Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models

Young-Jun Kweon
Graduate Student Reseacher The
University of Texas at Austin
fire264@mail.utexas.edu

and

Kara M. Kockelman
Clare Boothe Luce Assistant Professor of Civil Engineering
The University of Texas at Austin
kkockelm@mail.utexas.edu
(Corresponding Author)

The University of Texas at Austin
6.9 E. Cockrell Jr. Hall
Austin, TX 78712-1076
Phone: 512-471-0210
FAX: 512-475-8744

The following paper is a pre-print and the final publication can be found in
Accident Analysis and Prevention 35 (3):414-450, 2003.
Presented at the 81st Annual Meeting of the Transportation Research Board, January 2002

ABSTRACT

Traffic crash risk assessments should incorporate appropriate exposure data. However, existing U.S. nationwide crash data sets, the NASS General Estimates System (GES) and the Fatality Analysis Reporting System (FARS), do not contain information on driver or vehicle exposure. In order to obtain appropriate exposure data, this work estimates vehicle miles driven (VMD) by different drivers using the Nationwide Personal Transportation Survey (NPTS). These results are combined with annual crash rates and injury severity information from the GES for a comprehensive assessment of overall risk to different drivers across vehicle classes.

Data are distinguished by driver age, gender, vehicle type, crash type (rollover versus non-rollover), and injury severity. After correcting for crash exposure to drivers, results indicate that young drivers are far more crash prone than older drivers (per vehicle mile driven) and that drivers of sports utility vehicles (SUVs) and pickups are more likely to be involved in rollover crashes than those driving passenger cars. Although the results suggest that drivers of SUVs are generally much less crash prone than drivers of passenger cars, the rollover propensity of SUVs and the severity of that crash type offset the incident benefits for the younger drivers.

KEY WORDS

Crash risk, crash exposure, crash frequency, crash severity, sport utility vehicles, light-duty trucks

INTRODUCTION

Risk assessment of road users has been conducted in many ways, including estimation of crash rates by road type, injury severity by crash type, and frequency of speeding violations. The focus of the assessment may be on facilities, vehicles or travelers. Policy makers and roadway design engineers act to improve road safety by applying the knowledge gained in such studies.

Estimating crash rates is one of the most common ways to assess the risk of road users or road facilities. Rate calculation requires division of crash counts (i.e., crash frequency) by some measure of exposure (e.g., vehicle miles traveled [VMT]). As Hauer (1995) has noted, such normalization equalizes for differences in intensity of use, making safety comparisons more meaningful, and it helps identify differences between different populations' characteristic crash rates as a clue to causal factors.

As Evans (1991) and others have noted, one cannot draw reliable conclusions on safety issues without exposure information. Unfortunately, very few data sets provide adequate exposure information. However, several researchers have obtained surveys and/or estimates of such exposure. These are discussed here now.

Lourens et al. (1999) categorized the car drivers into five subgroups by annual mileage driven. They estimated crashes as a function of annual mileage interacted with gender, age, education level, and frequency of fines (or violations). After this implicit correction for annual mileage, the effects of gender and education level were found to not be statistically significant predictors of crash involvement. Young drivers exhibited the highest crash rates, and traffic violations were important predictors. Hu et al. (1998) also controlled for driving distance, but they focused on the medical condition of older drivers. They estimated that a gender effect for this age group of drivers was statistically insignificant. Such results would have been hidden from the modelers had driving exposure not be recognized.

Using a survey data of Ontario licensed drivers, Chipman et al. (1992) estimated driving time exposure and driving distance exposure, and they took the ratio of these to produce an average driving speed measure. All data were stratified by age (six levels), gender, and region (three levels); and three exposure measures were compared among the stratified subgroups. They argued that the fatal crashes should be analyzed separately from other crash types since their rates and involved factors are so distinctive, and they recognized that average driving speed probably relates to crash severity. The men in their data set drove 56% more distance than the women, but they only spent 35% more time driving than the women (suggesting higher speeds for men, and thus more dangerous exposure for men). As expected, teenage drivers (i.e., those under the age of 20) were found to be less exposed (about 23% less, in distance and in time) than the data set's "middle-aged" drivers (ages 25-59). And older drivers (age 60-69) drove about 33% less distance and 19% less time than the middle-aged drivers.

Davis and Gao (1993) tried to verify the assumption of random selection of crash victims, using induced exposure methods (where one examines ratios of cross-classified counts for high-risk populations and roadways). They parameterized relationships between induced exposure and contingency tables/cross-tabulations and then estimated the ratios of crash rates for different driver cohorts (e.g., male versus female, and middle-aged versus older drivers), based on an assumption of asymptotically normally distributed crash counts. An empirical Bayes method was used to spot the difference between crash rates of driver cohorts between two roadways. Through the contingency tables, they found that older drivers (aged 56 and over) were more likely to be involved in crashes when traversing certain roadways. Using Lyles et al.'s data

(1991) for crash rates, they found male drivers to be 40% more crash prone than female drivers. DeYoung et al. (1997) used a quasi-induced exposure method (which is similar to the induced method but does not correct for population representation) to estimate the exposure and fatal crash rates of suspended/unlicensed (S/R) drivers using data on two-vehicle crashes in the 1987-1992 Fatality Analysis Reporting System (FARS) data set. As it turned out, the quasi-induced exposure method did not predict the number of S/R and unlicensed drivers very well, in part because the purpose of this method is to estimate crash involvement rates, rather than the number of specific road users.

Using police reports from in Western Australia, Ryan et al. (1998) studied the relationship between crash risk and driver age. Their findings suggested that females exhibit higher crash involvement rates, across all age groups, a result often inconsistent with common perceptions, largely due to the neglect of driving exposure across genders in popular statistics. Drivers under the age of 20 were the most likely to be crash-involved, and drivers under age 25 exhibited very high rates as well. Age also exhibited a positive statistically significant relation to driver injury severity. Doherty et al. (1998) also found that drivers under age 20 were more likely to be crash-involved than all other age groups. Abdel-Aty et al. (1998) used a categorical method with 1994 and 1995 Florida accident databases to analyze the relationship between age and crash risk. Their results indicated that young drivers (aged 25 and under) and older drivers (over age 64) were more likely to be crash-involved, overall, but middle-aged drivers were more likely to be involved in crashes which involve alcohol or occur during rush hours. They also noted that older drivers are more likely to die in crashes, since they are relatively physically weak.

While the literature has uncovered a number of exposure differences across driver types, essentially no such research has considered the type of passenger vehicles being driven. Yet one may expect that passenger cars, vans, pickups, and sport utility vehicles (SUVs) perform very differently, in addition to having varying crash exposures. One particularly severe crash type is that involving a rollover. The National Highway Traffic Safety Administration (NHTSA) has found that the relationship between rollover risk and Static Stability Factor (SSF) (defined as a half the width of a vehicle divided by the height of its center of gravity) is negative. This is as expected: less stable vehicles (i.e., those with a lower SSF) include SUVs and pickups, and these are often perceived as more likely to roll over in single-vehicle crashes. The average SSF of the 100 vehicle models studied by NHTSA (2000) were 1.400 for passenger cars, 1.153 for vans, 1.087 for SUVs, and 1.170 for pickups. In controlling for exposure differences, this work is able to rigorously examine whether rollover crash rate differences exist across vehicle (and driver) types.

This research has made use of large U.S. data sets to provide robust estimates of crash rates by crash category, driver characteristics, and vehicle type. It also merges these results with crash-severity model estimates (as provided by Kockelman and Kweon [2001]) to estimate the probability of severe injuries and death to the different driver-vehicle cohorts, over 50,000 miles of driving. Such estimates are very valuable for public policy and driver assessments of driving risks.

THE DATA SETS

The 1995 Nationwide Personal Travel Survey (NPTS) and the General Estimates System (GES) data sets were used for this study. The NPTS data have been collected roughly every five years, but the 1995 NPTS data are the most recently available at this time.[1] And, although the

GES data is collected annually (the most recently available one in 1999), the 1995 data were used here, in order to correspond to the NPTS data.

1995 NPTS Data

The NPTS data set is the nation's best single course of a daily trips inventory. The 1995 dataset includes 42,033 households, 95,360 persons, and 75,217 vehicles. There are six physically separate files associated in the 1995 NPTS dataset, and four conceptually different levels (household, person, vehicle, and travel day) in the dataset. These separate files were appropriately combined with weighting factors to match this study's purposes.

Applying appropriate weighting factors is requisite for all analysts who intend to use survey data correctly. In the NPTS 1995 data set, households were over-sampled in add-on areas such as New York, Massachusetts and Oklahoma. There are four different weights for the different levels of analysis; these are household weights, person weights, travel day weights, and travel period weights. Since the purpose of using the NPTS data in this work was acquisition of drivers' travel distances, travel day weights were used when combining the files.

1995 GES Data

The National Automotive Sampling System (NASS) General Estimates System (GES) dataset, briefly the GES data, has been collected every year from 1988. The GES data are intended to be a nationally representative probability sample from the annual estimated 6.4 million police accident reports in the United States. The GES includes all types of crashes, including fatal crashes, injury crashes, and property-damage-only crashes. The 1995 GES data include 53,749 crashes, 95,803 vehicles, 140,512 persons, and 95,477 drivers. While these 53,749 crashes are .8% of the (estimated) U.S. total in 1995, and each GES crash observation comes with a national weighting factor. These were used to scale up the GES crash counts to a U.S. total (as shown in Table 2).

Extracting Necessary Information from the 1995 NPTS and GES

Since both the NPTS and GES datasets have several separate files, there was some work involved in merging and aggregating their data files. The person-level, vehicle-level, and travel-day-level files in the 1995 NPTS dataset were merged using travel-day weight factors (since VMT estimates were needed).

The merged data file was initially categorized across 48 cohorts (by age, gender, and vehicle type). Unfortunately, sufficient information on observations involving "unknown" and "other" vehicle types was not available, necessitating the removal of 12 cohorts. And the NPTS data set does not cover commercial vehicle operations, so the 6 cohorts involving heavy- and medium-duty trucks were removed. The remaining 30 cohort categories are defined by vehicle type, gender, and age. The 5 available vehicle types are those of passenger car, SUV, minivan (MVAN)[2], pickup (PU), and motorcycle (MC). SUVs are conventionally defined as light-duty vehicles that meet several clearance requirements (for off-road use)[3]. In the NPTS and GES data sets, pickup trucks are defined as a light conventional truck of pickup style cab weighing less than 10,000 lbs (with maximum payload). The three age divisions are comprised of young drivers (i.e., those less than 20 years of age), "middle-aged" drivers (i.e., those from 20 through 60 years of age), and older drivers (i.e., those over age 60).

Since this investigation's focus is on *driver* exposure and involvement, only trips records cases where the traveler was driving were extracted. The travel miles were then aggregated in these same categories, to provide an estimate of annual vehicle miles driven (VMD) for the national cohort. These numbers are provided in Table 1.

Three files in the 1995 GES data were merged into a single driver-level files; these are the vehicle, crash, and person files. And the national weight factors were applied to each, to reflect sampling biases. Imputed variables were used for all three factors where data were missing, which occurred in 4.5% of records for the case of age, 2.9% for the case of gender, and 2.2% for the case of vehicle body type. (All records were retained to permit a better estimate of the total, population counts.) The categorization and aggregation of the weighted number of crashes produced the final crash frequency data for the 30 cohorts; these counts are provided in Table 2. Crash counts were divided by VMDs in each of the cohorts for all crash types and rollover crashes; these crash rate values are provided in Tables 3 and 4.

Before presenting the results of this work, it should be noted that, *if* the weighted NPTS records do not capture total use of passenger cars, SUVs, pickups, minivans, and motorcycles in the U.S., Table 3 and 4's crash rate estimates are expected to be biased high. For example, the NPTS asks its sampled households' members to report any type of driving they do (even that driving which is part of their work), but the amount of commercial/heavy-duty-vehicle driving is clearly biased low in the NPTS data set. This work's focus is on vehicles driven by households; however, the NPTS's undersampling of VMD may occur here, to some extent.

RESULTS AND DISCUSSION

Crash rates of driver-vehicle cohorts are presented and discussed in this section. Crash rates for *all* types of crashes are discussed first; these are followed by an examination of rollover and non-rollover crash rates.

Vehicle Miles Driven (VMD) and Annual Crash Counts

Estimates of vehicle miles driven (for the year 1995, using the NPTS data set) are presented in Table 1, and estimates of annual crash counts (for the year 1995, using the GES data set) are provided in Table 2. No observations of motorcycle (MC) driving were made in the NPTS data set for two of the gender-age cohorts (though these were observed in crashes); thus, there are some missing cells in Table 1 (and in the resulting crash-rate tables).

As evidenced in Table 1, the total number of miles driven by young males is 46% more than that driven by young females. The same gap, in percentage terms, is visible for middle-age males and females, but widens to 94% for older persons. This widening may be attributed to cultural norms of the past, when such persons came of driving age.

In the category of pickup trucks (PU), men drive much more than women, logging 6 to 13 times more miles across the three age categories. In the middle-age driver cohorts, there is not much difference in total vehicle miles driven between men and women of passenger cars (PCs) or minivans (MVANs). However, this does not mean that women and men owning such vehicles drive similar distances: men, on average, drive more – because many of them are driving other vehicle types. Claiming almost 50% of the total VMD, the middle-age male cohort clearly drives the most.

Table 2 provides the annual crash counts (using 1995 data) for 30 age-gender-vehicle cohorts. Young and middle-aged male drivers experience almost 60% more vehicle crashes than female drivers. In the case of older drivers, the relative difference in the number of crashes rises to 70% – but this is against a 100% increase in exposure, as evidenced in Table 1.

In terms of the numbers of SUV-involved crashes, there are considerable differences between male and female drivers for each age cohort; and these differences become larger in crashes of pickup (PU) trucks. But the absolute numbers of vehicle crashes cannot provide solid evidence of safety problems associated with road users. An appreciation of relative crash frequencies requires recognition of exposure (VMD) data. This is done in the following section.

Crash Rates, All Crash Types

Dividing Table 2's cell values by those of Table 1 yields Table 3's estimates of crash rates. Using Table 3, an exposure-adjusted comparison of crash frequency or risk is now possible, between cohorts of interest.

Over all vehicle types, there are no substantial differences between the general crash rates for male and female drivers in the same age cohort. Young and middle-aged men are slightly more involved than their female counterparts, but older women are slightly more involved than older men. The additional driving experience (as evident in VMD) does not make male drivers less crash-involved; however, male drivers may be driving in more dangerous environments (e.g., when dark and rainy or at high speeds). And women may recognize their relative lack of driving experience, taking extra precautions. This may be true across vehicle types as well, resulting in a degree of incomparability across cohorts and cases.

Differences of interest also can be found across vehicle types in Table 3. Males are more crash-involved than females when driving passenger cars, but not when driving light-duty trucks (i.e., SUVs, pickups, and minivans). And drivers of passenger cars are more crash-involved than those driving light-duty trucks. They are more than twice as crash-involved as drivers of SUVs! Possible explanations for part of such striking distinctions include the following: light-duty truck (LDT) drivers may drive differently (e.g., more slowly) and/or be more risk averse; LDT drivers may be less likely to report crashes (e.g., in rural areas or with minor property damage); they may drive on less congested and/or safer roads during better conditions; they also may have better sight distances or more stable vehicles. Such data distinctions are generally not available in *both* the NPTS and GES data sets, but they are of interest to policy makers and drivers. (As an example of such work, Kockelman and Yong's [2001] multivariate models found that, *ceteris* *paribus*, larger, wealthier households in lower-density areas purchased more SUVs than other households. Smaller, less wealthy households in low-density areas were more likely to purchase pickups.)

Table 3 clearly also suggests that younger drivers are extremely involved in crashes (relative to their driving distances). This relative distinction between age groups (computed as the ratio of rates) is most acute for young drivers of SUVs and minivans. Motorcycle driving is a class of its own: its drivers' overall crash-rate estimates exceed those of other cohorts by four to fifteen times. While limited sample sizes make this cohort's estimates relatively variable, the trends are not inconsistent with expectations.

It should be noted that Table 2's counts and Table 3's rates include all crash types (e.g., head-on, rear-end, and rollover). Thus, they provide a rather general estimate of crash involvement and risks. Different crash types are more serious than others, as are the actions and responses of different driver types and their vehicles. The following examinations and their associated tables make such distinctions.

Rollover and Non-rollover Crash Rates

A common perception of light-duty trucks is that they are more likely to roll over. And rollovers are extremely high-risk crashes for death and severe injury (see, e.g., Kockelman and Kweon [2001]). As addressed in this paper's literature review, the static stability factors (SSFs) of LDTs are estimated to average 16.5 to 22.4 percent lower than those of passenger cars (NHTSA 2000). And SSFs tend to relate negatively to rollover risk. For these reasons, rollover and non-rollover crash rates are examined here, separately.

Table 4 provides estimates of rollover crash rates. As expected, SUV and pickup drivers are more likely to experience a rollover crash than their passenger-car- and minivan-driving counterparts. As one would also expect (given Table 2's results), younger drivers are the most involved in such crashes, and the magnitudes of difference are heightened – relative to their over-involvement in crashes of all types (shown in Table 3). Their rollover crash rates are over six times those of middle-aged drivers; in contrast, when considering all crash types, this difference was less than a factor of four. SUV driving offers the highest rollover rates for young males, but pickup driving offers the highest rates for young females.

While crash rates are an excellent indicator of crash involvement, they do not provide a very strong appreciation of crash severity. The probabilities of driver injury and death in such crashes are examined now, providing a better sense of crash "costs" to drivers.

Overall Injury Severity Probabilities

When assessing the driving risks of different drivers and vehicles, one may be most interested in the severity of the crash. If a type of vehicle performs very poorly in crashes, it may not matter much if it is not highly crash-involved (per mile driven); many people will not care to buy it and regulators may choose to restrict its sale. If a particular driver type survives crashes very well, such drivers are less likely to worry about having higher crash involvement rates. The provision of probability estimates for different drivers and vehicle types offers useful comparisons in this regard; such values encapsulate a variety of factors, giving a more comprehensive value to crash risks.

The results in this section are based on the crash rates estimated above and the driver-injury-severity models estimated by Kockelman and Kweon (2001) using ordered-probit models to predict four crash-severity levels – (i.e., no injury to driver, not-severe injury, severe injury, and fatality).[4]

Crash involvement rates and severe-injury and fatal-crash probabilities for each cohort are estimated assuming 50,000 miles of driving. This amount of driving represents on the order of five years driving for many drivers, but many more years driving for most young and older drivers. The calculations also assume that crash counts for individual drivers follow a Poisson distribution, with rates equal to those shown in Tables 3 and 4. The Poisson distribution can arise from a memoryless property of durations between crashes; such an assumption effectively implies that a driver's crash risk is uniform at all times. The Poisson distribution is quite common for crash-rate analysis (Evans 1991; Hauer 1997). However, it does imply that the variance of crash-involvement counts (from the sum of independent Poisson variables) equals their mean (within a specific cohort, as applied here). Since data on specific drivers' crash histories are not generally available (and are not in the GES data set), this assumption is not tested here.

The overall crash injury severity probability for each cohort is obtained by multiplying total Poisson-based crash involvement probabilities with the probabilities of various injury severity levels. All probabilities are calculated for the case of rollover and non-rollover crash types across 24 cohorts. Motorcycle (MC) crash probabilities are excluded here, since these were not estimated separately in Kockelman and Kweon's models (2001). However, it is expected that severe injury and fatal crash probabilities would be very high for this class of driver.

Overall severity crash rates are provided in Table 5 through 12, based on the following equation:

$$Rate_{ic} = TotalRate_i \times \Pr\left(CrashType\ c \mid Crash\right)_i \tag{1}$$

where i is an age-gender-vehicle cohort and c is the crash type. The probability of injury severity conditioned on a crash occurred is calculated using the ordered-probit model results and its formula is shown in Eq. (3). The figures provide crash rates for four different injury severity – no injury, non-severe injury, severe injury, and fatal injury – in rollover and non-rollover cases.

To acknowledge the fact that drivers may be involved in more than one crash of certain types, probabilities include the possibility of two and three crash events of each type, rather than just the probability of exactly one event. The following set of equations illustrate the type of equations used to calculate overall injury probabilities of the driver cohorts:

$$P_i^*(x;\lambda_i) = \frac{e^{-\lambda_i} \cdot \lambda_i^x}{x!} \tag{2}$$

where x is the number of crashes experienced, and λ_i is the rate of crash involvement (for every 50,000 miles driven). Eq. (2) assumes a Poisson process for crash involvement.

Once a crash has occurred, the ordered-probit probabilities rely on the following set of equations:

$$\begin{aligned}
P_i^{**}(0) &= \Pr(\text{No Injury} \mid \text{A crash occurred}) \\
&= \Pr(T_i = 0) = \Pr(T_i^* \le \psi_0) = \Pr(X_i\beta + \varepsilon_i \le \psi_0) \\
&= \Pr(\varepsilon_i \le \psi_1 - X_i\beta) = \Phi(\psi_1 - X_i\beta) \\
P_i^{**}(1) &= \Pr(\text{ Not Severe Injury} \mid \text{A crash occurred}) \\
&= \Pr(T_i = 1) = \Phi(\psi_1 - X_i\beta) - \Phi(\psi_0 - X_i\beta) \\
P_i^{**}(2) &= \Pr(\text{ Severe Injury} \mid \text{A crash occurred}) \\
&= \Phi(\psi_2 - X_i\beta) - \Phi(\psi_1 - X_i\beta) \\
P_i^{**}(3) &= \Pr(\text{ Fatal Injury} \mid \text{A crash occurred}) \\
&= 1 - \Phi(\psi_2 - X_i\beta)
\end{aligned} \tag{3}$$

where T_i is the observed, discrete injury level for driver/vehicle cohort i, T_i^* is the latent injury severity level, and ψ_n is an ordered probit's (latent) threshold between injury severity levels n and n+1. (For example, ψ_2 indicates a shift from severe injury to fatality.)

Drivers can experience more than one crash of a given type (except fatal crashes, *sans* reincarnation). Computation of overall rates of crash involvement that a driver experiences on average during a 50,000 miles driving period are relatively simple, as illustrated by Eq. (1). For probabilities, however, various equations combining multinomial outcomes with Poisson rates must be applied.

The overall injury severity probabilities are presented for severe injury and fatal crashes in more than one crash cases. In order to permit the possibility of different levels of injury severities in each crash, the multinomial (MN) probability conditioned on Poisson is used. The MN probability is associated with combo occurrences and can be written as:

$$MN\ probability = \frac{X!}{\prod_{j \in C} X_j!} \times \prod_{j \in C} p_j^{X_j} \tag{4}$$

where j is the crash type (such as one causing severe driver injury), C is a set of incidences of different types, X is the total number of occurrences experienced by the driver during that driving period (of 50,000 miles), X_j is the number of crash-type j occurrences, and p_j is the probability of crash type j's occurrence. As an example of this approach, if a driver experiences a total of three crashes, where one is severe and two are no injury, the probability of this example is the following:

$$MN\ probability\,(\text{1 is severe \& 2 involve no injury}) = \frac{3!}{1! \cdot 2!} \times p_{severe}^{1} \times p_{no\ injury}^{2} \tag{5}$$

The probability that a driver sustains a severe injury during the driving period of interest is the summation of all non-fatal crash combinations in a series of crashes. This probability is presented in Table 13. The combinations include: one crash with severe injury; two crashes, one with severe injury and the other with no or not-severe injury; two crashes, both with severe injury; three crashes, with severe injury in one and less-then-severe injury in the other two crashes; three crashes, with severe injury in two and less-than-severe in the other; three crashes, with severe injury in all three crashes; and so on. The resulting probability can be expressed as the following:

$$
\begin{aligned}
&\text{Prob(Crash at least once \& at least 1 crash is severe but none is worse)}_i \\
&= p_i^*(1) \times \frac{1!}{1!} p_i^{**}(3)^1 + p_i^*(2) \times \frac{2!}{1! \cdot 1!} p_i^{**}(3)^1 [1 - p_i^{**}(3) - p_i^{**}(4)] \\
&+ p_i^*(2) \times \frac{2!}{2!} p_i^{**}(3)^2 + p_i^*(3) \times \frac{3!}{1! \cdot 2!} p_i^{**}(3)^1 [1 - p_i^{**}(3) - p_i^{**}(4)]^2 \\
&+ p_i^*(3) \times \frac{3!}{2! \cdot 1!} p_i^{**}(3)^2 [1 - p_i^{**}(3) - p_i^{**}(4)]^1 + p_i^*(3) \times \frac{3!}{3!} p_i^{**}(3)^3 + \ldots
\end{aligned}
\tag{6}
$$

A geometric distribution, rather than a multinomial, is applied in the case of a driver's experiencing a fatal crash since this type of crash must occur last in any sequence of possible crashes. These probabilities are provided in Table 14. The fatal injury severity probability during any period (e.g., the driver's lifetime) can be written as:

$$
\begin{aligned}
&\text{Prob(Crashes \& only the last one is fatal)}_i \\
&= p_i^*(1) \cdot p_i^{**}(4) + p_i^*(2) \cdot p_i^{**}(4)[1 - p_i^{**}(4)] + p_i^*(3) \cdot p_i^{**}(4)[1 - p_i^{**}(4)]^2 + \ldots \\
&= \sum_{x=1} \left[p_i^*(x) \cdot p_i^{**}(4) \left\{1 - p_i^{**}(4)\right\}^{x-1} \right]
\end{aligned}
\tag{7}
$$

Thus, the above set of equations brings together crash involvement and crash severity probabilities.

Tables 5 and 6 provide estimates of rollover and non-rollover crash rates where the driver is not injured. These suggest that non-rollover, non-injury driver experiences are roughly 134 times more probable (using the average ratio of probabilities) than rollover, non-injury experiences. For purposes of comparison (and using results in subsequent tables), the average ratio of rates drops to 40, 15, and 6 when non-severe injury, severe injury, and fatal crashes are considered. Clearly, non-rollover crashes are much more likely than rollover crashes, but the probability of severe injury and death in a rollover crash is much higher.

As anticipated, Tables 5 and 6 confirm that young drivers are at relatively high risk for crashes. Females are estimated to be at lower risk for these two types of non-injury crash than males, for most vehicle types (and assuming 50,000 miles of driving). Females are at greatest risk for non-injury rollover crashes when driving SUVs and pickups, and for non-rollovers when driving passenger cars. In general, a female's probability of experiencing a non-injury rollover crash is less than that of males, but not when driving a pickup or minivan.

While crashes of all types are costly and emotionally, if not physically, painful, policy makers and drivers maybe often most interested in the probabilities of crashing and sustaining some sort of injury. Tables 7 and 8 present the rate estimates for non-severe injury crashes for drivers (per 50,000 miles driven); Tables 9 and 10 correspond to severe driver injuries, and Tables 11 and 12 present driver death rates. In all cases non-rollover crashes are more likely than rollover crashes, and, in most cases, this difference is by an order of magnitude. Yet the probabilities of such crashes are generally low. For example, a middle-aged woman exhibits a deadly crash rate of 0.048, per 50,000 miles of driving a passenger car. This is even lower if she is driving a light-duty truck. Of course, if she often also is a passenger, her risk of injury while traveling will rise (per time period).

According to Table 7, rollover crashes are more prevalent when driving an SUV than a passenger car. Female drivers are more likely to receive non-severe injuries than males in a pickup or minivan. In case of a rollover crash, female drivers are more apt to experience this type of crash in an SUV by all ages, and a passenger car by mid-aged or old drivers.

Table 8 suggests that passenger car drivers experience non-rollover crashes and sustain a non-severe injury more often than other vehicle drivers, except for young female (for her, a pickup is the most). And females are more likely to experience this type of crash than males in mid- or old age. This result is in general contrast to the results of Table 6, where men were at greater risk (of non-injury crashes). It may be that women are more likely to report to police officers that they have sustained a (non-severe) injury, or it may be that they injure more easily. It may also be that the types of crash in which they are involved differ; these details are not present in aggregate data like those presented here.

Tables 9 and 10 describe risks for sustaining severe injuries during 50,000 miles of driving. Young and older men appear to be at lower risk for this than women are except while driving a passenger car. It may be that females are less often wearing seatbelts or more likely to sustain a severe injury, for the same intensity of crash. However, in general, females are somewhat more crash-involved, for the same travel distance; their crash rates are higher than those of males in 11 of the 13 available cohort comparisons shown in Table 3.

To get a sense of the magnitudes of risk, one may observe that the severe-injury crash rate that a middle-aged woman driving a passenger car 50,000 miles will experience is about 0.613. Fortunately, the rates for death are roughly an order of magnitude lower, particularly for non-rollover crashes, as shown in Tables 11 and 12.

The differences in rates between rollover and non-rollover are reduced as the severity level is increased, since rollovers are very severe crashes (see, e.g., Kockelman and Kweon

[2001]). Unfortunately, young drivers are at relatively high risk of rolling over (Table 4). The resulting high risk to young drivers is an unfortunate reality.

Tables 11 and 12 suggest that female drivers (of all ages), young males, and older male drivers are more likely to die from a rollover while driving an SUV than a passenger car. But passenger car drivers are predicted to be at higher risk of death in non-rollover crashes than those in other vehicles. Summing the two tables' rates, the most death-prone driver cohort, for 50,000 miles of driving, involves young women driving SUVs, with a rate of 0.316 (0.190+0.126). The next more likely cohorts are young women in passenger cars (at 0.276) and in SUVs (0.252).

Why are women are at somewhat greater risk of dying (for the same distance driven)? As suggested before, it may be that women injure more easily for the same intensity crash; their somewhat higher crash involvements (as shown in Table 3) do not justify the increases in rates found in Tables 11 and 12, which typically are often on the order of two or three times more. It also may be that less experience driving (e.g., young and middle-aged men drive roughly 50% more than women, as shown in Table 1) translates to poorer response under crash circumstances. Higher death rates may also mean that such drivers pay less attention to driving or understand vehicle operations less. Unfortunately, the GES and NPTS data cannot really address these issues. Driving simulators or other data sets are necessary.

As previously mentioned, probabilities are provided for severe and fatal injury cases. Table 13 offers estimates of the probability that drivers receive at least one severe injury during the course of 50,000 miles of driving, and Table 14 provides estimates of the probability that drivers die during this period (in the last crash, of course). Both cases involve crashes of different severity, and all levels need to be added to get the probability, which may include an infinite number of crashes. Recognizing that the probability of experiencing five or more crashes (in the course of 50,000 miles) is negligible (for the average driver), the probabilities for multiple crashes in Tables 13 and 14 only have been computed for up to four crashes (and all their valid combinations).

Table 13 indicates that young drivers are at higher risk of receiving severe injuries than old or middle-aged drivers. Female drivers are more likely to suffer severe injury in crashes than are males (when driving all vehicle types except for passenger cars). Young and middle-aged males are more crash-prone when driving passenger cars than are their female counterparts. According to these results, a pickup places young female drivers at higher risk than a passenger car, and a minivan is the safest vehicle for a young driver. However, it's relatively rare for youths to drive minivan. However, neither of these two driver-vehicle combinations is very likely, in practice.

Table 14 provides estimates of the probability that drivers will die in the last of a series of (one or more) crashes that they may experience. Female drivers are at greater risk than males in SUVs, pickups, and minivans; but young and middle-aged males are more vulnerable in passenger cars. Note that these probabilities obscure the nature of the crash; rollover crashes are often fatal, and SUVs and pickups are more likely to roll than passenger cars.

CONCLUSIONS

Risk assessment of road users is an important area for investigation. The results of such work permit peoples' assessments of their own and others' driving safety. They are likely to impact vehicle ownership choices and driving behaviors. And they should be present in state and federal discussions of driving regulation – of both drivers and their vehicles. The work presented here illuminates many risk patterns across driver gender, driver age group, and vehicle type. Here, the computation of crash rates – through pairing of NPTS exposure data with GES crash data – offered valuable information. And the separate considerations of rollover and non-rollover crashes, as well as crash severity (as related to driver injury), provided additional conclusions.

These results suggest that policy makers may find it best to limit the driving of young persons, through raising the legal driving age, applying driving curfews, prohibiting freeway driving, and/or restricting such drivers to passenger cars. They also indicated that women are at greater risk of death from crashes than are men, for every mile driven; this result suggests that more attention to female driver education and/or female physical response under crash conditions. And automobile manufacturers may want to improve vehicle design features for higher-risk drivers. In contrast, older drivers appear to face relatively low risk of crash, for every mile driven. This may be due to personal compensation mechanisms, including use of slower speeds, avoidance of high-speed roadways, and/or avoidance of night-time driving conditions.

The results also highlight the need to control for driver exposure. Without pairing the GES crash data to exposure data, one might conclude that male drivers are over-involved in crashes and young drivers are under-involved. In reality, the differences are much lower between genders, but severe across age groups. The differences across vehicle types are also striking; light-duty trucks are more often involved in rollovers than are passenger cars, particularly for younger drivers, older drivers, and female drivers. However, they are substantially less involved in other crash types (except, in many cases, when driven by younger drivers).

As much as this work was able to conclude, important questions remain unanswered. For example, do driving conditions differ across vehicle and driver categories? Do some drivers drive more miles under dangerous (e.g., high-speed, rainy, or dark) conditions than others? And are high-risk drivers driving differently because they should (obscuring vehicle-related and other risk distinctions)? More extensive surveys of driver behavior and crash history would illuminate such information, providing more equitable comparisons and stronger conclusions. If, for example, middle-aged SUV drivers are highly risk averse and/or drive under uncongested but slow-speed conditions, their relatively low crash rates may not make for fair comparisons with the more common cohort of middle-aged passenger-car drivers. We hope that future work will illuminate any distinctions that exist.

ACKNOWLEDGEMENTS

The authors appreciate the administrative support of Annette Perrone and the financial support of the Luce Foundation and the University of Texas's Department of Civil Engineering.

ENDNOTES

[1] The 2001 National Household Travel Survey (NHTS), which is the combined data set of the NPTS and American Travel Survey (ATS), will be available in late 2002.

[2] The NPTS survey question of basic vehicle "type" does not discriminate between minivans and cargo vans. Thus, vehicle make and model information had to be matched to minivan codes to ensure consistency between the crash and use data sets.

[3] The "special features" enabling off-road use are four-wheel drive and at least four of the following five clearance characteristics: an approach angle of not less than 28 degrees, a breakover angle of not less than 14 degrees; a departure angle of not less than 20 degrees, a running clearance of not less than 8 inches, and front and rear axle clearances of not less than 7 inches each. (CFR 40CFR86.084-2)

[4] Kockelman and Kweon's (2001) injury severity models were obtained using the then most recently available crash data, from the 1998 GES. Vehicles and crash characteristics may have changed somewhat, between 1995 and 1998, affecting the parameter estimates of the severity models, but it makes good sense to use the more recent estimates since these are more applicable to today's crashes.

REFERENCES

Abdel-Aty, M. A., Chen, C. L., Schott, J. R., 1998. An assessment of the effect of driver age on traffic accident involvement using log-linear models. Accident Analysis and Prevention 30 (6), 851-861.

Chipman, M. L., MacGregor, C. G., Smiley, A. M., and Lee-Gosselin, M., 1992. Time vs. distance as measures of exposure in driving surveys. Accident Analysis and Prevention, 24 (6) 679-684.

Davis, G. A., Gao, Y., 1993. Statistical methods to support induced exposure analyses of traffic accident data. Transportation Research Record 1401, 43-49.

DeYoung, D. J., Peck, R. C., Helander, C. J., 1997. Estimating the exposure and fatal crash rates of suspended/revoked and unlicensed drivers in California. Accident Analysis and Prevention 29 (1), 17-23.

Doherty, S. T., Andrey, J. C., MacGregor, C., 1998. The situational risks of young drivers: The influence of passengers, time of day and day of week on accident rates. Accident Analysis and Prevention 30 (1), 45-52.

Evans, L., 1991. Traffic Safety and the Driver. Van Nostrand and Reinhold.

Federal Highway Administration, 1997. User's guide for the public use data files: 1995 nationwide personal travel survey. Publication# FHWA-PL-98-002.

Hauer. E., 1997. Observational Before-After Studies In Road Safety. Oxford, Pergamon.

Hauer, E., 1995. On exposure and accident rate. Traffic Engineering and Control 3 (3), 134-138.

Hu, P. S., Trumble, D. A., Foley, D. J., Eberhard, J. W., Wallace, R. B., 1998. Crash risks of older drivers: A panel data analysis. Accident Analysis and Prevention 30 (5), 569-581.

Kockelman, K. M., Kweon, Y.-J., 2001. Driver injury severity: An application of ordered probit models. Forthcoming in Accident Analysis and Prevention.

Kockelman, K. M., Zhao, Y., 2001. Behavioral Distinctions: The use of light-duty trucks and passenger cars. Journal of Transportation and Statistics 3 (3), 47-60.

Lourens, P. F., Vissers, J. A. M..M., Jessurun, M., 1999. Annual mileage, driving violations, and accident involvement in relation to drivers' sex, age, and level of education. Accident Analysis and Prevention 31 (5), 593-597.

Lyles, R., Stamatiades, N., Lighthizer, D., 1991. Quasi-induced exposure revisited. Accident Analysis and Prevention 23, 275-285.

National Highway Traffic Safety Administration, 2000. Consumer Information Regulations; Federal Motor Vehicle Safety Standards; Rollover Prevention. Docket# NHTSA-2000-6859.

National Highway Traffic Safety Administration (No date) NASS GES Analytical User's Manual 1988 – 1999 [Online]. Available: ftp://www.nhtsa.dot.gov/ges/Ges_Doc/ [2001, June 25].

Ryan, G. A., Legge, M., Rosman, D., 1998. Age related changes in drivers' crash risk and crash type. Accident Analysis and Prevention 30 (3), 379-387.

Table 1. Annual Vehicle Miles Driven (10^6 miles, 1995)

Age	Gender	PC	SUV	PU	MVAN	MC	Overall
Young	Male	31,506	3,467	11,422	1,034	97	48,426
Mid		504,795	90,401	230,551	60,032	2,133	917,284
Old		102,513	7,786	28,895	8,243	12	152,917
Young	Female	29,370	1,204	1,650	681	NA	33,119
Mid		465,469	49,621	40,631	59,370	53	627,506
Old		72,278	1,258	2,253	1,839	NA	78,677
Overall		1,205,931	153,736	315,401	131,199	2,294	1,857,930

Young = Age < 20, Mid = Ages from 20 to 60, and Old = Ages > 60 years. PC = Passenger car, SUV = Sport utility vehicle, PU = Pickup truck, MVAN = Minivan, and MC = Motorcycle.

Table 2. Annual Crash Counts (All Crash Types, 1995)

Age	Gender	PC	SUV	PU	MVAN	MC	Total
Young	Male	719,046	49,404	193,547	12,876	9,431	1,055,020
Mid		3,165,487	251,589	1,183,520	95,469	49,875	5,502,690
Old		504,866	15,810	128,722	13,296	3,173	724,237
Young	Female	569,590	25,489	44,597	9,200	605	670,192
Mid		2,846,540	147,951	211,846	111,731	2,540	3,509,659
Old		405,649	4,413	10,373	4,885	24	434,257
Total		8,211,178	494,656	1,772,605	247,458	65,648	11,896,055

Table 3. Crash Rates of Drivers (per 10^6 miles driven, 1995)

Age	Gender	PC	SUV	PU	MVAN	MC	Overall
Young	Male	22.82	14.25	16.95	12.46	97.66	21.79
Mid		6.27	2.78	5.13	1.59	23.39	6.00
Old		4.92	2.03	4.45	1.61	260.06	4.74
Young	Female	19.39	21.18	27.04	13.51	NA	20.24
Mid		6.12	2.98	5.21	1.88	48.38	5.59
Old		5.61	3.51	4.60	2.66	NA	5.52
Overall		6.81	3.22	5.62	1.89	28.62	6.40

Table 4. Rollover Crash Rates of Drivers (per 10^6 mile driven, 1995)

Age	Gender	PC	SUV	PU	MVAN	MC	Overall
Young	Male	0.661	1.449	1.028	0.389	72.2	0.987
Mid		0.105	0.124	0.157	0.024	22.5	0.184
Old		0.029	0.049	0.062	0.028	213.2	0.059
Young	Female	0.516	1.322	1.769	0.577	NA	0.627
Mid		0.068	0.111	0.224	0.037	39.3	0.084
Old		0.026	0.052	0.087	NA	NA	0.030
Total		0.104	0.155	0.196	0.036	26.2	0.162

Table 5. Rate of Rolling Over and Sustaining No Injury (per 50,000 miles driven)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	0.1657	0.4401	0.3170	0.1131
Mid		0.0257	0.0369	0.0475	0.0069
Old		0.0073	0.0149	0.0190	0.0082
Young	Female	0.0944	0.3011	0.4103	0.1249
Mid		0.0120	0.0248	0.0506	0.0078
Old		0.0047	0.0118	0.0201	NA

Table 6. Rate of Non-Rollover Crash Involvement and Sustaining No Injury (per 50,000 miles driven)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	13.62	8.62	10.80	7.96
Mid		3.74	1.77	3.34	1.02
Old		3.00	1.33	2.97	1.04
Young	Female	9.88	11.63	14.94	7.39
Mid		3.12	1.66	2.92	1.04
Old		2.91	2.02	2.66	1.51

Table 7. Rate of Rolling Over and Sustaining Non-Severe Injury (per 50,000 miles driven)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	0.2935	0.6439	0.4563	0.1733
Mid		0.0465	0.0551	0.0699	0.0108
Old		0.0130	0.0219	0.0275	0.0127
Young	Female	0.2202	0.5828	0.7813	0.2528
Mid		0.0287	0.0490	0.0985	0.0160
Old		0.0111	0.0229	0.0385	NA

Table 8. Rate of Non-rollover Crash Involvement and Sustaining Non-Severe Injury (per 50,000 miles driven)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	6.984	3.522	4.325	3.438
Mid		1.970	0.744	1.375	0.453
Old		1.549	0.548	1.200	0.454
Young	Female	6.971	6.615	8.336	4.426
Mid		2.256	0.969	1.668	0.639
Old		2.069	1.156	1.496	0.912

Table 9. Rate of Rolling Over and Sustaining Severe Injuries (per 50,000 miles driven)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	0.1597	0.2985	0.2087	0.0835
Mid		0.0257	0.0260	0.0325	0.0053
Old		0.0071	0.0102	0.0127	0.0061
Young	Female	0.1513	0.3415	0.4517	0.1540
Mid		0.0201	0.0292	0.0580	0.0099
Old		0.0077	0.0135	0.0224	NA

Table 10. Rate of Non-Rollover Crash Involvement and Sustaining Severe Injuries (per 50,000 miles driven)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	1.4172	0.6058	0.7334	0.6153
Mid		0.4074	0.1304	0.2376	0.0827
Old		0.3162	0.0948	0.2046	0.0816
Young	Female	1.7996	1.4494	1.8012	1.0089
Mid		0.5934	0.2162	0.3672	0.1485
Old		0.5372	0.2547	0.3251	0.2091

Table 11. Rate of Rolling Over and Sustaining a Fatal Injury (per 50,000 miles driven)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	0.0418	0.0664	0.0458	0.0193
Mid		0.0069	0.0059	0.0073	0.0013
Old		0.0019	0.0023	0.0028	0.0014
Young	Female	0.0503	0.0964	0.1258	0.0452
Mid		0.0068	0.0084	0.0164	0.0030
Old		0.0026	0.0038	0.0063	NA

Table 12. Rate of Non-Rollover Crash Involvement and Sustaining a Fatal Injury (per 50,000 miles driven)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	0.1410	0.0515	0.0616	0.0544
Mid		0.0413	0.0113	0.0203	0.0074
Old		0.0316	0.0081	0.0173	0.0073
Young	Female	0.2253	0.1551	0.1902	0.1121
Mid		0.0756	0.0236	0.0395	0.0168
Old		0.0676	0.0274	0.0345	0.0234

Table 13. Probability of Sustaining at Least One Severe Injury (in One or More Crashes, while driving 50,000 miles)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	7.144E-02	3.954E-02	4.224E-02	3.288E-02
Mid		2.105E-02	7.380E-03	1.288E-02	4.558E-03
Old		1.591E-02	5.068E-03	1.059E-02	4.530E-03
Young	Female	6.154E-02	5.336E-02	6.462E-02	3.832E-02
Mid		2.034E-02	7.801E-03	1.338E-02	5.323E-03
Old		1.807E-02	8.584E-03	1.104E-02	8.056E-03

Table 14. Probability of Sustaining a Fatal Injury (while driving 50,000 miles)

Age	Gender	PC	SUV	PU	MVAN
Young	Male	4.617E-03	2.719E-03	2.612E-03	2.273E-03
Mid		1.881E-03	6.249E-04	1.007E-03	4.128E-04
Old		1.431E-03	4.230E-04	8.164E-04	4.057E-04
Young	Female	4.206E-03	3.105E-03	3.382E-03	2.705E-03
Mid		1.814E-03	6.531E-04	1.056E-03	4.756E-04
Old		1.599E-03	6.862E-04	8.529E-04	7.346E-04