United States General Accounting Office
GAO

Report to Congressional Requesters

October 1994

HIGHWAY SAFETY

Factors Affecting Involvement in Vehicle Crashes

GAO/PEMD-95-3

GAO

United States
General Accounting Office
Washington, D.C. 20548

Program Evaluation and
Methodology Division

B-256555

October 27, 1994

The Honorable Ernest F. Hollings
Chairman, Committee on Commerce, Science, and Transportation
United States Senate

The Honorable Richard H. Bryan
Chairman, Subcommittee on Consumer
Committee on Commerce, Science, and Transportation
United States Senate

In our October 1991 report to you entitled Highway Safety: Have Automobile Weight Reductions Increased Highway Fatalities? (GAO/PEMD-92-1), we presented a number of findings regarding the relationship between car weight and safety. Among other things, we found the danger cited by some researchers and agency officials—that the increase in lighter cars on the highways since the 1970's would result in dramatically higher highway death tolls—to be overstated and to have excluded consideration of some important factors. One of these factors was the lowered threat, from reductions in both weight and force of impact, posed to other drivers on the road in multiple vehicle collisions.

At the time, we also reported that the safety effects of weight change or any other automotive design factor could be confounded by many other factors, chief among them driver attributes. For example, we discussed in qualitative terms how a driver's age could interact in different ways with car size attributes. If it is true that younger drivers drive smaller cars and also tend to drive more recklessly, then attributing the higher injury rates in smaller cars simply to car size or weight might be misleading. However, if it is true that elderly drivers drive larger cars and, if involved in a crash, are more likely to be injured than younger drivers, then larger cars may appear to be less safe than they really are.

Objectives, Scope, and Methodology

You have requested that we investigate these relationships more comprehensively and that we set the discussion of car size and safety into the larger context of the relative contributions to highway safety of driver attributes, vehicle characteristics, and their multiple interactions. Our response to your request involved an investigation into two distinct, and sometimes highly divergent, aspects of highway safety: crash involvement and crashworthiness. The study of crash involvement focuses attention on the factors likely to produce a crash. Crashworthiness, instead, examines the factors likely to produce serious injury, once a crash has occurred. The present report deals with crash involvement—that is, with the driver or vehicle characteristics that are related to the likelihood of a crash. The attributes we examined included driver age, gender and driving history, vehicle age and size (weight, wheelbase, and engine displacement). A companion report will examine crashworthiness: the factors that affect the likelihood of serious injury once a crash has taken place. A third report will examine the relationship between automobile crashworthiness and crash testing performed by the Department of Transportation.

In the present analysis, we have used a method known as "induced exposure" to estimate the likelihood of crash involvement. This approach assumes that not-at-fault drivers in two-vehicle accidents represent a random selection of drivers and vehicles on the road. The ratio of at-fault to not-at-fault drivers provides a measure of the relative involvement of drivers and vehicles in accident causation. We used a data base containing 340,000 records, with details on accidents reported in North Carolina in 1990, to produce ratios of at-fault to not-at-fault North Carolina drivers. Since these findings are based on data from only one state, they cannot be generalized to the nation. However, we did compare the North Carolina ratios to ratios we obtained from a Michigan data base and found the figures to be close and the trends quite similar. This finding is consistent with the logic of induced exposure—concerned with ratios of driver and vehicle characteristics rather than their absolute numbers—and suggests that the method may produce results that have more general applicability. (See appendix I for a discussion of the induced exposure approach and appendix II for descriptive statistics from North Carolina.)

Results in Brief

We found that, when other factors are controlled for, driver characteristics far outweigh vehicle factors in predicting crash involvement for passenger cars. For example, the odds of a 20-year-old driver being involved in a single-vehicle, nonrollover crash was over 4 times as great as that of a 50-year-old. By comparison, a 4,000-pound car was only 1.06 times as likely to be involved in this type of crash as a 2,000-pound car. Similarly, drivers with a history of previous traffic violations were more likely to be in a crash, and men were more likely to be in single-vehicle crashes than women. (Appendix III contains more detailed results of our passenger car analyses.)

A car's weight had little effect on the likelihood of a two-vehicle crash or a single-vehicle crash that did not involve a rollover. However, light cars were as much as three times as likely to be involved in single-vehicle rollover crashes as heavy cars. In other types of crashes, we found that car-size measures other than weight (wheelbase or engine size) were better predictors of crash involvement.

We found similar results when we applied our methodology to crashes involving light trucks and vans. A driver's age and violation history significantly affected the likelihood of crash involvement for these types of vehicles, as did vehicle age. In our analysis, however, driver gender did not contribute significantly to the prediction of light truck and van crashes in general (although it did in certain subcategories of these crashes). The vehicle weight of the light trucks or vans was only a marginally significant predictor. (Appendix IV contains more detailed findings.)

Our Analysis

Any investigation of crash involvement must include more than counts of units (vehicles or drivers). In order to calculate the relative odds of being in a serious crash, it is necessary (but not sufficient) to compute, for example, how many 1989 Ford Tauruses or how many 16-year-old males are involved in serious crashes in a given time period. Without knowing how many Tauruses or 16-year-old male drivers are on the road, we cannot conclude whether these cars or these drivers are more or less likely than other cars or drivers to be involved in crashes. We must, in other words, know their exposure to crashes. For example, consider that it is generally well known that, in absolute terms, elderly drivers are involved in fewer serious crashes than younger drivers. But they also drive fewer miles, and under less hazardous conditions, than younger drivers. In absolute terms, therefore, elderly drivers pose a rather small highway safety problem. When their relative exposure is considered, however, it turns out that, for the miles they drive, elderly persons are disproportionately involved in collisions, particularly two-vehicle collisions.

Crash exposure can be estimated in a number of ways. Vehicle exposure in a given year is frequently measured by the number of vehicles registered. Thus, in our previous report, we tracked the number of fatalities per 100,000 registered vehicles for different weight classes of cars. Driver exposure can also be represented by a single count of the number of licensed drivers in various categories (for instance, age groups or geographic regions).

Such direct measures of exposure have serious limitations, however. While we may know how many vehicles of a certain type are registered, we do not know how many miles (if any) and under what conditions they are driven or by whom they are driven. If large cars are driven more miles, and under more dangerous conditions, an estimate of crash involvement based simply on the number of crashes per registered vehicle or even—if such data were available—on crashes per mile driven would underestimate their exposure and their safety.

For this reason, some researchers have turned to methods of estimating exposure indirectly. For example, some calculate crash rates from a crash data base as the ratio of at-fault to not-at-fault drivers of a certain type (say, young females), arguing that the not-at-fault drivers serve as a representative sample of drivers on the road—or “exposed”—under the conditions represented by the data base. This method has the practical advantage of allowing exposure estimates to be derived from the same data base as the count of crashes and, arguably, the strategic advantage of being more sensitive to the variations of driver and vehicle characteristics than is possible with direct measures (see appendix I).

For this study, we employed such an indirect or “induced exposure” method. We applied this method to the police-reported crash data base of North Carolina for 1990 that was provided to us by researchers at the University of North Carolina Highway Safety Research Center.[1] This data base contains information on 183,616 crashes involving 484,258 individuals and 325,277 vehicles.[2] We supplemented the crash data base by merging with it information on the drivers’ history of previous traffic violations.

We performed separate logistic regression analyses of crash involvement corresponding to three types of crashes (two-vehicle, single-vehicle rollover, and single-vehicle nonrollover) and two types of vehicles—(1) passenger cars and (2) light trucks and vans. Sixty-six percent of the crashes in our analysis involved two vehicles, 29 percent were single-vehicle nonrollovers, and 5 percent were rollovers. (Although rollovers accounted for only a small proportion of crashes, this type of crash is second only to frontal impacts in terms of deaths and injury severity.) Sixty-eight percent of crashes involved passenger cars, 11 percent involved light trucks and vans, and 21 percent were between

---

[1] At-fault drivers were defined as the drivers in two-vehicle collisions for whom the police report indicated a violation. Collisions in which a violation was indicated for both drivers or for neither driver (approximately 10 percent of all two-vehicle accidents) were excluded from the analysis.

[2] Additional descriptive statistics on this data base are provided in appendix II. Because of missing data points, particularly on vehicle weight, as well as our restriction of the analysis to one- and two-car or light truck crashes, the effective data base for the individual analyses was substantially reduced. See appendixes III and IV.

cars and light trucks and vans. Appendix III presents the details of the analyses of passenger cars, appendix IV the light truck and van results. We present the main points here, first for passenger cars and, then, more briefly, for light trucks and vans.

Passenger Cars

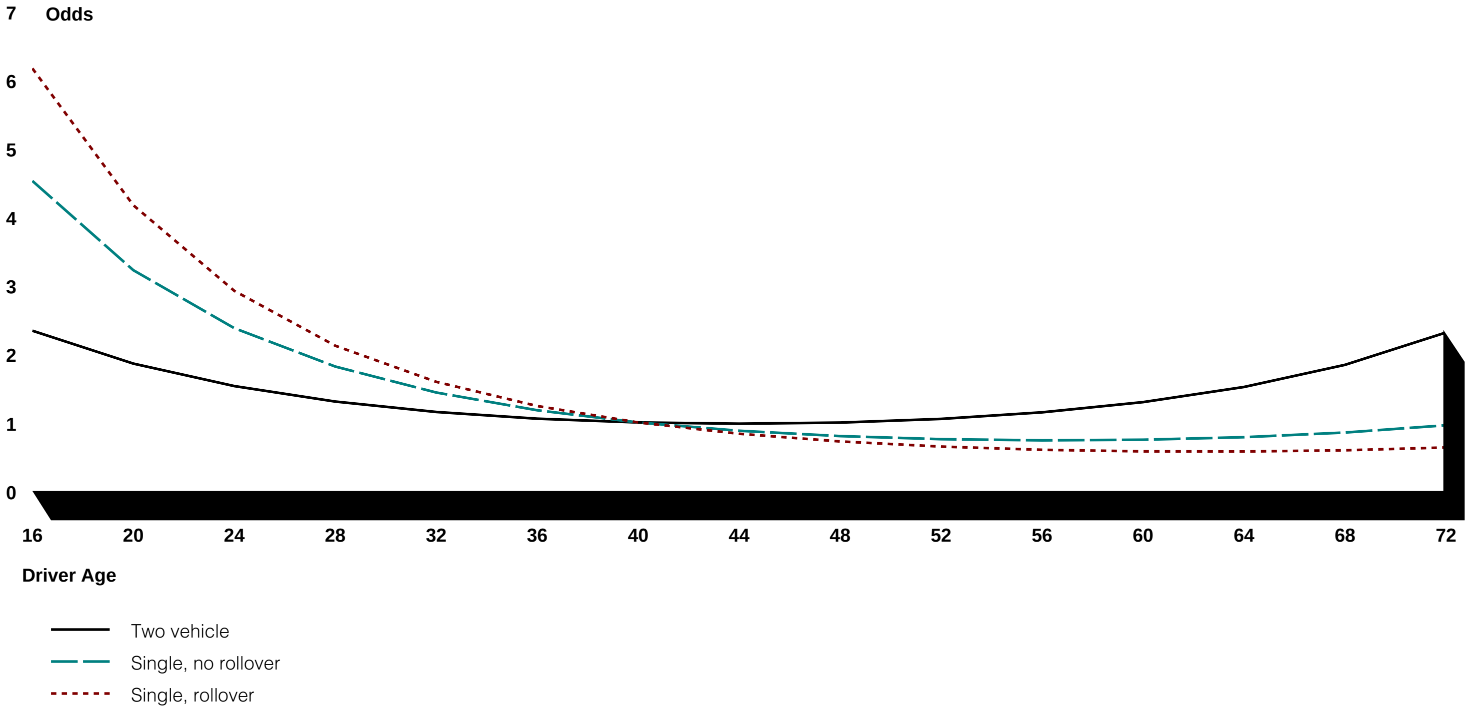
Driver Age

We found no straight-line relationship between a driver's age and crash involvement. In general, drivers under 25 were at greatest crash risk, followed by drivers over 65. The relationship was not the same for all crash types. A 16-year-old driver was over seven times more likely to be in a single-vehicle rollover crash, over five times more likely to be in a single-vehicle nonrollover crash, and more than twice as likely to be in a two-vehicle crash as was the safest driver overall—a 45-year-old.

Drivers least likely to be in a single-vehicle rollover crash were 62-year-olds. They were only one tenth as likely to be in such a crash as 16-year-olds. However, drivers in their mid-70s were about as likely as the 16-year-olds to become involved in a two-vehicle collision. As a driver's age approached 80 years, the likelihood of such involvement in a two-vehicle collision increased sharply.

This was not true, however, of single-vehicle crashes. Elderly drivers were more likely to be involved in single-vehicle nonrollovers than 40-year-olds only after age 74 and in single-vehicle rollovers only after age 86. Figure 1 summarizes the effects of age by comparing each age's odds of crash involvement in each crash type with those of a 40-year-old's.

Figure 1: Adjusted Odds Ratios Comparing Crash Involvement by Driver Age[a]



[a] The odds ratios were calculated using the coefficients from the logistic regression equations shown in appendix III, table III.2. The figures above compare the odds of crash involvement of drivers of different ages to a 40-year-old driver, assuming all other factors included in the equation are equal. The odds ratios tell how much more (or less) likely the outcome is in one group versus the comparison group of 40-year-old drivers. An odds ratio of 1.0 means that there is no difference between two groups in their odds of crash involvement. Values higher than 1.0 mean greater risk; values lower than 1.0 mean less risk.

Violation History

Driving history was a strong predictor of crash involvement for two-car and single-car crashes, ranking second only to driver age. A history of alcohol-related convictions was a particularly powerful predictor. For example, drivers with histories of nonalcohol traffic violations were only 1.15 times as likely to be involved in a single-car nonrollover crash as drivers with a “clean” history. However, drivers with a history of drunk driving were at least 3.7 times as likely as other drivers to be involved in such a crash.

For two-car crashes, driving history was also a significant but less powerful predictor. Drivers with prior alcohol violations were 2.1 times as likely to be involved in a two-vehicle collision as drivers with no prior violations and 1.6 times as likely as drivers with nonalcohol violations.

Driver Gender

As noted earlier, driver gender affected the likelihood of involvement in single-vehicle crashes only. Males were twice as likely as females to be involved in either type of single-vehicle crash. Female drivers were indistinguishable from male drivers in their likelihood of being involved in a two-car collision.

Vehicle Age

We introduced the age of vehicles into our model as a way of correcting for the possibility that we might confuse the safety effect of a vehicle size with that of its condition. As our earlier report found, passenger cars have become, on the average, much lighter than they were in the 1970's. Heavier cars, therefore, are more likely to be older cars and, presumably, to be in poorer condition. An analysis that did not control for this association would be in danger of overestimating the crash involvement of heavy cars.

Half of the cars in our data base were model year 1984 or newer, and 80 percent were built after 1978. We found that, regardless of size, newer cars were slightly less at risk for crash involvement. For example, if one car were 5 years older than another, the older car would have a risk 1.12 times that of the newer. We cannot tell, however, whether this difference stems from the deteriorated condition of the older car or the improved design of the newer car.

It should also be noted that (as the Department of Transportation (DOT) pointed out in its comments on a draft of this report) vehicle age may capture the effect of more than simply vehicle characteristics. Older cars may have more aggressive drivers and are more likely to be found in rural settings.

Car Weight

Car size can be expressed by different measures: wheelbase (the distance between the front and rear axles), track width (the distance between the left and right wheels), engine size, weight, and so on. Because all these variables tend to be very highly correlated with one another, it is frequently difficult to distinguish statistically their unique effects. It seems reasonable to believe that each of these factors has a differential effect on the likelihood of being involved in different types of crashes.[3] In one research report, for example, the National Highway Traffic Safety Administration (NHTSA) found that a combination of track width and center of gravity was the best predictor of vehicle rollover.[4]

We developed three sets of models corresponding to the three measures of car size readily available to us: wheelbase, weight, and engine displacement. The full results of these analyses are in appendix V. Here we are concerned with the relationship between car weight and crash involvement.[5] Figure 2 summarizes this relationship by comparing the odds for cars of different weights with the odds for a 2,678-pound car (the median car weight in the sample) of being in a crash (for each of the three crash types).

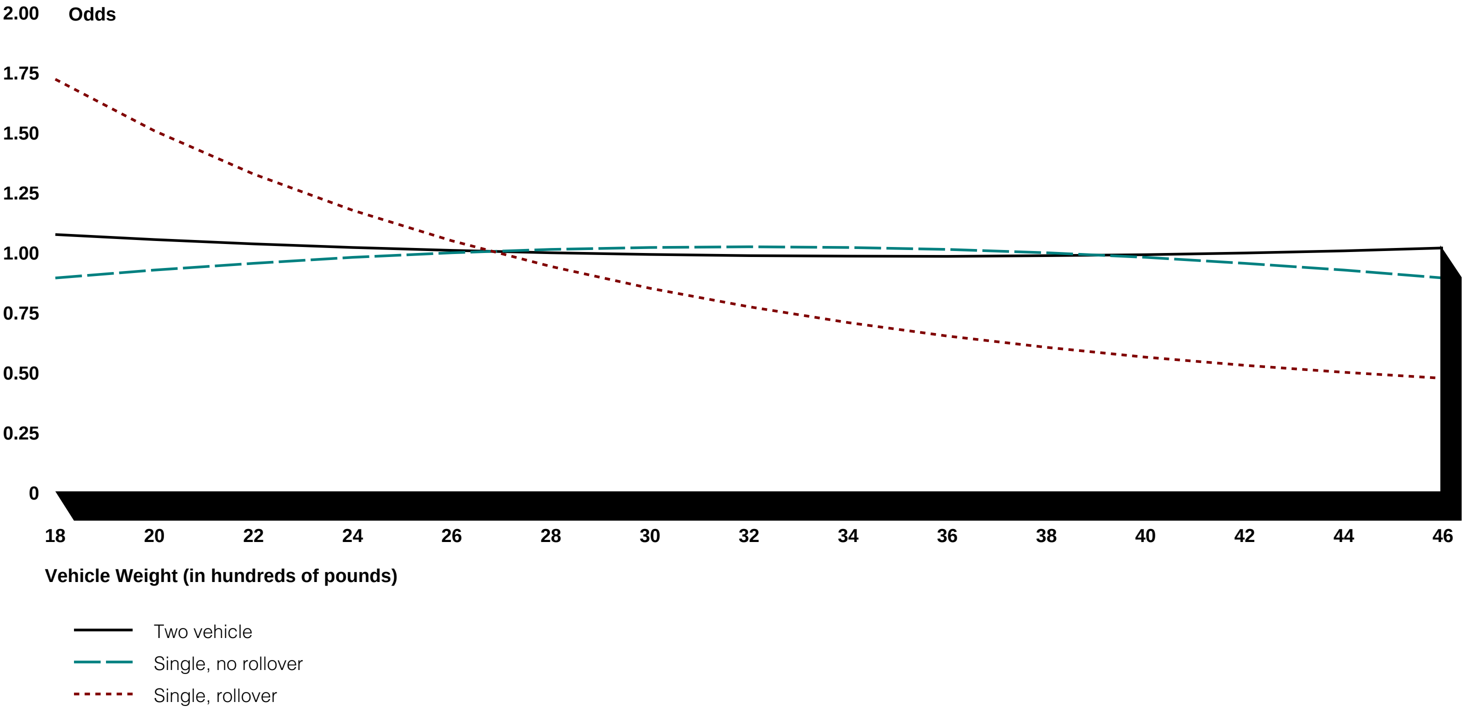
---

[3] Our separate analyses of the contributions of weight, wheelbase, and engine size lend support to this hypothesis. In predicting the likelihood of a single-vehicle nonrollover crash, engine size appeared to be most important; for single-vehicle rollovers, wheelbase was most important; and weight contributed more to the prediction of two-vehicle crashes than either of the other two. Overall, however, it is important to note that, relative to that of driver-related measures, the contribution of the vehicle size measures is substantially less (see appendix V).

[4] P. Mengert et al., Statistical Estimation of Rollover Risk, DOT-HS-807-489 (Washington, D.C.: National Highway Traffic Safety Administration, 1989).

[5] All vehicle size measures (wheelbase, weight, and engine displacement) were provided as part of the North Carolina data base and were derived from decoding vehicle identification numbers using R. L. Polk & Co.'s VINA program.

Figure 2: Adjusted Odds Ratios Comparing Crash Involvement by Car Weight[a]



[a] The odds ratios were calculated using the coefficients from the logistic regression equations shown in appendix IV, table IV.2. The figures above compare the odds of crash involvement of vehicles of different weights to a vehicle at the median weight of 2,678 pounds, assuming all other factors included in the equation are equal. The odds ratios tell how much more (or less) likely the outcome is in one group versus the comparison group of cars at the median weight. An odds ratio of 1.0 means that there is no difference between two groups in their odds of crash involvement. Values higher than 1.0 mean greater risk; values lower than 1.0 mean less risk.

For each crash type, weight had a statistically significant effect, but the effect was quite small for two-vehicle crashes and for single-vehicle crashes where a rollover did not occur. The odds ratio curves for these two crash types are almost mirror images of each other. The lightest and the heaviest cars were slightly more likely to be involved in two-vehicle crashes than were midweight cars and slightly less likely to be involved in single-vehicle nonrollover crashes.

The connection between car weight and rollover crashes, however, was substantially stronger. The lighter the car, the greater were its odds of rolling over. For example, the average 2,000-pound car was nearly three times as likely to be involved in a single-vehicle rollover crash as the average 4,500-pound car.

This finding needs some qualification. Factors other than car weight are probably more directly related to rollover propensity but, as we noted earlier, the high intercorrelation of the various measures of car size make the relationships difficult to disentangle statistically. When we used car size measures other than weight in our analyses, we found a stronger connection with rollover likelihood for wheelbase than for weight. (See appendix V.) Research by NHTSA has demonstrated that rollover propensity is related to several other vehicle factors, such as track width, weight distribution, and braking stability.[6]

Light Trucks and Vans

Our analysis of the crash involvement probability of light trucks and vans yielded many of the same findings as our analysis of passenger cars. Full details of the analysis are presented in appendix IV. Driver age and driving history remained by far the best predictors of crash involvement. Drivers involved in single-vehicle light truck crashes were one-and-a-quarter to one-and-a-third times more likely to be male. However, whereas for passenger cars driver gender appeared to be irrelevant to involvement in two-vehicle collisions, women were slightly but significantly more likely to be involved in a light truck two-vehicle collision than were men.

As with passenger cars, the vehicle factors were much less important than the driver factors. However, older light trucks were significantly more likely to be involved in all types of crashes. The relationship between light truck weight and crash involvement was weaker than for passenger cars. We found no relationship in two-vehicle crashes and only a marginally significant relationship in single-vehicle crashes. The connection between light truck weight and crash involvement was relatively strongest for rollover crashes. As with passenger cars, the lightest of these vehicles were more likely to roll over. However, all three alternative measures of size again contributed relatively little to predictions of crash involvement, and vehicle weight ranked either second or third among the size measures in all light truck models. (See appendix V.)

---

[6] We cited some of this research in earlier testimony before your committee when we reported that the greater likelihood of lighter cars to roll over could be offset by a very small increase in track width. See U.S. General Accounting Office, Automobile Weight and Safety, GAO/T-PEMD-91-2 (Washington, D.C.: April 11, 1991).

Summary

We developed models to predict the likelihood of crash involvement for passenger cars and for light trucks and distinguished between three different crash types: two-vehicle crashes, single-vehicle nonrollover crashes, and single-vehicle crashes involving a rollover. We used driver age, gender, and traffic violation history, as well as vehicle age and weight. The six models we developed, while varying somewhat from one another, provided a relatively consistent rank order of predictive importance for these factors.

Among our findings, the following five may be the most significant. First, information about the driver variables was much more important than information about the vehicle variables in any of our estimates of the relative likelihood of being in a traffic crash. Second, within the driver variables we considered, a driver's age was the strongest predictor, followed (in most models) by the driver's violation history and then by gender. Third, of the two-vehicle variables we considered, vehicle age contributed substantially more to our prediction of crash involvement than did vehicle weight. Fourth, we modified our models slightly by substituting other vehicle dimensions (wheelbase and engine size) for vehicle weight and found that, of our six specific crash type models, five were predicted better by these alternative measures than by weight.

Fifth and finally, we conclude that the induced exposure methodology we demonstrated in our analysis offers reasonable expectation of yielding results substantially more sensitive to the real world driving environment than can be achieved through currently available direct exposure methods without incurring prohibitive data collection costs.

Our work was performed in accordance with generally accepted government auditing standards.

We have provided draft copies of this report to NHTSA officials and discussed the study results with them. NHTSA also provided us with written comments on the draft report. These comments and our response are provided in appendix VI. We have incorporated their suggestions where appropriate. We plan no further distribution of this report until 30 days from the date of issue, unless you publicly announce its contents earlier. We will then send copies to the Secretary of Transportation. We will also make copies available to interested organizations, as appropriate, and to others upon request.

If you have any questions or would like additional information, please call me at (202) 512-3092. Other major contributors to this report are listed in appendix X.

Kwai-Cheung Chan
Director of Program Evaluation in
Physical Systems Areas

Contents

Tables

Figures

Abbreviations

DOT	Department of Transportation
GAO	General Accounting Office
NHTSA	National Highway Traffic Safety Administration
NPTS	National Personal Transportation Survey

Induced Exposure

The use of indirect methods to estimate the risk of being involved in a highway crash, variously referred to as “induced” or “quasi-induced” exposure, dates back at least to the 1960’s. The method is based on calculating the ratio of at-fault drivers or vehicles to not-at-fault drivers or vehicles in two-vehicle accidents contained in police accident reports. Its underlying assumption is that the not-at-fault drivers and vehicles constitute a representative sample of the drivers, vehicles, and driving conditions and their interactions for the geographical area being examined.

On the assumption that not-at-fault drivers represent the general population of drivers, the ratio of at-fault to not-at-fault drivers yields an estimate of the over- or underinvolvement of different levels of that dimension in highway crashes. R. W. Lyles et al. offer an example of estimating how much male drivers are overrepresented in interstate highway accidents.[1] In 1988, 11,335 pairs of drivers were involved in two-vehicle accidents in which fault was assigned on Michigan interstate highways. Of the at-fault drivers, 8,366 (73.8 percent) were male, whereas only 7,528 (66.4 percent) of not-at-fault drivers were male. Males were 1.1 times (73.8/66.4) overinvolved in interstate accidents relative to their presence on these highways. Females, however, represented 26.2 percent of at-fault drivers and 33.6 percent of not-at-fault drivers. Their “involvement ratio,” therefore, was 0.78 (26.2/33.6). Lyles et al. conclude, therefore, that when the calculation is adjusted for exposure, males caused interstate highway accidents at a rate 1.4 (1.1/0.78) times that of females.[2]

This indirect approach has a number of advantages. Foremost among them is the ability to define accident exposure in terms of any driver, roadway, or vehicle characteristic reported in the accident data base being used. For example, given a sufficiently large data base, a researcher could estimate the crash involvement risk of female drivers under 25 years of age on rural roads in the dark and could determine whether female drivers are more likely than males to become involved in accidents under such conditions.

---

[1] R. W. Lyles et al., “Quasi-induced Exposure Revisited,” Accident Analysis and Prevention, 23:4 (1991), 275-85.

[2] Another indirect measure of exposure has been used by Leonard Evans in examining the effects of car size. Instead of using the not-at-fault driver, Evans bases his exposure measure for single-vehicle crashes on the fatally injured pedestrian and estimates exposure for single-vehicle crashes from the ratio of driver fatalities to pedestrian fatalities. His reasoning is that pedestrian fatalities associated with a given type of vehicle or driver involved in a single vehicle accident will increase in relation to the number of vehicles or drivers of that type on the road. See Leonard Evans, Traffic Safety and the Driver (New York: Van Nostrand Reinhold, 1991).

Any attempt to measure exposure directly, at this level of detail, would be prohibitively expensive.

Two major uncertainties are associated with induced exposure measures, however. The first is common to any state or regional data base and involves whether the data being used to form estimates adequately represent other geographical areas and, hence, the universe to which they are being extrapolated. In the case of induced exposure, the concern is not that the absolute count within subcategories of drivers of vehicles may vary from state to state—that is, that there might be, for example, more light trucks or more elderly drivers in one state than another. (This is most likely the case.) Rather, we are concerned with the ratios of at-fault to not-at-fault drivers, however the absolute counts may vary geographically. It is assumed that these ratios would be less prone to substantial variation from one area to another. In other words, it is much less probable that light trucks or elderly drivers have different driving-related attributes from one state to another than that their numbers vary geographically.

Nevertheless, to test the seriousness of this concern, we compared by age and gender the accident involvement ratios we obtained from the 1990 North Carolina data base we used for our study with ratios we calculated from a data base of police-reported accidents in Michigan in 1987. In both cases, we looked strictly at two-vehicle accidents in which only one driver was considered at fault. The results are presented in table I.1.

Table I.1: Ratio of At-Fault to Not-at-Fault Drivers: Michigan 1987 and North Carolina 1990

	Michigan		North Carolina	
Age	**Male**	**Female**	**Male**	**Female**
Under 25	1.43	1.21	1.38	1.19
25-34	0.90	0.76	0.90	0.78
35-44	0.69	0.65	0.75	0.68
45-54	0.69	0.68	0.72	0.68
55-64	0.75	0.87	0.87	0.93
65-74	1.04	1.33	1.32	1.41
Over 75	2.25	2.73	2.92	2.91

While there is not absolute agreement between the involvement ratios derived from the two data bases (the greatest discrepancy being between the results for the oldest, male drivers), the figures are quite close and the trends are remarkably similar.

The second uncertainty associated with the use of induced exposure is potentially more serious and is less easily tested. Threats to the validity of these estimates have been suggested, in particular the possibility of systematic bias. It is possible that certain driver or vehicle types are more likely to be identified by police as being at fault in a two-vehicle accident. For example, in an ambiguous situation the police may be more inclined to place blame on a young driver. On a different plane, it is possible that the not-at-fault driver in a two-car accident is not totally without blame. This driver's ability to avoid accidents may be less than average; hence, he or she may be more "accident-prone." To the extent that accident-proneness exists, the not-at-fault population less than perfectly represents of the universe of drivers and vehicles.

The existence of such a bias cannot be tested directly, but indications of whether its existence effectively distorts the estimates derived from induced exposure methods can be tested both by comparisons internal to the data base and by comparing the estimates with those derived from direct exposure measurements. Lyles et al. used internal tests to determine whether discernible bias entered their estimates. They reasoned that, if not-at-fault drivers represent drivers on the road, we should find variations in their characteristics that are related to different driving conditions. For example, we know from direct observation that drivers on major freeways are less likely to be female than drivers on more local roads. Induced exposure findings should be consistent with observation and, in fact, Lyles et al. found that 63 percent of not-at-fault drivers on U.S.-numbered routes in Michigan were male, as opposed to 57 percent on local streets. Furthermore, male at-fault drivers should strike approximately the same proportion of male not-at -fault drivers as do female at-fault drivers. This turned out to be the case. On U.S.-numbered routes, male at-fault drivers struck male drivers 63 percent of the time and females 37 percent of the time. Female at-fault drivers struck male drivers 62 percent of the time and females 38 percent of the time.

Lyles et al. offer a series of similar crosschecks for the mutual independence of the at-fault and not-at-fault populations across a variety of conditions, including different roadway types, years, times of day, and driver age categories. While there were wide variations in the distribution of driver characteristics among different driving conditions, different subsets within the same condition yielded nearly identical estimates of exposure.

Comparisons with estimates formed from direct exposure methods are less straightforward. We can, for example, obtain the distribution of licensed drivers by gender from any state. However, we know that this provides a biased estimate of drivers on the road. Using the Department of Transportation's (DOT's) 1990 National Personal Transportation Survey (NPTS), we found that over 15 percent of all women licensed drivers over age 75 had not driven at all in the previous year (as contrasted with less than 1 percent of licensed women drivers between 25 and 34).

NPTS itself is perhaps our most comprehensive direct exposure source, and it is particularly valuable in discerning trends in travel habits in the United States over time. Yet, besides being subject to the weaknesses of human recollection, it is relatively insensitive to the quality of driving exposure. While its estimates of miles driven by respondents may be quite reliable, it cannot estimate the portion of miles driven under different conditions to the level of detail that, arguably, an induced exposure method can.

Nevertheless, comparisons of induced exposure results with those derived from more direct methods are informative. Accordingly, we compared our estimates of age and gender distribution with those from NPTS. The comparisons are presented in table I.2 in terms of the percentage of vehicle miles driven (from NPTS) and the percentage of exposure to accidents as derived from our data.

Table I.2: Comparative Estimates of Exposure From the NPTS of Vehicle Miles Traveled and Not-at-Fault Drivers From North Carolina, 1990

	Vehicle miles driven: NPTS		Induced exposure estimates	
Age	Male	Female	Male	Female
Under 25	8.01%	5.91%	14.01%	12.11%
25-34	18.30	10.14	13.81	13.25
35-44	16.69	9.46	10.40	10.01
45-54	10.46	4.78	6.36	5.51
55-64	6.87	2.92	4.58	3.48
65-74	3.29	1.61	2.90	2.05
Over 74	1.13	0.43	0.92	0.62
Overall	**64.75%**	**35.25%**	**52.98%**	**47.03%**

A comparison of the estimates of vehicle miles driven and of accident exposure illuminates the differences between the two measures. Put simply, not all miles are equal. It has been demonstrated that men tend to drive substantially more freeway miles than women and that freeway miles are the safest of all miles driven. These considerations are reflected in the substantial difference between the overall gender distribution estimates derived from the two measures.[3] While men may drive nearly twice as many miles as women (65 percent versus 35 percent of all miles; see table I.2), these are more often highway miles and, thus, are substantially safer, with the result that their accident exposure is only moderately higher than women's using the induced exposure approach. Similarly, young drivers drive fewer miles than middle-aged drivers, but their miles are considerably more dangerous both because of their timing (nights and weekends) and because of driver inexperience and risk-taking behavior. These relationships are shown in table I.2: using the induced exposure method, men age 45-54 represent 6 percent of the population at risk while men under 25 have twice the exposure (14 percent), whereas using NPTS the percentages are only slightly different.

In summary, the induced exposure method offers a means of estimating the relative risks of different types of drivers, vehicles, and driving conditions at a level of refinement that cannot be approximated in practice by any direct measurement technique. It yields summary estimates of exposure that differ from the global estimates of direct measures such as vehicle miles traveled, but the differences appear to be reasonable in view of the larger number of factors taken into consideration by the induced exposure method. Its estimates of relative risk appear to be quite stable across different geographic, driver, vehicle, and roadway conditions. In classic measurement terms, while the method's predictive validity has not been empirically demonstrated, evidence exists to support its reliability and construct validity. Its practical utility is beyond question.

---

[3] NPTS estimates do not exclude miles driven in commercial vehicles other than passenger cars, a fact that would also inflate the difference in estimates since presumably more men than women drive such vehicles.

The North Carolina Data File

The data set for our analysis was created from data tapes, provided by North Carolina's Division of Motor Vehicles, containing information on accidents in North Carolina for calendar year 1990. The information is derived from accident report forms filled out by investigating officers at accident scenes. The Highway Safety Research Center at the University of North Carolina added technical information concerning vehicles (such as vehicle weight, wheelbase, and engine size), which was obtained by decoding the vehicle identification numbers recorded on the accident forms. We also merged information, collected by the Division of Motor Vehicles, on drivers' violation histories.

The data file contained one record for each individual unit (vehicle, pedestrian, bicyclist, and so on) involved in the accident. Table II.1 provides counts of the types of individual records that were contained in the North Carolina file. Table II.2 provides the distribution of accident types in the data base. For the purposes of the current study, an accident was considered a single- or two-vehicle accident on the basis of the count of the number of in-motion, motorized vehicles involved. We excluded the accident category labeled "Other" in table II.2, which may contain single- or two-vehicle accidents if the type of vehicle involved was not reported or was a heavy truck, bus, or farm vehicle. The "Other" category also contains accidents with three or more in-motion vehicles.

Table II.1: Number of Records

Record type	Number
Moving vehicles	309,409
Parked vehicles	15,868
Pedestrians or bicyclists	3,747
Type not stated	11,118
Total	**340,142**

Table II.2: Types of Accidents

	Frequency	Percent	Cumulative percent
Single vehicle			
Nonrollover			
Passenger cars	33,680	18.3%	18.3%
Light trucks and vans	9,780	5.3	23.7
Rollover			
Passenger cars	5,398	2.9	26.6
Light truck and vans	2,435	1.3	27.9
Two vehicle			
Passenger cars	63,674	34.7	62.6
Trucks and vans	4,414	2.4	65.0
Cars, trucks, and vans	32,211	17.5	82.6
Other	32,024	17.4	100.0
Total	**183,616**	**100.0%**	

Logistic Regression Analysis for Passenger Car Crash Involvement

The outcome variable for the logistic regression equations was a dichotomous indicator of fault, coded “1” for at-fault and “0” for not-at-fault drivers. For all the equations presented here, both single- and two-vehicle accidents, the comparison group is not-at-fault drivers in two-vehicle accidents.[1] A driver was considered at fault if the investigating police officer checked one or more violations in the checklist provided on the North Carolina accident report form. (In two-vehicle accidents, cases were excluded if no violation was reported for either driver or if both drivers had violations.)

The independent variables included

- driver age, including a squared term to capture the curvilinear relationship between age and accident involvement;
- driver gender, with males coded “1” and females coded “0”;
- driver violation history, with four mutually exclusive categories: no previous traffic violations, one or more previous violations not involving alcohol, at least one alcohol violation (may also include nonalcohol violations), and violation history unknown (all out-of-state drivers and some North Carolina drivers are in this category). In the models shown, the three categories given are in contrast to the group having alcohol-related violations;
- vehicle age, last two digits of the vehicle model year;
- vehicle curb weight, expressed in hundreds of pounds and including a squared term to capture the curvilinear relationship between vehicle weight and accident involvement.

Table III.1: Number of Cases and Model Chi-Squares

			Single vehicle	
Accident type	**All**	**Two vehicle**	**Nonrollover**	**Rollover**
At fault	64,904	50,824	11,222	2,858
Not at fault	51,499	51,499	51,499	51,499
Total	116,403	102,323	62,721	54,357
Model chi-square	5,655.178	4,137.856	5,524.596	2,546.841
Model degrees of freedom	9	9	9	9

---

[1] Using not-at-fault drivers in two-vehicle accidents as the exposure group for single-vehicle accidents can be justified as long as there is a control for factors known to discriminate between the two accident types. Our models incorporated the strongest of the predictors of single- versus two-vehicle accidents in our data base: driver age and driver gender. See R. W. Lyles, “Quasi-induced Exposure: To Use or Not to Use?” presented at the Transportation Research Board annual meeting, January 10, 1994, p. 4.

Table III.2: Parameter Estimates

							Single vehicle					
	All			Two vehicle			Nonrollover			Rollover		
Variable	Coeff.	S.E.	Prob.	Coeff.	S.E.	Prob.	Coeff.	S.E.	Prob.	Coeff.	S.E.	Prob.
Constant	5.1503	.1518	.0000	4.2265	.1603	.0000	4.2161	.2652	.0000	5.2564	.4748	.0000
Driver age	–0.1080	.0019	.0000	–0.0970	.0020	.0000	–0.1245	.0036	.0000	–0.1374	.0074	.0000
Age squared	0.0012	.0000	.0000	0.0011	.0000	.0000	0.0011	.0000	.0000	0.0011	.0001	.0000
Male	0.1219	.0125	.0000	–0.0004	.0132	.9735	0.6126	.0232	.0000	0.6790	.0425	.0000
Nonalcohol violation[a]	–0.7582	.0391	.0000	–0.4929	.0425	.0000	–1.3094	.0509	.0000	–1.3960	.0823	.0000
No violation	–0.9993	.0388	.0000	–0.7515	.0422	.0000	–1.4489	.0505	.0000	–1.4180	.0817	.0000
Violation unknown	–0.6632	.0423	.0000	–0.2983	.0455	.0000	–1.6481	.0612	.0000	–1.7720	.1035	.0000
Vehicle year	–0.0227	.0014	.0000	–0.0193	.0015	.0000	–0.0332	.0025	.0000	–0.0257	.0044	.0000
Vehicle weight[b]	–0.0135	.0072	.0586	–0.0211	.0075	.0051	0.0448	.0130	.0006	–0.0974	.0256	.0001
Vehicle weight squared	0.0002	.0001	.1304	0.0003	.0001	.0105	–0.0007	.0002	.0027	0.0008	.0005	.0780

[a] Contrast group for violations is “Has alcohol violation.”

[b] Vehicle weight is calibrated in hundredweights.

Logistic Regression Analysis for Light Truck and Van Crash Involvement

The outcome variable for the logistic regression equations was a dichotomous indicator of fault, coded “1” for at-fault and “0” for not-at-fault drivers. For all the equations presented here, both single- and two-vehicle accidents, the comparison group is not-at-fault drivers in two-vehicle accidents.[1] A driver was considered at fault if the investigating police officer checked one or more violations in the checklist provided on the North Carolina accident report form. (In two-vehicle accidents, cases were excluded if no violation was reported for either driver or if both drivers had violations.)

The independent variables included

- driver age, including a squared term to capture the curvilinear relationship between age and accident involvement;
- driver gender, with males coded “1” and females coded “0”;
- driver violation history, with four mutually exclusive categories: no previous traffic violations, one or more previous violations not involving alcohol, at least one alcohol violation (may also include nonalcohol violations), and violation history unknown (all out-of-state drivers and some North Carolina drivers are in this category). In the models shown, the three categories given are in contrast to the group having alcohol-related violations;
- vehicle age, last two digits of the vehicle model year;
- vehicle curb weight, expressed in hundreds of pounds and including a squared term to capture the curvilinear relationship between vehicle weight and accident involvement.

Table IV.1: Number of Cases and Model Chi-Squares

			Single vehicle	
Accident type	**All**	**Two vehicle**	**Nonrollover**	**Rollover**
At fault	11,782	8,457	2,314	1,011
Not at fault	8,609	8,609	8,609	8,609
Total	20,391	17,066	10,923	9,620
Model chi-square	868.5668	553.830	772.251	624.847
Model degrees of freedom	9	9	9	9

---

[1] Using not-at-fault drivers in two-vehicle accidents as the exposure group for single-vehicle accidents can be justified as long as there is a control for factors known to discriminate between the two accident types. Our models incorporated the strongest of the predictors of single- versus two-vehicle accidents in our data base: driver age and driver gender. See R. W. Lyles, “Quasi-induced Exposure: To Use or Not to Use?” presented at the Transporation Research Board annual meeting, January 10, 1994, p. 4.

Table IV.2: Parameter Estimates

							Single vehicle					
	All			Two vehicle			Nonrollover			Rollover		
Variable	**Coeff.**	**S.E.**	**Prob.**	**Coeff.**	**S.E.**	**Prob.**	**Coeff.**	**S.E.**	**Prob.**	**Coeff.**	**S.E.**	**Prob.**
Constant	5.7770	.4990	.0000	3.9962	.5324	.0000	6.6988	.8479	.0000	7.4987	1.1664	.0000
Driver age	–0.1156	.0050	.0000	–0.1027	.0053	.0000	–0.1243	.0084	.0000	–0.1490	.0128	.0000
Age squared	0.0012	.0001	.0000	0.0012	.0001	.0000	0.0011	.0001	.0000	0.0013	.0002	.0000
Male	–0.0017	.0406	.9666	–0.0871	.0432	.0439	0.2296	.0724	.0015	0.3135	.1053	.0029
Nonalcohol violation[a]	–0.6092	.0748	.0000	–0.3487	.0825	.0000	–1.0389	.0999	.0000	–1.1448	.1317	.0000
No violation	–0.8209	.0739	.0000	–0.5608	.0817	.0000	–1.2493	.0991	.0000	–1.2542	.1308	.0000
Violation unknown	–0.5173	.0866	.0000	–0.1756	.0941	.0621	–1.1973	.1268	.0000	–1.2777	.1709	.0000
Vehicle year	–0.0233	.0034	.0000	–0.0191	.0036	.0000	–0.0370	.0056	.0000	–0.0363	.0080	.0000
Vehicle weight[b]	–0.0387	.0220	.0789	–0.0120	.0233	.6069	–0.0676	.0387	.0806	–0.1238	.0551	.0246
Vehicle weight squared	0.0006	.0003	.0501	0.0004	.0003	.2955	0.0008	.0006	.1587	0.0014	.0008	.0851

[a] Contrast group for violations is “Has alcohol violation.”

[b] Vehicle weight is calibrated in hundredweights.

Appendix V

Alternative Definitions of Vehicle Size

Models containing three alternative definitions of vehicle size were fitted to the data. These were vehicle weight (in hundreds of pounds), engine size (displacement expressed in cubic inches), and wheelbase (the distance between the axles in inches). To allow comparisons of their relative importance in predicting crash involvement, the improvement in goodness of fit for each model over the base model (including driver age, violation history, gender, and vehicle age) is presented in tables V.1 and V.2.[1]

As noted in the text, (1) with the exception of single-car rollovers, contributions to the model, though statistically significant in most cases, are small in comparison to most variables in the base model, and (2) different definitions are stronger depending upon the crash type being predicted.

Table V.1: Change in –2 Log Likelihood From Base Model Using Alternative Measures of Vehicle Size: Passenger Cars

	Change in –2 log likelihood	Degrees of freedom	Probability
All crash types			
Weight	9.607	2	.0082
Engine size	6.095	2	.0475
Wheelbase	8.159	2	.0169
Two-vehicle crashes			
Weight	9.953	2	.0069
Engine size	8.547	2	.0139
Wheelbase	0.009	2	.9922
Single-vehicle nonrollover crashes			
Weight	19.445	2	.0001
Engine size	30.922	2	.0000
Wheelbase	3.676	2	.1592
Single-vehicle rollover crashes			
Weight	218.057	2	.0000
Engine size	167.355	2	.0000
Wheelbase	276.602	2	.0000

---

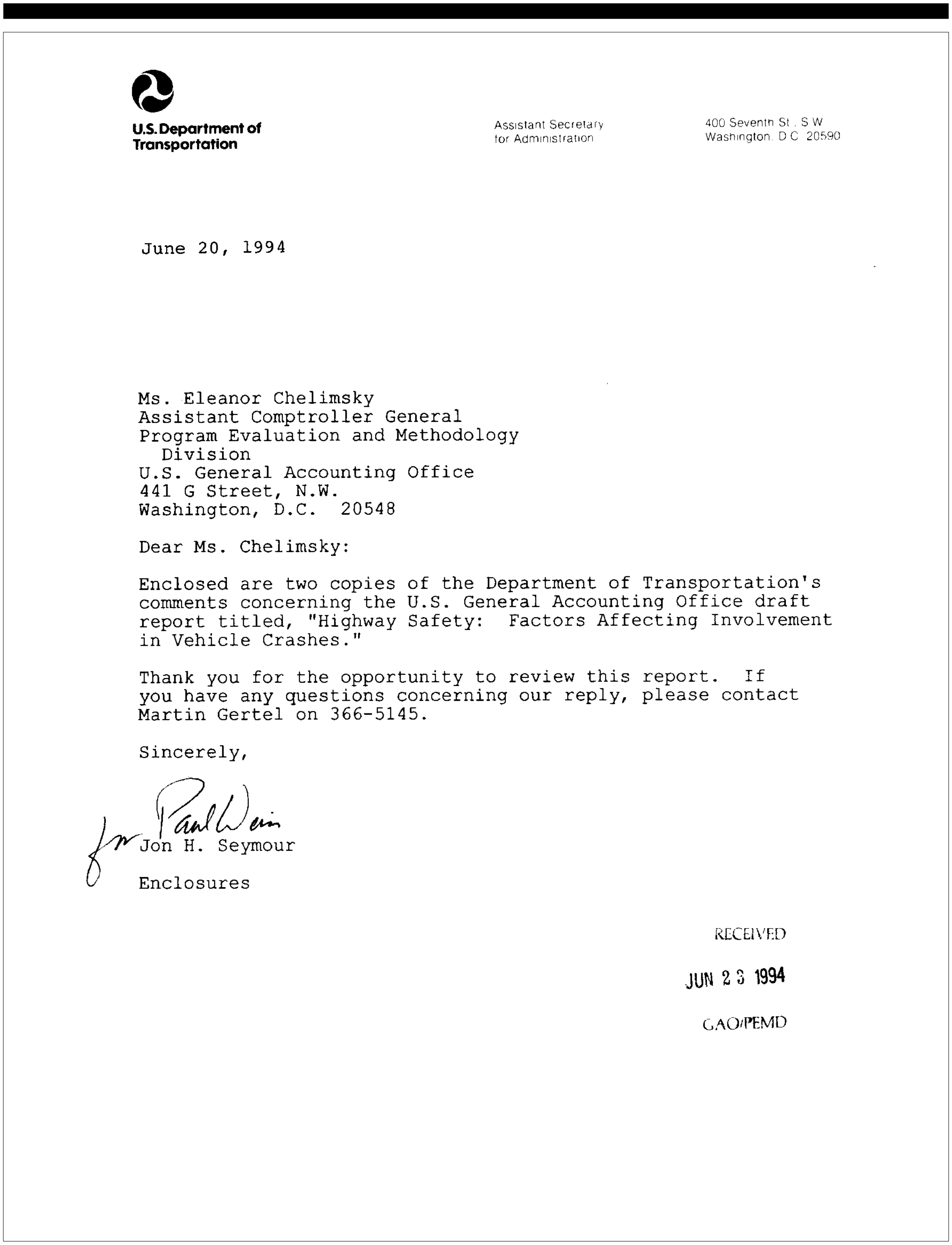
[1] See, for example, A. Agresti, Categorical Data Analysis (New York: Wiley, 1990), pp. 95-96.

Table V.2: Change in –2 Log Likelihood From Base Model Using Alternative Measures of Vehicle Size: Light Trucks

	Change in –2 log likelihood	Degrees of freedom	Probability
All crash types			
Weight	6.812	2	.0332
Engine size	7.556	2	.0229
Wheelbase	10.782	2	.0046
Two-vehicle crashes			
Weight	22.818	2	.0000
Engine size	22.144	2	.0000
Wheelbase	42.512	2	.0000
Single-vehicle nonrollover crashes			
Weight	12.078	2	.0024
Engine size	7.987	2	.0184
Wheelbase	12.951	2	.0015
Single-vehicle rollover crashes			
Weight	26.840	2	.0000
Engine size	35.571	2	.0000
Wheelbase	58.808	2	.0000

Comments From the Department of Transportation

Note: GAO comments supplementing those in the report text appear at the end of this appendix.



U.S. Department of Transportation

Assistant Secretary
for Administration

400 Seventh St., S.W.
Washington, D.C. 20590

June 20, 1994

Ms. Eleanor Chelimsky
Assistant Comptroller General
Program Evaluation and Methodology
Division
U.S. General Accounting Office
441 G Street, N.W.
Washington, D.C. 20548

Dear Ms. Chelimsky:

Enclosed are two copies of the Department of Transportation's comments concerning the U.S. General Accounting Office draft report titled, "Highway Safety: Factors Affecting Involvement in Vehicle Crashes."

Thank you for the opportunity to review this report. If you have any questions concerning our reply, please contact Martin Gertel on 366-5145.

Sincerely,

Jon H. Seymour

Enclosures

RECEIVED
JUN 23 1994
GAO/PEMD

DEPARTMENT OF TRANSPORTATION REPLY

TO

GENERAL ACCOUNTING OFFICE (GAO) DRAFT REPORT

ON

"Highway Safety: Factors Affecting Involvement in Vehicle Crashes"

SUMMARY OF GAO FINDINGS

The GAO draft report uses a method known as "induced exposure" to estimate the likelihood of crash involvement. This method is based on calculating the ratio of at-fault drivers or vehicles to not-at-fault drivers or vehicles in two-vehicle accidents described in police accident reports. Its underlying assumption is that the not-at-fault drivers and vehicles constitute a representative sample of the interactions among drivers, vehicles, and driving conditions for specific geographic areas. GAO used this method to develop models for predicting the likelihood of crash involvement for passenger cars and light trucks and to distinguish between three different crash types: two-vehicle crashes, single-vehicle nonrollover crashes, and single-vehicle crashes involving a rollover. The variables evaluated include driver age, gender, traffic violation history, as well as vehicle age and weight.

The draft report concluded that the induced-exposure methodology demonstrated in the analysis offers reasonable expectations of yielding results substantially more sensitive to the real world driving environment than can be achieved through currently available direct exposure methods without incurring prohibitive data collection costs. Specifically the draft concluded that driver-related variables were much more important than vehicle-related variables in estimating the likelihood of being in a traffic crash. Among the driver variables, a driver's age was a much stronger predictor followed in most models by the driver's violation history and then by gender. Within the vehicle variables considered, vehicle age contributed substantially more to predicting crash involvement than vehicle weight. In addition, substituting vehicle dimensions, particularly wheelbase and engine size, for vehicle weight provided a better predictor than vehicle weight in five of the six models.

DEPARTMENT OF TRANSPORTATION POSITION

See comment 1.

We have reviewed the subject report and are concerned that it oversimplified the analytical treatment of single-vehicle crash involvement. We maintain that the conclusions related to single-vehicle rollover crashes are incorrect and contradict the National Highway Traffic Safety Administration's (NHTSA) findings that result from an exhaustive five-year vehicle rollover analysis. As a result we are concerned that the report's conclusions could adversely affect NHTSA's rollover rulemaking efforts.

NHTSA Conducted Extensive Vehicle Rollover Analysis

Over the past five years NHTSA has been conducting extensive research into the issue of light-vehicle rollover. This research identified an exhaustive list of vehicle attributes related to stability and directional control, measured these attributes on a wide range of light-vehicles, developed models to estimate these characteristics, and conducted statistical modeling of the rollover phenomenon as a function of roadway, vehicle, and driver factors. A summary of previous research, as well as findings from the more recent efforts, have been made public through the use of a public docket (Docket 91-68, Notice 1) and notices in the Federal Register.

The vast majority of light-vehicle rollovers result from single-vehicle crashes. While rollover might appear to be a separate distinct type of crash, it is more appropriate to consider a rollover as a separate outcome of a single-vehicle crash. For example, when a moving vehicle leaves the roadway it is not predetermined that a rollover will or will not occur. The likelihood that a particular vehicle will roll over is a function of many factors including roadway geometry especially the shoulder architecture regarding tripping mechanisms and roadside grade, vehicle stability, and driver characteristics. Rural crashes are more likely to result in rollover, presumably due to the presence of more curves and grades. This important factor was not included in the draft report's analysis of crash involvement.

Roadway and Vehicle Characteristics Key to Single-Vehicle Rollover

See comment 2.

While driver characteristics play a role in rollover occurrence, our statistical models have strongly demonstrated that the roadway and vehicle are much more important than any driver characteristics. When the vehicle leaves the roadway, all of the factors present just prior to the vehicle's departure work to increase or decrease the chance that the vehicle will roll over. Once the departure has occurred the driver's characteristics are no longer of primary importance. Research has repeatedly demonstrated that the roadway and the vehicle's inherent stability characteristics are the prominent factors affecting vehicle rollover. Our recent research, which will shortly be released to the public, clearly demonstrates that statistical models without the contribution of vehicle stability measures do a much poorer job of predicting whether a vehicle rolls over.

We recommend that the draft report's analysis consider single-vehicle crash involvement including both rollover and nonrollover as a single group. This would undoubtedly yield results consistent with those already contained in the draft report, that driver age and violation history play the strongest roles in involvement in single-vehicle crashes, with little contribution from the various vehicle characteristics. However, it would eliminate the inaccuracy in the GAO draft report regarding the relative roles of drivers as compared to roadway and vehicle characteristics in contributing to vehicle rollover. A second set of models would need to be developed to address whether a single-vehicle crash becomes a rollover. For this part of the analysis, the work we have done over the past five years provides a solid basis for concluding that roadway and vehicle stability are much more important than driver characteristics in vehicle rollover occurrence. Adopting this approach would result in sound policy decisions and also good science.

SPECIFIC COMMENTS

The Department offers the following specific comments regarding the draft report.

Deleted.

- **Page 5, Table 1 and Page 7, Table 2:** The accompanying text could benefit by clarifying what outcome variable was analyzed to yield the results in these tables. It is not clear whether these tables represent the combination of involvement in at-fault two-vehicle crashes plus single-vehicle nonrollover plus single-vehicle rollover or other factors.

Now p. 2.

- **Page 6, paragraph 1:** It would be more accurate to describe the relative odds as "... the odds of a 20-year-old driver being involved in a single-vehicle, nonrollover crash was over 4 times *as great as* that of a 50-year-old" and not "... 4 times *greater than* ...." This terminology occurs throughout the draft report.

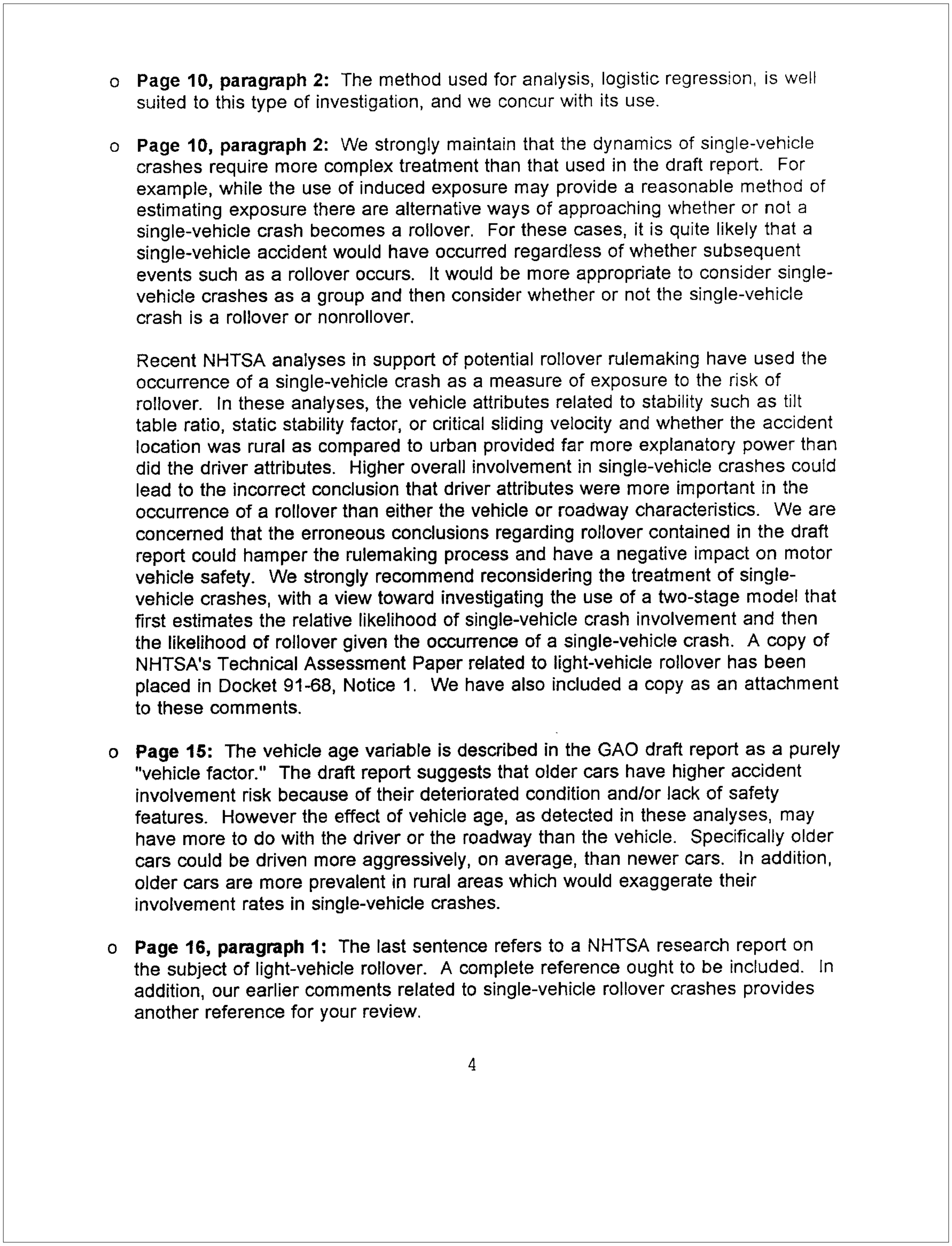
See p. 8.

- **Page 6, paragraphs 2 and 3:** The use of vehicle weight requires more detailed explanation. There are several choices for passenger car weight, including weight reported on the Fatal Accident Reporting System (FARS) file and weight as reported by R.L. Polk. In addition, the weight of light trucks is recorded on FARS as the Gross Vehicle Weight Rating (GVWR), which is not the actual weight of the vehicle. Light truck weights are particularly difficult to obtain as there are many possible truck configurations that result in different vehicle weights. The report could benefit by presenting greater detail regarding how vehicle weights were determined and which weights were used in the statistical models.

See p. 25.

See comment 2.

- **Pages 9-10:** It would be useful to the reader if a more detailed explanation were included to address how "fault" was determined. Clarification could be provided regarding whether fault was based solely on the issuance of a police citation or if additional information was used. For two-vehicle crashes, clarification could be provided regarding whether there was only one driver at-fault or if it was possible for both or neither driver to be at-fault. The determination of fault is critical to developing induced-exposure estimates. Because these estimates are the basis for the report's statistical analysis and conclusions it would be worthwhile to discuss the details of estimating exposure in the report. In addition there is some question about whether involvement in two-vehicle crashes provides a useful measure of exposure for single-vehicle crashes since the urban-rural exposure was not included in any of GAO's models. Rural areas tend to have more single-vehicle crashes due to lesser traffic density and more complex roadway geometry.



Now p. 4.

- Page 10, paragraph 2: The method used for analysis, logistic regression, is well suited to this type of investigation, and we concur with its use.

Now p. 4.

See comment 1.

- **Page 10, paragraph 2:** We strongly maintain that the dynamics of single-vehicle crashes require more complex treatment than that used in the draft report. For example, while the use of induced exposure may provide a reasonable method of estimating exposure there are alternative ways of approaching whether or not a single-vehicle crash becomes a rollover. For these cases, it is quite likely that a single-vehicle accident would have occurred regardless of whether subsequent events such as a rollover occurs. It would be more appropriate to consider single-vehicle crashes as a group and then consider whether or not the single-vehicle crash is a rollover or nonrollover.

  Recent NHTSA analyses in support of potential rollover rulemaking have used the occurrence of a single-vehicle crash as a measure of exposure to the risk of rollover. In these analyses, the vehicle attributes related to stability such as tilt table ratio, static stability factor, or critical sliding velocity and whether the accident location was rural as compared to urban provided far more explanatory power than did the driver attributes. Higher overall involvement in single-vehicle crashes could lead to the incorrect conclusion that driver attributes were more important in the occurrence of a rollover than either the vehicle or roadway characteristics. We are concerned that the erroneous conclusions regarding rollover contained in the draft report could hamper the rulemaking process and have a negative impact on motor vehicle safety. We strongly recommend reconsidering the treatment of single-vehicle crashes, with a view toward investigating the use of a two-stage model that first estimates the relative likelihood of single-vehicle crash involvement and then the likelihood of rollover given the occurrence of a single-vehicle crash. A copy of NHTSA's Technical Assessment Paper related to light-vehicle rollover has been placed in Docket 91-68, Notice 1. We have also included a copy as an attachment to these comments.

Now p. 7.

- **Page 15:** The vehicle age variable is described in the GAO draft report as a purely "vehicle factor." The draft report suggests that older cars have higher accident involvement risk because of their deteriorated condition and/or lack of safety features. However the effect of vehicle age, as detected in these analyses, may have more to do with the driver or the roadway than the vehicle. Specifically older cars could be driven more aggressively, on average, than newer cars. In addition, older cars are more prevalent in rural areas which would exaggerate their involvement rates in single-vehicle crashes.

Now p. 8.

- **Page 16, paragraph 1:** The last sentence refers to a NHTSA research report on the subject of light-vehicle rollover. A complete reference ought to be included. In addition, our earlier comments related to single-vehicle rollover crashes provides another reference for your review.

4

Deleted.

- **Page 16, footnote 2:** As stated earlier, we totally disagree with the characterization that, "*... relative to that of driver-related measures, the contribution of the vehicle size measures is trivial.*"

Now p. 11.

See comment 3.

- **Pages 19 and 20:** Although the induced-exposure method is excellent for measuring involvement risk in two-vehicle crashes, there are some shortcomings for single-vehicle crashes. A type of vehicle that is extensively used in rural areas such as the pickup truck will have relatively few two-vehicle crashes and many single-vehicle crashes. Since not-at-fault two-vehicle crash involvements are the measure of "exposure," these vehicles will seem to have a very high rate of single-vehicle crashes relative to "exposure." This could produce misleading results if the analysis mixes pickup trucks with other vehicles primarily used in urban areas, such as minivans. For this reason, we recommend that the analysis of trucks be split into three separate analyses: pickups, vans, and sport-utility vehicles.

Now p. 18.

- **Page I-1, paragraph 1**: There are a number of ways to "induce" exposure beyond the use of not-at-fault drivers. For example, Leonard Evans used drivers in pedestrian crashes as a measure of exposure in a study of vehicle weight and safety. Also single-vehicle crash occurrence has been used by a number of researchers including NHTSA as a measure of exposure to the risk of a vehicle rolling over. The draft report could expand its discussion of the induced exposure issue, bringing forth some of the controversies surrounding the use of this method.

Now p. 19.

See comment 4.

- **Page I-5, Table I.1:** This table compares relevant ratios between the North Carolina crash-involved drivers for 1990 used in the GAO report and the Michigan crash-involved drivers for 1987 not used in the report. While many of the ratios between the two states are comparable, this by itself does not demonstrate the ability to generalize the results to the nation as a whole. No discussion was provided about why 1987 Michigan data were presented, as opposed to including data from other years within North Carolina or data from additional states. It would be useful to demonstrate the similarity of these rates over additional states to support the statement in the draft report that the similarities between the data for North Carolina and Michigan raises the question of, "*... whether the findings will not, in fact, turn out to be representative nationally after all*," as appears on page 3 of the current report.

See comment 5.

- The report ought to emphasize that the accident propensity rates apply only to two-vehicle collisions where the driver is "at-fault," and not all two-vehicle collisions.

- Appendix I of the report ought to acknowledge the contributions of NHTSA's Ezio Cerrelli to the development of the induced-exposure method.

The following are GAO's comments on DOT's June 20, 1994, letter.

GAO Comments

1. We share with DOT the belief that every reasonable effort should be made to reduce the incidence of rollover crashes, which, as we noted in our report, are second only to frontal collisions in deadliness. We would be as concerned as DOT if our conclusion, that car size is substantially less predictive of rollover crashes than are driver characteristics, were misinterpreted to diminish the importance of efforts to reduce the rollover propensity of vehicles.

The analyses we performed differ significantly from NHTSA's rollover research, but our conclusions are not in conflict. Our concern was with the relative contribution of car size to crash involvement. We examined this relationship in a general model that combined crash types and then separately, using the traditional analytic taxonomy of multiple vehicle, single-vehicle nonrollover, and single-vehicle rollover crashes. NHTSA, in contrast, attempted to identify the factors that differentiated single-vehicle rollover from nonrollover crashes. The vehicle factors it examined included a number of constructs derived from laboratory measurements, such as tilt table ratio, side pull ratio, and critical sliding velocity. The single area of overlap between the NHTSA analyses and ours was in the inclusion of wheelbase in NHTSA's models and in one of our models.

It is not surprising, therefore, that we arrived at different conclusions regarding the importance of different vehicle characteristics relative to driver characteristics. Nevertheless, our findings also support the relatively greater importance of vehicle characteristics in rollover crashes than in other crash types. We found that lighter vehicles were more likely to be involved in single vehicle rollovers. We further found that wheelbase was a better predictor of rollover crashes than weight.

2. DOT made two suggestions for additional analyses to supplement our single-vehicle rollover model. First, agency researchers suggested that we include in our model some roadway characteristics that were beyond the scope of the research originally requested. They also suggested that we treat all single-vehicle crashes as one crash type and then perform a second-level analysis to identify the factors that distinguish between rollover and nonrollover crashes.

We performed these analyses and concluded that, while they provided important additional information about the dynamics of rollover crashes, they did not substantially alter our conclusions about the relative importance of the driver and vehicle characteristics we examined.

To respond to the first suggestion, we added two roadway variables to all our models: whether the roadway was curved or straight and whether the crash occurred in a rural or urban setting. The results of these analyses are provided in appendix VII (passenger cars) and appendix VIII (light trucks and vans). As anticipated, these roadway characteristics generally contributed significantly to the predictive power of the models. Single-vehicle crashes (rollover and nonrollover) are more likely to occur on rural and on curved roadways. Two-vehicle car crashes are less likely to occur on rural roads. The addition of these predictors, however, did not change the predominant importance of driver factors over vehicle weight in predicting crash involvement.

We also constructed a model combining both types of single-vehicle accidents. We included the results of this model in appendixes VII and VIII.

As DOT anticipated, the single-accident model yielded results consistent with our earlier findings, that driver age and violation history play the strongest roles in involvement in single vehicle crashes, with little contribution from the various vehicle characteristics. The aggregate model, however, also finds the contribution of weight to single-vehicle accidents nonsignificant. This is the net effect of the opposing influence of weight in rollover and nonrollover accidents. As our original analyses demonstrated, heavier cars are more likely to be involved in nonrollover crashes and less likely to be involved in rollovers.

We constructed a second set of models (one each for passenger cars and for light trucks and vans) to distinguish between rollover and nonrollover single-vehicle accidents. The results of these analyses are presented in appendix IX. The model produced the results anticipated by DOT—namely, that roadway characteristics and, to a lesser extent, vehicle weight are better predictors of whether a single-vehicle accident involves a rollover than the driver characteristics in our model, although the analysis did find that younger drivers were significantly more likely to be in a rollover than a nonrollover crash.

Like the analysis, the interpretation of the models must be in two stages. The findings suggest that driver characteristics predominate over vehicle characteristics in placing a vehicle in a likely single-vehicle crash situation.

Whether the resultant crash (if one does occur) involves a rollover is more determined by roadway and vehicle considerations.[1]

3. While DOT considers induced exposure an "excellent" method for measuring involvement risk in two-vehicle crashes, it expressed some cautions about its use for single-vehicle crashes. In particular, DOT suggested that the mix of light truck types (pickups, vans, and sport-utility vehicles) is different in urban and rural settings, and therefore three different light truck analyses should be performed. We agree that such analyses could provide valuable information, but we believe that they would unnecessarily expand the scope of this report. We reviewed the relative likelihood of fatal accidents in different types of light trucks and vans in a previous report.[2] The additional analyses we performed, at DOT's suggestion, that control for the urban and rural difference also address this concern. (See appendixes VII and VIII.)

4. DOT further suggested that we perform additional tests of the ability to generalize from the induced exposure method by comparing results from a larger number of state accident data bases. Our comparison of accident involvement ratios in North Carolina and Michigan (the two usable data bases readily available to us) was intended only to illustrate the relative consistency and reasonableness of the results obtained from applying this methodology. Many more such comparisons will need to be made before the exact parameters of the method's applicability can be defined. Nevertheless, the results obtained by different researchers over the years from this approach to defining exposure are a strong argument for its general utility.

DOT also suggested we include some additional details concerning our analyses and references to other related work performed by NHTSA and other researchers. We have incorporated these suggestions where appropriate.

5. This statement has been changed in the text. See page 2.

---

[1] The NHTSA methodology makes the simplifying assumption that all the single-vehicle accidents in an accident data base would have occurred whether or not the vehicle rolled over. This is clearly not the case; for example, a vehicle that left the road may have recovered without incident if it had not rolled over, or it may have collided with a tree before rolling over. From the available state accident data bases, it is impossible to determine which situations would not have resulted in a crash if the vehicle had not rolled over. Unfortunately, the induced exposure method cannot answer this question, since its reference crash type is a two-vehicle crash.

[2] U.S. General Accounting Office, Highway Safety: Fatalities in Light Trucks and Vans, GAO/PEMD-91-8 (Washington, D.C.: November 1990).

Logistic Regression Analysis Including Roadway Characteristics: Passenger Cars

The outcome variable for the logistic regression equations was a dichotomous indicator of fault, coded “1” for at-fault and “0” for not-at-fault drivers. For all the equations presented here, both single- and two-vehicle accidents, the comparison group is not-at-fault drivers in two-vehicle accidents. A driver was considered at fault if the investigating police officer checked one or more violations in the checklist provided on the North Carolina accident report form. (In two-vehicle accidents, cases were excluded if no violation was reported for either driver or if both drivers had violations.)

The independent variables included

- driver age, including a squared term to capture the curvilinear relationship between age and accident involvement;
- driver gender, with males coded “1” and females coded “0”;
- driver violation history, with four mutually exclusive categories: no previous traffic violations, one or more previous violations not involving alcohol, at least one alcohol violation (may also include nonalcohol violations), and violation history unknown (all out-of-state drivers and some North Carolina drivers are in this category). In the models shown, the three categories given are in contrast to the group having alcohol-related violations;
- vehicle age, last two digits of the vehicle model year;
- vehicle curb weight, expressed in hundreds of pounds and including a squared term to capture the curvilinear relationship between vehicle weight and accident involvement;
- rural location, coded “1” for rural locations and “0” for mixed or urban locations;
- curved roadway, coded “1” if curved, “0” otherwise.

Table VII.1: Number of Cases and Model Chi-Squares

Accident type	Rollover and nonrollover combined	No rollover	Rollover	Two vehicle
N of cases, not-at-fault	52,915	52,915	52,915	52,915
N of cases, at fault	14,490	11,563	2,927	52,287
Total N of cases	67,405	64,478	55,842	105,202
Model chi-square	20,519.696	15,418.326	9,470.896	4,270.173
Model degrees of freedom	11	10	11	11

Table VII.2: Parameter Estimates

	Single vehicle											
	Rollover and nonrollover combined			No rollover			Rollover			Two vehicle		
Variable	**Coeff.**	**S.E.**	**Prob.**	**Coeff.**	**S.E.**	**Prob.**	**Coeff.**	**S.E.**	**Prob.**	**Coeff.**	**S.E.**	**Prob.**
Constant	3.4290	.2732	.0000	2.9800	.2420	.0000	1.9702	.5592	.0004	4.1323	.1559	.0000
Driver age	–0.1147	.0037	.0000	–0.1142	.0039	.0000	–0.1130	.0084	.0000	–0.0976	.0019	.0000
Age squared	0.0010	.00004	.0000	0.0010	.00005	.0000	0.0008	.0001	.0000	0.0011	.00002	.0000
Male	0.5699	.0241	.0000	0.5625	.0255	.0000	0.6598	.0492	.0000	–0.0008	.0130	.9482
Nonalcohol violation	–1.2270	.0549	.0000	–1.2302	.0567	.0000	–1.2287	.1031	.0000	–0.4938	.0416	.0000
No violation	–1.4549	.0545	.0000	–1.4515	.0563	.0000	–1.4474	.1026	.0000	–0.7469	.0413	.0000
Violation unknown	–1.5815	.0647	.0000	–1.5962	.0675	.0000	–1.5783	.1241	.0000	–0.2967	.0446	.0000
Vehicle year	–0.0279	.0025	.0000	–0.0277	.0025	.0000	–0.0161	.0050	.0013	–0.0175	.0014	.0000
Vehicle weight	–0.0103	.0134	.4422	0.0068	.0021	.0010	–0.1228	.0291	.0000	–0.0244	.0074	.0009
Vehicle weight squared	0.0001	.0002	.5347	[a]	[a]	[a]	0.0011	.0005	.0249	0.0004	.0001	.0017
Rural location	1.5257	.0234	.0000	1.3386	.0245	.0000	2.7020	.0649	.0000	–0.0372	.0142	.0090
Curved roadway	1.8721	.0257	.0000	1.7959	.0271	.0000	2.3021	.0469	.0000	0.0319	.0227	.1588

[a] Quadratic term removed since main effect only achieves significance without squared term.

Logistic Regression Analysis Including Roadway Characteristics: Light Trucks and Vans

The outcome variable for the logistic regression equations was a dichotomous indicator of fault, coded “1” for at-fault and “0” for not-at-fault drivers. For all the equations presented here, both single- and two-vehicle accidents, the comparison group is not-at-fault drivers in two-vehicle accidents. A driver was considered at fault if the investigating police officer checked one or more violations in the checklist provided on the North Carolina accident report form. (In two-vehicle accidents, cases were excluded if no violation was reported for either driver or if both drivers had violations.)

The independent variables included

- driver age, including a squared term to capture the curvilinear relationship between age and accident involvement;
- driver gender, with males coded “1” and females coded “0”;
- driver violation history, with four mutually exclusive categories: no previous traffic violations, one or more previous violations not involving alcohol, at least one alcohol violation (may also include nonalcohol violations), and violation history unknown (all out-of-state drivers and some North Carolina drivers are in this category). In the models shown, the three categories given are in contrast to the group having alcohol-related violations;
- vehicle age, last two digits of the vehicle model year;
- vehicle curb weight, expressed in hundreds of pounds;
- rural location, coded “1” for rural locations and “0” for mixed or urban locations;
- curved roadway, coded “1” if curved, “0” otherwise.

Table VIII.1: Number of Cases and Model Chi-Squares

	Single vehicle			
Accident type	Rollover and nonrollover combined	No rollover	Rollover	Two vehicle
N of cases, not-at-fault	8,971	8,971	8,971	8,971
N of cases, at-fault	3,444	2,392	1,052	8,777
Total N of cases	12,415	11,363	10,023	17,748
Model chi-square	4,141.020	2,679.975	2,548.760	557.139
Model degrees of freedom	10	10	10	10

Table VIII.2: Parameter Estimates

	Single vehicle											
	Rollover and nonrollover combined			No rollover			Rollover			Two vehicle		
Variable	Coeff.	S.E.	Prob.	Coeff.	S.E.	Prob.	Coeff.	S.E.	Prob.	Coeff.	S.E.	Prob.
Constant	4.5531	.5437	.0000	4.3259	.5912	.0000	2.1671	.9021	.0163	3.4486	.3375	.0000
Driver age	–0.1247	.0085	.0000	–0.1168	.0092	.0000	–0.1427	.0144	.0000	–0.1021	.0052	.0000
Age squared	0.0011	.0001	.0000	0.0010	.0001	.0000	0.0012	.0002	.0000	0.0012	.00006	.0000
Male	0.1943	.0718	.0068	0.1903	.0788	.0158	0.2148	.1195	.0722	–0.0827	.0423	.0505
Nonalcohol violation	–0.9934	.1044	.0000	–0.9901	.1108	.0000	–1.1573	.1645	.0000	–0.3686	.0807	.0000
No violation	–1.2685	.1034	.0000	–1.2789	.1099	.0000	–1.3483	.1626	.0000	–0.5664	.0799	.0000
Violation unknown	–1.1464	.1288	.0000	–1.1814	.1389	.0000	–1.1957	.2030	.0000	–0.1972	.0921	.0322
Vehicle year	–0.0316	.0056	.0000	–0.0335	.0061	.0000	–0.0151	.0093	.1061	–0.0172	.0035	.0000
Vehicle weight[a]	–0.0189	.0042	.0000	–0.0153	.0046	.0009	–0.0392	.0071	.0000	0.0122	.0026	.0000
Rural location	1.6660	.0525	.0000	1.4001	.0565	.0000	2.6262	.1179	.0000	–0.0277	.0328	.3987
Curved roadway	1.7116	.0551	.0000	1.6115	.0598	.0000	2.0296	.0831	.0000	0.0033	.0519	.9489

[a] Quadratic term for weight removed since main effect only achieves significance without squared term.

Predicting Rollover Versus No Rollover Among Single-Vehicle Accidents

The outcome variable for the logistic regression equations was a dichotomous indicator of vehicle rollover, coded “1” for rollover and “0” for nonrollovers.

The independent variables included

- driver age, including a squared term to capture the curvilinear relationship between age and accident involvement;
- driver gender, with males coded “1” and females coded “0”;
- driver violation history, with four mutually exclusive categories: no previous traffic violations, one or more previous violations not involving alcohol, at least one alcohol violation (may also include nonalcohol violations), and violation history unknown (all out-of-state drivers and some North Carolina drivers are in this category). In the models shown, the three categories given are in contrast to the group having alcohol-related violations;
- vehicle age, last two digits of the vehicle model year;
- vehicle curb weight, expressed in hundreds of pounds and including a squared term to capture the curvilinear relationship between vehicle weight and accident involvement.
- rural location, coded “1” for rural locations and “0” for mixed or urban locations;
- curved roadway, coded “1” if curved, “0” otherwise.

Table IX.1: Number of Cases and Model Chi-Squares

Accident type	Passenger cars	Light trucks and vans
N of cases, rollover	2,927	1,052
N of cases, no rollover	11,563	2,392
Total N of cases	14,490	3,444
Model chi-square	1,127.879	210.994
Model degrees of freedom	10	10

Table IX.2: Parameter Estimates

	Passenger cars			Light trucks and vans		
Variable	**Coeff.**	**S.E.**	**Prob.**	**Coeff.**	**S.E.**	**Prob.**
Constant	0.3351	.5040	.5061	–0.4513	.8560	.5980
Driver age	–0.0093	.0019	.0000	–0.0300	.0136	.0271
Age squared	[a]	[a]	[a]	0.0002	.0002	.2468
Male	0.0289	.0469	.5379	0.0041	.1171	.9721
Vehicle year	0.0018	.0046	.6954	–0.0056	.0088	.5293
Nonalcohol violation	–0.0327	.0819	.6899	–0.0752	.1355	.5788
No violation	–0.0230	.0810	.7763	–0.0601	.1351	.6566
Violation unknown	–0.0176	.1063	.8688	0.0397	.1819	.8273
Vehicle weight	–0.1597	.0272	.0000	–0.0132	.0067	.0490
Vehicle weight squared	0.0018	.0005	.0001	[a]	[a]	[a]
Rural location	1.3257	.0654	.0000	1.2492	.1212	.0000
Curved roadway	0.4415	.0444	.0000	0.3572	.0777	.0000

[a] Quadratic term removed since main effect only achieves significance without squared term.

Major Contributors to This Report

Program Evaluation and Methodology Division

Robert E. White, Assistant Director
Beverly A. Ross, Project Manager
Martin T. Gahart, Project Adviser

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20884-6015

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000
or by using fax number (301) 258-4066.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (301) 258-4097 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

United States
General Accounting Office
Washington, D.C. 20548-0001

Bulk Mail
Postage & Fees Paid
GAO
Permit No. G100

Official Business
Penalty for Private Use $300

Address Correction Requested