	-0.012
OtherRelations_m	0.092	-0.036	-0.062	-0.139	-0.455	0.415	0.210	0.157	-0.653	-0.193	-0.173	-0.068	-0.132	-0.049	0.001	-0.011	-0.010

Table 1(c): Unrotated Factor Matrix

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17
NewPosition_m	0.153	-0.061	0.373	0.631	-0.223	-0.076	-0.164	0.049	0.022	0.026	-0.393	0.423	-0.117	-0.054	0.002	0.005	0.001
ShippingDeadline_m	0.249	0.814	-0.008	-0.029	0.042	-0.140	0.060	-0.089	-0.052	-0.057	-0.148	-0.036	-0.150	0.357	-0.244	0.027	0.030
EXPWorkSchedule_m	0.375	0.129	-0.040	-0.269	-0.305	0.220	0.353	0.245	0.385	0.461	-0.185	0.100	0.174	0.075	0.043	0.005	0.005
Quotas_m	0.085	0.790	-0.027	0.061	0.080	-0.115	0.069	-0.180	-0.272	-0.063	0.041	0.217	0.333	-0.040	0.261	-0.025	-0.016
ExtraLoads_m	0.536	-0.165	-0.558	0.324	-0.031	-0.167	-0.084	0.112	-0.106	-0.005	-0.063	-0.057	0.331	-0.117	-0.270	0.076	0.009
Demoted_m	0.174	-0.074	0.542	0.393	-0.053	-0.211	0.232	-0.049	-0.274	0.412	0.165	-0.362	0.055	0.026	-0.028	-0.063	-0.006
SelfInducedOther_m	0.215	-0.009	0.214	-0.337	0.256	0.113	-0.591	0.353	-0.298	0.305	0.108	0.190	0.054	0.077	-0.060	-0.017	-0.002
OtherPressure_m	0.105	0.048	0.086	0.057	-0.638	0.251	-0.440	-0.425	0.174	-0.020	0.270	-0.033	0.118	0.098	-0.046	-0.022	-0.004
LoadPressure_m	0.752	-0.110	0.401	0.031	0.164	0.193	0.129	0.098	0.102	-0.315	0.177	0.043	0.099	-0.022	0.011	-0.021	0.128
LoadPressureIndicator_m	0.844	-0.046	0.319	-0.027	0.131	0.088	0.095	0.106	0.135	-0.297	0.059	-0.016	0.038	0.058	-0.046	0.036	-0.141
WorkFatigueCount	0.766	-0.171	-0.026	-0.237	-0.078	-0.274	-0.023	-0.278	-0.106	0.159	0.043	0.038	-0.164	-0.043	0.138	0.283	0.017
ScheduledExtensions_m	0.601	-0.059	0.058	-0.452	-0.110	-0.391	-0.158	-0.183	0.024	-0.056	-0.329	-0.128	-0.008	-0.149	-0.005	-0.231	0.005
UnscheduledExtensions_m	0.561	-0.193	-0.434	0.133	0.106	0.051	0.234	-0.195	-0.082	0.167	0.321	0.350	-0.196	0.045	-0.038	-0.180	-0.012
ShortNoticeTrips_m	0.552	0.056	-0.394	0.377	0.045	0.095	-0.289	0.246	0.075	-0.005	-0.085	-0.315	-0.107	0.168	0.292	-0.048	0.007
FillInTrips_m	0.222	0.775	-0.007	0.078	-0.029	0.173	-0.084	0.140	0.137	0.106	0.165	-0.107	-0.215	-0.406	-0.078	0.012	-0.004
UnpaidLoading_m	0.141	-0.059	0.019	0.066	0.484	0.581	-0.040	-0.495	-0.038	0.169	-0.327	-0.106	0.046	-0.049	-0.036	0.016	-0.002
OtherRelations_m	0.174	-0.051	-0.074	-0.159	-0.495	0.430	0.211	0.153	-0.612	-0.172	-0.149	-0.056	-0.089	-0.031	0.001	-0.005	-0.002

Table 2: Correlation Matrix

Pearson Product-Moment Correlation

	New…	EXP…	Quot…	SelfI…	SelfI…	Othe…	Load…	Sche…	Unsc…	Rota…	Unpa…	Othe…	Ship…	Extr…	Dem…	Shor…	FillIn…	Over…
NewPositi…	1.000																	
EXPWork…	-0.010	1.000																
Quotas_m	-0.002	-0.002	1.000															
SelfInduc…	0.130	-0.008	-0.002	1.000														
SelfInduc…	-0.012	0.020	-0.003	-0.010	1.000													
OtherPre…	0.103	0.087	-0.002	-0.007	-0.012	1.000												
LoadPres…	0.144	0.224	-0.003	0.086	0.177	0.027	1.000											
Scheduled…	-0.007	0.178	-0.002	-0.006	0.143	0.065	0.442	1.000										
Unschedul…	-0.005	0.149	-0.001	-0.004	-0.008	-0.005	0.306	0.110	1.000									
RotatingS…	-0.007	0.178	-0.002	-0.006	0.143	0.065	0.442	1.000	0.110	1.000								
UnpaidLoa…	-0.006	-0.007	-0.001	-0.005	0.047	-0.006	0.102	-0.005	0.123	-0.005	1.000							
OtherRela…	-0.009	0.124	-0.002	0.119	0.060	0.094	0.132	0.059	0.077	0.059	-0.007	1.000						
ShippingD…	-0.004	0.115	0.577	-0.003	-0.005	-0.004	0.187	0.172	-0.002	0.172	-0.003	-0.004	1.000					
ExtraLoa…	0.084	0.071	-0.001	-0.004	0.056	-0.005	0.243	0.225	0.427	0.225	-0.004	0.077	-0.002	1.000				
Demoted_m	0.235	-0.002	-0.000	-0.002	-0.003	-0.002	0.163	-0.002	-0.001	-0.002	-0.001	-0.002	-0.001	-0.001	1.000			
ShortNoti…	0.111	0.094	-0.002	-0.007	0.115	0.051	0.321	0.145	0.280	0.145	0.162	0.046	0.142	0.565	-0.002	1.000		
FillInTrips…	-0.005	0.162	0.408	-0.004	0.062	0.092	0.128	-0.004	-0.003	-0.004	-0.004	-0.005	0.470	-0.003	-0.001	0.200	1.000	
OverDispa…	0.005	0.343	0.132	-0.027	-0.015	0.227	0.489	0.270	0.190	0.270	0.030	0.264	0.229	0.137	0.132	0.195	0.170	1.000

Table 3: WorkPressureTotalD

Type of analysis: OLS ANOVA
 Factors all fixed effects, discrete variables
Partial (Type 3) Sums of Squares
2284 total cases of which 221 are missing

Table 1(a): ANOVA

Analysis of Variance For	WorkPressureTotalD_m
2284 total cases of which 221 are missing	

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	25.6423	25.6423	27415	≤ 0.0001
NewPosition	1	16.0064	16.0064	17113	≤ 0.0001
EXPWorkSchedule	1	22.2996	22.2996	23842	≤ 0.0001
SelfInducedIllegal	1	11.4715	11.4715	12265	≤ 0.0001
SelfInducedOther	1	33.4902	33.4902	35806	≤ 0.0001
OtherPressure	1	16.9224	16.9224	18093	≤ 0.0001
UnpaidLoading	1	8.52203	8.52203	9111.3	≤ 0.0001
OtherRelations	1	19.3985	19.3985	20740	≤ 0.0001
Quotas	1	0.655784	0.655784	701.13	≤ 0.0001
ExtraLoads	1	4.36563	4.36563	4667.5	≤ 0.0001
Demoted	1	0.927028	0.927028	991.13	≤ 0.0001
ShippingDeadline	1	1.56763	1.56763	1676.0	≤ 0.0001
LoadPressureIndicator	1	27.0088	27.0088	28876	≤ 0.0001
ShortNoticeTrips	1	9.03415	9.03415	9658.9	≤ 0.0001
FillInTrips	1	4.01509	4.01509	4292.7	≤ 0.0001
HurryingW	1	19.0272	19.0272	20343	≤ 0.0001
Error	2047	1.91460	0.000935		
Total	2062	432.358			

$R^2 = 0.996$

Table 1(b): Results for factor NewPosition

Coefficients of: WorkPressureTotalD_m on NewPosition

Level of N_m	Coefficient	std. err.	t Ratio	prob
0	-0.5009	3.829e-3	-130.8	≤ 0.0001
1	0.5009	3.829e-3	130.8	≤ 0.0001

Table 1(c): Results for factor EXPWorkSchedule

Coefficients of: WorkPressureTotalD_m on EXPWorkSchedule

Level of E_m	Coefficient	std. err.	t Ratio	prob
0	-0.5188	3.360e-3	-154.4	≤ 0.0001
1	0.5188	3.360e-3	154.4	≤ 0.0001

Table 1(d): Results for factor SelfInducedIllegal

Coefficients of: WorkPressureTotalD_m on SelfInducedIllegal

Level of S_m	Coefficient	std. err.	t Ratio	prob
0	-0.5007	4.522e-3	-110.7	≤ 0.0001
1	0.5007	4.522e-3	110.7	≤ 0.0001

Table 1(e): Results for factor SelfInducedOther

Coefficients of: WorkPressureTotalD_m on SelfInducedOther

Level of S_m	Coefficient	std. err.	t Ratio	prob
0	-0.4992	2.638e-3	-189.2	≤ 0.0001
1	0.4992	2.638e-3	189.2	≤ 0.0001

Table 1(f): Results for factor OtherPressure

Coefficients of: WorkPressureTotalD_m on OtherPressure

Level of O_m	Coefficient	std. err.	t Ratio	prob
0	-0.4972	3.697e-3	-134.5	≤ 0.0001
1	0.4972	3.697e-3	134.5	≤ 0.0001

Table 1(g): Results for factor UnpaidLoading

Coefficients of: WorkPressureTotalD_m on UnpaidLoading

Level of U_m	Coefficient	std. err.	t Ratio	prob
0	-0.5008	5.246e-3	-95.45	≤ 0.0001
1	0.5008	5.246e-3	95.45	≤ 0.0001

Table 1(h): Results for factor OtherRelations

Coefficients of: WorkPressureTotalD_m on OtherRelations

Level of O_m	Coefficient	std. err.	t Ratio	prob
0	-0.4972	3.452e-3	-144.0	≤ 0.0001
1	0.4972	3.452e-3	144.0	≤ 0.0001

Table 1(i): Results for factor Quotas

Coefficients of: WorkPressureTotalD_m on Quotas

Level of Q_m	Coefficient	std. err.	t Ratio	prob
0	-0.5209	0.0197	-26.48	≤ 0.0001
1	0.5209	0.0197	26.48	≤ 0.0001

Table 1(j): Results for factor ExtraLoads

Coefficients of: WorkPressureTotalD_m on ExtraLoads

Level of E_m	Coefficient	std. err.	t Ratio	prob
0	-0.4965	7.267e-3	-68.32	≤ 0.0001
1	0.4965	7.267e-3	68.32	≤ 0.0001

Table 1(k): Results for factor Demoted

Coefficients of: WorkPressureTotalD_m on Demoted

Level of D_m	Coefficient	std. err.	t Ratio	prob
0	-0.5027	0.0160	-31.48	≤ 0.0001
1	0.5027	0.0160	31.48	≤ 0.0001

Table 1(l): Results for factor ShippingDeadline

Coefficients of: WorkPressureTotalD_m on ShippingDeadline

Level of S_m	Coefficient	std. err.	t Ratio	prob
0	-0.4863	0.0119	-40.94	≤ 0.0001
1	0.4863	0.0119	40.94	≤ 0.0001

Table 1(m): Results for factor LoadPressureIndicator

Coefficients of: WorkPressureTotalD_m on LoadPressureIndicator

Level of L_m	Coefficient	std. err.	t Ratio	prob
0	-0.4962	2.920e-3	-169.9	≤ 0.0001
1	0.4962	2.920e-3	169.9	≤ 0.0001

Table 1(n): Results for factor ShortNoticeTrips

Coefficients of: WorkPressureTotalD_m on ShortNoticeTrips

Level of S_m	Coefficient	std. err.	t Ratio	prob
0	-0.4998	5.086e-3	-98.28	≤ 0.0001
1	0.4998	5.086e-3	98.28	≤ 0.0001

Table 1(o): Results for factor FillInTrips

Coefficients of: WorkPressureTotalD_m on FillInTrips

Level of F_m	Coefficient	std. err.	t Ratio	prob
0	-0.4926	7.519e-3	-65.52	≤ 0.0001
1	0.4926	7.519e-3	65.52	≤ 0.0001

Table 1(p): Results for factor Hurrying

Coefficients of: WorkPressureTotalD_m on Hurrying

Level of HrW	Coefficient	std. err.	t Ratio	prob
0	-0.5235	3.670e-3	-142.6	≤ 0.0001
1	0.5235	3.670e-3	142.6	≤ 0.0001

Results of Analysis of LTCCS

Variables that have economic origins and might contribute to assignment of the critical reason (ACR) for the critical event but are not contributory include AdvanceNoticeD, ExtentNotice, PayContingency(Driver), HourlyPay, DisciplinedLateD, IDROnTimePerformance, White, RotatingShift, O-O, DriverUnloadThisTrip, WaitLoadThisTrip, OnCallThisTripD, OverDispatchD (t-statistic is 1.76 and probability of 0.185).

One decision on the use of variables deserves special attention. As discussed extensively in Appendix 1 and as critiqued in advance by the NRC letter report, the compensation variables are so poorly executed that most of the true information is lost and analysts are left with very scant information (the public data set provides virtually no accurate or useful information). Sound Science tried to repackage the compensation data to make some sense, but errors in data collection and interpretation require that such repackaging trades off information on one dimension for information loss on another dimension. No data were collected or reported on compensation level on any of the dimensions previously studied for FMCSA (Belzer, Rodriguez, and Sedo 2002).

It appears that two formulations of pay method, using the non-public interview data, provide the best proxies for compensation. The most comprehensive proxy is DrivePayThisTrip (see Appendix 1, footnote 77), from the non-public IntvwDrDriver relation, which creates an ordinal variable with three components ranging from the least contingent (hourly pay) to the most contingent (percentage and "by the load"). This variable, when tested in the logistic regression model, has an F-ratio of 2.5686 and a p-value of 0.1095, slightly below the common lowest threshold level for significance used; this hints at validity but does not confirm it. MileagePayThisTrip, from the non-public IntvwCarrier relation, merely introduces a "dummy" variable for mileage pay (coded as 1) in which both hourly pay and percentage/"by the load" pay are coded as zero. This pay method, reported in Table 4, is significant at the 0.07 level with a higher F-ratio; not perfect, but good considering the flaws in the data. MileagePayThisTrip (Driver), created from the non-public IntvwDrDriver relation, provides a measure that is slightly more significant than the one created from the carrier interview relation. The difference between these variables might come from the formulations (a similar ordinal index created from data from the IntvwCarrier relation is insignificant; mileage indicator variables from two different underlying sources have similar significance in the model) or from the underlying concepts measured or proxied here, but with these data we cannot judge. This analysis uses the measure with the greatest significance because Sound Science believes that the major problem is noisy data created by imprecise measurements.

Control variables that might be expected to contribute to ACR but do not are not reported here; they are statistically insignificant. They include "White" (an indicator variable created from EthnicOrigin); Owner-Operator; all variables related to whether drivers loaded or unloaded; and all variables related to whether the company pays drivers to load and/or unload. Because the responses to questions regarding drivers are paid for loading and unloading are contingent on whether they did so this trip, and because this contingency reduces the *n*, these also are insignificant.

Finally, Table 6 shows a logistic regression model with the WorkPressureTotal variables entered individually. This formulation may lose validity simply because the indicators for the components of this index are so rare, and the sum of their appearance, which can only be determined by using the index, may be the important factor. Table 7 shows the reduced form of this regression model with insignificant variables removed in a stepwise fashion, with least significant variables removed first until the point that all variables are at least significant to the 0.10 level.

Table 4: Industrial Relations/Economics Contributors to Assigned Critical Reason; Using WorkPressure and Aggression Count Indices

Type of analysis: Logistic
Factors

Name	F/R	Kind
WorkPressureTotal	Fix	Cont
AggressionCount	Fix	Cont
Fatigue	Fix	Disc
ClassYears[3]	Fix	Cont
IDRSafetyBonus	Fix	Disc
HoursDriving	Fix	Cont
MileagePayThisTrip(Driver)	Fix	Disc

Partial (Type 3) Sums of Squares

General Results
2284 total cases of which 1514 are missing

Iteration	LogLikelihood	Convergence
1	-470.318	0.081357
2	-470.316	0.003172
3	-470.316	0.000009

[3] Conventional empirical analysis would expect ClassYearsSquared should tighten prediction, but this variable is not significant: Sum of Squares: 2.35687; Mean Square: 2.35687; F-ratio: 2.3660; Prob: 0.1244 . For this reason, the model does not include this variable. See Table 5 for reported results.

Table 4(a): ANCOVA

Analysis of Covariance For	AssignedCriticalReason
2284 total cases of which 1514 are missing	

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	24.3421	24.3421	24.443	≤ 0.0001
WorkPressureTotal	1	11.1730	11.1730	11.219	0.0008
AggressionCount	1	16.2121	16.2121	16.279	≤ 0.0001
Fatigue	1	28.7747	28.7747	28.894	≤ 0.0001
ClassYears	1	9.41141	9.41141	9.4504	0.0022
IDRSafetyBonus	1	9.83685	9.83685	9.8776	0.0017
HoursDriving	1	11.5259	11.5259	11.574	0.0007
MileagePayThisTrip(Driver)	1	3.84467	3.84467	3.8606	0.0498
Error	762	758.855	0.995873		
Total	769	847.286			

$R^2 = 0.147$[4]

Table 4(b): Coefficients of: AssignedCriticalReason on WorkPressureTotal

Covariate	Coefficient	std. err.	Wald	prob
WorkPressureTotal	0.6434	0.1925	11.17	0.0008

Table 4(c): Coefficients of: AssignedCriticalReason on AggressionCount

Covariate	Coefficient	std. err.	Wald	prob
AggressionCount	1.529	0.3796	16.21	≤ 0.0001

Table 4(d): Coefficients of: AssignedCriticalReason on Fatigue

Level of Fatigue	Coefficient	std. err.	Wald	prob
0	-0.8431	0.1572	28.77	≤ 0.0001
1	0.8431	0.1572	28.77	≤ 0.0001

Table 4(e): Coefficients of: AssignedCriticalReason on ClassYears

Covariate	Coefficient	std. err.	Wald	prob
ClassYears	-0.0229	7.480e-3	9.411	0.0022

Table 4(f): Coefficients of: AssignedCriticalReason on IDRSafetyBonus

Level of IDRSafetyBonus	Coefficient	std. err.	Wald	prob
0	0.3278	0.1045	9.837	0.0017
1	-0.3278	0.1045	9.837	0.0017

[4] Derived by calculating the predicted values from the logistic regression and then calculating the correlation between the predicted values and the dependent variable. The Pearson Product-Moment Correlation, 0.383, is squared to obtain the R^2.

Table 4(g): Coefficients of: AssignedCriticalReason on HoursDriving

Covariate	Coefficient	std. err.	Wald	prob
H_m	-0.0902	0.0266	11.53	0.0007

Table 4(h): Coefficients of: AssignedCriticalReason on MileagePayThisTrip(Driver)

Level of MileagePayThisTrip(Driver)	Coefficient	std. err.	Wald	prob
0	0.1863	0.0950	3.845	0.0499
1	-0.1863	0.0950	3.845	0.0499

Table 5: Industrial Relations/Economics Contributors to Assigned Critical Reason; Including ClassYearsSq

Type of analysis: Logistic

Factors

Name	F/R	Kind
WorkPressureTotal	Fix	Cont
AggressionCount	Fix	Cont
Fatigue	Fix	Disc
ClassYears	Fix	Cont
ClassYearsSq	Fix	Cont
IDRSafetyBonus	Fix	Disc
HoursDriving	Fix	Cont
MileagePayThisTrip(Driver)	Fix	Disc

Partial (Type 3) Sums of Squares

2284 total cases of which 1514 are missing

Iteration	LogLikelihood	Convergence
1	-469.179	0.064497
2	-469.171	0.004895
3	-469.171	0.000030
4	-469.171	0.000000

ANCOVA

Analysis of Covariance For	AssignedCriticalReason
2284 total cases of which 1514 are missing	

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	24.5309	24.5310	24.626	≤ 0.0001
WorkPressureTotal	1	10.4150	10.4150	10.455	0.0013
AggressionCount	1	16.6288	16.6288	16.693	≤ 0.0001
Fatigue	1	29.1529	29.1528	29.266	≤ 0.0001
ClassYears	1	6.30588	6.30588	6.3303	0.0121
ClassYearsSq	1	2.35687	2.35687	2.3660	0.1244
IDRSafetyBonus	1	9.85796	9.85796	9.8962	0.0017
HoursDriving	1	11.3722	11.3722	11.416	0.0008
MileagePayThisTrip(Driver)	1	3.91476	3.91476	3.9299	0.0478
Error	761	758.059	0.996136		
Total	769	848.263			

$R^2 = 0.151$[5]

Table 5(a)

Coefficients of: AssignedCriticalReason on Const

Level of Const	Coefficient	std. err.	Wald	prob
()	0.2326	0.3401	0.4679	0.4940

Table 5(b)

Coefficients of: AssignedCriticalReason on WorkPressureTotal

Covariate	Coefficient	std. err.	t Ratio	prob
WorkPressureTotal	0.0982	0.0291	3.371	0.0008

Table 5(c)

Coefficients of: AssignedCriticalReason on AggressionCount

Covariate	Coefficient	std. err.	t Ratio	prob
AggressionCount	0.2953	0.0636	4.640	≤ 0.0001

Table 5(d)

Coefficients of: AssignedCriticalReason on Fatigue

Level of Fatigue	Coefficient	std. err.	t Ratio	prob
0	-0.1694	0.0285	-5.955	≤ 0.0001
1	0.1694	0.0285	5.955	≤ 0.0001

[5] Derived by calculating the predicted values from the logistic regression and then calculating the correlation between the predicted values and the dependent variable. The Pearson Product-Moment Correlation, 0.383, is squared to obtain the R^2.

Table 5(e)

Coefficients of: AssignedCriticalReason on ClassYears

Covariate	Coefficient	std. err.	t Ratio	prob
ClassYears	-0.0118	4.517e-3	-2.609	0.0093

Table 5(f)

Coefficients of: AssignedCriticalReason on ClassYearsSq

Covariate	Coefficient	std. err.	t Ratio	prob
ClassYearsSq	1.802e-4	1.153e-4	1.563	0.1184

Table 5(g)

Coefficients of: AssignedCriticalReason on IDRSafetyBonus

Level of IDRSafetyBonus	Coefficient	std. err.	t Ratio	prob
0	0.0713	0.0215	3.315	0.0010
1	-0.0713	0.0215	-3.315	0.0010

Table 5(h)

Coefficients of: AssignedCriticalReason on HoursDriving

Covariate	Coefficient	std. err.	t Ratio	prob
HoursDriving	-0.0178	5.314e-3	-3.354	0.0008

Table 5(i)

Coefficients of: AssignedCriticalReason on MileagePayThisTrip(Driver)

Level of MileagePayThisTrip(Driver)	Coefficient	std. err.	t Ratio	prob
0	0.0411	0.0198	2.075	0.0383
1	-0.0411	0.0198	-2.075	0.0383

Table 6: Industrial Relations/Economics Contributors to Assigned Critical Reason; WorkPressureFactors added individually
Dependent variable: AssignedCriticalReason

Type of analysis: OLS ANCOVA[6]
Factors

Name	F/R	Kind
HurryingW	Fix	Disc
NewPosition_m	Fix	Disc
ShippingDeadline_m	Fix	Disc
EXPWorkSchedule_m	Fix	Disc
Quotas_m	Fix	Disc
ExtraLoads_m	Fix	Disc
Demoted_m	Fix	Disc
SelfInducedIllegal_m	Fix	Disc
SelfInducedOther_m	Fix	Disc
OtherPressure_m	Fix	Disc
LoadPressureIndicator_m	Fix	Disc
ShortNoticeTrips_m	Fix	Disc
FillInTrips_m	Fix	Disc
UnpaidLoading_m	Fix	Disc
OtherRelations_m	Fix	Disc
RotatingShift_m	Fix	Disc
AggressionCount_m	Fix	Cont
ClassYears_m	Fix	Cont
ClassYearsSq	Fix	Cont
IDRSafetyBonus_m	Fix	Disc
HoursDriving_m	Fix	Cont
MileagePayThisTrip(Driver)_m	Fix	Disc
Fatigue_m	Fix	Disc

Partial (Type 3) Sums of Squares
General Results
2284 total cases of which 1514 are missing

[6] ANCOVA produces very similar results to logistic regression. The number of indicator variables is so great in this model that it fails to converge, so the iterative stepwise exclusion of insignificant variables is based on the ANCOVA results.

ANCOVA

Analysis of Variance For	AssignedCriticalReason
2284 total cases of which 1514 are missing	

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	170.187	170.187	786.47	≤ 0.0001
HurryingW	1	0.677144	0.677144	3.1292	0.0773
NewPosition_m	1	0.633347	0.633347	2.9268	0.0875
ShippingDeadline_m	1	0.000106	0.000106	0.00049	0.9823
EXPWorkSchedule_m	1	0.002514	0.002514	0.01162	0.9142
Quotas_m	1	0.039057	0.039057	0.18049	0.6711
ExtraLoads_m	1	0.008725	0.008725	0.04032	0.8409
Demoted_m	1	0.029783	0.029783	0.13763	0.7107
SelfInducedIllegal_m	1	0.054709	0.054709	0.25282	0.6152
SelfInducedOther_m	1	0.223746	0.223746	1.0340	0.3096
OtherPressure_m	1	0.669908	0.669908	3.0958	0.0789
LoadPressureIndicator_m	1	0.344924	0.344924	1.5940	0.2072
ShortNoticeTrips_m	1	0.285064	0.285064	1.3173	0.2514
FillInTrips_m	1	0.022464	0.022464	0.10381	0.7474
UnpaidLoading_m	1	0.087566	0.087566	0.40466	0.5249
OtherRelations_m	1	0.207939	0.207939	0.96093	0.3273
RotatingShift_m	1	0.000420	0.000420	0.00194	0.9649
AggressionCount_m	1	4.10896	4.10896	18.988	≤ 0.0001
ClassYears_m	1	1.33945	1.33945	6.1899	0.0131
ClassYearsSq	1	0.486178	0.486178	2.2467	0.1343
IDRSafetyBonus_m	1	2.29759	2.29759	10.618	0.0012
HoursDriving_m	1	2.60144	2.60144	12.022	0.0006
MileagePayThisTrip(Driver)_m	1	0.778988	0.778988	3.5999	0.0582
Fatigue_m	1	7.29991	7.29991	33.734	≤ 0.0001
Error	746	161.430	0.216394		
Total	769	191.813			

Table 7: Industrial Relations/Economics Contributors to Assigned Critical Reason; Insignificant components of "WorkPressureFactors" removed iteratively
Dependent variable: AssignedCriticalReason

Type of analysis: Logistic
Factors

Name	F/R	Kind
HurryingW	Fix	Disc
NewPosition	Fix	Disc
OtherPressure	Fix	Disc
LoadPressureIndicator	Fix	Disc
ShortNoticeTrips	Fix	Disc
AggressionCount	Fix	Cont
ClassYears	Fix	Cont
ClassYearsSq	Fix	Cont
IDRSafetyBonus	Fix	Disc
HoursDriving	Fix	Cont
MileagePayThisTrip(Driver)	Fix	Disc
Fatigue	Fix	Disc

Partial (Type 3) Sums of Squares Design Help
General Results
2284 total cases of which 1514 are missing

Iteration	LogLikelihood	Convergence
1	-468.551	————————
2	-465.460	0.128249
3	-465.379	0.022646
4	-465.379	0.001261
5	-465.379	0.000007

Table 7(a): ANCOVA

Analysis of Covariance For	AssignedCriticalReason
2284 total cases of which 1514 are missing	

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	25.9094	25.9094	25.906	≤ 0.0001
HurryingW	1	3.75630	3.75630	3.7559	0.0530
NewPosition	1	3.11458	3.11458	3.1142	0.0780
OtherPressure	1	3.60288	3.60288	3.6024	0.0581
LoadPressureIndicator	1	2.97011	2.97011	2.9698	0.0852
ShortNoticeTrips	1	2.36793	2.36793	2.3677	0.1243
AggressionCount	1	16.2455	16.2455	16.244	≤ 0.0001
ClassYears	1	6.80414	6.80414	6.8033	0.0093
ClassYearsSq	1	2.65486	2.65486	2.6545	0.1037
IDRSafetyBonus	1	9.87739	9.87739	9.8762	0.0017
HoursDriving	1	13.0045	13.0045	13.003	0.0003
MileagePayThisTrip(Driver)	1	3.44807	3.44807	3.4477	0.0637
Fatigue	1	36.8092	36.8092	36.805	≤ 0.0001
Error	757	757.090	1.00012		
Total	769	850.765			

$R^2 = 0.154$[7]

Table 7(b): Coefficients of: AssignedCriticalReason on HurryingW

Level of HurryingW	Coefficient	std. err.	Wald	prob
0	-1.069	0.5515	3.756	0.0526
1	1.069	0.5515	3.756	0.0526

Table 7(b): Coefficients of: AssignedCriticalReason on NewPosition

Level of NewPosition	Coefficient	std. err.	Wald	prob
0	-0.7557	0.4282	3.115	0.0776
1	0.7557	0.4282	3.115	0.0776

Table 7(c): Coefficients of: AssignedCriticalReason on OtherPressure

Level of OtherPressure	Coefficient	std. err.	Wald	prob
0	-1.060	0.5582	3.603	0.0577
1	1.060	0.5582	3.603	0.0577

[7] Derived by calculating the predicted values from the logistic regression and then calculating the correlation between the predicted values and the dependent variable. The Pearson Product-Moment Correlation, 0.383, is squared to obtain the R^2.

Table 7(d): Coefficients of: AssignedCriticalReason on LoadPressureIndicator

Level of LoadPressureIndicator	Coefficient	std. err.	Wald	prob
0	-0.4780	0.2774	2.970	0.0848
1	0.4780	0.2774	2.970	0.0848

Table 7(e): Coefficients of: AssignedCriticalReason on ShortNoticeTrips

Level of ShortNoticeTrips	Coefficient	std. err.	Wald	prob
0	0.7262	0.4720	2.368	0.1239
1	-0.7262	0.4720	2.368	0.1239

Table 7(f): Coefficients of: AssignedCriticalReason on AggressionCount

Covariate	Coefficient	std. err.	Wald	prob
AggressionCount	1.535	0.3808	16.25	$\leq$ 0.0001

Table 7(g): Coefficients of: AssignedCriticalReason on ClassYears

Covariate	Coefficient	std. err.	Wald	prob
ClassYears	-0.0551	0.0211	6.804	0.0091

Table 7(h): Coefficients of: AssignedCriticalReason on ClassYearsSq

Covariate	Coefficient	std. err.	Wald	prob
ClassYearsSq	8.720e-4	5.352e-4	2.655	0.1032

Table 7(i): Coefficients of: AssignedCriticalReason on IDRSafetyBonus

Level of IDRSafetyBonus	Coefficient	std. err.	Wald	prob
0	0.3308	0.1052	9.877	0.0017
1	-0.3308	0.1052	9.877	0.0017

Table 7(j): Coefficients of: AssignedCriticalReason on HoursDriving

Covariate	Coefficient	std. err.	Wald	prob
HoursDriving	-0.0979	0.0271	13.00	0.0003

Table 7(k): Coefficients of: AssignedCriticalReason on MileagePayThisTrip(Driver)

Level of MileagePayThisTrip(Driver)	Coefficient	std. err.	Wald	prob
0	0.1774	0.0955	3.448	0.0633
1	-0.1774	0.0955	3.448	0.0633

Table 7(l): Coefficients of: AssignedCriticalReason on Fatigue

Level of Fatigue	Coefficient	std. err.	Wald	prob
0	-0.9546	0.1573	36.81	$\leq$ 0.0001
1	0.9546	0.1573	36.81	$\leq$ 0.0001

Discussion
Analysis of the economic contributors to large truck crash causation suggest that economic variables related to driver work pressures and compensation significantly predict truck crashes. Recall the model formulated above:

AssignedCriticalReason (ACR) = $\alpha + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7 + \beta_8 + \varepsilon$

where α = constant, β_1 = WorkPressureTotal, β_2 = AggressionCount, β_3 = Fatigue, β_4 = ClassYears, β_5 = ClassYearsSq, β_6 = SafetyBonus, β_7 = HoursDriving, β_8 = MileagePayThisTrip(Driver), and ε = error. Theory, previous research, and confirmatory logistic regression and factor analysis supports

Logistic regression, shown in Table 5, demonstrates the following significant effects. The LogLikelihood of this model is -469.171, which is significant, and the R^2 = 0.151. All variables are significant individually except for ClassYearsSq, which approaches significance with a t-statistic of 1.563 and a probability that this ClassYearsSq provides the expected diminution of the ClassYears effect over time of 12%.

ACR = $0.2326 + (0.0982)\beta_1 + (0.2953)\beta_2 + (0.1694)\beta_3 + (-0.0118)\beta_4 + (1.802e^{-4})\beta_5 + (-0.0713)\beta_6 + (-0.0178)\beta_7 + (-0.0411)\beta_8 + \varepsilon$

Interpretation of the foregoing results indicates the extent to which the presence of "1" in the dependent variable (that is, the data collectors' assignment of the critical reason for the critical event to the truck) is associated with the independent variables that predict this assignment. That is, every assignment of the critical reason to the truck is associated positively and significantly with WorkPressureTotal, AggressionCount, and Fatigue, according to the coefficients reported above. Every assignment of the critical reason to the truck is associated negatively and significantly with ClassYears, SafetyBonus, HoursDriving, and MileagePay, according to the coefficients reported above.

Logistic regression on WorkPressureTotal shows the factors that contribute to the validity of the index. Weaknesses in the data collection, compilation, reporting, and database construction have made it extremely difficult to identify significant factors predicting crashes, but this analysis suggests areas for future research and analysis. The low R^2 of both models shows that even though significant predictors have been found, the model overall predicts only about 15% of the variance, suggesting that most of the economic factors predicting crashes remain unknown.

Logistic regression shows that work pressure significantly contributes to crashes. The strength of the work pressure factor is evident in the model that uses the single constructed factor, which is highly significant. In the model that decomposes this factor, HurryingW, NewPosition, OtherPressure, and LoadPressureIndicator separately are significant at the 0.05 and 0.10 level and ShortNoticeTrips is nearly significant (which is why it is included in the output). IDRSafetyBonus, HoursDriving, and MileagePayThisTrip(Driver) are significant predictors as well (see Table 7). Like the work pressure variables, these compensation and work intensity proxies show that these factors are important predictors of truck crashes. However, the fact that the factors may be important as a pathway through which crashes can be predicted. In other words, while the work pressure variables appear to be relatively weak predictors individually, when combined into the single highly significant WorkPressureTotal construct suggests that the overall work-pressure factor is highly predictive of crashes (a two-stage relationship). Finally, Fatigue and AggressionCount concepts also have strong predictive value although it is not clear from the data just how closely fatigue and aggression follow from work-related constructs.

Conclusion

Previous research has shown that human capital factors are very important predictors of truck crashes and human capital value is reflected in compensation levels (Rodriguez et al. 2003; Rodriguez, Targa, and Belzer 2006), but the weaknesses in the LTCCS data set are such that careful additional research would be necessary to try to determine whether path modeling of these factors, using multiple-stage factor models, might reveal patterns of significance. This analysis supports the previous findings that work organization, economic pressure, and compensation directly affects safety. In the logistic regression used for this analysis, the use of factors embedded within the regression – factors created, like in a two-stage model – from independent variables that are not generally significant individually (see Tables 6 and 7) but which are highly significant when combined into a single factor (confirmed by factor analysis in Table 1 and OLS ANOVA in Table 3), suggests that these contributing factors are significant but only when considered as a system, as in an employment relationship in organized activity within a market.

Economic theory predicts that driver quality – including safety – is directly related to the human capital of the worker, and compensation is the ideal predictor of human capital. In addition, pressure exerted on the worker either by his/her employer, his/her customer, or by the collective pressure of the market itself, clearly affects the driver's safety. Though the causes of fatigue and driver aggressiveness are not determined here, they may be associated with work-related factors. Efforts on the part of regulators to create and reinforce an environment conducive to reducing the stresses that cause fatigue and aggressiveness will help to reduce crashes as well. This exploratory analysis of the Large Truck Crash Causation Study data therefore confirms that these factors are strongly related to truck safety, but further data collection designed to remedy the flaws in the LTCCS is needed to examine these factors in more depth.

One would be tempted to conclude that economic factors predict only 15% of truck crash probability, but this would be incorrect because the data weaknesses have presented the greatest barrier to this analysis. A more conservative statement is that analysis of these particular data show that 15% of truck crash probability can be predicted using LTCCS. A carefully designed survey in which key economic factors that might predict crashes are identified and collected would provide the soundest basis on which to determine the extent to which such factors contribute to truck crashes. This would require developing a research design for such a study and collecting data that such a design identifies as possible factors. Data collected this way would allow a statistical analysis of industrial relations and economic factors predicting truck crashes.

References

Belzer, Michael H. 2000. Sweatshops on Wheels: Winners and Losers in Trucking Deregulation. Oxford, UK and New York, NY: Oxford University Press.

———. 2004. Trucking Industry Benchmarking Program: Partners in Safety (P04-0632). In Annual Meeting of the Transportation Research Board. Washington, DC: TRB Task Force on Trucking Industry Research.

Belzer, Michael H., Kenneth L. Campbell, Stephen V. Burks, Dale Ballou, George Fulton, Donald Grimes, and Kristen A. Monaco. 1999. Hours of Service Impact Assessment. Ann Arbor, MI: University of Michigan Transportation Research Institute and Federal Highway Administration, Office of Motor Carriers and Highway Safety.

Belzer, Michael H., Daniel A. Rodriguez, and Stanley A. Sedo. 2002. Paying for Safety: An Economic Analysis of the Effect of Compensation on Truck Driver Safety. Washington, DC: United States Department of Transportation, Federal Motor Carrier Safety Administration.

Blower, Daniel, and Kenneth L. Campbell. 2005. Methodology of the Large Truck Crash Causation Study. In Large Truck Crash Causation Study Analysis Series. Washington, DC: Federal Motor Carrier Safety Administration.

Council, Forrest M., Michael H. Belzer, John R. Billing, Kenneth L. Campbell, James W. Dally, Anne T. McCartt, Hugh W. McGee, A. James McKnight, Jack Stuster, Steven J. Vaughn, and Frank R. Wilson. 2003. Letter Report to Federal Motor Carrier Safety Administration Administrator Annette M. Sandberg. Washington, DC: Transportation Research Board.

Donaldson, Gerald A. 2005. The Large Truck Crash Causation Study: Fundamental Problems and Issues. In 2005 International Truck and Bus Safety and Security Symposium. Alexandria, VA, USA.

Hedlund, James, and Daniel Blower. 2006. Using LTCCS Data for Statistical Analyses of Crash Risk. In Large Truck Crash Causation Study Analysis Series. Washington: Federal Motor Carrier Safety Administration.

Lazear, Edward P. 1990. "Pensions and Deferred Benefits as Strategic Compensation." Industrial Relations Vol.29, No.2:pp 263-80.

Mayhew, Claire, and Michael Quinlan. 1997. Trucking Tragedies: Why Occupational Health and Safety Outcomes are Worse for Subcontract Workers in the Road Transport Industry. Sydney, NSW, Australia: Worksafe Australia.

———. 2000. Occupational Health and Safety Amongst 300 Long Distance Truck Drivers: Results of An Interview-Based Survey. In Report of Inquiry into Safety in the Long Haul Trucking Industry.

———. 2006. "Economic pressure, multi-tiered subcontracting and occupational health and safety in Australian long-haul trucking." Employee Relations Vol.28, No.3:pp 212-29.

Quinlan, Michael. 2001. Report of Inquiry into Safety in the Long Haul Trucking Industry. Sydney, NSW, Australia: Motor Accidents Authority of New South Wales. Original edition, Appendix 3: Occupational Health and Safety Amongst 300 Long Distance Truck Drivers: Results of an Interview-Based Survey (not attached).

Quinlan, Michael, Claire Mayhew, and Richard Johnstone. 2006. "Trucking Tragedies: The Hidden Disaster of Mass Death in the Long-Haul Road Transport Industry". In Working Disasters: The Politics of Recognition and Response, edited by E. Tucker: Baywood Publishing.
Quinlan, Michael, Lance Wright, and National Transport Commission. 2008. Safe Payments: Addressing the Underlying Causes of Unsafe Practices in the Road Transport Industry. Melbourne: National Transport Commission.
Rodriguez, Daniel A., Marta Rocha, Asad J. Khattak, and Michael H. Belzer. 2003. "Effects of Truck Driver Wages and Working Conditions on Highway Safety: Case Study." Transportation Research Record Vol.Freight Policy, Economics, and Logistics; Truck Transportation, No.1833:pp 95-102.
Rodriguez, Daniel A., Felipe Targa, and Michael H. Belzer. 2006. "Pay Incentives and Truck Driver Safety: A Case Study." Industrial and Labor Relations Review Vol.59, No.2:pp 205-25.
Williamson, Ann M., and Anne-Marie Feyer. 1992. Stage 1: The industry perspective. In Strategies to combat fatigue in the long distance road transport industry. Australia: Federal Office of Road Safety.

Appendix 1
Variable Evaluation

Overall Note
We followed the requirement indicated on page 5 of the User's Manual:
"The RATWeight variable should be used to filter out invalid cases. Only cases with a positive RATWeight are valid for the purpose of the study." While 2284 cases exist in the data set (cases defined as crashes by vehicle, or 2284 vehicles), we set 206 of these cases to missing because RATWeight = 0.

We also note the following from page 1 of the Codebook:
"The cases with zero weights either are from the pilot phase of the project or are cases that were sampled but later found to not meet the selection criteria. They are included for clinical analysis or as anecdotal data. They should not be included in estimates of nationally-representative results."

We also followed the requirement noted at the bottom of page 6 of the "LTCCS Analytical User's Manual" to remove invalid cases:

"3) Select only cases with values for RATWeight > 0. Please note: this step is optional. Including this step will filter out large trucks which have zero weighting. There were 107 such cases, which were surveyed as part of the study but given a weight of zero because they did not meet the study criteria for one reason or another. These cases were still included in the study dataset, however, as additional qualitative data."

We confirmed independently that we must throw out 107 cases for this reason. This is not optional for a quantitative analysis. We discovered many problems, including multiple entries for many of these variables, with no way to distinguish acceptable from unacceptable entries.

The analysis will not use RATWeight, however, because I understand that the weights on which these were based are approximately 35 years old and apply only to cars because they were created to weight automobile traffic use in the original NASS sample. One cannot assume these are correct for trucks, especially 35 years later, since truck travel follows patterns that may be quite different. In addition, a close look at the data suggests that some zones were much more likely than others to cooperate and some researchers appear to have left more gaps in data collected on certain variables. For these reasons, this analysis will only analyze the raw data.

CrashAssessment Relation (n=2284)

CaseID

VehicleNumber[1]

- ACRCriticalEvent[2]
- CriticalEventCat
- ACRReason[3]
- ReasonCat

DriverAssessment Relation (n=2284)[4]

CaseID

VehicleNumber

- Truck[5]

---

[1] CaseID and VehicleNumber must be in all data sets as it links the sets together.

[2] This could be a dependent variable, but the critical event is assigned to 80% of the trucks involved in these crashes. This is an unbalanced dependent variable

[3] This will be a dependent variable. The critical reason is assigned to 50% of the trucks in these crashes, making it a balanced dependent variable. "The critical reason is the immediate reason for this event and is often the last failure in the causal chain (i.e. closest in time to the critical precrash event)", according to the *LTCCS Analytical User's Manual*.

[4] I have collapsed variables in red font in the bulleted list below into the derived variables listed here. There are only 40 positive responses to these questions across all of these variables and the originals have been deconstructed into indicator variables that show only the presence of each coded effect, commingling "absent" and the absence of information. I don't think it is cost-effective to merge these individual variables into the master relation.

[5] "Truck" is defined as follows, per page 5 in the Manual. I excluded 79 because it is defined as "unknown truck type" and includes "light", while 78 is more precise. An alternative analysis would exclude medium-duty trucks define "truck" as greater than 8,850 kg (19,512 pounds), because it more likely reflects the general definition of a CMV operated by a driver who has a CDL. This study appears to have considered this factor less relevant, however, since categories do not follow those requirements.

60 Step van
61 Single unit straight truck(4500kg<GVWR<=8850kg)
62 Single unit straight truck(8850kg<GVWR<=12000kg)
63 Single unit straight truck (GVWR > 12,000kg)
64 Single unit straight truck (GVWR unknown)
66 Truck-tractor (Cab Only, or any trailing units)

- ADRWorkSchedule
- Fatigue[6]
- Upset
- Hurrying
- WorkPressureCount (continuous variable)
- WorkPressureCountD (continuous derived variable)[7]
- WorkPressureTotalD (continuous derived variable)[8]

---

67 Truck-tractor with no cargo trailer
68 Truck-tractor pulling one trailer
70 Truck-tractor (unknown if pulling trailer)
74 Medium/heavy Pickup (>=4,536kg)
78 Unknown medium/heavy truck type

It is not clear why the documentation defines truck as excluding the following:
69 Truck-tractor pulling two or more trailers. From page 5 of the Analytical Users Manual (http://152.122.44.126/ltccs/data/documents/LTCCS_Manual_Public.pdf): Select only cases with values for GVEBodyType = {60, 61, 62, 63, 64, 66, 67, 68, 70, 74, 78}.

[6] **Driver Fatigue**
Definition: This variable assesses driver fatigue at the time of the crash. The assessment is based on an evaluation of the driver's current and preceding sleep schedules, current and preceding work schedules, and a variety of other fatigue-related factors including recreational and non-work activities.
Source: Determined by the Case Reviewer using all available information inputs. The primary data source here is the driver interview, however, due to the inaccuracies inherent in these data, the Case Reviewer should compare driver responses with other data sources including log book entries, time stamped fuel and toll receipts, carrier records, and other interview sources to determine the veracity of the driver responses. The final assessment of fatigue involvement is made from all of these sources and may include the on-site assessments of the NASS Researcher.
Cross Reference: Elaborates on Overview.DriverFatigue. Related to the IntvwDrFatigue data set, values will differ in part due to Interview vs. Researcher determined values.
Variable Name: Fatigue
Attribute Codes
Code Meaning
0 Driver not fatigued
1 Driver fatigued
8 No driver present
9 Unknown

[7] The sum of 'NewPosition', 'ShippingDeadline', 'EXPWorkSchedule', 'Quotas', 'ExtraLoads', 'Demoted', 'SelfInducedIllegal', 'SelfInducedOther', 'OtherPressure', 'Upset', 'Hurrying', 'Emotional'. It is unclear how "Upset", "Hurrying", "Emotional", or "OtherPressure" can contribute to work-related pressure, except insofar as a work-based source has been identified (which is not the case with these variables).

- WorkPressureTotalD_Binary[9]
    - NewPosition
    - ShippingDeadline
    - EXPWorkSchedule[10]
    - Quotas
    - ExtraLoads
    - Demoted
    - SelfInducedIllegal
    - SelfInducedOther
    - OtherPressure
    - LoadPressure
    - LoadPressureIndicator (derived)
    - RotatingShift
    - ShortNoticeTrips
    - FillInTrips
    - UnpaidLoading
    - OtherRelations
    - OtherPressure
    - Hurrying
- ComfortCount (continuous variable)

DriverDecisionAggression Data Set (n=2284)

---

[8] This derived variable, created by Sound Science, represents occupational concepts related to the work process. It is the sum of 'NewPosition', 'ShippingDeadline', 'EXPWorkSchedule', 'Quotas', 'ExtraLoads', 'Demoted', 'SelfInducedIllegal', 'SelfInducedOther', 'OtherPressure', 'LoadPressureIndicator' (under pressure to accept scheduled and unscheduled loads, loads proffered on short notice or when over legal driving hours), 'ScheduledExtensions', 'RotatingShift', 'ShortNoticeTrips', 'FillInTrips', 'UnpaidLoading', and 'OtherRelations' from the DriverAssessment Relation. This variable is necessary because the data reported in the LTCCS is very limited on these concepts and many of the results do not meet the "smell test" for accuracy.

[9] For purposes of analysis, this variable counts 147 nonzero factors; 1,931 respondents reported no such factors. In the analysis, a binary variable has been created for any positive values, and the model includes any respondents who report any work pressures. I also made an arithmetic transformation, adding 1 to this variable so that the zero values (indicating no presence of work pressure factors) are transformed to one and all other values increased by one. This removes any effect of zero in the calculations but preserves the ordinality of the variables, which is all this measures.

[10] Scheduled and Unscheduled Extensions strongly predict EXPWorkSchedule in logistic regression so they are excluded here.

CaseID
VehicleNumber

- Speeding
- Tailgating
- Misjudgment
- Approach
- ADATravelSpeed (continuous variable)
- Assumption
- Evasion
- ADAOtherFactor
- ManeuverCount (continuous variable)
- Undertaking
- WrongTurnLane
- IllegalUTurn
- RanLights
- WrongWay
- OtherManeuver
- AggressionCount (continuous variable)[11]
- SpeedingBehavior
- TailgatingBehavior
- Weaving
- LightViolations
- RapidAcceleration
- Honking
- Flashing
- ObsceneGestures (all values are "0" or "no")
- BlockingOthers
- OtherAggression
- AggressionReason

DriverSleep Data Set (n=2284)
CaseID
VehicleNumber

- LastSleepHours
- LastSleepHoursD (continuous non-ordinal derived variable)
- LastSleepStart
- LastSleepStartD (continuous non-ordinal derived variable)

[11] Computed from the following variables in the DriverDecisionAggression data set: SpeedingBehavior, TailgatingBehavior, Weaving, LightViolation, RapidAcceleration, Honking, Flashing, ObsceneGestures, BlockingOthers, OtherAggression.

- LastSleepEnd
- LastSleepEndD (continuous non-ordinal derived variable)
- HoursSinceSleep
- HoursSinceSleepD (continuous non-ordinal derived variable)
- MainSleepHours
- MainSleepHoursD (continuous non-ordinal derived variable)
- MainSleepStart
- MainSleepStartD (continuous non-ordinal derived variable)
- MainSleepEnd
- MainSleepEndD (continuous non-ordinal derived variable)
- HoursDriving
- HoursDrivingD (continuous non-ordinal derived variable)
- ADSHoursOnDuty
- ADSHoursOnDutyD (continuous non-ordinal derived variable)
- WeekLongest
- WeekLongestD (continuous non-ordinal derived variable)
- WeekAverage
- WeekAverageD (continuous non-ordinal derived variable)
- Rotation
- HoursWorked
- HoursWorkedD (continuous non-ordinal derived variable)[12]
- ADSLongestDay
- ADSLongestDayD (continuous non-ordinal derived variable)[13]
- ADSShortestDay
- ADSShortestDayD (continuous non-ordinal derived variable)[14]

---

[12] Note that manual indicates that the response "00:00" for this variable about hours worked the day of the crash indicates that it means "no sleep," and that it is physically impossible to have a crash when not working even one minute of the day. Since this is meaningless, I have set to missing.

[13] "This is the number of hours the driver slept in the 7-day interval preceding the crash that represents his/her longest interval of daily sleep." How is it possible for longest interval of daily sleep in 7 days to be zero? Twenty-five drivers, according to this data, did not sleep at all in preceding 7 days. It appears that many case measures for LongestDay, ShortestDay, and possibly others are faulty in the data because they claim zero (00:00) for both longest day's sleep and shortest day's sleep. Since these could only be true if driver had no sleep at all for the week, and since such an occurrence is virtually impossible, I will have to invalidate all such responses even if they are true (i.e., that the driver had no hours of sleep).

[14] Underlying variable has both 00:00 and 24:00. The latter returns "0" for a value when using "Seconds" transformation.

- AverageDay
- AverageDayD (continuous non-ordinal derived variable)[15]
- LastWeekHours
- LastWeekHoursD (continuous non-ordinal derived variable)[16]

GeneralVehicle (n=2284)

CaseID

VehicleNumber

- GVEBodyType
- GVETruck (derived)[17]
- GVE CDL Truck (derived)[18]

---

[15] Note that 26 cases indicate average length of daily sleep in past 24 is zero. These all are faulty data entry and I have set them to zero. These may be calculated from 00:00 entries for ShortestDay and LongestDay.

[16] Code is: "if left('LastWeekHours', 2) = 99 then • else left('LastWeekHours', 2)&"."&(right('LastWeekHours', 2)/60*100)" for derived variable.

[17] Definition:

60 Step van
61 Single unit straight truck(4500kg<GVWR<=8850kg)
62 Single unit straight truck(8850kg<GVWR<=12000kg)
63 Single unit straight truck (GVWR > 12,000kg)
64 Single unit straight truck (GVWR unknown)
66 Truck-tractor (Cab Only, or any trailing units)
67 Truck-tractor with no cargo trailer
68 Truck-tractor pulling one trailer
70 Truck-tractor (unknown if pulling trailer)
74 Medium/heavy Pickup (>=4,536kg)
78 Unknown medium/heavy truck type

It is not clear why the documentation defines truck as excluding the following:
69 Truck-tractor pulling two or more trailers.
From page 5 of the Analytical Users Manual
(http://152.122.44.126/ltccs/data/documents/LTCCS_Manual_Public.pdf):
Select only cases with values for GVEBodyType = {60, 61, 62, 63, 64, 66, 67, 68, 70, 74, 78}. This derived variable uses exactly the definition provided in the Manual.

[18] These vehicles fall as close to the definition of "commercial motor vehicle" as is possible within this data set. The law defines a CMV as follows, but the data set defines the nearest category as >12,000 kg, which is slightly greater than this threshold. The mandate of the LTCCS, however, defined a "truck" as greater than 10,000 pounds.

Part 383: Commercial Driver's License Standards; Requirements and Penalties

- GVEVehicleClass
- GVEVehicleClassTruck (derived)
- GVEVehicleClassCDL (derived)[19]

<http://www.fmcsa.dot.gov/rules-regulations/administration/fmcsr/fmcsrruletext.asp?section=383.5#49CFR383.5>
§383.5 Definitions.

- Commercial driver's license (CDL) means a license issued by a State or other jurisdiction, in accordance with the standards contained in 49 CFR Part 383, to an individual which authorizes the individual to operate a class of a commercial motor vehicle.
- Commercial motor vehicle (CMV) means a motor vehicle or combination of motor vehicles used in commerce to transport passengers or property if the motor vehicle-
  (a) Has a gross combination weight rating of 11,794 kilograms or more (26,001 pounds or more) inclusive of a towed unit(s) with a gross vehicle weight rating of more than 4,536 kilograms (10,000 pounds); or
  (b) Has a gross vehicle weight rating of 11,794 or more kilograms (26,001 pounds or more); or
  (c) Is designed to transport 16 or more passengers, including the driver; or
  (d) Is of any size and is used in the transportation of hazardous materials as defined in this section.

Definition: (63) Single unit straight truck (GVWR > 12,000kg); (66) Truck-tractor (Cab Only, or any trailing units); (67) Truck-tractor with no cargo trailer; (68) Truck-tractor pulling one trailer; (69) Truck-tractor pulling two or more trailers; (70) Truck-tractor (unknown if pulling trailer).

[19] CDL truck defined for “vehicle class” variable:
63 Single unit straight truck (GVWR > 12,000kg)
66 Truck-tractor (Cab Only, or any trailing units)
67 Truck-tractor with no cargo trailer
68 Truck-tractor pulling one trailer
69 Truck-tractor pulling two or more trailers
70 Truck-tractor (unknown if pulling trailer)
78 Unknown medium/heavy truck type

These vehicles fall as close to the definition of "commercial motor vehicle" as is possible within this data set. These are what I would consider "medium/large", as in the miscellaneous category coded "78". The law defines a CMV as below, but the data set defines the category as >12,000 kg, which is slightly greater than this threshold.

Part 383: Commercial Driver's License Standards; Requirements and Penalties

---

http://www.fmcsa.dot.gov/rules-regulations/administration/fmcsr/fmcsrruletext.asp?section=383.5#49CFR383.5

§383.5 Definitions.
Commercial driver's license (CDL) means a license issued by a State or other jurisdiction, in accordance with the standards contained in 49 CFR Part 383, to an individual which authorizes the individual to operate a class of a commercial motor vehicle.

Commercial motor vehicle (CMV) means a motor vehicle or combination of motor vehicles used in commerce to transport passengers or property if the motor vehicle-

(a) Has a gross combination weight rating of 11,794 kilograms or more (26,001 pounds or more) inclusive of a towed unit(s) with a gross vehicle weight rating of more than 4,536 kilograms (10,000 pounds); or
(b) Has a gross vehicle weight rating of 11,794 or more kilograms (26,001 pounds or more); or
(c) Is designed to transport 16 or more passengers, including the driver; or
(d) Is of any size and is used in the transportation of hazardous materials as defined in this section.

These vehicles fall as close to the definition of "commercial motor vehicle" as is possible within this data set. The law defines a CMV as below, but the data set defines the category as >12,000 kg, which is slightly greater than this threshold.

Part 383: Commercial Driver's License Standards; Requirements and Penalties
http://www.fmcsa.dot.gov/rules-regulations/administration/fmcsr/fmcsrruletext.asp?section=383.5#49CFR383.5

§383.5 Definitions.
Commercial driver's license (CDL) means a license issued by a State or other jurisdiction, in accordance with the standards contained in 49 CFR Part 383, to an individual which authorizes the individual to operate a class of a commercial motor vehicle.

Commercial motor vehicle (CMV) means a motor vehicle or combination of motor vehicles used in commerce to transport passengers or property if the motor vehicle-
(a) Has a gross combination weight rating of 11,794 kilograms or more (26,001 pounds or more) inclusive of a towed unit(s) with a gross vehicle weight rating of more than 4,536 kilograms (10,000 pounds); or

- EthnicOrigin
- CMVCrashes (continuous variable)[20]
- CMVViolations (continuous variable)[21]
- NonCMVCrashes (continuous variable)[22]
- NonCMVViolations (continuous variable)[23]

(b) Has a gross vehicle weight rating of 11,794 or more kilograms (26,001 pounds or more); or
(c) Is designed to transport 16 or more passengers, including the driver; or
(d) Is of any size and is used in the transportation of hazardous materials as defined in this section.

[20] Manuals provide conflicting documentation. Public Manual and Fed Manual say:
88 No driver present
97 No official records found
Codebook says
88 No official records found
97 No driver present
98 Not a CMV driver
99 Unknown

[21] Manual says the values are:
88 No driver present
97 No official records found
98 Not a CMV driver
Codebook says the values are:
88 No official records found
97 No driver present
98 Not a CMV driver

[22] Manual says the values are:
88 No driver present
97 No official records found
Codebook says the values are:
88 No official records found
97 No driver present
98 Not a CMV driver
99 Unknown

[23] Manual says the values are:
88 No driver present
97 No official records found
Codebook says the values are:
88 No official records found
97 No driver present

- TotalCrashes (continuous variable)[24]
- GVETotalViolations (continuous variable)[25]
- PrevViolations (continuous variable)[26]
- MCMIScrashes (continuous variable)[27]
- MCMIScmvViolations (continuous variable)[28]

---

98 Not a CMV driver

[24] Manual says the values are:
88 No driver present
97 No official records found
Codebook says the values are:
88 No official records found
97 No driver present
98 Not a CMV driver
99 Unknown

[25] Manual says the values are:
88 No driver present
97 No official records found
Codebook says the values are:
88 No official records found
97 No driver present
98 Not a CMV driver
99 Unknown

[26] Manual says the values are:
88 No driver present
97 No official records found
Codebook says the values are:
88 No official records found
97 No driver present
98 Not a CMV driver

[27] Manual says the values are:
88 No driver present
97 No official records found
98 Not a CMV driver
Codebook says the values are:
88 No official records found
97 No driver present
98 Not a CMV driver

[28] Manual says the values are:
88 No driver present
97 No official records found

- MCMIScmvNonViolations (continuous variable)[29]

IntvwCarrier Data Set (n=1244)

CaseID

VehicleNumber

- BussType
- BussTypeD[30]
- CAICarrierType
- CAIStatus[31]
- OperType
- Owner
- O-O (owner-operator indicator variable)[32]
- TruckOper (continuous variable)[33]
- TractorOper (continuous variable)[34]

---

98 Not a CMV driver

Codebook says the values are:

88 No official records found
97 No driver present
98 Not a CMV driver

[29] Manual says the values are:

88 No driver present
97 No official records found
98 Not a CMV driver

Codebook says the values are:

88 No official records found
97 No driver present
98 Not a CMV driver

[30] Business Type "unknown" set to missing.

[31] I don't understand the definition because the "intent of the shipper" does not define the carrier type legally. The carrier must be defined as an interstate carrier if it has authority to operate out of state of domicile, which triggers a number of legal requirements. A shipper does not define this for the carrier. An exception may exist for cities, like Kansas City, with a greater metropolitan area that crosses state lines (I frankly am not up on the regulatory details here) and where the state boundary is virtually invisible in a practical sense.

[32] Owner-operator indicator variable created by Sound Science. Coded as O-O when truck is "Leased (30 days+) from owner/operator" (2) or "Short term rental from owner/operator" (4).

[33] Straight trucks.

- PowerUnits (continuous derived variable created by Sound Science)[35]
- TrailerOper (continuous variable)[36]
- CarrierExp (originally in YYMM form; continuous variable)
- SubjectSafety
- FederalSafety
- StateSafety
- Fulltime (continuous variable)[37]
- Parttime (continuous variable)
- AdvanceNotice[38]
- AdvanceNoticeD[39]
- ExtentNotice (continuous variable)
- SecondJobQuery
- SecondJob[40]

---

[34] I created a variable "PowerUnits" that sums up the straight trucks and tractors. Twenty-two of the cases involve carriers that operate no power units at all, according to these data. How can a carrier operate no power units unless it is just a broker? If it is, it is not a carrier and these cases should be treated as "missing". I can find no variable that indicates the number of "owner-operators" or contractors a carrier uses.

[35] Sum of straight trucks and tractors operated.

[36] There is no codebook entry for this variable so I cannot verify, but 14 cases indicate 8887 trailers operated and 46 cases indicate 9999 trailers operated. These numbers are way different from adjacent counts by a factor of 14-46, and the other trailer counts anywhere near that range of number of trailers round off to the 100. I believe these all are mis-coded entries (should have been 99999) and should be missing. Note also that the codes for 8887 and 9999 do not even exist in the manual.

[37] Quite surprising that only 27 carriers employ zero full-time drivers. Did they code owner-operators as "not applicable?" This seems unlikely also since only 13 cases are coded "not applicable", but it is hard to determine how they code owner-operators.

[38] Main coded responses were "yes", "no", and "Trips scheduled well ahead (fixed schedule)", which also is "yes". For some reason, the contractor coded overlapping responses. The following variable, ""ExtentNotice" should cover the third category. As usual, "other (specify)" is not specified and there is no suggestion regarding the possible meanings of "other" in a yes/no question. The finding, however, is that 7.3% of respondents report short notice trips (though "short notice" is undefined); 12.3% of those for whom there is a valid response report short notice trips.

[39] For this derived variable I collapsed category 1 (yes) and 3 ["Trips scheduled well ahead (fixed schedule)"] together. Since "other" is not specified, I set it to missing.

[40] Carriers were aware of 20 drivers who have second jobs. For this variable I set "yes" = 1 and "no" = 0, rendering the remaining as missing.

- HoursWorked7Days (continuous variable)[41]
- SecondJobCons[42]
- TripLog[43]
- MonitorHour[44]
- MinimumAge (continuous variable)[45]
- YearsExpDriv (continuous variable)[46]
- YearsExpVeh (continuous variable)
- PassedTraining
- BackgroundCheck
- DrivingTimeComp[47]

[41] Of carriers aware that drivers have second jobs, only 10 carriers report hours worked on 10 of the 20 drivers who reported having second jobs. This suggests that half of the carriers aware of drivers' second jobs do not account for work time in those second jobs when assigning work themselves, which very possibly would push those drivers over hours, but unreported and unrecorded.

[42] For this variable I set "yes" = 1 and "no" = 0, rendering the remaining as missing.

[43] For this variable I set "yes" (1) = 1, "no" (2) = 0, and "no, exempt" (3) = 0, rendering the remaining as missing

[44] Definition of "monitor" unclear and undefined.

[45] Median minimum age reported is 27, which frankly is implausible unless firms that have higher maturity standards have systematically higher crash probability. The fact that the "maximum" required minimum experience is 35 years, accompanied by the fact 11 others reported extremely high minima (7 years or more), suggests that some respondents understood the question not as a minimum but rather as the experience of the individual involved in the crash or something else, like the average experience at the firm.

[46] The fact that the results for "YearsExpDriv" and "YearsExpVeh" are virtually identical suggests that the reported responses are suspect.

[47] This is so faulty it really will be hard to analyze at all. "This variable establishes the basis used by the carrier to compensate its drivers for time spent driving the vehicle." This is a general carrier-level statement and if some trips are compensated on an hourly basis and some on a mileage basis, the carrier reasonably would say "both" (extend that to percentage as well). No drivers can be paid *for driving time* both by the hours and by the mile, but carriers report that 43 drivers (3.5% of those for which responses were collected) were paid both miles and hours for driving, which is meaningless; 19 drivers (01.5%) supposedly are paid "all of the above" – hourly, mileage, and percentage – which is entirely absurd. Presumably these responses must be excluded. In addition, 173 (13.9%) are coded "other" and since this is unspecified, those cases are not usable either. "Other" might mean "pay by the load", a common confusing response that reflects some arbitrary payment schedule unrelated to miles or freight revenue, or it could be "pay by

- DrivingTimeCompD[48]
- DrivingTimeCompHourly[49]
- DrivingTimeCompMileage[50]
- DrivingTime%[51]
- LoadUnloadComp[52]
- LoadUnloadCompDummy[53]
- WaitingTimeComp[54]
- WaitingTimeCompDummy[55]
- AdminFtnsComp
- AdminFtnsCompDummy[56]

---

cwt [hundredweight]", which is less common. Despite the flaws in this variable, it is incomprehensible that the public data file claims that zero drivers' pay is "variable" by any conventional meaning of the word, but the term "variable" is undefined in the documentation.

[48] Derived variable created by Sound Science for which responses are coded 'DrivingTimeComp' = 1 then "Hourly"; 'DrivingTimeComp' = 2 then "Mileage"; 'DrivingTimeComp' = 4 then "Percentage"; 'DrivingTimeComp' = 3 or 5 then "Miscoded"; 'DrivingTimeComp' = 8 then "Other", and other codings "Missing".

[49] Indicator variable created by Sound Science for which responses are coded Hourly = 1 and both Percentage and Mileage coded 0, with all other responses coded as missing.

[50] Indicator variable created by Sound Science for which responses are coded Mileage = 1 and both Percentage and Hourly coded 0, with all other responses coded as missing.

[51] Indicator variable created by Sound Science for which responses are coded Percentage= 1 and both Mileage and Hourly coded 0, with all other responses coded as missing.

[52] Responses are by the hour, flat rate, and "other", which is not specified. Unfortunately, though this response is 8% of the total, the lack of specificity requires that it be treated also as missing (1/3 of all responses). This is a general carrier-level statement and if some drivers are compensated to load and some are not, under varying circumstances, responses will be non-exclusive and will vary and be not applicable to the trip associated with the crash.

[53] Created by setting compensation by hour and by flat rate to "1", not compensated to "0", and the remainder to "missing".

[54] Responses are by the hour, flat rate, and "other", which is not specified. Unfortunately, though this response is 7% of the total, the lack of specificity requires that it be treated also as missing (nearly 1/3 of all responses).

[55] Created by setting compensation by hour and by flat rate to "1", not compensated to "0", and the remainder to "missing".

- CAIOntimePerformance
- CAISafetyBonus
- OtherSpecialPayment
- LengthEmployed[57]
- QualUptoDate
- MedicalCert
- CertCurrent
- SecondJob[58]
- HrsWkd7Days (continuous variable)
- TripType[59]
- IntendedDuration (continuous variable)[60]
- LogExempt
- PaymentMethod[61]

---

[56] Responses are by the hour, flat rate, and "other", which is not specified. Unfortunately, though this response is 4.7% of the total, the lack of specificity requires that it be treated also as missing (more than 1/4 of all responses).
[57] Thirteen cases indicate that driver has been employed zero months. This is difficult to interpret since they must have been employed for some time to have had the truck crash; documentation provides no clue on interpretation. I have set these observations to "missing."
[58] Twenty carriers report awareness that their driver involved in the crash had or has a second job but only ten of them report the number of hours the driver worked. In contrast, 48 drivers report that they have second jobs ("Moonlight") but only 16 of them report their hours to the primary employer and they average 19.1 hours/week on these second jobs (the same as the ten reporting carriers indicate).
[59] The responses for this variable may not have been clear to the carrier since their definitions, as indicated in the documentation, are somewhat self-contradictory. The definitions are: 1 Scheduled trip (Advance notification); 2 Unscheduled trip (<8 hours advance notification); 3 Unscheduled trip (>8 hours advance notification). A scheduled trip therefore is defined as one with advance notification and an unscheduled trip also is defined as one with advance notification, with categories of notification greater and less than 8 hours (notification of exactly 8 hours is left undefined). The only logical option is to create a dummy variable for "Advance Notification", where 1, 2, and 3 are equal to 1 and no advance notification would be equal to zero, but the documentation does not specify this option. The variable, therefore, is entirely useless.
[60] Fourteen carriers report an intended duration of zero days and hours. Since this response is logically inconsistent with having made a trip and having had a crash, I set these responses to missing.
[61] See footnote 38. This response applies to this trip only. This is so faulty it really will be hard to analyze at all. "This variable establishes the basis used by the carrier to

- PaymentMethodD[62]
- Contingency(Carrier)[63]
- DriverLoad[64]
- DriverLoadD and DriverLoadUnloadThisTrip
- DriverLoadThisTrip[65]
- DriverUnloadThisTrip[66]
- DriverLoad&UnloadThisTrip[67]
- CompensatedLoad[68]

---

compensate the driver for driving time for this trip." No drivers can be paid *for driving time* both by the hours and by the mile, but carriers report that 29 drivers (2.3% of those for which responses were collected) were paid both miles and hours for driving, which is meaningless; 5 drivers (0.4%) supposedly are paid "all of the above" – hourly, mileage, and percentage – which is entirely absurd. Presumably, analysis must exclude these responses. In addition, 151 (12.2%) are coded "other" and since this is unspecified, those cases are not usable either. "Other" might mean "pay by the load", a common confusing response that reflects some arbitrary payment schedule unrelated to miles or freight revenue, or it could be "pay by cwt [hundredweight]", which is less common. Despite the flaws in this variable, it is incomprehensible that the public data file claims that zero drivers' pay is "variable" by any conventional meaning of the word, but the term "variable" is undefined in the documentation.

[62] This is an indicator variable I created that collapses plausible categories of hourly, mileage, and percentage into categorical indicators and then sets everything else to missing. This response applies to this trip only.

[63] Ordinal variable created by Sound Science to proxy contingency of compensation. Based on 'DrivingTimeComp'. 1 = hourly; 2 = miles and hours (a concept that reflects miscoding, as discussed above); 3 = Mileage; 4 = Percentage.

[64] The response "9" is undefined in the documentation. This response applies to this trip only. 168 responses are coded as 9, which is undefined in the manual. I treated them as "unknown." This is a general carrier-level statement and if some drivers are compensated to load and some are not, under varying circumstances, responses will be non-exclusive and will vary and be not applicable to the trip associated with the crash. Indeed, I have to assume that responses indicating that the driver loaded or unloaded are exclusive of each individual process, and responses indicating that driver loaded and unloaded include both processes.

[65] Indicator variable if driver loaded.

[66] Indicator variable if driver unloaded.

[67] Indicator variable if driver both loaded and unloaded.

- WaitLoad[69]
- CompensatedWait[70]
- CAIOnSchedule
- DelayReason
- CAIHoursOnDuty (continuous variable)[71]
- CAIHrsOnDuty (continuous derived variable)[72]
- HrsPriorDay (continuous variable)
- LongestDay7Days (continuous variable)
- ShortestDay (continuous variable)[73]
- ShortestDay7Days (continuous variable)
- CAIAverageDay (continuous variable)[74]

---

[68] This response applies to this trip only. 329 carriers reported in the previous question that drivers performed loading, unloading, or both. 317 carriers responded to the question regarding whether they were paid, with approximately 80% of those responding indicating that they pay the drivers. These responses are suspect since all available previous evidence suggests the proportion paid explicitly for this activity may be the reverse of that reported here. Also, it is not clear from the documentation that the question asked if the drivers were paid explicitly for this time. Most motor carriers interpret unpaid load/unload time as part of the job and therefore incorporated within the rate.

[69] This response applies to this trip only.

[70] This response applies to this trip only.

[71] Nine cases are coded as "zero" hours on duty at the time of the crash. An additional 41 cases document that drivers were on duty for less than one hour at the time of the crash, for a total of 50 drivers on duty for less than one hour. Weekly hours for at least seven of the drivers are coded as "unknown" (does "weekly hours" not include the day of the crash?) and more than 100 apparently are coded with some known number of hours the day of the crash and "unknown" work time for the week. In addition, 104 drivers are coded as having been on duty for exactly one hour at the time of the crash. It is impossible that the driver had been on duty for zero hours at the time of the crash unless these cases pertain to team drivers in the sleeper berth, but since that person isn't the "driver" and there must be a driver for each case, these cases must be coded as missing. I can not determine boundaries for data entry error on this variable.

[72] See footnote 53. This variable has been converted to hours. Also, I have set zero hours to missing.

[73] Manual says 8887 = Not applicable but this code is absent. The code in the data is 9797, which I assume means the same thing.

[74] While I have concerns about reported zero hours of work in past seven days, in some of the average hours responses, the average number of hours reported worked is zero, while the total reported hours in the past 7 days is a positive number. With many other

- AverageDay (continuous variable)
- CAITotalHours (continuous variable)[75]
- TotalHrs7Days (continuous variable)

IntvwDrDriver Data Set (n=1908)
CaseID
VehicleNumber
- TruckYears (continuous variable)
- ClassYears (continuous variable)
- TimeSinceTraining (continuous variable in "years/months" form; converted)
- NonCDLStatus
- CDLClass
- CDLStatus
- CMVCitations (continuous variable)
- MovingCitations (continuous variable)
- CommCrashes (continuous variable)
- NonCommCrashes (continuous variable)
- TripPayment[76]

---

variables showing implausible zeroes, one suspects that some researchers entered "0" instead of missing, in one form or another, but these data entry errors could have happened anywhere in the LTCCS.

[75] These are immensely flawed data. Five cases report zero hours during the past seven days and they all contradict the variable CAIHoursOnDuty, the variable documenting the number of hours the driver was on duty at the time of the crash. One case documents that the driver was on duty one hour in the last seven days but had zero hours on duty at the time of the crash.

[76] Twenty-two drivers report being paid by the mile and by the hour *for driving time*. This is impossible, as indicated in footnote above with respect to carrier reporting. I am not sure how to deal with this but there appear to be no such responses. In addition, pay "by the load" tells us only that the driver's pay is contingent on factors other than labor time, but it doesn't tell us more than that. We know that "load" pay is based on some hidden factors, such as mileage and revenues, but none are specified. "Other" also is not specified so it provides no information (impossible to determine whether the information provided is incorrect, as for the "miles and hours" information above). One alternative would be to code "miles and hours" as miles on the assumption that the data collector and the "researchers" did not understand fundamental driver pay concepts and the mention of "miles" indicates that the drives were paid mileage for driving time, but this requires an assumption. The fact that more than 50% of all responses are missing creates major issues of bias; it is impossible to determine which way the bias goes.

- DrivePayThisTrip[77]
- PayContingency(driver)[78]
- Moonlight[79]
- ReportMoonlight[80]
- HoursPrecedingSeven (continuous variable)
- HoursTypicalSeven (continuous variable)
- IDROnTimePerformance
- IDRSafetyBonus
- OtherPayment

IntvwDrFatigue Data Set (n=1903)

CaseID

VehicleNumber

- OverHours[81]
- OverHrsD[82]
- OverDispatch
- ThisTrip
- OverDispatchThisTrip[83]
- ThreatenedOverHours
- DisciplinedLate

---

[77] Variable created by Sound Science that implements the above changes and correctly names the concept reflected by the question. 'TripPayment': Hourly = 1; Mileage = 2; Percentage and "by the load" = 3. This is the best that can be done with flawed data.

[78] Ordinal variable: 1=Hourly; 2=Miles; 3=Percentage and By the load (the latter is less precise than percentage but probably is based on freight rate and miles hauled, as is percentage).

[79] Note that 71 drivers report that they have second jobs, 3.55 times as many drivers as carriers that report that they know that drivers have second jobs; 27 of the 71 say they report to carriers.

[80] This variable is miscoded because it indicates that 96.4% of the responses are unknown, though most of them are inapplicable because the driver does not have a second job.

[81] I am not sure whether driver can declare themselves "exempt" from HOS, or whether the researchers concluded they were exempt. 20.5% of all drivers involved in crashes and reporting an answer to this question in the LTCCS are coded "exempt." Since "exempt" is not really an answer to the question, I coded this "missing" for purposes of analysis. However, since the response provides an insignificant result and removes 185 cases, I have not included it in analysis.

[82] Indicator variable that considers "exempt" to be missing.

[83] Sound Science derived variable.

- OnCallThisTrip[84]
- ADVNotice (continuous variable)
- SleepInterrupted[85]
- Loading
- LoadType
- EffortType
- Effort-Manual[86]
- Compensated
- WaitedLoad
- WaitCompensated
- WaitDelay (continuous variable; converted from HHMM to minutes)

IntvwDrSleep Relation (n=1840)
CaseID
VehicleNumber

- LastLocation
- LastLength (converted to hours)[87]
- LastEnd (converted to hours)
- LastSinceD_m (converted to hours)[88]

---

[84] The data documents zero drivers as working on "on-call status" generally (though all are coded as "unknown"). But this variable, which should be a subset of the former one, shows that 137 were on-call for this trip and 1449 were not, with the remaining unknown or not applicable.

[85] **Did This Call Interrupt A Sleep/Rest Period?**
Definition: This variable establishes whether or not the call-in to work interrupted the driver's sleep or rest period.
Variable Name: SleepInterrupted
Attribute Codes
Code Meaning
1 Yes (specify)
2 No
7 No driver
8 Not applicable
9 Unknown

[86] Indicator variable separating out manual from other loading.

[87] I set zero values to missing, since the length of last sleep that is zero sleep is undefined. The self-report sleep durations are pretty suspect. One such value shows the driver sleeping 16 hours, and if it is accurate at all it probably represents the time logged off.

[88] I don't know how to interpret hours since last sleep when reported hours is zero; this seems undefined. An additional 40 cases report hours since last sleep as less than one. This at least is defined but note that the lowest reported hours since last sleep is 5 minutes.

- Less4Location
- Less4Length (converted to hours)[89]
- Less4Start (converted to hours)[90]
- Less4End (converted to hours)[91]
- Last24Hours (converted to hours)[92]
- Longest (converted to hours)[93]
- Shortest (converted to hours)[94]
- Average (converted to hours)[95]

---

[89] Interesting to note that for 30 of these drivers whose last hours of sleep were less than 4 hours, the hours of last main sleep were less than 8 hours, and 9 report 4 hours or less. Also, the length of last sleep for one person was zero, which is really contradictory; I set that to missing. (note to self: create a variable adding up LastLength and Less4Length.)

[90] The same driver whose reported last hours of sleep are "0000" also started his last sleep at "0000". This means that this is a miscoding so I set that case to missing on this variable as well (though two other legitimate entries of "0000" exist in this variable).

[91] The same case that starts last sleep at 0000 ends at 0000, so I marked it as missing. (note to self: create a variable differencing Less4Start and Less4End.)

[92] While I think having zero hours of sleep in the past 24 is extreme but plausible (6 drivers report zero hours and one reports 6 minutes), I think scientific research would show that it is virtually impossible for someone to exceed 12 hours of actual sleep even if catching up on a great sleep deficit (this literature can be checked). This variable indicates 48 drivers got 12-24 hours of sleep, with one reporting 24. It is hard to determine where real data leaves off and reporting error begins. It is important to know on this and all other questions regarding sleep time whether the LTCCS researchers used log books or actual statements from drivers to determine the answers to these questions, and whether they verified the data collected against other sources. Some of the responses suggest they did not.

[93] Implausible that longest length of sleep in 7 days was zero hours. Set to missing. Note also that 45 drivers report 12-24 hours of sleep, which is really high and implausible. Close investigation shows that the case for which the response is "0000" also shows this same response in multiple variables, supporting hypothesis that these are coding errors. We do not know where coding errors leave off and substantive data begin.

[94] Same person who shows longest sleep is 0000 also shows shortest sleep at 0000. This is a data entry error. Also, it is pretty hard to believe that 31 drivers can report having slept 10 hours or more during their SHORTEST sleep in the last 7 days.

[95] The same driver that reports 0000 as shortest and as longest also reports 0000 as average. Set to missing. Shockingly, 85 drivers report 10 hours or more as average sleep hours during the week; this is really implausible. Also, I have real concerns about how LTCCS researchers determined sleep intervals across the board, and the extremity of the mode (650) at 8 hours being more than three times the nearest other reported number of

- Intervals[96]
- ShiftRotate
- RelateToWork[97]
- IDSShortestDay (converted to hours)[98]
- IDSLongestDay (converted to hours)[99]

---

hours (7) and being more than 6 times the next nearest reported number of hours (9), and the fact almost all of the responses are reported in whole hours (not fractions) strongly suggests that either they used logged off-duty time (conveniently, 8 hours at the time by regulation) as a proxy for actual sleep (thus the variables should be called "LoggedAverage" and so forth) or that drivers answered logged hours or researchers took whatever drivers logged as "sleep", rather than true hours of sleep. These data therefore should be considered highly suspect.

[96] The choices here are day, night, and a mix of day and night. Three responses are "other", which I cannot define, since the foregoing selections include the whole set of possible responses. I set those 3 cases to missing.

[97] Sound Science indicator variable derived from "RelateTo", where 1 = relate to work; else 0.

[98] A reported shortest work day at 24 hours in the last 7 is impossible, though other "shortest days", such as 20, are implausible as well. I set the 2400 hour workday at missing. 82 drivers report the shortest day as zero, but the question arises whether one can have any work day of zero hours; a zero hour work day is not a work day. Perhaps this means that these drivers did not work during the week but this is not clear from the documentation and would confound any analysis of working time. I set those to missing also for that reason and I can run the data both incorporating the notion that 82 drivers simply did not work during the preceding week and alternatively assuming that this just means they were off the previous week, which could contribute to crashes. This can be changed if documentation is found that explains how researchers got these figures. In addition, the fact that many of the cases in which the shortest day is "0000" also show that the longest day is "0000", further causing one to wonder whether this is data entry error rather than true information on the driver. Finally, the concept of "day of work" is not defined in the documentation. Does this mean period defined as a shift of work, or single work trip bracketed between two minimum eight hour breaks? Does it mean total working time logged in a calendar day? Without better definition it is extremely difficult or impossible to interpret responses to this question or related questions.

[99] 22 drivers report a longest work day of zero hours. A zero hour work day cannot be a work day so I set these also to missing and I can run the data both incorporating the notion that 22 drivers simply did not work during the preceding week and alternatively assuming that this just means they were off the previous week, which could contribute to crashes. Perhaps this means that these drivers did not work during the week but this is

- AverageWorkDay (converted to hours)[100]
- TotalHours (converted to hours)[101]
- SinceLastDayonDuty[102]
- OnWaking
- Home (converted to hours)[103]
- Road (converted to hours)[104]
- AwakeFeeling

---

not clear from the documentation and would confound any analysis of working time. This can be changed if documentation is found that explains how researchers got these figures.

[100] Note that coding for "not applicable" does not match coding in variable, though it does match the coding for preceding variable (probably a coder inconsistency). Also, two cases claim average working days of 24. Since an average working day of 24 would mean working the full 168 hours, which only can happen in ironic humor, I am setting these to missing. Also, consistent with the preceding, I am reporting both an average work day of zero hours (24 cases) set to missing as well as zero hours, as reported, because this means they have to have been off all week and thus not only do they not have working days but they do not have a working week at all. These all may be data collector errors or may be coding errors, but in any case they were never "proofed" and cannot be used without further explanation from researchers who supervised the collection of the data.

[101] This is hours worked during last seven days preceding crash. Hours for two cases are reported at 240 and 290, respectively, and there are only 168 possible hours in a week. I set these to missing. In addition, 24 cases report zero hours during the past seven days. I do not know how to determine, without doing a great deal of work, whether these are coding errors or true information. I report descriptive data both ways but will run regressions with the zero values unless clarification is obtained.

[102] Documentation says this is in "HHMM" form, but refers to measure in terms of days. The actual measure in the original variable clearly is in days, in simple form. "Not applicable" and "Unknown" both appear as different values from what appears in the documentation. Note also that the complete variable name is "Number Of Days On Duty Since Last Day Off" [sic], while the variable name in the set conveys the opposite meaning.

[103] I set the response "0000" to missing for the single driver reporting zero hours typically at home. The response for "Road" variable is also zero, which leave no sleep ever typically. This suggests the response should be "not applicable". Presumably the 11 drivers who report typical sleep on the road as zero do not typically make over-the-road runs.

[104] I set the response "0000" to missing for the case that also reports that "Home" variable is also zero, which leave no sleep ever typically. This suggests the response should be "not applicable".

- StartDriveFeeling
- Since8Driving (converted to hours)[105]
- Since8onDuty (converted to hours)[106]
- Last24Worked (converted to hours)[107]

IntvwDrTrip Data Set (n=1845)

CaseID

VehicleNumber

- StartDate (continuous variable in YYYY-MM-DD format; converted)
- StartTime (continuous variable in HH:MM format; converted)
- WorkSchedule
- IDTOnSchedule

[105] It is not clear to me that someone could report zero hours on duty at the time of the crash because zero is a non-positive number and the documentation says they are working at the time of the crash, so responses of zero should be set to missing. I have no a priori reason to exclude responses close to zero but they seem pretty implausible; 15 drivers report working ten minutes or less (not including the 8 drivers who claim to be working yet on duty for zero time) at the time of the crash. The language also is ambiguous relative to whether this is applicable to those whose job is entirely driving, though I assume it is. "This variable applies only to drivers who drive as part of their job and are working at the time of the crash. Particularly, this is the number of hours that the driver has beendriving since he/she has had a break of at least 8 hours."

[106] It is not clear to me that someone could report zero hours on duty at the time of the crash because zero is a non-positive number and the documentation says they are working at the time of the crash, so responses of zero should be set to missing. I have no a priori reason to exclude responses close to zero but they seem pretty implausible; 7 drivers report working ten minutes or less (not including the 8 drivers who claim to be working yet on duty for zero time) at the time of the crash. The fact that both Since8Driving and Since8onDuty show 8 drivers reporting zero, and 7 of the 8 in each case report zero for both, makes one think that at least for these 7, this is a coding error.

[107] It is unclear from the documentation how 97 drivers can have not worked at any time during the 24 hours preceding the time of the crash, since they had to be working to have had a crash. If the documentation said "24 hours preceding the time the driver came on duty on the day of the crash" or something like that, it would make sense, but I can't see how zero is a plausible response. I set these responses to missing until I can obtain clarification. It also is not clear how drivers can have been on duty for more than 24 hours during the preceding 24 hours. While it is possible for a driver to have been on duty for the preceding 32, 33, 40, or 60 hours before the crash (5 cases), I will set these also to missing pursuant to clarification. The variable only permits a 24-hour maximum response.

- OneWayDistance (continuous variable)[108]
- IntendedDayDistance (continuous variable)

Overview (n=2284)

CaseID

VehicleNumber

- OVEConfiguration
- StraightTruck[109]
- ArticulatedTruck[110]
- Truck
- Movement
- OVECriticalEvent
- InvolvedInCriticalEvent[111]
- OVEReason (Critical Reason For The Critical Precrash Event; see also Crash Assessment Relation)
- AssignedCriticalReason[112]
- OVERightOfWay
- DriverFatigue[113]

---

[108] While neither OneWayDistance nore IntendedDayDistance are necessary for this analysis, it is important to note that in some cases (especially at the lower mileage levels) the intended day distance is much greater than the overall OneWayDistance. It is possible that the question is ill-conceived for this purpose because anyone running a serial run (peddle run of some kind, either short or long distance) may not be able to answer the "OneWayDistance" question meaningfully.

[109] Sound Science created derived variable in which all vehicles identified as "straight trucks" are identified as "1" and all those not identified as straight trucks are "0", with other responses coded as missing.

[110] Sound Science created derived variable in which all vehicles identified as "articulated trucks" are identified as "1" and all those not identified as articulated trucks are "0", with other responses coded as missing.

[111] Categorical variable created by Sound Science to which to assign to those vehicles coded as involved in critical event. If 'OVECriticalEvent' = 125 then 0; else "1"

[112] Categorical variable created by Sound Science to which to assign to those vehicles coded as assigned critical reason for critical event. If' OVEReason'=0 then 0; else if 'OVEReason'="999" then "*"; else "1".

[113] THIS IS A MISCODED AND FAULTY VARIABLE AND SHOULD NOT BE USED. For some unclear reason, this version of "Fatigue", drawn from the "Fatigue" variable in the Driver Assessment relation, has different results. This variable indicates 198 fatigued and 1880 not fatigued, with 206 missing cases. The prior Fatigue variable shows 198 fatigued drivers and 1265 not fatigued, with 15 drivers coded as "not present"

- CarrierEmployer

SaferAuthorityStatus (n=605)
CaseID
VehicleNumber
- CommonStatus
- ContractStatus
- BrokerStatus[114]

SaferCarrier Data Set (n=796)[115]
CaseID
VehicleNumber
- SCAStatus[116]
- OperationClassification_m[117]
- CarrierOperationDescription
- CarrierOperationIndicator[118]
- CargoCarried[119]

---

and 600 coded as "unknown." If researchers use this variable they will err by counting these extra missing responses as "not fatigued" and make a big systematic error.

[114] Note that documentation provides apparent multiplication of response fields, so analysis would be incorrect without combining them by assumption of meaning:

Group	Count
ACTI	11
ACTIVE	113
INAC	3
INACTIVE	30
NONE	384
UNK	19

[115] Manual says the following variables are in this data set but I cannot find them there: AuthorizedForHire, ExemptForHire, PrivateProperty, PrivatePassengersBusiness, PrivatePassengersNonBusiness, Migrant, USMAIL, FederalGovernment, StateGovernment, LocalGovernment, IndianTribe, Other, Unknown. They may all be included in "OperationClassification" but these classifications are so tangled that they cannot be used without major programming effort.

[116] "99" undefined for one case (not documented in Manual).

[117] The main categories should be For Hire, Not For Hire ("private carrier"), and Government, yet this variable reports out 23 unconsolidated categories, the differences among which are not clarified by the documentation.

[118] I created this variable to combine Intrastate HazMat and Intrastate Non-HazMat into "Intrastate".

- CountDrivers (continuous variable)[120]
- CountPowerUnits (continuous variable)
- MileageYear (continuous variable)[121]
- Mileage (continuous variable)[122]

SaferCrashSummary Data Set (n=2784)

CaseID
VehicleNumber
SCSYear (additional linking variable for merging relations)[123]

- SCSFatalities
- SCSInjuries
- SCSTowaway

SaferDriverInspection Data Set (n=795)

CaseID
VehicleNumber

- SDIDriverDOB (in simple year form)[124]

---

[119] Cargo carried might be quite important, but the information on cargo has been combined into a single variable with 56 categories, all of which duplicate others because of substantial data entry and coding errors. For example, 18 trucks carry "beverages" and 2 trucks carry "beverages, beverages". These unexplainable and incomprehensible coding errors appear for every single commodity entered except oil field equipment, utility, and unknown. For these reasons, this variable cannot be used.

[120] Documentation states that "999999999999" is "unknown" but this number does not appear in the variable. However, the number "1E+12" appears 25 times and I believe it, rather than the foregoing number, represents unknown values. Two carriers report zero drivers, which conflicts with reporting of values in variables discussed above, but I will leave those zeroes in the variable.

[121] Documentation says that MileageYear (the year for which the carrier's mileage is reported) is "999999999999" when it is "Unknown", but the values that appear to be missing are indicated as "1E+12", as above.

[122] Documentation says that Mileage (the number of miles operated by the carrier) is "999999999999" when it is "Unknown", but the values that appear to be missing are indicated as "1E+12", as above.

[123] For all data in this relation, I calculated the mean rate of fatalities, injuries, and towaways per year for each carrier, thus compressing the information from multiple years into a single variable representing the average annual rate over the years covered in the data set.

- NumOfInspections_m
- SDITotalViolations_m
- TotalOOS_m
- NumOfInspections_m
- SDITotalViolations_m
- TotalOOS_m

SaferInspectionsSummary Data Set (n=796)
CaseID
VehicleNumber
- DriverInspections (continuous variable)
- DriverOutOfService (continuous variable)
- DriverOutOfServicePerc (continuous variable between 0 and 1)[125]
- DriverViolationsAverage (continuous variable)
- VehicleInspections (continuous variable)
- VehicleOutOfService (continuous variable)
- VehicleOutOfServicePerc (continuous variable between 0 and 1)[126]
- VehicleViolationsAverage (continuous variable)
- SISTotal (continuous variable)[127]

SafeStat (n=1882)[128]
CaseID
VehicleNumber
- Score (continuous variable)[129]
- AccidentSEA (continuous variable)[130]

---

[124] The driver date of birth file is too inconsistent to use. For unexplained reasons, the date of birth does not seem to be associated consistently with the driver. For this reason, an important predictor, the age of the driver, must be excluded from analysis.
[125] "Percentage" numbers, supposed to be between 0 and 1, are not reported in percentages or fractions between 0 and 1 in data set.
[126] "Percentage" numbers, supposed to be between 0 and 1, are not reported in percentages or fractions between 0 and 1 in data set.
[127] Total number of inspections for each carrier in 24 months preceding this collection.
[128] Manual and code book provide no guidance with respect to the meaning of these indices, or how any missing cases might be coded. Given that only 182 (9.7% of these cases) have an SEA Score, it is hard to believe that the Accident, Driver, Vehicle, and Safety SEAs have all valid cases. I presume the zero values for these cases are actually missing data, but I don't know.
[129] Is it possible to have a SafeStat Score of zero? Two cases are zero.
[130] Is it possible to have a AccidentSEA Score of zero? 561 cases are zero.

- DriverSEA (continuous variable)[131]
- VehicleSEA (continuous variable)
- SafetySEA (continuous variable)[132]
- STACategory

TruckUnits Data Set (n=2179)

CaseID

VehicleNumber

- NumberOfUnits_m
- TUNBodyType
- CargoType

[131] Is it possible to have a DriverSEA Score of zero? 284 cases are zero.
[132] Is it possible to have a SafetySEA Score of zero? 1882 cases are zero