Effects of Truck Driver Wages and Working Conditions on Highway Safety: Case Study

Article *in* Transportation Research Record Journal of the Transportation Research Board · January 2003

DOI: 10.3141/1833-13

4 authors:

Daniel Rodriguez
University of California, Berkeley

Marta Sousa Rocha
University of Lisbon

Asad J. Khattak
University of Tennessee

Michael H. Belzer
Wayne State University

Paper number 03-2125

The effects of truck driver wages and working conditions on highway safety: A case study

Daniel A. Rodríguez[†], Marta Rocha, Asad J. Khattak, Michael H. Belzer

Daniel A. Rodríguez
Assistant Professor
Department of City and Regional Planning
New East Hall, CB #3140
University of North Carolina, Chapel Hill
Chapel Hill, NC 27599-3140
Telephone: 919-962-4763
Fax: 919-962-5206
E-mail: danrod@unc.edu

Marta Rocha
Department of City and Regional Planning
New East Hall, Room 318
University of North Carolina, Chapel Hill
Chapel Hill, NC 27599-3140
E-mail: mrocha@unc.edu

Asad J. Khattak
Department of City and Regional Planning
New East Hall, CB #3140
University of North Carolina, Chapel Hill
Chapel Hill, NC 27599-3140
E-mail: khattak@email.unc.edu

Michael H. Belzer
College of Urban Labor and Metropolitan Affairs
Office 3198 FAB
Wayne State University
656 W. Kirby St.
Detroit, MI 48202
E-mail: Michael.H.Belzer@wayne.edu

[†] Corresponding author

Abstract

The role that human capital and occupational factors play in influencing driver safety outcomes has gained increased attention from trucking firms and policy-makers. This paper examines the role of these factors, in addition to demographic factors, in influencing crash frequency at the driver level. A unique driver-level dataset from a large truckload firm collected over a period of 26 months is used for estimating regression models of crash counts. Based on estimates from a zero inflated Poisson regression model, results suggest that human capital and occupational factors, such as pay, tenure at the job, and percent of miles driven during winter months, have a significantly better explanatory power of crash frequency than demographic factors. Taking into account both the zero-inflation and the count model, results suggest that higher pay rates and getting a pay raise are related to lower expected crash counts and to a higher probability of having no crashes at all, all else held equal. Although the data for the study come from a single firm, the evidence provided is a first step in examining the structural causes of unsafe driving behavior, such as driver economic rewards, and crash outcomes. These results can motivate other firms in modifying operations and driver hiring practices. They also support the need for a broader examination of the relationship between driver compensation and driver safety.

Keywords: truck driver safety, compensation, economic rewards, count model

INTRODUCTION

Although the involvement of large trucks in fatal crashes in the U.S. has dropped substantially over the last decade when measured per unit of travel, the public health burden of large truck crashes, as measured by deaths per 100,000 population, has not improved over time because of the large increase in truck mileage (1). Crashes involving trucks impose costs on truck drivers, road users, trucking firms, shippers, and the public. In 2000 about 5,211 people died and 140,000 were seriously injured in large truck-related crashes (2). With trucking operations accounting for almost one third of the total freight ton-miles traveled, and expected to grow in the future, trucking safety continues to demand heightened attention from researchers and policy-makers.

The role that truck driver occupational and behavioral factors play in crash involvement is a particular area that has received increased attention by researchers and policy-makers. Research examining potential modifications to the hours-of-service regulations and their enforcement (3, 4), detecting and measuring driver alertness and fatigue (5, 6), and understanding driver speeding behavior (7-9), underscore the increasing awareness regarding the importance of driver behavioral factors for trucking safety.

Even though an acute research focus on particular truck driver behaviors that increase crash risk is useful, it is at least equally important to confront the factors that motivate such behaviors. At the individual level, such factors include scheduling and operational pressures, pay rate, pay method, and personal characteristics, among others. Indeed, over the past decade several studies (3, 10-12) have raised questions about the role that occupational and human capital factors play in truck crashes. Therefore, improving our understanding of the importance of the structural causes of certain behaviors, such as compensation and driver economic rewards, and their relationship to driver safety outcomes can lead to appropriate policy responses at the firm and government levels. Insurance companies, trucking firms, shippers, and regulators have an interest in developing such understanding in order to improve the safety outcomes of trucking operations.

In this paper, we examine how compensation and work conditions are associated with frequency of truck crashes at the driver level by estimating count regression models while controlling for driver socio-demographic characteristics and individual exposure. As such, the research focuses on better understanding the occupational factors that can lead to reducing the risk of truck involved collisions as well as in quantifying the relationships between crash frequency and available explanatory variables at the driver level of analysis. To this end, we use a proprietary, driver-level dataset from JB Hunt, one of the largest truckload firms in the U.S. Even though the results are not expected to be fully representative of the population of for-hire truckload drivers, we find that JB Hunt drivers are comparable to other drivers in terms of demographic and occupational characteristics. Given this, the relationships among the variables may be representative of similar relationships in other firms. Results can be of immediate use to individual trucking firms and can stimulate policy-maker's interest in extending this single-firm study to a more general case. The next section presents a review of the literature focusing on occupational factors and driver safety, followed by a detailed explanation of the data, the crash modeling performed, and the results and implications.

LITERATURE REVIEW

Human capital theory suggests that variations in human capital across individuals and firms explain differences in labor force outcomes, such as productivity and safety (for the theoretical foundations see 13, 14). Greater job experience, for example, is expected to be related to greater safety. Similarly, because pay can be considered partly a proxy for different levels of human capital, it is expected to be related to better employee outcomes. In a competitive market, higher pay would allow firms to attract and retain drivers with certain characteristics, which will lead to better safety records.

The association between driver behavior, driver pay, and driver characteristics (education, skills, and other experience) suggested by human capital theory has been tested empirically in several studies of the trucking industry. Krass (15) detects a significant inverse relationship between wages and crash risk for the period after economic deregulation of the trucking industry. The work of Hirsch (16) suggests that a substantial fraction of driver wage differences may account for human capital differences among drivers. A recent study concludes that trucking industry compensation and human capital characteristics appear to be more significant determinants of safety than demographic variables (17). Even though other studies have also supported a connection between driver safety and human capital characteristics and driver compensation (10-12, 18), only a handful have examined this relationship explicitly and, to our knowledge, none has focused on crash frequency.

Other research has also examined the link between compensation and occupational factors such as working conditions and driver fatigue. Fatigue is arguably one of the most important risk factor that emerges from analyzing the role of occupational factors in driver safety (4, 6, 8, 19, 20). Most studies of fatigue have examined the causes and the extent of fatigue in truck drivers (for current reviews see 21-23) and more recently the link between fatigue and crash risk (10, 24). McCartt et al. (23, 25) find that drivers perceive the scheduling of loads (measured as driving hours and waiting time for loads) as a significant factor that contributes to fatigued driving. Similarly, after conducting focus groups to examine the factors related to truck crashes, Chatterjee et al. (26) conclude that direct pressure from dispatchers forces drivers to work long hours under unsafe conditions. Lin et al. (27) rely on operational data from another large national less-than-truckload (LTL) carrier to find that total driving time has a greater effect on crash risk than either time of day or driving experience. Not surprisingly, occupational factors also have been associated with illegal substance use (11, 24, 28) and a higher propensity to speed (7, 29).

Related to fatigue is driver (and firm) non-compliance with established hours-of-service regulations, which limit the amount of driving time of truck drivers. Using self-reported data of 498 long-distance drivers, Beilock (3) estimates that 26% of schedules given to drivers result in violations to existing service hours regulations assuming that the average speed limits do not exceed legal limits. In a 1992 survey, Braver et al. (12) found that drivers who violated the hours-of-service rules were more likely to report that they have fallen asleep at the wheel. Clearly, non-compliance with hours of service regulations is also related to negative safety outcomes.

In summary, the literature suggests that certain occupational and human capital factors are related to unsafe driving behaviors and crash outcomes. Tight schedules, fatigue, increasing demands on drivers, and low pay are positively correlated to crash occurrence, although it is unclear if these correlations are carried through to crash frequency. Given the apparent link between level of driver pay and driver safety, one expects that firms would raise pay in order to skim the cream of the trucking labor market. However, this does not seem to be occurring in most cases. Furthermore, on average, earnings of truck drivers and the quality of driving jobs continue to erode, especially among non-union drivers (30). Although speculative, one explanation may be that motor carriers do not perceive that the safety benefits of higher pay offset the increased costs to firms. This points to our limited knowledge about the relationship between pay and driver behavior, and underscores the identified need to develop relevant research that can inform policy-makers and firms. As such, we address the paucity of empirical work regarding the study of crash frequency for truck drivers, while contributing to the understanding of the complex relationship between driver pay and driver safety.

DATA

A unique longitudinal dataset was used to analyze the association between truck drivers' socio-economic and occupational factors and crash involvement and frequency. The dataset is rich because it contains human resources, operations, and safety data for 11,540 unscheduled over-the-road dry-van tractor-trailer drivers of JB Hunt, a major U.S. for-hire truckload company, over a period of 26 calendar-months. The drivers were observed for two periods of 13 months each beginning in September of 1995 and ending in March 1998, with an interval of five months between October 1996 and February 1997 during which no data were collected. Some drivers are observed for a single month while others are observed for the entire 26 months, a characteristic accounted for in the modeling approach discussed below. On average, each driver is observed for 9.2 months.

The end of the first time period (October 1996) coincides with the announcement of changes in the firm's human resource practices designed to improve driver safety, which was then implemented at the beginning of the second period (February 1997). Of particular interest to this study are significant increases in driver per-mile compensation. Only the subset of drivers who were hired before the pay increase announcement and remained with the firm until the pay raise became effective experienced a pay increase. Pay for new hires was substantially higher after Hunt implemented the new pay policy, so drivers who joined the firm after the pay raise were hired at a higher base pay but did not experience a pay increase.

The longitudinal nature of the data allows us to study the dynamics of truck driver involvement in crashes. The dataset also is unique because it relies on data recorded by the firm, and not on driver recall, a known source of bias present in prevailing survey data. The richness of the data therefore facilitates an in-depth investigation of driver human capital and occupational factors for crash-involved *and* crash-uninvolved drivers, a characteristic commonly unavailable in other crash databases. Despite the longitudinal nature of the data at the driver-month level, we aggregate it at the driver level and conduct the analysis at such level for two reasons. First, we are mainly interested in driver-level attributes such as human capital, which vary more from subject to subject than from month to month. Second, unobserved variables such as vehicle and environmental factors would likely bias the coefficients estimated if we performed this analysis at the crash-level. This would occur if unobserved variables influencing driver safety were correlated with demographic and occupational variables. A driver's age and rate of pay, for example, might correlate with exposure to interstate highways or with the type of vehicle driven. Because type of vehicle and exposure to interstate highways may influence a driver's crash risk, not accounting for their effect can yield biased coefficients for age and pay or both. The failure to include roadway and environmental factors might also be a source of bias.

Data validity

A comparison with other sources of information about the TL sector provides information regarding the degree to which the dataset allows cautious generalizations to other firms in this sector (Table 1). The first source of data for comparison is a survey conducted the University of Michigan Trucking Industry Program (UMTIP) (for details, see 31). The numbers reported cover 233 full-time drivers who are employees and are paid by the mile. Owner operators and those drivers who are paid hourly are excluded from the figures presented. The second data source is a survey of firms included in the National Survey of Driver Wages published by Signpost, Inc. Signpost surveys approximately 200 truckload firms of various sizes. Most major TL carriers are represented and the set includes a sample of medium sized and smaller carriers. The figures presented are for 102 firms with mileage-paid employee drivers and which responded to the UMTIP survey of Signpost respondent firms regarding their pay practices for non-driving time (for a description of this survey, see 31). The third source of data presented includes figures estimated from a 1999 survey conducted for the Truckload Carriers Association (32).

{Insert table 1 here}

One major difference between Hunt and the first two sources of data is in the average length of each dispatch. This may be the result of the firm's reliance on rail transportation for hauling freight over long distances or due to particular characteristics of this individual firm's freight business. However, Hunt's figures are more similar to those reported by the Truckload Carriers Association (TCA) survey. The other major difference is the average tenure of each driver at the firm. Hunt's average tenure during the time the data was collected is significantly less than what is suggested by the UMTIP survey or the TCA survey (although this may be a measurement artifact, as measurement methods differ). Compensation or demographic characteristics are very similar.

It is important also to highlight that using firm-specific data has some shortcomings. Most prominent is that the results apply exclusively to the population of drivers belonging to the firm. As a result, any inferences about other truckload drivers are limited. This limitation should be viewed in the context of the relative unavailability of driver-level demographic and occupational data to researchers, which may explain the paucity of research on this topic. Researchers have primary data in a limited number of studies (for example see 24, 33, 34) or have collaborated with firms to examine their human resources and operations data (for example see 27, 35). Even when following the former approach, the ability to make general statements remains an issue. Truck stop surveys, for example, may oversample truckload for-hire carriers and over-the-road drivers. Similarly, self-reports about illegal behaviors such as speeding behavior or violation of the hours-of-service rules can result in known response bias in those surveys.

Variables observed

The outcome variable for our analysis is the total number of all crashes (at fault/not at fault) for each driver involving $500 or more of actual or estimated damages and that were recorded during the period of observation. We tested other crash cost cutoff points for the outcome variable such as $200, $1000 and $2000, and found no significant changes in the results. The dollar figure for the crash is the firm's estimated or actual cost associated with each crash (including bodily injury, property damage, and recording costs to all parties involved, but excluding potential changes in insurance costs) or the firm's actuarial estimates of the cost based on data for past crashes with similar characteristics. Non-casualty costs to the firm (such as productivity losses or the cost of recruiting or hiring new employees) and social costs such as losses to third parties (negative externalities) are excluded from these figures. Drivers have an average of 0.38 crashes during the observed period. Of course, this masks the fact that the majority of drivers (77%) do not record crashes during the period in which they were observed.

Independent variables include measures of working conditions, driver's demographic characteristics, two explicit measures of human capital and compensation variables, which we use as proxies for unobserved human capital characteristics. Human capital variables explicitly included in the model are driver tenure with the firm when first observed (years) and the race of the driver. Because tenure is not allowed to vary, the effect of "learning on the job" is not captured by this variable. Instead, we argue that the total number of miles driven captures such effect because the higher the number of miles driven, the higher the driving experience being acquired. The compensation variables included are driver pay rate when hired (cents per mile) and the percent pay raise received at the beginning of the second time period, if applicable. Based on the evidence provided by the review of the literature, we expect drivers' rate of pay to be negatively associated with their expected crash count, as higher pay will attract drivers with higher human capital. We also hypothesize that the higher the percent raise, the lower the expected crash count. This incentive effect is due to the anticipated impact of higher pay on individual behavior – effectively making unsafe behavior more costly to the individual –.

It is possible that the coefficient for percent pay raise also captures a second effect associated with the specific drivers who get a pay increase. To illustrate this potential bias, consider the drivers who experienced a pay raise. These drivers remained with the firm until the announcement of the pay raise and thus they were the only ones able to enjoy such raise. If drivers who tended to remain with the firm also happen to be safer drivers – a perfectly plausible supposition – their safety would in part be responsible for the fact that they got a pay increase. Thus, the causality for the percent pay raise variable would be muddled for this subset of drivers. We address this potential shortcoming by including two additional dummy variables. The first variable (*Cross*) indicates that a driver was hired before the pay raise and effectively received a pay increase thus capturing the potential influence of a driver's safety record on getting the pay increase. The second variable (*After)* indicates drivers who were hired after the pay raise. This variable measures the effect of human capital characteristics not captured by pay or other observed driver characteristics, and assuming that the firm is able to observe such characteristics. The default category corresponds to drivers hired and who left that company before the pay raise.

Variables that measure working conditions we include are total number of miles driven during the time each driver is observed (in millions), percentage of total miles driven during winter months (defined as December through March), and the total number of dispatches recorded. All else held equal, we expect that a higher number of dispatches involves a higher expected crash count due in part to the fact that each dispatch may be associated with more unpaid and unproductive waiting time and more frequently pulling-in and out of traffic conflict zones such as docks and urban areas. "Miles driven" is treated as an exposure variable, at the same time that the percentage of miles driven during the winter captures possible seasonal effects of the weather on crash risk.

Demographic characteristics we include are age when first observed, sex, and marital status. Descriptive summaries of the data show that the average driver age is 39.69 years and 48% of drivers are married (Table 2). Mirroring the industry, drivers tend to be mostly male (96%) and white (77%). The average pay rate at the time of hire was \$0.30 per mile and the pay increase averaged across all drivers is 9%. The latter figure substantially understates the pay raise because only the drivers working with the firm when the wage raise went into effect (24%) received the pay increase. Neither drivers who were hired during the first period and left before the pay raise became effective, nor drivers who were hired at a higher rate during the second period receive a pay raise. Among drivers receiving a pay raise, the average increase is 39.5%. Finally, the average miles driven for each driver is 70,000 miles although these vary considerably depending on how long a driver is observed. Finally, 38% of all the miles occurred during a winter month.

{Insert table 2 here}

CRASH COUNTS MODELS

To examine the impact of compensation, work conditions, and driver demographic characteristics on crash frequency while controlling for driver exposure, we estimated a model where crash frequency is the dependent variable, and demographic and occupational variables are the independent variables. When "count" dependent variables are treated as continuous variables, estimates derived using ordinary least squares regression can be inefficient, inconsistent, and biased (36). We therefore applied regression models such as Poisson and negative binomial models which are the most appropriate modeling technique for data that have a large number of zeroes and a lower number of positive integer variables (37, 38). In our case, we observed zero crashes for 77.2% of drivers, one crash for 12.9%, two crashes for 6.4%, three crashes for 2.7%, and four or more crashes for less than 1 percent of drivers.

Count models share similarities with linear regression models. For example, both attempt to explain variation in the dependent variable with a set of independent variables. They use a multivariate approach to isolate unique effects on the dependent variable. However, they also have some significant differences. Foremost is that count models assume that the expected number of counts have a certain probabilistic distribution (e.g., Poisson, negative binomial) with a conditional mean that depends on the independent variables observed. As such, results of count regression models provide two related pieces of information. On the one hand, they provide information regarding the influence of the independent variables on the expected count for each individual. On the other hand, count models also provide information on the distribution of counts for each individual. In our case, therefore, a count model will allow us to predict how the expected crash count varies for each individual with changes in an independent variable. It also allows us to explore how the probability of zero crashes, one crash, two crashes, and so on, changes for each individual as compensation or work conditions vary. We rely on both pieces of information provided by the models, since we believe they provide useful information for researchers and policy-makers.

Poisson regression and negative binomial models arguably are the most popular count regression models. The key distinction between the two is that the Poisson model requires that the mean of crash counts equal its variance, while the latter allows for differences between the mean and the variance of crash counts. Such difference in the mean and variance may be the result of unobserved heterogeneity across drivers. Because heterogeneity can also cause excess zeros, and in certain cases will always do so (39), we also examine the appropriateness of using zero-inflated models. Indeed, Lee et al. (37) encourage the application of zero-inflated models in the presence of unobserved heterogeneity. Zero inflated models assume that there are drivers that will always have a crash zero count and other drivers for whom the crash frequency can vary. Thus the zero inflated models identify both processes separately, with a binary function such as a logit or probit equation for determining always-zero cases and a Poisson or negative binomial equation for modeling the counts. This added flexibility of modeling crash counts may assist in crash prediction, and their results can suggest many specific relationships between independent variables and crash rates. For details on the derivation of these count models, see Long (36).

RESULTS

Our research approach was to select the best fit among a number of count regression models as determined by visual inspection of predicted versus observed counts, likelihood ratio tests, non-parametric tests, and prior theory (where appropriate). Once we selected a preferred model, we evaluated the contribution of demographic and occupational characteristics to crash frequency. In addition to the different model specifications, we explored alternatives to incorporating the impact of miles driven (exposure) on crash rates.

Preferred model selection

Following the suggestion from Lee et al. (37) an examination of the empirical frequency distribution of crashes suggests that the variance of crashes exceeded the mean and that there were a relatively high number of zero cases. This evidence was indicative, although not conclusively, of a poor fit for the Poisson model. Using the independent variables described in Table 2, we fit to the data four models: Poisson, negative binomial, and their zero-inflated counterpart models. In addition, the natural log of the exposure variable (miles driven) was included in each model with its coefficient constrained to one. This implied an assumption that the estimated crash rate increases linearly with exposure.

To determine the best-fitting model we estimate a Poisson model to use as a baseline against which other models could be tested. The three additional count models (negative binomial, zero-inflated Poisson, and zero-inflated negative binomial) were examined against this baseline model. Likelihood ratio tests were used to determine the preferred model when these models were nested (as with the Poisson and the negative binomial model). When one model is not nested within the other, we use non-parametric statistical test for comparing their fit as proposed by Vuong (for details see 40).

Consistent with our expectations, results suggested that the zero-inflated Poisson model was the preferred model. The preferred model was identified in two steps. First, the overdispersion parameter in the zero-inflated negative binomial specifications was not significantly different from zero (0.468) leading to its rejection over the Poisson model (results not shown). Second, the Vuong statistic comparing the zero-inflated Poisson model to the Poisson model (15.45) suggested that the former had better fit than the latter. Visual inspection of observed minus predicted frequency by type of count model confirmed that the zero inflated Poisson model fit the data best. Both the negative binominal and the Poisson models tended to underpredict zero counts and overpredict counts greater than zero. Fitting the same models for different crash cost cutoff points (no cutoff point, $200, $1000 and $2000) instead of the $500 initially specified also resulted in the selection of the zero-inflated Poisson model as the preferred regression model for this data. Furthermore, the estimated coefficients for the results using different crash cost cutoff points for the dependent variable did not vary significantly from those discussed in the next section.

Although we have no structural reason to believe that a zero inflation process occurs in the data (i.e., that certain truck drivers are risk-free of crash involvement while driving at least one mile), we specified a model that included a zero-inflation process. This is because Cameron and Trivedi (39) suggest that in many cases unobserved heterogeneity can result in excess zeroes in addition to overdispersion. Thus, specifying the model with a zero-inflation process was the result of practical rather than theoretical considerations. For completeness, the estimated coefficients for the inflation equation cannot be interpreted separately from the coefficients of the count equation. The practical implication is that the net effect of both equations on crash count should be considered at all times because failing to do so would have resulted in biased estimates.

Estimated influence of human capital and driver occupational factors on crash frequency
Results for the preferred count model are provided in Table 3, while results for the always-zero logit equations accompanying each count model are provided in Table 4. The left set of columns in Table 3 (model 1) shows the preferred model specification with only demographic characteristics as independent variables and with total miles driven as an exposure variable. The model explains 0.99% of the log-likelihood of the constant-only model (1- (-9,347.292/ -

9,437.056)). The importance of the unobserved heterogeneity created by excluding the human capital and occupational variables is reflected in the bias implicit in the estimated demographic variable coefficients. Inclusion of the occupational and human capital variables (Model 2) increases the share of the log-likelihood explained to 7.84%, all else held equal. This confirms the relevance of human capital and occupational factors in explaining crash frequency at the driver level, and indicates that information collected regarding crashes should include some level of occupational and human capital data for the involved driver. Despite the statistical significance of model 2, it has a substantial amount of unexplained variance. This is not entirely surprising given the stochastic nature of crash involvements and the relatively large sample size. However, several additional factors may contribute to improving the explanatory power of the model, such as additional driver-level factors (e.g. driving ability and ability to tolerate fatigue), vehicle factors (vehicle condition), other occupational factors (time spent waiting for loads, or loading and unloading, regularity of schedule, and hours worked/awake), and environmental factors (weather and quality of roads).

{Insert table 3 here}

In Table 3, the right-hand side column for the preferred model, labeled factor change, shows the change in the expected crash count given a change of one unit in the independent variable, holding all other variables constant. The factor change in expected crash count is calculated as the exponential of the estimated coefficient for each variable. This column is useful because, unlike linear regression, the coefficients estimated in count regression models do not indicate the effect of a unit change in the *j*th independent variable. Similarly, the partial derivative of the function with respect to the *j*th independent variable (known popularly as the marginal effect) cannot be used to estimate such an effect because the function is not linear. Instead, the factor change provides information about how to relate each variable with the expected crash count.

{Insert table 4 here}

The coefficient estimated for the pay rate variable in the count equation for Model 2 suggests that for every additional cent per mile a driver is paid, the expected crash count decreases by 8.15% (1-exp(-0.085)). Evaluated at the mean pay rate of $0.30, this translates into an elasticity of crash count with respect to pay rate of –2.47. However, because the pay coefficient in the always-zero equation has a positive sign, the estimated effect of the count equation underestimates the importance of pay rate. Taking into account both equations jointly, the coefficients suggest that one-cent higher pay increase is related to a 2.22 % lower probability of observing one or more crashes. Figure 1 shows how the probability of having a zero crash count varies as driver pay rate increases. The probability in the y-axis accounts for the probability resulting from the inflation model and the count model. As a proxy for unobserved human capital characteristics of drivers, this result implies that higher pay is associated with better driving records. Unobserved human capital characteristics include driving experience, other work experience, and driver character and disposition, among others. Interestingly, the coefficient for the dummy variable *After* indicates is not statistically significant. This suggests that drivers who were hired after the pay raise were not inherently safer than drivers hired before the pay raise, once we control for the human capital, occupational, and demographic variables noted before.

{Insert Figure 1 here}

Similar to results for pay rate, the coefficient for the percent pay raise variable in the count model and the inflation model jointly suggest that for every additional percentage point in the percent pay raise variable, drivers have a 0.23 percent lower expect crash frequency. The coefficient for percent pay raise suggests that there appears to be a relevant motivational effect related to the pay increase that led drivers to have better safety records. The coefficient for the dummy variable *Cross*, which accounts for the causality issue of who got a pay raise is positive but not statistically significant. Although speculative, it is possible that the link between percent pay raise and crash risk may be mediated by a variable such as intent to quit. At higher pay levels, drivers may be less likely to want to quit and therefore this accumulated experience may reflect positively in their driving record.

For driver occupational factors, we find that the higher the miles driven during winter months the higher the expected crash count. This probability, however, is moderated by the coefficient of the zero-inflation equation. A similar moderation effect is detected for dispatches, where the zero-inflation equation suggests that more dispatches are associated with a higher probability of a crash, but the count equation suggests the opposite. The effect captured by the zero-inflation equation is stronger than the effect captured by the count equation, suggesting overall that higher dispatches are associated with a lower probability of remaining crash-free.

Finally, it is useful to standardize the independent variables and re-estimate the preferred model in order to identify the variables having the strongest influence on crash count. After this re-estimation, model fit and the statistical significance of the independent variables remains identical but the newly estimated coefficient should be interpreted in units of standard deviation. Results of the count-model only suggest that driver pay, number of dispatches, and age are the strongest predictors of each driver's crash count. Thus, for example, an increase in one standard deviation of pay rate (from 30.3 to 37 cents/mile) is associated with a decrease in expected crash count of 43.6% A comparison with the standardized coefficient results of the always-zero equation reveals some insightful differences. In the always-zero equation, occupational variables related to driving activity such as miles, dispatches, and miles driven during winter become stronger predictors than in the count model. Similarly, tenure becomes a stronger predictor, but pay decreases in importance. These results seem to suggest that even though driver pay does seem to be correlated with the occurrence of a crash, the correlation becomes stronger when crash frequency (as opposed to crash occurrence) is examined as the dependent variable. A similar argument applies to percent pay raise, which was not significant in the always-zero equation and is significant in the count equation.

Estimated influence of demographic factors on crash frequency

While our primary interest lies with the impacts of human capital and occupational factors on driver safety, it is also useful to examine the estimated influence of control variables on crash frequency. Obtaining reasonable results with respect to other variables lend credence to this exercise. Consistent with prior literature (41-43), the coefficients estimated for the age variable suggest that the expected crash count diminishes as driver age increases, but at a decreasing rate.

Taking into account both equations simultaneously suggests that married individuals are 7.07 percent less likely to have any crashes than non-married individuals. The count-only model equation coefficient suggests that the expected crash count for married drivers is 0.89 times the count of non-married drivers, although this is an underestimate because the zero-always equation is not being taken into account. Finally, looking at both models simultaneously we find that the probability of having no crashes for females is 6.9% lower than for males. This contrasts with recent research (44) finding that, for the population at large, there is no difference between crash involvement rates by sex after controlling for annual miles driven.

As with the age variable, the impact of tenure at the firm when first observed on expected crash count is also quadratic, a result that is consistent with the prior research (e.g., 27). The estimated coefficient for tenure and its square suggest that the probability of having a zero count is highest when the driver has been with the firm for 5.81 years (Figure 1, right panel). In contrast, the net crash count effect of a unit increase from the mean for dispatches or for the percent miles driven during winter is ambiguous because the effect of these two variables in the count equation and the always-zero equation are in opposing directions. Finally, a somewhat surprising result relates to the two control variables introduced to address potential causality problems with the percent raise variable. Coefficients for both dummy variables suggest that there is no difference in the expected crash count between drivers hired before and after the pay raise, all else held equal. This does not mean, however, that there was no safety improvement before and after the pay raise. The coefficients for the compensation variables suggest the opposite: that being hired at a higher pay or getting a pay raise contributed to improving the overall crash record of drivers.

CONCLUSIONS

Recent research has shown that human capital and occupational factors are important predictors of driver crash involvement as well as of unsafe behaviors such as speeding and violation of the hours-of-service rules. Using a different methodological approach and a unique dataset, the results of this research support such evidence by showing that such factors are also important predictors of frequency of crash involvement. In particular, higher pay rates and pay raises are related to lower expected crash counts and to a higher probability of zero crash counts, all else held equal.

The results strengthen the limited empirical evidence linking structural occupational factors of drivers – such as economic rewards – with safety outcomes, and extend it by examining their effects on crash occurrence and frequency of occurrence. The effect of tenure and age on expected crash count exhibits the expected nonlinear form, thereby suggesting that factors that keep drivers at jobs can also contribute to better safety outcomes. Although the evidence provided here is not definitive, it may suffice to motivate changes in human resources and hiring practices for some firms. Policy-makers can use this research to support efforts to investigate trucking labor markets using a broader framework that traces the impact of individual carrier compensation on the truck driver labor market and other related labor markets.

Potential limitations of this study include the firm-specific nature of the data, which constrains the extent to which inferences about the truckload sector can be made, and that several variables, such as driving experience and prior driving record, are not available for all observations. However, comparisons between the current data and three surveys of the truckload sector suggest that the average characteristics of the drivers in this study seem comparable with those of the sector. More importantly, the unavailability of certain variables can introduce bias to the coefficients of certain observed variables. In practical terms, however, variables such as age, marital status, and tenure at the firm are expected to be reasonable proxies for unobserved variables such as driving experience. A natural extension of this study is to include disaggregate data from other firms in order to understand the unique contribution of firm characteristics, such as financial performance and size, to driver safety. Such an approach will also provide the possibility of determining the extent to which pay solely is a proxy for human capital characteristics or if it plays a broader role in motivating employees. At the driver level, other extensions of this study include marrying these types of disaggregate, driver-level data sources with other driver-related variables such as driving hours and non-driving work hours. Similarly, the combination of driver-level data and firm-level data to study the effects of policies such as driver education, compensation, or operating policy that varies by firm.

From the modeling perspective, zero-inflated models improve model fit but reduce the interpretability of the results. Unbiased estimates of expected crash count and of the probabilities of particular crash counts other than zero are not readily available for these models because of the nature of the zero-always inflation equation. Although Poisson and negative binomial models are more practical and provide conceptual simplicity, we find that the more sophisticated zero-inflated models provide enough information to yield useful results while reducing the possibility of introducing bias in the estimated coefficient. The amount of unexplained variation points to the fact that other factors than the ones included in these models are important to explain truck crashes. A comprehensive model should include firm characteristics, vehicle factors, environmental factors and roadway factors along with driver factors. The lack of available data on all of these variables remains an important problem. Despite this, the results provide continued indication of the relevance of accounting for occupational and human capital variables in the study of truck driver safety. Failing to do so may results in biased results of limited value to policy-makers and researchers.

ACKNOWLEDGEMENTS

The authors are grateful for the generous support of the Alfred P. Sloan Foundation, the Trucking Industry Program, and the Southeastern Transportation Center, and for JB Hunt’s assistance in allowing us access to the data. We also are grateful for feedback from anonymous reviewers.

References

1. Lyman, S. and E. Braver, Occupant Deaths in Large Truck Crashes in the United States: 25 Years of Experience, *Accident Analysis and Prevention*, forthcoming, pp. 1-9.

2. US Department of Transportation, *Large Truck Crash Facts*, Federal Motor Carrier Safety Administration, 2000.

3. Beilock, R., Schedule-Induced Hours-of-Service and Speed Limit Violations Among Tractor Trailer Drivers, *Accident Analysis and Prevention*, vol. 27, 1995, pp. 33-42.

4. Arnold, P. K., L. R. Hartley, A. Corry, D. Hochstadt, F. Penna, and A. M. Feyer, Hours of Work and Perceptions of Fatigue Among Truck Drivers, *Accident Analysis and Prevention*, vol. 29, 1997, pp. 471-477.

5. Wierwille, W. W. and L. A. Ellsworth, Evaluation of Driver Drowsiness by Trained Raters, *Accident Analysis and Prevention*, vol. 26, 1994, pp. 571-581.

6. Wylie, C. D., J. C. Shultz, M. M. Miller, and R. R. Mackie, *Commercial Motor Vehicle Driver Fatigue and Alertness Study Project Report*. Report No. FHWA-MC-97-002, October 1996.

7. Beilock, R., R. Capelle, and E. Page, Speed and Training Factors Associated with Heavy Truck Accidents, *Transportation Quarterly*, vol. 43, 1989, pp. 571-589.

8. Golob, T., "Driver Behavior of Long Distance Truck Drivers: The Effects of Schedule Compliance on Drug Use and Speeding Citations," presented at 74th Transportation Research Board Meeting, Washington, D.C., 1995.

9. Walton, D., Examining the Self-enhancement bias: Professional Truck Driver's Perceptions of Speed, Safety, Skill and Consideration, *Transportation Research Part F: Traffic Psychology and Behavior*, vol. 2, 1999, pp. 91-113.

10. National Transportation Safety Board, *Factors That Affect Fatigue in Heavy Truck Accidents Vol. 1: Analysis*. NTSB/SS-95/01. National Transportation Safety Board, Washington, D.C. 1995.

11. GAO, *Freight Trucking: Promising Approach for Predicting Carriers' Safety Risks*. GAO/PEMD-91-13. United States General Accounting Office (US GAO), Washington, D.C. April, 1991.

12. Braver, E. R., C. W. Preusser, D. F. Preusser, H. M. Baum, R. Beilock, and R. Ulmer, Long Hours and Fatigue: A Survey of Tractor-trailer Drivers, *Journal of Public Health Policy*, 1992, pp. 341-366.

13. Becker, G., *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press, 1964.

14. Becker, G., Investment in Human Capital: A Theoretical Analysis, *Journal of Political Economy*, vol. 70 (supplement), 1962, pp. 9-44.

15. Krass, A., The Impact of the US Motor Carrier Act of 1980 on Road Safety in California: An Econometric Policy Evaluation, *Logistics and Transportation Review*, vol. 29, 1993, pp. 179-92.

16. Hirsch, B., Trucking Deregulation and Labor Earnings: Is the Union Premium a Compensating Differential, *Journal of Labor Economics*, vol. 11, 1993, pp. 279-301.

17. Monaco, K. and E. Williams, Assessing the Determinants of Safety in the Trucking Industry, *Journal of Transportation and Statistics*, vol. 3, 2000, pp. 69-80.

18. Beilock, R., "Schedule Tightness and Work Hours of Long-Distance Tractor Trailer Drivers in the United States," presented at Canadian Transportation Research Forum, Victoria (BC), 1994.

19. Sucharov, L. J., F. Saccomanno, M. Yu, and J. H. Shortreed, "Effect of Driver Fatigue On Truck Accident Rates.," in *Urban Transport and the Environment for the 21st Century*, L. J. Sucharov, Ed. Boston: Computational Mechanics Publications, 1995, pp. 439-446.

20. Feyer, A.-M., A. M. Williamson, R. Jenkin, and T. Higgins, *Strategies to Combat Fatigue in the Long Distance Road Transport Industry - Bus and Coach Perspective*. CR 122. Australian National Occupational Health and Safety Commission, Sydney., Canberra 1993.

21. Lyznicki, J., T. Doege, R. Davis, and M. Williams, Sleepiness, Driving and Motor Vehicle Crashes, *Journal of the American Medical Association*, vol. 279, 1998, pp. 1908-1913.

22. Hakkanen, H. and H. Summala, Sleepiness at Work Among Commercial Truck Drivers, *Sleep*, vol. 23, 2000, pp. 49-57.

23. McCartt, A., J. W. Rohrbaugh, M. C. Hammer, and S. Z. Fuller, Factors Associated With Falling Asleep at The Wheel Among Long-Distance Truck Drivers, *Accident Analysis and Prevention*, vol. 32, 2000, pp. 493-504.

24. Hensher, D., R. Daniels, and H. Battellino, "Safety and Productivity in the Long Distance Trucking Industry," presented at Proceedings, 16th Australian Research Board Conference, Perth, 1992.

25. McCartt, A., M. C. Hammer, and S. F. Fuller, "Understanding and Managing Fatigued Driving: Study of Long-Distance Truck Drivers in New York State," presented at International Large Truck Safety Symposium, Knoxville, Tennessee, 1997.

26. Chatterjee, A., E. Cadotte, H. Sink, M. Venigalla, and G. Gaides, *Driver-Related Factors Involved With Truck Accidents*. . Institute for Transportation Research and Education, Raleigh, NC 1994.

27. Lin, T. D., P. Jovanis, and C. Z. Ynag, Modeling the Safety of Truck Driver Service Hours Using Time-Dependent Logistic Regression, *Transportation Research Record*, 1993, pp. 1-10.

28. Hensher, D. and H. C. Battelino, “Long-Distance Trucking: Why Do Truckers Speed?,” presented at Australasian Transport Research Forum, 1990.

29. Hensher, D., H. Battellino, J. Gee, and R. Daniels, *Long Distance Truck Drivers On-road Performance and Economic Reward*. Australia, Institute of Transport Studies, Sydney University, December 1991.

30. Belzer, M., *Sweatshops on Wheels: Winners and Losers in Trucking Deregulation*. New York: Oxford University Press, 2000.

31. Belzer, M., D. Rodriguez, and S. Sedo, *Paying for Safety: An Economic Analysis of the Effect of Compensation on Truck Driver Safety*. Report to the Federal Motor Carrier Safety Administration under Prime Contract Number DTFH 61-98-C-0061. Wayne State University, Washington D.C. April 2002.

32. Martin Labbe Associates, *1999 Dry Van Driver Survey*. Prepard for Truckload Carriers Association. Martin Labbe Associated, Ormond Beach, FL 1999.

33. Hertz, R. P., Hours of service violations among tractor-trailer drivers, *Accident Analysis and Prevention*, vol. 23, 1991, pp. 29-36.

34. Belman, D. L. and K. A. Monaco, The Effects of Deregulation, De-unionization, Technology, and Human Capital on the Work and Work Lives of Truck Drivers, *Industrial & Labor Relations Review*, vol. 54, 2001, pp. 502-524.

35. Jovanis, P. and H. Chang, Disaggregate Model Of Highway Accident Occurrence Using Survival Theory, *Accident Analysis and Prevention*, vol. 21, 1989, pp. 445-458.

36. Long, S., *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage, 1997.

37. Lee, A. H., M. R. Stevenson, K. Wang, and K. K. W. Yau, Modeling Young Driver Motor Vehicle Crashes: Data With Extra Zero, *Accident Analysis and Prevention*, vol. 34, 2002, pp. 515-521.

38. Miaou, S. P., The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions, *Accident Analysis and Prevention*, vol. 26, 1994, pp. 471-482.

39. Cameron, A. C. and P. K. Trivedi, *Regression Analysis of Count Data*. Cambridge: Cambridge University Press, 1998.

40. Vuong, Q. H., Likelihood Ratio Tests For Model Selection and Non-nested Hypotheses, *Econometrica*, vol. 57, 1989, pp. 307-333.

41. Campbell, K., Fatal Accident Involvement Rates by Driver Age for Large Trucks, *Accident Analysis and Prevention*, vol. 23, 1991, pp. 287-295.

42. Blower, D., *The Accident Experience of Younger Truck Drivers*. GLCTTR 81-96/01. University of Michigan Transportation Research Institute/Great Lakes Center for Trucking and Transit Research, Ann Arbor, MI 1996.

43. National Highway Traffic Safety Administration, *Large-Truck Accident Causation*. United States Department of Transportation, Research and Development, National Center for Statistics and Analysis, NHTSA, Washington D.C. 1982.

44. Lourens, P. F., J. A. M. M. Vissers, and M. Jessurun, Annual Mileage, Driving Violations and Accident Involvement in Relation to Drivers' Sex, Age and Level of Education, *Accident Analysis and Prevention*, vol. 31, 1999, pp. 593-597.

List of Tables

List of Figures

Table 1. Comparison between JB Hunt data and other data sources for the truckload sector

Variable	JB Hunt	UMTIP driver survey	Signpost and UMTIP firm survey	Truckload Carriers Association
Age (years)	39.69	42.18	n.a.	41.0
Race (1 = Non-white)	22.7%	14%	n.a.	n.a.
Married	48%	69%	n.a.	n.a.
Tenure at firm when first observed (yrs)	1.20	3.46	n.a.	4.2
Base pay (cents/mi)	30.28	28.6	28.6	n.a.
Miles per dispatch	575.8	858.0	905.9	686.0

Table 2. Summary statistics and variable explanation (N=11,540)

Variable label	Explanation	Mean	St. Dev.	Min	Max
DEMOGRAPHIC					
Age	Mean driver age	39.69	10.14	20	76
Female	= 1 if female driver, = 0 otherwise	0.04	0.19	0	1
Married	= 1 if married, = 0 otherwise	0.48	0.50	0	1
COMPENSATION					
Pay	Pay (cents/mile) when hired	30.28	6.73	16	49
%Raise	Percentage pay raise	9.16	19.45	0	123.53
HUMAN CAPITAL					
Race	= 1 if non-white, = 0 otherwise	0.23	0.42	0	1
Tenure	Tenure at firm when first observed (years)	1.20	2.16	0.08	19.17
OCCUPATIONAL					
Miles	Million miles driven during observed period	0.07	0.08	0	0.5
Mile_win	Percentage of miles driven during winter season (December – March)	0.38	0.33	0	1
Dispatch	Total number of dispatches during observed period	128.63	138.86	1	940
OTHER CONTROL VARIABLES					
Cross	=1 if driver was hired before the pay raise period *and* received a pay raise, = 0 otherwise	0.24	0.42	0	1
After	= 1 if driver was hired after the pay raise occurred, = 0 otherwise	0.36	0.48	0	1
DEPENDENT VARIABLE					
Crash_ct	Crash count per driver	0.38	0.81	0	8

Table 3. Zero-inflated Poisson crash models

Variable	Model 1 Coefficient	Model 1 T-statistic	Model 2 Coefficient	Model 2 T-statistic	Factor change	Standardized Coefficients
Constant	4.761***	19.660	6.087***	18.150		2.690
DEMOGRAPHIC						
Age	-0.103***	-8.650	-0.038***	-3.100	0.963	-0.383
Age^2	0.001***	7.530	0.001***	4.020	1.001	0.484
Female	0.571***	5.460	0.189*	1.690	1.208	0.035
Married	-0.197***	-4.550	-0.112***	-2.430	0.894	-0.056
COMPENSATION						
Pay			-0.085***	-8.890	0.919	-0.572
%Raise			-0.008***	-3.940	0.992	-0.163
HUMAN CAPITAL						
Race			0.160***	3.310	1.173	0.067
Tenure			-0.056*	-1.720	0.946	-0.121
Tenure^2			0.006*	1.860	1.006	0.111
OCCUPATIONAL						
Miles	1.000		1.000			
Mile_win			0.417***	3.570	1.517	0.136
Dispatch			-0.003***	-13.350	0.997	-0.446
OTHER CONTROL VARIABLES						
Cross			0.136	1.190	1.146	0.058
After			0.144	1.050	1.155	0.069
Log-L full model	-9347.292		-8,394.167			
Log-L constant-only (1)	-9437.056		-9,052.593			
LR test (model 2 vs. 1)			-1906.25			
Rho-square	0.99%		7.84%			
N	11,540		11,540			
Vuong statistic	16.650***		15.450***			

*** Significant at a 99% confidence level
** Significant at a 95% confidence level
* Significant at a 90% confidence level
(1) Log-likelihood of constant-only model varies from model 1 to model 2 because the likelihood at convergence of the zero-always logit model differs for both models.

Table 4. Logit zero-always inflation equations (1)

1 = non-crash state	Model 1		Model 2		
Variable	**Coefficient**	**t-statistic**	**Coefficient**	**t-statistic**	**Standardized Coefficients**
Constant	-0.117	-0.770	-1.073***	-2.660	0.187
DEMOGRAPHIC					
Age	-0.002	-0.660	-0.010***	-2.450	-0.103
Female	0.487***	2.690	0.494***	2.550	0.092
Married	0.257***	3.220	0.239***	2.850	0.120
COMPENSATION					
Pay			0.043***	2.820	0.290
%Raise			0.004	1.050	0.084
HUMAN CAPITAL					
Race			-0.296***	-3.140	-0.124
Tenure			0.142***	2.370	0.306
Tenure^2			-0.011*	-1.710	-0.203
OCCUPATIONAL					
Miles			6.812***	4.140	0.563
Mile_win			0.805***	4.230	0.262
Dispatch			-0.005***	-4.780	-0.712
OTHER CONTROL VARIABLES					
Cross			-0.225	-1.030	-0.096
After			0.158	0.730	0.076

*** Significant at a 99% confidence level
** Significant at a 95% confidence level
* Significant at a 90% confidence level
(1) The inflation process models the probability of remaining in non-crash state (zero-always), which is the opposite of the count model. As such, the coefficients are expected, although not required, to have different signs in the inflation model than in the count model.

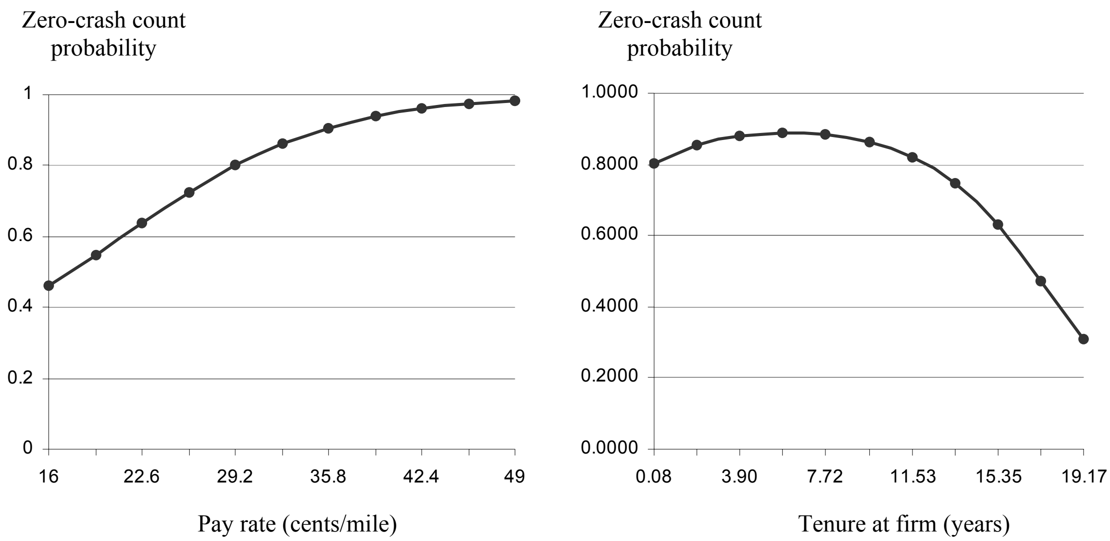

Figure 1. Estimated variations in the probability of having no crashes with changes in pay rate (left panel) and tenure (right panel)