UMTRI-2011-12 MARCH 2011

INDEPENDENT REVIEW: STATISTICAL ANALYSES OF RELATIONSHIP BETWEEN VEHICLE CURB WEIGHT, TRACK WIDTH, WHEELBASE AND FATALITY RATES

PAUL E. GREEN
LIDIA P. KOSTYNIUK
TIMOTHY J. GORDON
MATTHEW P. REED

UMTRI-2011-12

Independent Review:
Statistical Analyses of Relationship between Vehicle Curb Weight, Track Width, Wheelbase and Fatality Rates

Paul E. Green
Lidia P. Kostyniuk
Timothy J. Gordon
Matthew P. Reed

The University of Michigan
Transportation Research Institute
Ann Arbor, MI 48109-2150
U.S.A.

March, 2011

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UMTRI-2011-12		

4. Title and Subtitle	5. Report Date
Independent Review: Statistical Analyses of Relationship between Vehicle Curb Weight, Track Width, Wheelbase and Fatality Rates	March, 2011
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Green, Paul E., Kostyniuk, Lidia P., Gordon, Timothy J., Reed, Matthew P.	UMTRI-2011-12
9. Performing Organization Name and Address	10. Work Unit no. (TRAIS)
The University of Michigan Transportation Research Institute 2901 Baxter Road Ann Arbor, Michigan 48109-2150 U.S.A.	065442 11. Contract or Grant No. DTNH22-10-C-00185
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
U.S. Department of Transportation National Highway Traffic Safety Administration 1200 New Jersey Ave, SE Washington, D.C. 20590	Special report 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

NHTSA selected the vehicle footprint (the measure of a vehicle's wheelbase multiplied by its average track width) as the attribute upon which to base the CAFE standards for model year 2012-2016 passenger cars and light trucks. These standards are likely to result in weight reductions in new light duty vehicles. As part of its regulatory analysis, the government would like to estimate the effect of the new CAFE standards on safety in terms of crash injuries and fatalities. A number of fairly comprehensive statistical papers have been published analyzing associations between fatality/injury rates and vehicle weight, track width, and wheelbase. Many of the papers arrive at conclusions that are inconsistent.

This report is a review of papers analyzing associations between crash/fatality outcome and vehicle weight and size. The various studies are based on different data sources, model assumptions, and methodologies. The authors of these studies represent a mix of those in government, research institutes, and academia, and have a broad range of professional backgrounds and philosophies. The goal of this report is to provide an independent review of the papers and to critically assess the methods and conclusions presented. The review is independent in the sense that it was conducted by a third party without any interest in the reported outcome. This review focuses on issues such as multicollinearity, data sources, the use of logistic regression, and induced exposure methods. Comments and suggestions are also made with regard to methods used in the various papers.

17. Key Words	18. Distribution Statement
Statistical review, fatality rates, vehicle size, vehicle weight	Unlimited

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	84	

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

Table of Contents

List of Tables

Independent Review: Statistical Analyses of Relationship between Vehicle Curb Weight, Track Width, Wheelbase and Fatality Rates

1. Executive Summary

In 1997, the National Highway Traffic Safety Administration (NHTSA) published a report on the relationships between vehicle size and fatality risk in passenger cars and light trucks.[5] The report was very thorough and detailed. Data derived from various sources were combined into an impressive database, and the data were analyzed extensively using various statistical methods. A 100-pound reduction in the average weight of passenger cars was associated with an estimated increase of 302 fatalities per year. However, a 100-pound reduction in the average weight of light trucks was associated with an insignificant decrease of 40 fatalities. Thus, a significant increase in fatalities was found for weight reduction in passenger cars, but no significant effect was found for light trucks.

In 2002, Dynamic Research, Inc. (DRI) published their findings on the effects of vehicle weight on fatality risk in passenger cars and light trucks.[14] The data sources and statistical methodology used in the DRI report were similar to those used in the NHTSA 1997 report. In fact, the methods chosen were specifically designed to follow those in NHTSA's report. For a 100-pound reduction in the average weight of passenger cars and light trucks, DRI found no overall significant change in fatalities. Throughout the report, statements were made suggesting general good agreement between NHTSA's and DRI's results. Yet, NHTSA's final conclusions suggested a significant result for passenger cars, while DRI's conclusions did not.

It appears that two independent research organizations, using similar data sources and statistical methodology, arrived at different conclusions concerning the overall net change in fatalities. First, the data were not exactly the same. The State data used in the two studies were not precisely from the same states, and the DRI report used more recent data. But if the methodology is robust, and the methods were applied in a similar way, small changes in data should not lead to different conclusions. The main conclusions and findings should be reproducible.

A more plausible explanation for the different results is not that the data were different, but that the statistical methodology was too ambitious. While all the methods presented were designed to improve the estimation process, it could be that certain adjustments and intermediate steps only served to make the estimation process unstable and subject to extra uncertainty. For example, in the two-step aggregate linear regression, results from the Step 1 regression were used to adjust inputs into the Step 2 regression. In the Step 2 regression, additional adjustments were made to force age and gender coefficients to equal the sum of their respective coefficients from two other regression models. These two other models, one logistic and one linear, were fit to induced exposure data taken from a collection of states. All of these intermediate steps and adjustments likely increased the chance of introducing extraneous error into the final conclusions.

Perhaps it is not too surprising that two researchers would arrive at different conclusions under these circumstances. It is recognized that estimating the overall change in fatalities associated with changes in vehicle weight is a difficult task. In fact, the objective of the study is very broad. Estimating the change as it pertains to the entire United States using available data sources likely requires some ambitious assumptions and complex modeling. However, one of the goals in statistical modeling is to find the simplest model with the fewest number of parameters that explains the data well. Such a model will typically lead to improved inference in terms of tighter confidence intervals and hypothesis tests with more power.

In 2003, NHTSA abandoned much of the methodology of the 1997 report, and published updated findings on the association between vehicle weight and fatality risk.[6] In place of the two-step linear regression method, logistic regressions were fit. Curb weight was entered into each model as a two-piece linear variable to account for differences in lighter and heavier vehicles. Based on the new methodology, NHTSA found a greater increase in fatalities associated with a reduction in curb weight for passenger cars than in the 1997 report. In addition, unlike the 1997 report, a significant increase in fatalities was associated with a reduction in curb weight for light trucks. Keeping the models simpler may have led to improved inference.

In the 1997 report and others, six crash types were considered: principal rollover, hit fixed object, hit pedestrian/bicycle/motorcycle, car-to-heavy truck, car-to-car, and car-to-light truck. It is possible that the statistical methodology was too complicated and the number of crash types was too few. Crashes resulting in fatalities tend to be severe high-energy crashes. So the three single-vehicle crash types seem to be well-specified. However, the three multiple-vehicle crash types seem to be too general. Many of the high-impact crashes in the FARS data are opposite direction or head-on crashes. Similarly, FARS data should support analysis of side impact and rear-end crashes. Would the statistical design lead to improved inference if the multiple-vehicle crash types were extended to include these additional ones? It appears that the simpler logistic models incorporated into the 2003 NHTSA report improved inference. Possibly, focusing on additional multiple-vehicle crash types would as well, by reducing variability in the more broadly defined ones.

In 2003, DRI updated the results in their 2002 report.[15] One of the objectives of the 2003 report was to not only estimate the effect of a reduction in curb weight, but to also estimate separate effects of reductions in wheelbase and track width. Some of the results were based on the methods used in the NHTSA 1997 report, but some of the results were based on new methods introduced by DRI. For example, a two-stage logistic regression model was introduced for separating out effects due to vehicle crashworthiness, compatibility, and crash avoidance. However, the two-step aggregate linear regression method, originally proposed by NHTSA in their 1997 study, was retained by DRI for modeling induced exposure involvements per vehicle registration year. Therefore, as in the earlier report, the 2003 DRI model retained complexity with final results depending on the output from the two-stage logistic model and the two-step aggregate linear model. Unlike the 1997 NHTSA study and the 2002 DRI study which were based on similar methods, the methods used in the 2003 reports from both organizations differed substantially.

One of the biggest differences between the 2003 reports from both organizations is that DRI included the three predictor variables curb weight, track width, and wheelbase in the same regression models. As shown in several NHTSA and DRI reports, these three variables tend to have strong positive correlations. It is well-known that inclusion of highly correlated variables generally leads to multicollinearity, which can result in unstable estimation of parameters. If predictor variables are highly correlated and have a strong positive association with the response, those variables are potential surrogates for one another. When entered into separate models one at a time, they generally have strong associations in the same direction. However, if entered together in the same model, the potential exists for the magnitudes of the parameter estimates and associated standard errors to change significantly.

The author of the NHTSA 1997 report was well-aware of the effects of multicollinearity when curb weight, track width, and wheelbase were entered together in the same model. Table 1 shows regression coefficients from fitting logistic regression models with predictor variables entered separately and together for the principal rollover crash type. The measure of risk is fatalities in the crash, relative to induced exposure. When predictor variables were entered separately, each suggested a significant increase in fatality risk associated with a reduction in the measure under investigation.

When the variables were entered together, classic symptoms of multicollinearity became evident. The coefficients for the size variables, track width and wheelbase, were in the *right* direction, but the magnitudes increased considerably. Furthermore, the coefficient for the weight variable changed sign and had a large magnitude. Given the results in Table 1, would it be reasonable to suggest that a 100-pound reduction in curb weight is associated with a reduction in fatality risk while holding track width and wheelbase fixed?

Table 1 Comparison of Regression Coefficients for Weight, Track Width, and Wheelbase when Entered Separately and Together for the Rollover Crash Type (NHTSA 1997)

	Separately		Together	
Measure of Size (Case Car)	Effect per 100 Pound or 1 Inch Reduction (%)	$\chi2$	Effect per 100 Pound or 1 Inch Reduction (%)	$\chi2$
Weight	+ 2.48 per 100	6.5	- 11.10 per 100	not reported
Track Width	+ 10.80 per inch	31.3	+ 18.90 per inch	not reported
Wheelbase	+ 2.96 per inch	17.2	+ 5.34 per inch	not reported

It is interesting that in the 2003 DRI report that includes all three predictor variables, the authors conclude that overall, curb weight reduction tends to decrease the overall number of fatalities, but typical corresponding reductions in wheelbase and track width tend to increase fatalities by a nearly equal amount, and that the overall net change is not statistically significant. It appears that the conclusions presented by DRI in 2003 coincide in some sense with those that resulted in the presence of multicollinearity shown in Table 1.

After observing the effects of including all three predictor variables in the same model, the author of the NHTSA 1997 report made the following comments:

> Couldn't a better case be made by putting all three parameters in the same regression? The problem, of course, is that they are highly intercorrelated: among these 1985-93 passenger cars, the correlation coefficients are .86 for curb weight with track width, .89 for curb weight with wheelbase and .79 for track width with wheelbase. When they are entered simultaneously (C4), it leads to typical "wrong signs" and meaningless results: the "effect" for curb weight is a very large 11.1 percent per 100 pounds, in the wrong direction, while the effects for track width and wheelbase, while in the right direction, are double the values in C2 and C3. At least, the results are so obviously wrong that the analyst will not be tempted to rely upon them. (Kahane, 5, p. 46, first paragraph)

Regression is not designed to separate out the effects from highly correlated variables. It does not engage in *intelligent* variable selection. No distinction is made between curb weight, track width, and wheelbase, other than they are three predictor variables being included in the same model. Note that in this problem, there are not just two highly correlated predictors, there are three. When two columns of a design matrix in a regression model are close to being linear combinations of one another, the design is *ill-conditioned*, and the estimation process is unstable. The variance inflation factor (VIF), referenced in some of the reports, is commonly used to measure collinearity among predictors.

The high correlation observed in these three variables may be an artifact of the use of historical data. In the future mix of vehicles that make up the on-road fleet in the United States, the observed correlation may decrease. However, considering the effect that multicollinearity has on the estimation process when fitting regression models, the practice of including variables that are known to be correlated should be guarded against. For each particular regression, one remedy is to use the one variable with the strongest association.

After the 2003 NHTSA and DRI reports were made publicly available, additional reports and documents were published by both organizations. Most of them focused on responding to criticisms and defending results published in earlier reports. In 2004, DRI published a report reviewing results in the 1997 and 2003 NHTSA reports, along with the DRI 2002 and 2003 reports.[16] Also, in 2004, NHTSA responded to three criticisms from outside sources in Docket NHTSA-2003-16318-16.[7] In 2005, DRI defended their findings and responded to comments made in the NHTSA docket.[17] In 2009, DRI presented comments on what they considered to be misstatements or misinterpretations in regard to a proposed rulemaking procedure.[18] In 2010, NHTSA made available certain pages of a final regulatory impact analysis about relationships between fatality risk, mass, and footprint.[8]

In some of the reports after 2003, DRI presented results suggesting that they could reasonably approximate some of NHTSA's findings if certain data and model assumptions were made. The assumptions were based on the use of specific data years, logistic regression models, and restricting the analysis to 4-door non-police cars. Similarly, NHTSA presented results suggesting that they could reasonably approximate some of DRI's findings if they included track width and wheelbase variables into their models, in addition to curb weight. NHTSA used its 2003 database and methods that were slightly different than DRI's.

In general, we believe that simpler is better. Simple and parsimonious models generally lead to improved inference, as long as the data and model assumptions are appropriate. In that regard, the disaggregate logistic regression model used by NHTSA in the 2003 report seems to be the most appropriate model. In the context that it was used, it is a valid exposure-based risk model for the analysis of rates. In some sense, it could be regarded as too simple, as described below. However, we believe that it can be used to find general associations between fatality risk and mass, and that the general directions of the reported associations are correct. The two-stage logistic regression model in combination with the two-step aggregate regression used by DRI seems to be more complicated than is necessary based on the data being analyzed. Summing regression coefficients from two separate models to arrive at conclusions about the effects of reductions in weight or size on fatality rates seems to add unneeded complexity to the problem.

Finally, a few comments are made regarding the use of induced exposure and logistic regression. The NHTSA and DRI reports both relied on the method of *induced exposure*. Induced exposure vehicles are generally the non-culpable vehicles in two-vehicle crashes and were derived from various State data files. In the absence of a traditional exposure measure, such as vehicle miles traveled (VMT), induced exposure is a surrogate that represents the denominator of a rate. Admittedly, there are no other sources of exposure data available that are recorded at the level required to analyze fatality rates in the studies reviewed. In the NHTSA 2003 report, a novel approach was used whereby vehicle registration data and odometer readings were used to apportion vehicle miles traveled to each induced exposure crash. In the absence of viable alternatives, the approach seems logical. However, there is a concern that the method could introduce bias in certain situations. For example, non-culpable vehicles tend to have very different speed distributions than vehicles involved in fatal crashes. The authors of the studies seem to be aware of these and other differences and attempts were made to adjust for potential bias. The use of induced exposure that is limited to certain states is likely to be an issue for further investigation as long as other sources of exposure such as VMT remain unavailable.

Another question of interest is whether disaggregate logistic regression is an appropriate model for analyzing fatality risk. Both NHTSA and DRI used this model in one form or another in several reports. Logistic regression is not one of the standard exposure-based risk models for analyzing rates. However, when rates are very small, as is the case when fatalities are relatively rare and the induced exposure denominators are large, the model approximates the Poisson log-linear model for rates, which is a standard exposure-based risk model. However, in practice the Poisson model is generally too simple for use in observational studies.

As stated above, we feel that the model is adequate but that it may be too simple. We claim that simple is good, as long as the data and model assumptions are appropriate. Likelihood-based tests, derived from fitting logistic and Poisson models, tend to be significant even when results show small effects, as long as sample sizes are large enough. Construction of confidence intervals and tests of hypotheses depend on specification of a model that accommodates the variation in the data. The study under consideration is an observational one using various sources of data, and it could be argued that the logistic model is somewhat misspecified. In the presence of extra-variation, standard errors tend to be too small and significance can be overstated. A more robust model would at least adjust standard errors to account for the extra-variation often encountered in observational studies. In Section 3.2 and 3.4, alternative models and methods are described that could be used to account for the extra-variation that was likely present in the data analyzed.

In addition to the NHTSA and DRI reports, several other papers were written about the effects of vehicle weight and size on safety. Wenzel [19, 20] and Wenzel and Ross [21, 22, 23], published a series of papers addressing associations between crash risk, weight, and size. Much of their work focused on certain passenger car and light truck model types. While the papers contribute to understanding some of the relationships between risk, weight, and size, the statistical methods presented appear to be too simple to adequately describe associations with a great degree of precision. No doubt, some of the papers describe findings that are generally in the *right* direction. However, least squares linear regression models, without modification, are not exposure-based risk models and are generally not used to analyze fatality or casualty risk. For the most part, inference drawn from these models tends to be weak since they do not account for differences in exposure measures in the denominators of the rates. The R-squared measures describing overall fit that are presented are not the preferred measures in a rates analysis. Estimated relative risks are more useful for assessing the effects of size and weight variables on fatality or injury risk.

Two papers by J.P. Research [4,11] and one paper by Nusholtz et al. [10] were reviewed based on underlying engineering principles for vehicles involved in frontal crashes. The 2009 J.P. Research paper focused on the difficulties associated with separating out the contributions of weight and size variables when analyzing fatality risk. This paper properly recognized the problem arising from multicollinearity. The authors also include a clear explanation of why fatality risk is expected to increase with increasing mass ratio. The positive fatality rate increases associated with a 100-pound weight reduction in vehicle weight estimated by Kahane and JP Research are broadly more convincing than the 6.7 percent reduction of fatalities reported by DRI.[17]

For the Nusholtz et al. paper, the focus is again on frontal crashes, but now restricted to a population of passenger cars only. Although limited in scope, their model addresses the question of whether vehicle size can reasonably be the dominant vehicle factor for fatality risk. It is found that changing the mean mass of the vehicle population (leaving variability unchanged) has a stronger influence on fatality risk than corresponding (feasible) changes in mean vehicle dimensions. If one accepts the methodology, there is an unequivocal conclusion that reducing vehicle mass while maintaining constant vehicle dimensions will increase fatality risk, and this conclusion is robust against realistic changes that may be made in the force vs. deflection characteristics of the impacting vehicles.

Finally, two papers by Robertson, one a commentary paper, and the other a peer-reviewed journal paper were reviewed.[12,13] Considering the title of the commentary paper, *Blood and Oil: Vehicle Characteristics in Relation to Fatality Risk and Fuel Economy,* an agenda in favor of lighter vehicles can be inferred. Some of the claims in the paper appear to be overstated. One of the claims is that half the deaths involving passenger cars, vans, and SUVs could have been prevented if all vehicles had crashworthiness and stability equal to those of the top rated vehicles. Considering the complex nature of the events associated with fatal crash involvement, and the simple statistical models upon which the result is based, this is a very ambitious claim. Other claims are that fatality rates would have been reduced by 28 percent and fuel use reduced by 16 percent if vehicle weights had been reduced to the weight of vehicles with the lowest weight per size. Intermediate results and more documentation would help the reader determine if these claims are valid. Separate models are not fit according to crash type, and passenger cars, vans, and SUVs are included in the same model. The second paper follows on from the first paper except that curb weight is not fit and fuel economy is used as a surrogate. The effects of electronic stability control (ESC) are a major focus of the second paper.

2. Introduction

In December 2007, Congress passed the Energy Independence and Security Act (EISA) that required NHTSA to set "attribute-based" Corporate Average Fuel Economy (CAFE) standards, in which a manufacturer's compliance obligation depends on the mix of vehicles they produce for sale. NHTSA selected the vehicle footprint (the measure of a vehicle's wheelbase multiplied by its average track width) as the attribute upon which to base the CAFE standards for MYs 2012-2016 passenger cars and light trucks. These standards are likely to result in weight reductions in new passenger cars and light trucks.

As part of its regulatory analysis, the government would like to estimate the effect of the new CAFE standards on safety in terms of crash injuries and fatalities. One approach is to use relationships between fatality and injury rates and weight or size attributes such as curb weight, track width, and wheelbase from past statistical analyses and apply them to the future fleets. A problem with this approach, however, is that although a considerable number of studies on this topic have been published, their results are not consistent. Some studies report an increase in fatalities with vehicle weight and others report a decrease. Still other studies point out that other elements of vehicle design are better related to fatality rates than weight. The inconsistency of results from these studies is not surprising, in that the assumptions, databases, statistical methods, and variables vary considerably across the studies. Another problem with this approach is that statistical analyses of historic data capture the relationships between vehicle characteristics and safety from the time in which the data were generated. Innovations in materials, changes in vehicle design, more crash avoidance technology, and advances in occupant protection systems will influence fatality and injury risks in vehicles of the future. Thus, it is important that methods for estimating future vehicle safety do not rely strictly on past historic relationships, but also consider changes in vehicle design and technology.

Recognizing these problems, and wishing to be able to estimate the effect of the new CAFÉ standards on safety, NHTSA sought an independent review of a set of statistical analyses of relationships between vehicle curb weight, the footprint variables (track width, wheelbase) and fatality rates from vehicle crashes. The purpose of this review is to examine analysis methods, data sources, and assumptions in a set of previous statistical studies, with the objective of identifying the reasons for the differences in results. Another objective is to examine the suitability of the various methods for estimating the fatality risks of future vehicles.

The University of Michigan Transportation Research Institute (UMTRI) undertook this assignment. We reviewed a set of papers, reports, and manuscripts provided to us by NHTSA (see list in Appendix A) and examined the statistical analyses of relationships between crash or fatality rates and vehicle properties such as curb weight, track width, wheelbase and other variables.

First, we wish to acknowledge the effort undertaken by the authors of the reviewed reports who addressed the effects of weight and size on fatality risk for passenger cars and light trucks. This is a very difficult topic to tackle, with many sources of uncertainty that typically arise in an observational study. We recognize that the researchers devoted much time and energy arriving at their conclusions, and it is clear that much thought went into developing the methods, considering the limited data sources available for analysis.

The well-known statistician George Box is often credited with the quote: “All models are wrong, some are useful.” Box was likely referring to the idea that statistical models are based on underlying assumptions, and that validity of the inference and conclusions drawn from a particular model depends on the underlying assumptions that must be made before any statistical analysis can begin. These assumptions often have to do with choosing a particular probability distribution that represents the physical mechanism that generated the study data to be modeled.

In addition to the statistical model and its underlying assumptions, statistical analysis also depends on quality and choice of the data, how the data are sampled, sample size, types of bias, and design of the experiment which may include some form of randomization. These decisions are generally made before data analysis begins and are chosen according to certain criteria, such as increasing the power of statistical tests of hypotheses. Thus, applied statistics is an art form, and various choices and decisions are required by the investigators. For this reason, the statistical community occasionally comes under criticism, especially when different investigators arrive at different conclusions about the same research topic.

This report summarizes our review of the studies examined and is organized as follows. The next section reviews a series of reports from 1997 to 2010 by Kahane of NHTSA. Section 4 reviews a series of reports by Van Auken and Zellner from Dynamic Research, Inc. (DRI). Section 5 reviews a series papers by Wenzel and Ross, and Section 6 reviews two papers by J. P. Research and one paper published by three authors from Daimler Chrysler Corporation. Section 7 is devoted to the review of two papers by Robertson. Conclusions and final comments appear in the last section.

3. Review of the National Highway Traffic Safety Administration (NHTSA) Reports

3.1 The NHTSA 1997 Report

Relationships between Vehicle Size and Fatality Risk in Model Year 1985-93
Passenger Cars and Light Trucks [5]
DOT HS 808 570
January 1997
C. Kahane

Summary

The objective of the 1997 report was to estimate the relationship between curb weight and the fatality risk, per million vehicle exposure years, for model year 1985-93 passenger cars and light trucks based on their crash experience in the United States from 1989 through 1993. The goal was to find the net effect on society. That is, fatality risk includes fatalities to all occupants of motor vehicles, pedestrians, and bicyclists. Estimates were obtained for six crash types:

- Principal rollovers
- Collisions with objects
- Collisions with pedestrians, bicycles, or motorcycles
- Collisions with heavy trucks (GVWR greater than 10,000 pounds)
- Collisions with passenger cars
- Collisions with light trucks (pickups, SUVs, or vans)

Based on the methods of the study, Table 2 shows the estimated net fatality change associated with a 100-pound weight reduction for passenger cars. Overall, the estimated net change was an increase in 302 fatalities with confidence bounds suggesting the result was significant.

Table 2 Effect of 100 Pound Weight Reduction for Passenger Cars (light truck weights unchanged), NHTSA 1997 [5]

Crash type	Fatalities in 1993 Crashes	Effect of 100-Pound Weight Reduction	Net Fatality Change
Principal rollover	1,754	4.58%	+80
Hit object	7,456	1.12%	+84
Hit ped/bike/motorcycle	4,206	-0.46%	-19
Hit big truck	2,648	1.40%	+37
Hit another car	5,025	-0.62% NS	-31
Hit light truck	5,751	2.63%	+151
Overall	26,840	1.13%	+302
2-sigma confidence bounds			(214, 390)
3-sigma confidence bounds			(170, 434)

Similarly, based on the methods of the study, Table 3 shows results for light trucks. Overall, the estimated net change was a decrease in 40 fatalities; however, in this case confidence bounds suggest the result was not significant.

Table 3 Effect of 100 Pound Weight Reduction for Light Trucks (car weights unchanged), NHTSA 1997 [5]

Crash type	Fatalities in 1993 Crashes	Effect of 100-Pound Weight Reduction	Net Fatality Change
Principal rollover	1,860	0.81% NS	+15
Hit object	3,263	1.44%	+47
Hit ped/bike/motorcycle	2,217	-2.03%	-45
Hit big truck	1,111	2.63%	+29
Hit passenger car	5,751	-1.39%	-80
Hit another light truck	1,110	-0.54%NS	-6
Overall	15,312	-0.26%	-40
2-sigma confidence bounds			(-100, 20)
3-sigma confidence bounds			(-130, 50)

Chapters 2 and 3 of the NHTSA report present logistic regression analyses of fatalities per 100 induced-exposure crashes, based on accident data from 11 States. In Chapter 4, induced exposure crashes per 1000 vehicle years were analyzed using aggregate weighted least squares (WLS) linear regressions in two steps. The analyses are based on data from 11 States and Polk registration data. In Chapters 5 and 6 of the NHTSA report, FARS data and Polk registration data for the entire United States were analyzed to estimate fatality rates per million vehicle years. These analyses were also performed using WLS aggregate linear regressions in two steps. The primary findings of the 1997 NHTSA report were presented in Chapters 5 and 6.

Data

It is recognized that creating the database for the analyses in this study was a formidable task. Data were derived from various sources.

- FARS – fatality case involvements
- State data from 11 states – induced exposure involvements
- R.L. Polk data – vehicle registrations
- Other sources – curb weight, track width, wheelbase

These are most likely the best sources of data available for conducting this study. It is well-known that good exposure data are not recorded at the level needed for the analyses presented. It appears that these data were appropriate for answering the research questions under investigation, and that the data reduction techniques applied to FARS and State data were reasonable. Coding of State data can vary between states. As a note, other databases such as GES and NMVCCS have an accident type variable that makes classification of crash types relatively straightforward compared to FARS.

Review of Chapter 3 in the NHTSA 1997 Report

Logistic regression models were used to estimate fatality risk per 1000 induced-exposure crashes according to curb weight, track width, wheelbase, and other control variables. The design is similar to a case-control study in which cases are assigned the value 1 and controls are assigned the value 0. For this study, the cases were fatal involvements and the controls were induced exposure involvements. Parameter estimates in logistic regression models have interpretations as log odds ratios. For example, when curb weight is included as a predictor variable, the model can be used to estimate the change in the odds of fatality when curb weight decreases by 100 pounds. The relative risk, which is a ratio of rates, is the usual exposure-based risk measure used for analyzing rates. However, the disaggregate logistic model used here should provide a good measure of *risk*, even though the measure being produced is a ratio of odds. This is true as long as fatalities are rare relative to induced exposure (when rates are small), which is the case for the data being analyzed.

Some discussion is provided about the importance of the control variables age and gender. Based on diagnostic plots presented earlier in the report, Kahane created a type of interaction variable between age and gender. The variable incorporates information about gender and ages 35, 45, and 50. This appears to be an appropriate procedure to adjust for age and gender in the model. It is also good to center the variables, as was done.

An example logistic regression is presented for passenger cars in rollovers (Section 3.3, page 41). There were 971 principal rollover crashes resulting in 1,036 fatalities. Since there were more fatalities than crashes, there were multiple fatalities in some vehicles. One of the assumptions of logistic regression is that observations are independent. Treating multiple fatalities in the same vehicle as separate observations ignores the correlated outcome and increases the sample size. The resulting effect is that standard errors of parameter estimates tend to be too small and significance can be overstated. For example, the chi-square value attached to the curb weight coefficient is 6.499 with an associated p-value of 0.0108. While this result is significant at the 0.05 level, it would not be at the stricter 0.01 level. In addition, the reported standard error is likely ambitious and too small. It is hard to know exactly what effect the correlated outcomes have on the final results, except that the p-value would be greater than 0.0108.

Logistic regressions were fit for other crash types and results are reported in Table 3-2. The discussion of correlated outcomes in the preceding paragraph could be relevant to some of the findings. It would depend on how many vehicles were involved in crashes with multiple fatalities. For some of the crash types such as frontal-fixed object (chi-square=6.53) and pedestrian/bicycle/motorcycle (chi-square=3.45) adjustment of standard errors due to correlated outcomes could lead to different conclusions.

<u>Inclusion of Curb Weight, Track Width, and Wheelbase</u>

One of the most interesting aspects of Table 3.2 as it relates to this study is that Kahane considered including the variables curb weight, track width, and wheelbase both separately and together in the principle rollover model (C1-C4, p.45). When each variable was entered one at a time, the effect of a reduction in 100 pounds of weight, or a 1 inch reduction in either track width or wheelbase, increased fatality risk. However, when all three variables were included together in the same model, a reduction in curb weight suggested decreased fatality risk, while reductions in both track width and wheelbase suggested increased fatality risk. Kahane made the following comments:

> Couldn't a better case be made by putting all three parameters in the same regression? The problem, of course, is that they are highly intercorrelated: among these 1985-93 passenger cars, the correlation coefficients are .86 for curb weight with track width, .89 for curb weight with wheelbase and .79 for track width with wheelbase. When they are entered simultaneously (C4), it leads to typical "wrong signs" and meaningless results: the "effect" for curb weight is a very large 11.1 percent per 100 pounds, in the wrong direction, while the

> effects for track width and wheelbase, while in the right direction, are double the values in C2 and C3. At least, the results are so obviously wrong that the analyst will not be tempted to rely upon them. (Kahane, 5, p. 46, first paragraph)

These comments imply that Kahane had encountered the issue of multicollinearity early on in the 1997 report and was well-aware of it. Furthermore, his statements indicated that he would not rely on such results.

Regressions for light trucks in Section 3.5 proceed in a similar manner as those for passenger cars.

In multi-vehicle crashes, the use of standard logistic regression is more complicated. In this situation correlated outcomes result due to occupants in the same vehicle, and vehicles in the same crash. This is a concern for crash types such as car-to-car, truck-to-truck, and car-to-truck involvements. In this case, it appears that Kahane identifies one passenger vehicle as the “case” vehicle for inclusion in the regression model; however, he recognizes the preferred method is to analyze the effects of the weight, driver age, etc. for *both* vehicles, and defers to Sections 3.6 – 3.8 where this is done. (Kahane, 5, p. 47, last paragraph)

In Section 3.6, regressions of car-to-car crashes are performed where pairs of vehicles in crashes are modeled. Again, it appears that each fatal occupant in either car was entered in the regression. For example, if there were two fatalities in the case car and one fatality in the other car, three separate observations were created for entry into the logistic regression model. Note that these three observations are correlated since they represent two occupants in the same vehicle, and three occupants in the same crash. Treating these fatalities as independent observations in logistic regression violates the independence assumption, since they are not independent. Again, the result is that standard errors of parameter estimates tend to be too small and significance is overstated. The degree of overstatement depends on the number of crashes with multiple fatalities, which cannot be determined from information in the report.

Review of Chapter 4 in the NHTSA 1997 Report

The objective of this chapter was to estimate the extent of size-related bias in fatality rates relative to induced exposure. The strategy was to model induced-exposure rates as a function of vehicle weight, controlling for driver age and gender. If the induced exposure rate is constant across vehicle weights, then induced exposure may be considered an unbiased surrogate for exposure. Polk data were collected from the same 11 States used for induced exposure data. Plots in Figure 4-1 through Figure 4-5 show that for various vehicle types, rates tend to decrease with curb weight, except for vans.

In Section 4.4 regression analyses were conducted with the log rate as the dependent variable. The numerator of the rate was induced exposure crashes and the denominator was vehicle registration years. Weighted Least Squares (WLS) regressions were performed on aggregated data in two stages. In the first regression, the log rate was regressed on vehicle age, state, and calendar year. This regression was used to provide weights for induced exposure crashes in the second regression.

The first regression was weighted by vehicle registration years, but no explanation for weighting by the denominator was given. The Poisson log-linear model is a standard model for the analysis of rates, where counts in the numerator are assumed to follow the Poisson distribution, and log exposure in the denominator is assumed fixed and treated as an *offset*. The model, fit by the method of maximum likelihood, leads to parameter estimates that have interpretations as log relative risks (RRs). For data collected in an observational setting, and not from a controlled experiment, data are often more variable than assumed by Poisson sampling. The Poisson distribution has only one parameter, and the mean is restricted to equal the variance. This restriction generally leads to standard errors of parameter estimates that are too small, especially for large samples. For this reason, researchers have considered alternative models for analyzing rates, such as negative binomial regression, random effects models, or even Bayesian models.

Kahane uses normal theory regression to model rates. This makes good sense, especially because the normal model has two parameters – a location parameter for modeling the mean log rate, and a scale parameter for adjusting standard errors. Unlike the one-parameter Poisson model, the two-parameter normal model estimates the mean and variance independently, and standard errors of parameter estimates can be inflated to account for extra variation. Tests of hypotheses and confidence intervals depend on estimation of the scale parameter.

A standard WLS model for the analysis of rates, however, uses the counts in the *numerator* as weights. This model is asymptotically equivalent to the Poisson model estimated by maximum likelihood. As long as counts in the numerator are sufficiently large, results will be similar. In addition, the WLS model adjusts standard errors due to estimation of the scale parameter. Kahane uses the denominator (vehicle registration years) as weights which tend to be much larger than the counts in the numerator. What is the rationale for weighting by the denominator?

In the Step 2 WLS regression, the numerator of the rate is adjusted based on results from the Step 1 regression, and the regression is again weighted by vehicle registration years. The Step 2 regression includes curb weight, driver age and gender, and other control variables as predictors.

The purpose of these aggregated WLS regressions was to adjust for biases introduced by using induced exposure as the measure of exposure. On page 78, the uncorrected results of Chapter 3 were compared to the corrected results using the two-step regression method. The adjusted amounts were 0.27 percent per 100 pounds for cars, and 2.50 percent for trucks. The result for light trucks was much larger.

Review of Chapter 5 in the NHTSA 1997 Report

In the second paragraph of Chapter 2 on page 15, the author states that the primary findings of the 1997 NHTSA report are those of Chapters 5 and 6. In Chapter 5, fatalities are expanded to include fatalities from all states, and exposure is expanded to include vehicle registration years in all states. Induced exposure averages from 11 states were used to estimate driver age and sex. States were clustered into five groups according to fatality rate to control for state group. It was assumed that the distributions of driver age and gender in the 11 States were representative, in relative terms, of the general driving public in the United States. Induced exposure crashes were weighted to give each of the 11 States a contribution proportional to its share of vehicle registrations. Before regressions were performed, a series of unadjusted log fatality rates were plotted against weight for passenger cars and light trucks.

Two-step aggregate linear regressions were performed. Step 1 regressed log rate on vehicle age, state group and calendar year. Vehicle registration counts were adjusted by the results in Step 1. Step 2 regressed log rate on curb weight or track width and remaining variables. Many of the variables were averages derived from induced exposure results. Both regressions were weighted by vehicle registration years (the denominator of the rate). An explanation by the author for using these weights would be helpful since WLS regression models for rates typically weight by the numerator (the count) of the rate. [see, for example, 1, p. 600-604]

In the aggregation procedure, an attempt was made to ensure at least 5 expected fatalities per cell. Some zero cells were encountered. One wonders if the results were sensitive to some sparse cells.

In the Step 2 regressions, the author notices that something went wrong. Many coefficients were either insignificant or in the wrong direction. The problem was attributed to correlation between curb weight and the driver age variables. A method for treating driver age and gender as exogenous variables was described in which they would be held fixed. Coefficients estimated for the age and gender variables from Chapter 3 and 4 would be summed and would be forced into the Step 2 regression. This would be accomplished by adjusting registration years a second time. At this point, it appears the model has become more complicated than is necessary.

Review of Chapter 6 in the NHTSA 1997 Report

A type of exogenous age and sex method is used to force the regressions to perform in a particular way. There are also various kinds of adjustments being made to the denominator of the rate. It appears several adjustments are being made. It is very difficult for the reader to keep track of all the adjustments, weighting factors, and special considerations being given certain variables. The model should be as simple as possible in order to understand the effects of curb weight or track width on fatality rates. It appears that unsatisfactory results were found, and now an ad-hoc procedure has been implemented to explain shortcomings of the modeling procedure. It is difficult for the reader to follow the reasoning.

3.2 The NHTSA 2003 Report

Vehicle Weight, Fatality Risk and Crash Compatibility of
Model Year 1991-99 Passenger Cars and Light Trucks [6]
DOT HS 809 662
October 2003
C. Kahane

Summary

The 2003 statistical analysis of MY 1991-99 vehicles in CY 1995-2000 crashes supersedes NHTSA's 1997 report. Logistic regressions were used to estimate fatality rates per billion miles. Crash fatality rates included fatalities to occupants of the case vehicle, occupants of the other vehicles it collided with, and any pedestrians. Unlike the 1997 report, the 2003 report provides separate estimates for passenger cars and LTVs (pickup trucks, SUVs, minivans, full-sized vans) according to two weight groups as shown below:

Vehicle types:

- Passenger cars 4-door, non-police (< 2,950 pounds)
- Passenger cars 4-door, non-police (>= 2,950 pounds)
- LTVs < 3,870 pounds
- LTVs >= 3,870 pounds

The six fundamental crash types are the same ones considered in the 1997 report.

Crash types:

- Principal rollover
- Fixed object
- Ped/bike/motorcycle
- Heavy truck
- Car
- Light truck

Table 4 (reproduced from Kahane, [6]), shows the average fatality increase per 100-pound reduction in LTVs according to each crash type. Interval estimates that contain zero are judged to be insignificant. For light trucks weighing 3,870 pounds or more, results show that increases in fatalities for the single-vehicle crashes were generally offset by decreases in fatalities for the multiple-vehicle crashes. The overall effect for the heavier LTVs was insignificant.

Table 4 Fatality Increase per 100-Pound Weight Reduction, Light Trucks [6]
(Baseline=CY 1999 total fatalities, MY 1996-99/CY 1996-2000 fatality distribution)

Light Trucks Weighing 3,870 Pounds Or More		Effect (%) of 100-Pound Reduction		Annual Net Fatality Change	
Crash type	Annual Baseline Crash Fatalities	Point Estimate	Interval Estimate	Point Estimate	Interval Estimate
Principal rollover	2,183	2.56	(0.81, 3.94)	56	(18, 86)
Fixed Object	2,639	3.06	(1.41, 4.34)	81	(37, 115)
Ped/bike/motorcycle	2,043	0.13	(-1.56, 1.45)	3	(-32, 30)
Heavy truck	860	0.62	(-1.61, 2.48)	5	(-14, 21)
Car	5,186	-0.68	(-1.79, 0.06)	-35	(-93, 3)
Light truck < 3,870	1,010	-1.50	(-3.20, -0.17)	-15	(-32, -2)
Light truck 3,870+*	784	-3.00	(-6.40, -0.34)	-24	(-50, -3)
Overall	**14,705**	**0.48**	**(-1.06, 1.64)**	**71**	**(-156, 241)**

Light Trucks Weighing Less Than 3,870 Pounds		Effect (%) of 100-Pound Reduction		Annual Net Fatality Change	
Crash type	Annual Baseline Crash Fatalities	Point Estimate	Interval Estimate	Point Estimate	Interval Estimate
Principal rollover	1,319	3.15	(0.64, 4.30)	42	(8, 57)
Fixed Object	1,687	4.02	(1.71, 4.97)	68	(29, 84)
Ped/bike/motorcycle	1,148	1.24	(-1.26, 2.38)	14	(-14, 27)
Heavy truck	584	5.91	(3.10, 7.36)	35	(18, 46)
Car	2,062	1.13	(-0.92, 1.82)	23	(-19, 38)
Light truck < 3,870	247	6.98	(1.92, 9.32)	17	(5, 23)
Light truck 3,870+*	1,010	3.49	(0.96, 4.66)	35	(10, 47)
Overall	**8,057**	**2.90**	**(0.73, 3.67)**	**234**	**(59, 296)**

*Assumes both light trucks in the collision were reduced by 100 pounds.

In the lower portion of the table, overall results were significant for LTVs weighing less than 3,870 pounds. The estimated increase in fatalities was 234 with a confidence interval of (59, 296). For every crash type there was a positive estimate, although two of them were judged to be insignificant.

Results for passenger cars are shown in Table 5 (reproduced from Kahane, [6]). Much stronger effects were produced for passenger cars than for LTVs. For cars weighing 2,950 pounds or more, there was a significant estimated increase in fatalities for each crash type, except for ped /bike/motorcycle. Overall, the increase was estimated at 216 fatalities.

Table 5 Fatality Increase per 100-Pound Weight Reduction, Passenger Cars [6]
(Baseline=CY 1999 total fatalities, MY 1996-99/CY 1996-2000 fatality distribution)

Cars Weighing 2,950 Pounds or More		Effect (%) of 100-Pound Reduction		Annual Net Fatality Change	
Crash type	Annual Baseline Crash Fatalities	Point Estimate	Interval Estimate	Point Estimate	Interval Estimate
Principal rollover	715	4.70	(2.40, 7.00)	34	(17, 50)
Fixed Object	2,822	1.67	(0.63, 2.71)	47	(18, 76)
Ped/bike/motorcycle	1,349	-0.62	(-1.83, 0.59)	-8	(-25, 8)
Heavy truck	822	2.06	(0.67, 3.45)	17	(6, 28)
Car < 2,950	1,342	1.59	(0.70, 2.48)	21	(9, 33)
Car 2,950+*	677	3.18	(1.40, 4.96)	22	(9, 34)
Light truck	3,157	2.62	(1.74, 3.50)	83	(55, 110)
Overall	**10,884**	**1.98**	**(1.19, 2.78)**	**216**	**(129, 303)**

Cars Weighing Less Than 2,950 Pounds		Effect (%) of 100-Pound Reduction		Annual Net Fatality Change	
Crash type	Annual Baseline Crash Fatalities	Point Estimate	Interval Estimate	Point Estimate	Interval Estimate
Principal rollover	995	5.08	(0.87, 7.55)	51	(9, 75)
Fixed Object	3,357	3.22	(0.25, 4.45)	108	(8, 149)
Ped/bike/motorcycle	1,741	3.48	(0.22, 5.00)	61	(4, 87)
Heavy truck	1,148	5.96	(2.50, 7.68)	68	(29, 88)
Car < 2,950	934	4.96	(-0.72, 7.16)	46	(-7, 67)
Car 2,950+*	1,342	2.48	(-0.36, 3.58)	33	(-5, 48)
Light truck	4,091	5.63	(2.85, 6.67)	230	(117, 273)
Overall	**13,608**	**4.39**	**(1.66, 5.25)**	**597**	**(226, 715)**

*Assumes both cars in the collision were reduced by 100 pounds.

The lower portion of the table shows that the largest increase in fatalities was associated with cars weighing less than 2,950 pounds. Every point estimate was positive and all but two of the confidence intervals were significant. The two crash types that were insignificant were the car-to-car involvements. The 2003 report estimates a substantially larger fatality increase per 100-pound weight reduction than the NHTSA 1997 report.

Results were also presented for fatality rates by vehicle type, and an investigation into car-light truck compatibility was conducted.

Data

The database combines information from FARS, R.L. Polk registration data, State crash data, NASS CDS, and other sources. It appears that the database was created at the vehicle level. That is, each record corresponds to one vehicle. Curb weight by make-model and model year was derived from seven sources. Fatalities and many control variables were taken from the FARS data. Induced exposure crashes were derived from State data files which contain information about age and gender.

- FARS 1995-2000 – case fatalities
- R.L. Polk registration data – vehicle registrations
- State crash data from 8 states – induced exposure
- NASS CDS – odometer readings
- Curb weight derived from seven sources

One of the major obstacles to performing a rates analysis in transportation-related studies is that there are few sources of exposure data that form the denominator in a rate. Vehicle miles traveled (VMT) and Average Annual Daily Traffic (AADT) are two common measures of exposure used when available. It is well-known that no established exposure-based data files exist at the level required for the analysis being conducted. The author uses an approach by considering induced exposure vehicles, which are non-culpable vehicles in multiple-vehicle crashes.

For a vehicle with particular make-model, body style, model year, and calendar year (eg. in 1998 a 1997 Ford Taurus), the author uses vehicle registration data and induced exposure crashes from eight states to calculate the ratio of the number of registered vehicles to induced exposure crashes in each state. The resulting ratio measure is called *vehicle years*. Based on registrations for that particular vehicle in the entire United States, vehicle years are weighted to represent national totals. Therefore, each vehicle type, based on make-model, body style, model year, and calendar year, is assigned a measure of exposure that represents the number of registered vehicles per induced exposure crash in the United States. Vehicle miles of travel are also apportioned to each induced exposure crash, based on annual mileage by vehicle age and class using NASS CDS data.

Considering the lack of publicly available sources of exposure data needed in a study of this kind, the method of induced exposure used by the author seems appropriate. Although many sources of data were combined to create the database used in this study, it is assumed that the data were robust for the regression analyses performed.

Review of Chapter 3 in the NHTSA 2003 Report: Passenger Cars

The analysis was limited to 4-door cars, excluding police cars. Some concerns regarding 2-door sporty or muscle cars were expressed. This appears to be reasonable since exclusion of those vehicles could reduce unexplained variation in results. In addition, Kahane found no significant association between annual mileage and curb weight for 4-door non-police cars.

Before proceeding with regression, exploratory plots were made showing relationships between log fatality rates and curb weight overall, and by the six crash types (Figures 3-1 through 3-7). These plots are compelling because they were based on aggregated data where fatalities and vehicle years were summed into cells. The aggregated fatality rates were generally stable because each point in the plot was based on possibly hundreds, if not thousands, of fatalities. In addition, the denominators of the rates were also sums of many vehicle years. Therefore, rates such as these generally have small variances and are reliable estimates of risk. The trends in these plots are clear. In every crash type, for curb weight up to about 3,500 pounds, crash rates decrease as curb weight increases. In most of the crash types, the trend is linear and decreasing, even after 3,500 pounds. For passenger cars, these plots of aggregated crash rates provide strong evidence that crash rates decrease as curb weight increases.

After demonstrating these strong associations in the various plots, it is highly unlikely that controlling for additional variables such as driver age, driver gender, rural/urban area of the crash, speed limit, time of day, or others, will alter the basic associations. It is likely that these control variables will affect the association between fatality risk and curb weight to some degree, but it is unlikely that adjustment for control variables will render curb weight insignificant or reverse the direction of association. Usually, only an extreme confounding variable that is strongly associated with both fatality risk and curb weight could alter the basic relationships shown.

Additional plots restricted to female drivers between 30 and 49 years old were also shown to eliminate certain age and sex effects. These plots are also very compelling because they are plots of aggregated rates and tend to confirm that the fatality rate decreases as curb weight increases. This is the case in all six crash types.

In Section 3.3, variables were screened for inclusion in the models. Age/gender interaction terms were created that were centered around 50 years. Centering predictor variables is a good idea for reducing correlation among the predictors.

In Section 3.4, logistic regressions were fit. The design can be viewed as a case-control study in which each fatality is a case and assigned the value 1, and each induced exposure involvement is a control and assigned the value 0. This design is similar to the one used in the NHTSA 1997 report, except that induced exposure has been weighted to vehicle registration years.

In footnote 21, comments are made that standard errors of regression coefficients are too small due to the weights being applied to induced exposure crashes.[6, p.77] While the weight factor being applied to induced exposure crashes is discussed in terms of the effect on standard errors, a weight factor is also applied to the number of occupant fatalities in a "case" vehicle. For example, if two fatalities occurred in the same vehicle, that vehicle is assigned a weight of two. As discussed in the review of the NHTSA 1997 report, one of the assumptions of logistic regression is that the observations are independent. Assigning weights based on the number of fatalities is equivalent to duplicating observations that are not independent. This procedure also tends to lead to standard errors of parameter estimates that are too small, resulting in overstatement of significance.

In Sections 3.5 and 3.8, lengthy discussions are provided about sources of uncertainty unaccounted for. Attempts are even made to adjust model-based standard errors arising from induced exposure data collected from just 8 of the States, and the idea of *self-selection* in which it is assumed that better drivers select heavier cars.

The first example presented using logistic regression is for vehicles in fixed object crashes. A weight factor is used if more than one fatality occurred in the same vehicle. Since fixed object crashes are single-vehicle crashes, the effect of correlated observations and the artificially increased sample size may not be as severe as in multiple-vehicle crashes. However, for the fixed-object regression, there were 10,569 fatalities in 9,537 crashes.

Table 6 provides an example that assumes exactly two fatalities in vehicles with multiple fatalities. In that case, there would be 1,032 duplicate records and twice that many, or 2,064 pair-wise fatalities, that are not independent. The percentage of correlated fatalities in the regression model in that case is 2,064/10,569, or almost 20 percent. Not all vehicles with multiple fatalities would have exactly two fatalities, but the percentage of correlated observations in the regression model would very likely be greater than 15 percent. Does this create a big enough concern such that the coefficients in the logistic regression model would change direction? Most likely this is not a serious concern, but one could argue that the model is somewhat misspecified, and not accounting for correlated observations is another source leading to standard errors that are too small.

The ordinary logistic regression model is a valuable tool in this study and it appears to be capturing the general trends and associations between fatality risk and curb weight. However, considering the various limitations described above, one could argue that it is too simple for modeling variation in the particular observational study in which it is being used.

Table 6 Example of Duplicate Records for Fatalities in Single-Vehicle Crashes

Crash	Fatals	
1	1	1
2	1	1
3	1	1
.	.	.
.	.	.
1,032	1	1
1,033	1	.
.	.	.
.	.	.
9,536	1	.
9,537	1	.
Total	9,537	1,032

With respect to the use of a 2-piece linear variable for curb weight, this appears to be appropriate based on the diagnostic plots showing possible differences in fatality risk for smaller and larger passenger cars. Negative coefficients suggest increased fatality risk with reduced curb weight. The centering of the curb weight variable about 2,950 pounds has the effect of reducing correlation between this variable and other predictor variables in the model. Kahane makes qualifying statements regarding significance, study design, and control variables. Considering this modeling procedure overall, it appears the trends reported are in the right direction, correct, and fairly robust.

Regressions are fit in a similar manner for the single-vehicle crash types principal rollover and pedestrians/bicyclists/motorcyclists. Being single-vehicle crashes, fatalities in the same crash may not be a serious issue. However, the example given in Table 6 suggests that even for single-vehicle crashes, a fairly substantial percentage of the fatalities are likely correlated. In general, however, methodology and results appear to be reasonably correct.

The multiple-vehicle crash types are car-to-truck, car-to-car, and car-to-light truck. For these crash types, analysis may be more complicated because there are occupants in vehicles and vehicles in crashes. Thus, another source of correlation between fatality outcomes arises. For the car-to-truck crash type, it seems that the same methods used for the single-vehicle crash types were used. That is, weights were assigned to case cars with multiple fatalities. The number of collisions is 4,556 with 5,467 fatalities and using the same type of argument shown in Table 6, it is conceivable that approximately 30 percent of the fatal outcomes entered into the regression were not independent. The modeling procedure does not address fatalities in the heavy truck, but it is likely for this crash type most fatalities were in the car.

For the car-to-car crash type, the issue of correlated outcomes and multiple counting of fatalities may be more severe than for the single-vehicle crashes. The method appears to designate one vehicle as the "case" car which is entered into the regression, and the crash partner as the "other" vehicle which is not entered into the regression. However, if two vehicles involved in the same crash are MY 1991-99 cars, one car is designated the case car, the other vehicle is designated the other car, but then the roles of these vehicles are reversed, and cars in the same crash can be counted more than once. A weight factor of one is applied to each case car to mitigate the effects of over counting, but multiple fatalities in the same car are not counted. Fatalities can be to the driver of the case car or any of its occupants, but only one fatality gets counted. Thus, 13,513 cars were entered into the regression, but the number of cars in the same crash is not reported. A note is made that regression methods for analyses of two-car crashes using curb weight for both vehicles is presented in Section 6.6.

For the car-to-light truck crash type, the method reverts back to counting individual fatalities in the case cars since it is reported that 12,119 case cars provided 14,518 fatalities. The rationale provided for counting fatalities is that most were occupants of the case cars. For fatal records, this regression assigns a weight equal to the number of fatalities in the case car. Again, fatal outcomes in the same vehicle are not independent observations. Logistic regression assumes independent observations.

The rest of Chapter 3 is devoted to discussion of results, possible effects of driver quality issues, adjustment of interval estimates due to various sources of uncertainty, and other topics. In an observational study such as this, which combines data from various sources, applies weight factors to fatalities in the same vehicle, and uses data from 8 states for induced exposure crashes, it is recognized that calculation of confidence intervals is a formidable task. Use of such intervals should be done with caution and the intervals should only serve as guidelines for assessing significant findings. Using intervals that are plus or minus 2.57 standard errors around point estimates are equivalent to approximate 99 percent confidence intervals if standard errors from only the logistic regression model are used. This provides wider intervals than the usual plus or minus 1.96 factor commonly used, and may be viewed as an adjustment to standard errors that are known to be too small.

In general, results presented using logistic regression models have likely captured valid trends between fatality rates and vehicle curb weight. In this review, it is the opinion that the data collected and the methods presented are appropriate and valid. However, all models are wrong to some degree, and some models are better than others. For the logistic regressions used, each observation in the data file was a vehicle record. In other words, the data were recorded at the vehicle level. But the goal of the analysis was to relate total fatality risk, or occupant fatality risk to vehicle curb weight. Should the data file have been recorded at the occupant level, and should the regression model have been developed to analyze the data at the occupant level?

Researchers familiar with analyzing the FARS data are well-aware that data are recorded in a hierarchical format with separate files for variables recorded at the crash level, the vehicle level, and the occupant level. That is, the FARS database is a collection of files with an Accident file, a Vehicle file, and a Person file, among others. The Accident file captures variables common to the crash such as time of day, rural/urban, and other roadway and environmental conditions. The Vehicle file captures vehicle level variables such as body type, make/model, manner of collision, and so on. The Person file captures information such as age, sex, and injury status of occupants or non-motorists involved in the crash. The various files can be merged by certain key variables that identify the state and accident number, and the vehicle number.

Since the FARS database is a collection of files with information recorded about fatal crash involvements, these crashes tend to be high energy impacts that result in a fatal outcome for at least one person, and likely serious injury to any other persons involved. Single-vehicle crashes may be characterized by running off the road or crashing into a fixed object. These crashes may also end in rollover. Multiple vehicle crashes can be opposite direction crashes such as head-on involvements. Many of the fatal crashes are also side impact crashes.

Based on the preceding discussion, it seems that an argument can be made for analysis of fatality outcome at the person level. Casualties inside the same car and the same crash are more correlated than casualties involved in a different car or crash. The natural structure of crash data consists of correlated observations with nested crash-car-occupant levels. Already, it has been stated several times that the standard logistic regression model assumes independent observations. In addition, the logistic model in the NHTSA study used data recorded at the vehicle level, and since curb weight is a car characteristic, that feature was collapsed to the level of fatalities and replicated across all fatalities in the same car.

The logistic regression model falls into the class of generalized linear models (GLMs).[2,3,9] The generalized linear mixed model (GLMM) is an extension of the GLM that includes random effects.[1] The GLMM can be used to take into account the crash-car-occupant correlation structure found in crash databases such as FARS. An example of the logistic regression model with random effects is

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \alpha + \sum_{m=1}^{M} \beta_m X_{mi} + \sum_{n=1}^{N} \beta_n X_{nij} + \sum_{q=1}^{Q} \beta_q X_{qijk} + \varepsilon_i + \varepsilon_{ij}$$

where p_{ijk} is the probability that occupant k in vehicle j involved in crash i was a fatality. Crashes are indexed by i which are assumed to be independent observations. The variables $X_{mi}, X_{nij}, X_{qijk}$ are crash (eg. road surface, time of day, rural/urban), vehicle (eg. curb weight, footprint), and occupant (eg. age, sex), predictor variables, respectively, with corresponding fixed effect model parameters $\beta_m, \beta_n, \beta_q$. The crash and vehicle random effects are $\varepsilon_i \sim N(0, \sigma_c^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_v^2)$, respectively, and they are assumed to be independent.

Estimation of the random effects allows taking into account the differences in each crash and each car and permits modeling of correlations between occupants of the same car and between cars of the same crash. This random effects model would be fit to a database constructed at the person level and would provide adjusted standard errors relative to the ones produced by the ordinary logistic model. Since the goal of the analysis is to estimate the net change in fatalities to society as a whole, an analysis at the person level could be considered.

3.3 Response to Docket Comments on NHTSA Technical Report

Vehicle Weight, Fatality Risk and Crash Compatibility
of Model Year 1991-99 Passenger Cars and Light Trucks [68 FR 66153]
Docket No. NHTSA-2003-16318 [7]
C. Kahane

Summary

NHTSA responds to three principal criticisms of the NHTSA 2003 report:

1. The first criticism is that the analyses only considered the relationship of fatality risk to vehicle mass. It did not consider track width and wheelbase.

2. The second criticism argues that vehicle “quality” has a much stronger relationship with fatality risk than vehicle mass. The belief is that lighter cars have higher fatality risk, on average, because they are usually the least expensive cars, and in many cases, the poorest quality cars.

3. The third criticism questions the accuracy and robustness of the report’s calculation of a “crossover weight” above which weight reductions have a net benefit rather than harm when all road users are taken into account.

In their response, NHTSA disagreed with the first two comments and presented regression analyses to defend their results. With respect to the third comment, NHTSA agreed that the “crossover weight” was not accurately known at the time, and believes the safety implications were overstated in the comments.

Review of the First Criticism

Even in the NHTSA 1997 report, the effects of correlation between curb weight, track width, and wheelbase were apparent, and the author correctly did not include them in the same regression model. Kahane explains the effects of multicollinearity well, and describes the reasons why it is dangerous to include all three variables together [page 3, paragraph 2]. All three variables should not be included in the models. Table 1 and Table 2 of the response report show the adverse effects of multicollinearity.

Restricting the analysis to 4-door non-police cars seems reasonable, given that two-door muscle cars tend to have a short wheelbase relative to their weight and high fatality rates.

Review of the Second Criticism

Ross and Wenzel argue that the historical trend of lower fatality rates in heavier cars may be due to the higher quality of those cars, not their mass. Kahane shows that when price is added into the NHTSA model that includes other control variables, such as age, gender, and other control variables, the weight-safety relationships does not change considerably. Ross and Wenzel did not adjust for age and gender. Sales price is only a surrogate for quality.

Judging this second criticism is more difficult. The regression models described are not *causal* models. They can only be used to determine if associations exist between dependent and independent variables. The NHTSA model is quite exhaustive in the sense that it includes many relevant control variables. In certain cases, sales price may be a good predictor of fatality risk. Kahane provides a good example based on head-on collisions, and shows that the relative price of the two vehicles has little or no effect on the relative fatality risk.

Review of the Third Criticism

Only general discussion is given to the issue of crossover weight. William E. Wecker Associates, consultant to General Motors, identifies additional sources of variability beyond those considered in the NHTSA report. Broad comments are made here and definitive results are not provided.

3.4 The NHTSA 2010 Report

Relationships Between Fatality Risk, Mass, and Footprint
in Model Year 1991-1999 and Other Passenger Cars and LTVs [8]
March 24, 2010
C. Kahane

A request was made by NHTSA to review the 2010 report, along with two other independent reviewers, in a particular format that was designed to address “specific charge questions”. The review and responses to the specific questions are reproduced below, exactly as they appeared in the original document.

Background

Footprint is a measure of a vehicle’s size, defined roughly as the wheelbase times the average of the front and rear track widths. Footprint-based standards are intended to discourage downsizing by giving a higher mpg target to smaller footprint vehicles. As a consequence of technologies that are available for improving fuel economy and footprint-based standards, it is important for NHTSA to consider the potential effects of reductions in mass on fatality rates, while holding footprint constant. In 1997 and 2003, NHTSA published statistical analyses of historical crash data that estimated the effects of vehicle curb weight on fatality rates. Analysis of historical data suggests that reductions in mass are generally associated with commensurate reductions in track width and wheelbase. NHTSA has performed new statistical analyses of its historical database of passenger cars and LTVs (light trucks and vans), assessing relationships between fatality risk, mass, and footprint.

Before presenting findings and conclusions, Kahane makes several comments with respect to the statistical methods used and the use of historical data. First, regression models are fit to assess the effects of mass reduction on fatality risk while maintaining footprint. Historical correlations between the weight and size variables raise concerns about multicollinearity and its effect on statistical estimation. Second, the analyses are “cross-sectional” and do not apply to a specific make and model, but to all vehicles on the road. Finally, mass reductions in historical data might not be consistent with future mass reductions.

The immediate purpose of the report is to develop the four inputs to the Volpe model that predicts safety effects of the modeled mass reductions in MY 2012-2016 cars and LTVs over the lifetime of those vehicles. The four numbers are the overall percentage increases or decreases per 100-pound mass reduction while holding footprint constant for cars<2,950, cars>= 2,950, LTVs<3,870, and LTVs>=3,870. All show positive increases in fatalities, except for the heavier LTVs. NHTSA reports a regression scenario, an upper estimate scenario, and a lower estimate scenario.

Specific Charge Questions:

1. Are the analytical methods and data used to estimate relationships between fatality risk, mass, and footprint appropriate?
2. Is the organization of the document appropriate and does it present the material in a clear and concise manner?
3. In your opinion, what are the weakest and strongest parts of the technical report? Please make suggestions on how the weakest parts of the report can be strengthened.

Response to Question 1:

<u>Data</u>

The procedure used to collect and prepare the data sources for subsequent analyses appears to be very appropriate. It is assumed that creation of the database was one of the most time-consuming and ambitious tasks of the study. The most relevant data sources used in the study are listed below:

- FARS 1995-2000 – case fatalities
- State crash data – induced exposure
- R.L. Polk registration data – vehicle registrations
- NASS CDS – odometer readings
- Curb weight and footprint derived from various publications

FARS data were used to collect information about fatal involvements. It is well-known that exposure data recorded at the level required for this study are not available, so state data were used to collect information about induced exposure crashes. State files have variables on driver age, driver gender, and so on. The induced exposure vehicles were those in multiple-vehicle crashes in which it could be determined that the vehicle was not at fault. Polk registration data were used to allocate each induced exposure vehicle its fair share of the nation’s vehicle registration years. NASS CDS data which records information on odometer readings were used to estimate vehicle miles traveled (VMT) from vehicle registration years. Size and weight information were derived from several publications. The data sources used are likely the best ones available for answering the research questions being considered in this study. The database created appears to be an impressive collection of files from appropriate sources.

Analytical Methods

Standard logistic regression models were fit to estimate the effect of a 100-pound reduction in curb weight on fatality risk while maintaining footprint and controlling for driver age, driver gender, time of day, road type, and other variables. The analyses considered six crash types separately: first-event rollovers, collisions with fixed objects, pedestrians-bicyclists-motorcyclists, heavy trucks, other passenger cars, and LTVs. Curb weight was entered as a two-piece linear variable to capture effects for lighter and heavier vehicles. The methods were designed to estimate societal fatality rates including fatalities to all persons involved. Kahane acknowledges that future vehicle design is likely to take advantage of safety-conscious technologies that could reduce risk associated with lighter vehicles in the historical analyses.

Due to the lack of exposure data needed for the calculation of rates, the method of induced exposure was used. Induced exposure crash involvements were the non-culpable vehicles in two-vehicle crashes and were taken from State data files. In a retrospective (historical data) case-control study using logistic regression, fatalities can be viewed as the cases, and induced exposure vehicles can be viewed as the controls. Regression parameters in logistic models have natural interpretations as odds ratios on the log scale. In this model for example, the regression coefficient for curb weight can be used to estimate the change in the odds of a fatality given a 100-pound reduction in curb weight. More traditional models for the analysis of rates include Poisson log-linear models, negative binomial regression, and random effects models. In these models, parameter estimates generally have interpretations as relative risks on the log scale.

However, for a rare outcome, such as fatalities, relative to large exposure, the logistic model parameters should be good approximations to those in one of the more traditional exposure-based risk models. Therefore, disaggregate logistic regression should be appropriate as a risk-based-model for rates, even though parameters are usually interpreted as log odds ratios.

The bigger concern here, however, is the estimation of model standard errors. Likelihood-based tests tend to be significant even when results show small effects, as long as the sample size is large enough. Construction of confidence intervals and tests of hypotheses depend on specification of a model that accommodates the variation in the data. This study is an observational one using various sources of data, and it could be argued that the logistic model is somewhat misspecified. In the presence of extra-variation, standard errors tend to be too small and significance can be overstated. A more robust model, such as a random effects logistic model, would at least adjust standard errors to account for the extra-variation often encountered in studies such as this one. This is a rather technical comment and is offered only as a suggestion. The comment is made because in reading the NHTSA 2003 report, it appears that Kahane was well-aware that standard errors were too small, and adjustments were made after fitting a logistic model. Fitting a more robust model in the beginning might preclude the need for adjustments after the model is fit.

Here is another example of how the logistic model might be somewhat misspecified for this problem, and how it could lead to standard errors that are too small. The study focused on estimation of societal fatality rates, so one might assume that each record in the data file referred to one person. However, it appears that each record in the data file was a vehicle, and not a person. Kahane makes a point that fatal involvements were weighted by the number of fatalities in the crash. One of the assumptions of logistic regression is that the observations are independent. What effect does weighting each involvement have on the conclusions? In single-vehicle crashes, the effect may not be too great. It would depend on the number of fatalities in each vehicle. However, in multiple-vehicle crashes, the problem is more complicated. In multiple-vehicle crashes, there are occupants in vehicles and vehicles in crashes. This crash-vehicle-occupant hierarchy gives rise to correlated outcomes in the crash and in the vehicles. In reading the 2003 report, it appears that Kahane has mitigated some of this concern by defining a "case" car and an "other" car in which data on the "other" car were not recorded. However, case cars, which are model year 1991-99 cars, contribute multiple records to the data file since their roles as "case" and "other" cars get reversed. Significant findings from the fit of a logistic regression model would tend to be overstated. However, if findings are *strongly* significant using logistic regression, they would likely remain significant, but to a lesser degree, even in a more robust model, such as a random effects model.

Considering the various data sources and the design of this study, it is recognized that estimation of standard errors, and the related tasks of constructing confidence intervals and performing tests of hypotheses, is formidable. The comments above are general in nature and are intended to stimulate discussion about using logistic regression. For example, was conditional logistic regression for matched pairs considered? In that model, each case vehicle would be matched to one or more induced exposure vehicles on certain key variables such as make/model or age and gender of the driver.

Discussion of Multicollinearity

It appears that one of the biggest issues regarding this work is the historical correlation between curb weight, track width, and wheelbase. Including three predictor variables in a regression model that are highly correlated can have adverse effects on the fit of the model, especially with respect to the parameter estimates, as outlined by Kahane on page 479, paragraph 4. The correlations with curb weight are reported as 0.796 for track width, 0.868 for wheelbase, and 0.893 for footprint. The strong positive correlation between these predictor variables suggests that if separate regression models were fit with each predictor one at a time, regression coefficients would all have the same sign.

If the three variables are entered together in the same model, and one or more coefficients change sign, then making inference about the effects of certain predictors on the response variable can be misleading. For example, if the estimate for curb weight changes sign, making inference about changes in curb weight while holding track width and wheelbase constant could be regarded as overly ambitious. We seek the simplest model with the fewest parameters that explain the data well. Including highly correlated predictor variables in the same regression model leads to over-fitting and unstable estimation. Inference in the presence of multicollinearity should be judged with great concern.

Overall results should not be greatly affected since adding more variables to a model will improve the fit to some degree, even if the additional variables are not significant. Table 2-2 shows results based on the fit of a model that includes all three predictor variables. It is not surprising that *combined* results shown in Table 2-2 agree fairly well with those presented in Table 2-1 in which only curb weight was fit. Note that the combined effect of reducing wheelbase by 1.01" resulted in a *reduction* of 127 fatalities. Has the coefficient for wheelbase changed sign in most of these regressions? Can we now infer that reducing wheelbase generally results in a reduction of fatalities while holding curb weight and track width fixed?

In an attempt to alleviate issues associated with multicollinearity, Kahane combines wheelbase and track width into a footprint variable. In this case, overall results in Table 2-4 show fatality increases for reductions in both curb weight and footprint. It appears that Kahane has centered the curb weight variable around 2,950 pounds, as documented on page 490. Centering variables is a tool often used to alleviate the effects of multicollinearity. It tends to reduce correlation among parameter estimates. Were the track width and wheelbase variables also centered about their mean values? Or in this case, was the footprint variable centered about its mean value? It might be useful to examine the variance-covariance matrix of the parameter estimates with and without centering predictor variables to see the effects that centering has on the correlation between parameter estimates.

In an observational study such as this, it seems that the best option available is to fit the most parsimonious models that find basic trends and associations, and to report those trends and associations. Those results tend to be the most compelling because they demonstrate direct relationships between size and weight and fatality risk.

Response to Question 2:

The document is very well-written. The author writes very clearly and it is straightforward to understand how the data were collected and how the methods were used in the various analyses. The author is very thorough and the work is very detailed, but sometimes highlighting the main points and conclusions would help the reader. This is not to suggest that leaving out important material is advocated, but only that highlighting the major conclusions and not dwelling on minor points could improve readability.

Response to Question 3:

Differences in results produced by NHTSA and DRI are discussed in the report. These differences are likely very controversial and seem to center around the issue of multicollinearity. One of the strongest parts of the report is the reproduction of results similar to those of DRIs in which the predictor variables curb weight, track width, and wheelbase were entered into a regression model together. The methods were slightly different than DRI's, and the data used were from NHTSAs 2003 study. However, the main point of the analysis was to demonstrate that results similar to those of DRI could be reproduced – fewer fatalities per 100 pound reduction of curb weight while holding track width and wheelbase fixed. Kahane's explanation describing the different results was very convincing, both in terms of the effects of multicollinearity and in terms of the use of a two-step regression procedure. In the NHTSA 2003 report, Kahane abandoned much of the methodology described in the NHTSA 1997 report upon which some of the DRI results were based. NHTSA also believes two-step regression weakens relationships between curb weight and dependent variables.

One of the weaknesses of the paper is that the great majority of the work was devoted to analyses of passenger cars and much less to LTVs. In fact, when including footprint into the model along with curb weight, the overall results still suggest increased fatalities with reduction in curb weight. However, it appears that the effect for curb weight has changed sign in the presence of footprint. Isn't this the same concern that was expressed in the analyses for passenger cars with respect to multicollinearity? Kahane stresses caution due to the potential effects of multicollinearity, but the result leaves the reader wondering to some extent if valid inference can be made about curb weight while holding footprint constant.

Various complications were addressed related to the analyses of LTVs. A discussion was provided that described many LTVs as “niche” vehicles. Design features between light pickup, heavy pickup, SUVs, and minivans can be very different. Kahane considers exclusion of high-CG SUVs and discusses the possibility of separate analyses for different vehicle types. Considering that LTVs pose a potential greater challenge than passenger cars due to design features, should more consideration be given to LTVs? Fatality rates were analyzed per billion miles rather than vehicle registration years since it was found that annual mileage for LTVs varies by the type, size, and mass of the LTV. This appears to represent another difference between analyses for passenger cars and LTVs.

Alcohol-Related Fatalities

Examination of NHTSA’s own 2005 publication *Traffic Safety Facts*, suggests about 20 percent of drivers in fatal crashes had BAC greater than 0.08 g/dl. Would this have any effect on results? It seems that the induced exposure vehicles would not have BAC recorded since those vehicles were for the most part not at fault in the crash. Would deleting alcohol-related crashes from the FARS data change any results? Judging by the care that was taken to include all relevant variables in this study, alcohol-related fatalities were likely considered by the author at some point in time, but perhaps a decision was made that including these fatalities in the models would have little effect on the outcome.

Accident Type and Vehicle Type

Six accident types were considered separately in this study. The single-vehicle crashes seem to be fairly well covered by first-event rollover, hit fixed object, and pedestrian/motorclist/bicyclist. However, car into car and car into light truck, for example, cover many crash types. In addition, fatal crashes tend to be high-energy, high-impact type crashes. Would there be any benefit in considering more specific crash types such as head-on, side impact, or rear-end separately? This might help reduce variability due to different crash types.

Passenger cars and LTVs were the only vehicle types considered, even though two-piece linear effects were fit for lighter and heavier vehicles. Would it be feasible to consider more types of LTVs? Considering the discussion about LTVs as “niche” vehicles, another way to reduce variability and focus in on the effects of size and weight could be to consider more vehicle types.

- Light pickups
- Heavier pickups
- Minivans (if possible)
- SUVs

Residual Analysis, Detection of outliers, Possible Interactions

Have residuals been investigated for outlying observations? Identification of outliers could lead to a better understanding of the effects of size and weight on fatality rates in certain conditions. Examination of outliers could give an indication why certain observations are not fitting well and could lead to formulation of more research questions. Admittedly, the age/gender variables represent an interaction between age and gender. But other than that, no interaction terms were fit in any of the models. Are any interaction terms significant that were not included in the model? Interaction terms might also aid in interpretation and improve model fit.

Decile analyses were performed in an attempt to control for footprint. This appears to be similar to a technique of stratification. It appears to be beneficial as a diagnostic tool. Similarly, effects of footprint were assessed holding curb weight fixed.

4. Review of the DRI Papers

4.1 The DRI 2002 Report

An Assessment of the Effects of Vehicle Weight on Fatality Risk
in Model Year 1985-98 Passenger Cars and 1985-97 Light Trucks [14]
DRI-TR-02-02
February 2002
R.M. Van Auken and J.W. Zellner

Summary

In this report, the effects of vehicle weight were assessed in terms of the net change in the number of U.S. fatalities based on 1995-99 data involving 1985-98 passenger cars and 1985-97 light trucks. The authors claim that this report is an update of results presented by Kahane in NHTSA's 1997 report using the same methodology. [5] Estimates were obtained for the same six crash types:

- Principal rollovers
- Collisions with objects
- Collisions with pedestrians, bicycles, or motorcycles
- Collisions with heavy trucks (GVWR greater than 10,000 pounds)
- Collisions with passenger cars
- Collisions with light trucks (pickups, SUVs, or vans)

The major findings of the DRI 2002 report are shown in Table 7 and Table 8. Results suggest that a 100 pound weight reduction overall would have a small and insignificant effect on fatalities. The claim is that a 100 pound reduction in weight would increase fatalities in certain types of crashes, and would decrease fatalities in other types of crashes, effectively cancelling both effects and resulting in insignificant overall changes. While the NHTSA 1997 report found similar and insignificant effects for a 100-pound weight reduction in light trucks, that report estimated a significant increase of about 300 fatalities for passenger cars. The results for light trucks agree – those for passenger cars do not.

Table 7 Estimated Effect of a 100-Pound Passenger Car Weight Reduction on 1999 US Fatalities, DRI 2002 [14]

Crash Type	Fatalities in 1999 Crashes	Effect of 100 Pound Weight Reduction	Net Fatality Change	One Standard Deviation
Principal rollover	1,663	3.77%	63	11.6
Hit object	7,003	0.03%	2	17.4
Hit ped/bike/motorcycle	3,245	-2.39%	-77	9.3
Hit big truck	2,496	1.20%	30	9.6
Hit passenger car	4,047	-2.42%	-98	19.4
Hit light truck	6,881	1.67%	115	21.0
Overall	**25,335**	**0.13%**	**34***	**37.9****
3-sigma confidence bounds			**(-80, 148)**	

*Overall is calculated from the net fatality changes before rounding to the nearest integer value.

**Standard deviation for "overall" is the square root of the sum of the squares of the 6 individual standard deviations.

Table 8 Estimated Effect of a 100-Pound Light Truck Weight Reduction on 1999 US Fatalities, DRI 2002 [14]

Crash Type	Fatalities in 1999 Crashes	Effect of 100 Pound Weight Reduction	Net Fatality Change	One Standard Deviation
Principal rollover	2,605	1.42%	37	13.3
Hit object	3,974	1.23%	49	12.4
Hit ped/bike/motorcycle	2,432	-0.79%	-19	8.5
Hit big truck	1,506	1.50%	23	8.6
Hit passenger car	6,881	-1.55%	-106	16
Hit light truck	1,781	-1.06%	-19	12.1
Overall	**19,179**	**-0.19%**	**-36**	**29.6**
3-sigma confidence bounds			**(-125, 53)**	

*Overall is calculated from the net fatality changes before rounding to the nearest integer value.

**Standard deviation for "overall" is the square root of the sum of the squares of the 6 individual standard deviations.

DRI claims that sensitivity analyses suggest that results are similar to those of Kahane.

Data

Data used in this study parallel those used in the NHTSA 1997 study. There were some differences in data and model years. FARS data covered calendar years 1995-99. Case vehicles were 1985-98 MY passenger cars and 1985-97 MY light trucks. In addition State data, which supplied information on induced exposure crashes, were derived from 7 states, and not 11 as in the NHTSA report.

- FARS – fatality case involvements
- State data from 7 states – induced exposure involvements
- R.L. Polk data – vehicle registrations
- Other sources – curb weight, track width, wheelbase

Methods

The methods used in this report were designed to closely follow those presented by Kahane in NHTSA's 1997 study. First, logistic regression was used to assess fatality risk per induced exposure crash using State data for passenger cars and light trucks. Then, in an attempt to estimate size-related bias in fatality rates relative to induced exposure, aggregate linear regression was used to assess induced exposure crash risk per vehicle year. Weighted Least Squares (WLS) regressions were fit on aggregated data from the States in two steps. These regressions modeled the log induced exposure rate as the dependent variable. The regressions were weighted by the denominators (vehicle registration years). In the Step 2 regression, the numerator of the rate (induced exposure) was adjusted based on results from the Step 1 regression. Finally, fatalities were expanded to include fatalities from all states, and exposure was expanded to include vehicle registration years from all states. For national estimates, the two-step aggregate regression was applied in the same manner used when analyzing State data. It appears that much of the coding of predictor variables in the various regression models followed the conventions established by Kahane.

General Comments on the DRI 2002 Report

For the logistic regressions, except for minor differences, the authors report good agreement with Kahane. This applies to the models fit for passenger cars and light trucks. Tables 3.6 and 3.8 provide summaries comparing results from the two studies.

According to the results from aggregate linear regression models, the authors again report generally good agreement with the NHTSA 1997 report for both regression steps. To assess the effects of a 100-pound weight reduction on fatality risk, regression coefficients were summed from the logistic models and the WLS models. For passenger cars and light trucks, the authors suggest comparable results with Kahane.

In Section V, an approach was presented for assessing effects of reduction in vehicle weight on fatality risk in all states, not just seven. As in Kahane, states were classified into five groups. Two-step regressions were performed for passenger cars and light trucks. Results were presented for the rollover crash type. A method using exogenous coefficients for driver age and gender was introduced after observing excessive correlation between curb weight and driver age in the aggregated data.

In Section VI, exogenous control for driver age and gender was applied. Following Kahane's approach, the aggregate linear regressions were performed in two steps. The Step 1 results were unchanged. The Step 2 results used exogenous driver age and gender coefficients from previous regressions using data from seven states.

Sensitivity tests were conducted to assess effects on changes in fatalities due to exogenous driver and gender coefficients, exclusion of 4-door sedans and hatchbacks, and exclusion of all but pickup trucks. According to the driver age and gender sensitivity tests, it was concluded that results were generally robust to the small expected changes in these variables. The sensitivity to excluding sporty cars suggested an increase in the estimated weight effect. For regression limited only to pickups, Kahane found an increase in estimated fatalities due to a 100 pound weight reduction, but Van Auken and Zellner found less sensitivity. Checks on rates with zero counts in the numerator were also investigated with little effect found. These rates were adjusted by adding small positive constants to cells with zero counts.

Specific Comments on the 2002 DRI Report:

The purpose of this paper was to assess effects of reduced curb weight on fatality risk for passenger cars and light trucks using the methods presented in Kahane's 1997 paper. [5] Throughout the report, statements were made that results were generally in good agreement with those produced by Kahane. If that were the case, then why did the 1997 NHTSA report estimate a significant increase of about 300 fatalities for a 100-pound reduction in curb weight for passenger cars? The findings in the two reports agree fairly well with respect to insignificant findings for a 100-pound weight reduction for light trucks. But the two reports disagree with respect to findings for passenger cars.

One reason for the different results could be that the databases were not the same. Many data sources were combined and each one suffers from data quality to some extent. However, a more fundamental reason is that there were just too many analyses going on here. Before reaching the final conclusions, there were too many intermediate steps and adjustments being made, any one of which had the potential to add extra uncertainty and variation into the estimation process. The problem is difficult enough as it stands.

In general, the goal in statistical modeling is to find the simplest model with the fewest number of parameters that explains the data well. The two-step regression model, which was used to improve the estimation procedure, may have actually introduced additional error into the estimation process. Results from the Step 1 regression, which suffer from some sources of uncertainty even if the R-squared is very high, were used as inputs into the Step 2 regression. If the two-step regression model has been used successfully in other similar problems, it might be a good idea to provide references.

Note that problems encountered with the aggregate WLS model led to consideration of exogenous control variables for age and gender. Coefficients in the final models were forced to equal functions of logistic model and aggregate model coefficients estimated from State data. Here is another example of using estimates with uncertainty as inputs into another model.

Finally, it is hard to know what effect induced exposure involvements had on the final results. Given that good sources of exposure data are not available at the level required for a study of this nature, induced exposure appears to be a reasonable alternative. However, the goals of this study were very broad, and estimating the net effect on fatalities in the US associated with a 100-pound reduction in curb weight is a formidable task. It is likely that the use of induced exposure data and how it can be improved will continue to be a source of discussion for some time.

4.2 The DRI 2003 Report

A Further Assessment of the Effects of Vehicle Weight and Size Parameters on Fatality Risk in Model Year 1985-98 Passenger Cars and 1985-97 Light Trucks [15]
DRI-TR-03-01
January 2003
R.M. Van Auken and J.W. Zellner

Summary

This report extends results in the DRI 2002 report to include effects of vehicle weight and size on fatality rates. In particular, the effects of passenger car and light truck curb weight, wheelbase, and track reduction were assessed in terms of the net change in the total number of fatalities nationwide. The same six crash types described in the 2002 DRI report were considered.

The main findings of the DRI 2003 report are shown in Table 9. The results suggest that overall, curb weight reduction tends to decrease the overall number of fatalities, but typical corresponding reductions in wheelbase and track width tend to increase fatalities by a nearly equal amount, and that the overall net change is not significant at the 0.05 level. For example, if there had been a 100-pound passenger car and light truck weight reduction, and wheelbase and track width were held fixed, the net result in 1999 would have been a reduction of 799 fatalities with a corresponding 95 percent confidence interval of (-1115, -483).

Table 9 Estimated Effects of a 100-Pound Vehicle Weight and Corresponding Wheelbase and Track Reduction on 1999 US Fatalities, Based on Data for 7 States, DRI 2003 [15]

Vehicle Type	Fatalities in 1999 US Crashes	Estimated Net Change in 1999 US Fatalities due to							
		100-Pound Curb Weight Reduction		Typical Corresp Wheelbase Reduction*		Typical Corresp. Track Reduction**		Combined Weight and Size Reductions	
		Est.	2σ	Est.	2σ	Est.	2σ	Est.	2σ
Passenger cars	25,335	**-580**	**(260)**	**368**	**(174)**	**191**	**(134)**	-21	(340)
Light Trucks	19,179	**-219**	**(179)**	**174**	**(81)**	**106**	**(104)**	61	(222)
Total	37,633	**-799**	**(316)**	**542**	**(192)**	**297**	**(170)**	40	(406)
± 2 sigma confidence bounds		(-1115, -483)		(350, 734)		(127, 467)		(-366, 446)	

*Typical wheelbase reduction is 1.01 in for passenger cars and 1.21 in for light trucks
**Typical track reduction is 0.34 in for passenger cars and 0.57 in for light trucks
Bold numbers are significant at the 0.05 level

Data

The data used in this study appear to be the same data that were analyzed in the DRI 2002 study.

- FARS – fatality case involvements
- State data from 7 states – induced exposure involvements
- R.L. Polk data – vehicle registrations
- Other sources – curb weight, track width, wheelbase

A discussion of “total number of fatalities” (last paragraph, p.13) is given under heading C. which is titled, “State Non-Fatal Accident Data Reduction and Accident Classification.” It is assumed that nonfatal data were collected using state data.

Methods

In Section G (p.19), the Step 1 regression model is explained. Now it is clear why nonfatal accident (A) data were collected. The logistic regression model was separated into two components

$$\log\left(\frac{F}{IE}\right) = \log\left(\frac{F}{A}\right) + \log\left(\frac{A}{IE}\right) \qquad (1)$$

where A represents vehicles in nonfatal crashes collected from state data, IE designates induced exposure crashes, and F designates fatality. This model deviates from Kahane's model since it includes an intermediate step. The Kahane model fit

$$\log\left(\frac{F}{IE}\right)$$

directly, without consideration of the nonfatal crash (A) component. Predictor variables are attached as a linear function of regression parameters to the right side of the model equations.

A simultaneous logistic regression for a two-stage risk model is used where data for three regressions are stacked on top of each other (see matrix equation 20, p.25). The three regressions correspond to the F/A, A/IE, and F/IE models. Is this model developed by the authors, or has this model been used before? If so, are there any references indicating under what circumstances this model has been applied successfully? For instance, what are the properties of the estimators? Are they biased?

The simultaneous logit model looks similar to the Seemingly Unrelated Regression (SUR) method of Zellner (A. Zellner, not J.W. Zellner) whereby regression data for separate regressions are stacked and all parameters in the various models are estimated simultaneously.[24] However, SUR has an implied correlation structure through a covariance matrix between the response variables in the separate regressions. Can it be assumed observations in the logit model are independent after stacking the data? Although separate intercepts are fit for the three regression equations, the predictor variables in the third regression F/IE appear to be a linear combination of the F/A and A/IE predictors. There is mention of a $\sqrt{2}$ correction factor, but is this sufficient?

In this study, in addition to curb weight, wheelbase and track width were entered into the model. For passenger cars, the correlation between curb weight and wheelbase was found to be 0.878. For curb weight and track width the correlation is 0.765. For light trucks, the correlations with curb weight were 0.625 and 0.749.

Clearly, as shown by the authors, the three variables weight, wheelbase, and track width are highly correlated. These positive correlations indicate that if these variables were entered into regressions one at a time, the coefficients should all have the same sign. However, as shown in Table 3.9 for car-car crashes, the coefficient attached to curb weight has a positive coefficient, but the coefficients attached to wheelbase and track width are negative. This appears to be the result of confounding where the predictor variables are associated among themselves and with the dependent variable. Multicollinearity between the three predictors caused some of the estimates to change signs. When all three were entered into the regressions simultaneously, were they centered around their average values to reduce the effects of collinearity?

In Table 3.10 the *overall* effects when combined may agree with results in previous reports because multicollinearity generally will not affect the fitted values, but the individual parameters may be in the wrong direction. Adding variables to a regression model does not make the fit any worse. The question is whether inclusion of additional variables improves the fit significantly and aids in interpretation. If the curb weight parameter estimate is the one that has reversed sign in the regressions for the six crash types, making inference about reductions in curb weight while holding wheelbase and track width fixed should be cautioned against. It is the opinion in this review that the change in sign of the curb weight estimate is due to multicollinearity.

The last column of Table 3.11 based on the two-stage logistic regression gives very different results than Table 3.6 from the DRI 2002 report based on single-stage logistic regression. For all crash types, except hitting a big truck, the estimated effects of a 100 pound weight reduction suggests reduced fatalities per induced exposure crash. However, in the 2002 report, for all crash types except ped/bike/motorcycle and hit passenger car, the estimated effects of a 100 pound weight reduction suggests an increased fatality rate. The discrepancies between the two results appear to be due to inclusion of all three size and weight variables in the 2003 DRI report.

The same comments apply to the analyses for light trucks shown in Table 3.13. If the estimate for curb weight is the one that changed sign, then making statements about reductions in curb weight while holding track width and wheelbase fixed should be reconsidered very carefully. The results in Table 3.13 are quite different from those presented in Table 3.8 in the 2002 DRI report.

In Section IV the two-step aggregate linear regression is described using data from seven states. The log rate with induced exposure in the numerator and vehicle registration years in the denominator is the dependent variable. No comment is made as to whether this is a weighted regression using the denominator as weights. Weighted regressions were fit in previous reports by Kahane (1997) and DRI (2002). Results from the Step 1 regression were used to adjust the induced exposure crashes for input to the Step 2 regression. The size and weight variables are only fit in the Step 2 regression. The Step 1 results are not shown since they are the same as those in the DRI 2002 report.

The variables curb weight, track width, and wheelbase, are entered as predictors in the Step 2 regression. Again, the opposite signs of regression coefficients attached to these variables is concerning. These variables were shown to be highly correlated, suggesting that if they were entered into separate regressions one at a time, parameter estimates should have the same sign. Yet, when the three variables are entered together, the estimate for curb weight is negative, while estimates for track width and wheelbase are positive. Then inference is made suggesting that heavier cars are less likely to be involved in induced exposure crashes than lighter cars with the same wheelbase and track. Due to the correlation between the size and weight variables, such statements should be carefully reconsidered. And again, including additional variables in a regression model will not make the *overall* fit any worse, so it is not surprising that the *net effect* of increasing both weight and size is consistent with the DRI 2002 report which considered weight only (14, last paragraph, p.45). Were the size and weight variables centered around their means to reduce potential effects of multicollinearity?

Considering the Step 2 regression for light trucks, the coefficient for wheelbase is not significant. Is wheelbase strongly significant when fit without curb weight and track width? As with the comments above for passenger cars, based on the data used, it appears that any one of the three size and weight variables is a candidates as a surrogate for the other two.

The method for assessing effects of a 100 pound weight reduction, controlling for vehicle size on the risk of fatality per vehicle registration year, is to sum coefficients for curb weight from the logistic regression and the aggregate linear regression. Results presented in Table 4.3 (p.50) disagree considerably from those in Table 4.5 (p.46) in the 2002 report. It appears the differences are due to controlling for wheelbase and track, which are correlated with weight.

In Section V, results are extended to make inference about fatality rates per vehicle registration year not only in seven states, but in all states. For passenger cars, Table 5.1suggests that car curb weight reduction, while controlling for wheelbase and track, significantly reduces fatality risk, but reductions in wheelbase or track significantly increase fatality risk. As described above, these statements should be judged with caution. Similar statements are made regarding curb weight in Table 5.2 for certain crash types. For light trucks, it is also argued that curb weight reduction leads to reduced fatality rates, but reductions in track width and wheelbase increase fatality rates. The claim is that overall, these effects tend to cancel.

A series of sensitivity tests were performed to check assumptions about induced exposure and nonfatal data collected from seven states, exclusion of one state at a time, and exclusion of sporty vehicles. The authors report that the methods presented were robust to small departures in assumptions made.

Comments on 2003 DRI paper:

The first DRI paper in 2002 reported results that were in general agreement with results reported in Kahane's 1997 paper. The methods of the two papers were very similar. Methodology of the DRI 2003 paper deviates from that in the original report in several ways. Instead of fitting a logistic regression model to fatal and induced exposure data, an intermediate step was applied where state nonfatal data were used. A two-stage disaggregate logistic regression model was fit that considered the log of fatal outcomes to nonfatal outcomes (F/A), and the log of nonfatal outcomes to induced exposure crashes (A/IE). The models were fit to data that were stacked and parameters were estimated simultaneously. It would be useful to see references to previous work where this method had been applied successfully.

Another difference is that the 2003 DRI report did not use exogenous age and gender variables in a two-step aggregate linear regression to arrive at final effects of curb weight on fatality risk. Final results were estimated by combining effects from the two-stage logistic regression that modeled fatality rates per induced exposure and the two-step linear regression that modeled induced exposure rates per vehicle registration year.

The most important difference, however, is the inclusion of size variables track width and wheelbase, in addition to curb weight in the various regression models. It was shown that the three variables were highly correlated. One could assume that if the variables were fit one at a time in separate regression models that they would all have the same sign. However, it appears that the coefficient for curb weight, in at least some of the regressions, changed sign. The authors then conclude that curb weight reduction tends to decrease the overall number of fatalities, but typical corresponding reductions in wheelbase and track width tend to increase fatalities by a nearly equal amount, and that the overall net change is not significant. Making inference from coefficients that change sign from their original direction can be misleading.

Furthermore, the observation that the net, or overall effects, agree in some sense with the 2003 report should not be surprising since addition of variables in a model will always improve the fit by some amount, even if the additional variables are insignificant. It should also be noted that confidence intervals reported from the models used in an observational study such as this are ambitious, and should only serve as guidelines since there are many sources of variation that cannot be accounted for.

4.3 The DRI 2004 Report

A Review of the Results in the 1997 Kahane, 2002 DRI, 2003 DRI, and 2003 Kahane Reports on the Effects of Passenger Car and Light Truck Weight and Size on Fatality Risk [16]
DRI-TR-04-02
March 2004
R.M. Van Auken and J.W. Zellner

Summary

This report is a review of the 1997 and 2003 NHTSA studies and the 2002 and 2003 DRI studies. The 2003 DRI study departed from previous studies in the sense that, in addition to weight, the effects of wheelbase and track width were considered. The 2004 DRI report presents potential reasons why NHTSA's and DRI's results were different. The 2003 DRI study found that curb weight reduction would be expected to decrease the overall number of fatalities and that wheelbase and track reduction would be expected to increase the number of fatalities.

There is some discussion regarding Kahane's failed regressions, but the source of those problems was removed in the 2003 report. DRI shows that if exogenous variables from logistic regression were used and, in addition, analysis was restricted to 4-door non-police cars, results similar to Kahane's can be reproduced. Again, the DRI finding that curb weight reduction would be expected to decrease the overall number of fatalities, but wheelbase and track reduction would be expected to increase the number of overall fatalities is likely attributable to the presence of collinearity among the three predictor variables. Often tables were presented in this report without supporting text or documentation describing the results in the tables.

Specific Comments on the 2004 DRI Report:

It is likely that the inconsistent results were due to differences in statistical methodology, not to differences in data. From the 1997 to 2003 NHTSA reports, Kahane simplified the statistical approach, thereby leading to improved inference. In 2003, Kahane abandoned much of the methodology used in the 1997 report. The DRI 2003 report, on the other hand, retained much of Kahane's original ideas, added more complexity to the models, and in addition, included highly correlated variables in regression models. The added complexity involves a three-stage model consisting of two logistic models and a two-step aggregate linear model. There was no discussion of collinearity anywhere in this report

4.1 The DRI 2005 Report

Supplemental Results on the Independent Effects of Curb Weight, Wheelbase, and Track on Fatality Risk in 1985-1998 Model Year Passenger Cars and 1985-1997 Model Year LTVs [17]
DRI-TR-05-01
May 2005
R.M. Van Auken and J.W. Zellner

Summary

In this report, results are presented indicating that different independent effects of vehicle weight and size on fatality risk are observable in the 1995 to 1999 calendar year data for 1985 to 1988 model year passenger cars and 1985 to 1997 model year light trucks. The results indicate that vehicle weight reduction tends to decrease fatalities, but vehicle wheelbase and track reduction tends to increase fatalities.

Sensitivity results indicate that by restricting the data to 1991 to 1998 model year 4-door only non-police passenger cars, the results for the combined effect of weight, wheelbase, and track reduction are similar to those obtained by NHTSA using similar data and methods. This report is largely a response to the discussion presented by Kahane in Docket NHTSA-2003-16318-16. [7]

The data used in this report are the same as the data used in the DRI 2003 report, with some minor exceptions when relating findings to the NHTSA 2003 report.

5. Review of the Wenzel and Ross Papers

5.1 The Wenzel and Ross 2005 Paper

The Effects of Vehicle Model and Driver Behavior on Risk [21]
Accident Analysis and Prevention, 37, p. 479-494
T. Wenzel and M. Ross

The Wenzel and Ross papers focus on the relationship between driver fatality rates and vehicle type, particularly on vehicle model.[21,22,23] Risk of fatality is defined as driver deaths per year per million registered vehicles. This risk is divided into risk to drivers in subject vehicle (for all types of crashes) and risk to other drivers (risk to others) in two-car crashes. Combined risk is the sum of the two risks. The authors note that their definition of risk is "risk as driven", because the risk of a serious crash is influenced by driver and environmental factors, and the risk of a fatality is influenced by safety belt use, vehicle design, and driver frailty.

In this paper, Wenzel and Ross used data on driver fatalities from FARS from 1997-2001. Vehicle registrations for vehicle year makes and models from RL Polk for January, 2000, and January, 2002 were used to estimate vehicle registrations for the years 1997 through 2001. Risks were calculated for 92 popular models with at least 0.4 million registrations years over the 5-year period. The vehicles were grouped into car categories (subcompact cars, compact cars, midsize cars, large cars, import luxury cars, sports cars) and light truck categories (minivans, SUV, compact pickups, and full-size pick-ups by tonnage).

Risks as functions of variables including curb weight, vehicle type and model, resale price, interior volume, and capacity were explored through scatter plots and simple linear regressions between risk and one dependent variable at a time. Risk to drivers in roll-over crashes by vehicle type was also examined. The effect of driver behavior by vehicle type was explored by comparing the proportions of fatalities attributable to young males (<age 26 years), elderly (> 64 years), and bad drivers (defined as drivers who had been drinking, had no driver's license, were recorded as reckless in the subject crash, or had these events in his/her driving record in the last 3 years). The effect of the driving environment on driver risk was examined through the proxy of population density of the county in which the crash occurred.

The results of the study indicate a wide range of risk to drivers of cars. The risk to drivers in the subject vehicle increased as mass (curb weight) decreased. Risk for cars of the same mass varied greatly, and overall curb weight alone was found to be a modest predictor of risk. Stratifying by broad manufacturer category (US cars, Japanese/German, Korean) increased the correlation between risk and curb weight. For example, in the linear regression for risk to drivers and curb weight for all cars the $R^2 = 0.40$, while for Japanese/German models the $R^2 = 0.73$.

When exploring the risk relationship with other variables, Wenzel and Ross found the resale value of a vehicle at five years to be related to risk, with considerably higher correlation (R^2

=.82) than curb weight (R^2 = 0.40). They speculate that this may be because of the quality of design and manufacture. However, an alternate explanation could be that higher-value cars are driven by safer drivers.

Among the different types of cars, risk to drivers was the highest for sports cars and lowest for luxury imports. Because risk to drivers in subcompacts was very broad, it was divided into two categories: “bottom-of-market” subcompacts and the rest of the subcompacts. The bottom-of-market subcompact models had high levels of risks to drivers that were just below that of sports cars. The remaining subcompacts had much lower risk to drivers. The risk to other drivers for car models was the highest for large cars and sports cars and lowest for the sub compacts in the low risk group. Minivans had the lowest risk to their drivers, similar to that of import luxury cars but higher risk to others than cars. Wenzel and Ross state that this is due to their higher mass and front end relative to cars.

Among light trucks, risk to drivers of SUVs varied considerably. Bottom-of-market smaller SUVs had high risk to their driver, while other smaller (not bottom-of-market) SUVs had low risk to drivers. Among larger SUVs, those with body on frame construction have higher risk to their drivers and also higher risk to other drivers, while the unibody construction SUVs have much lower risks both to their drivers and other drivers. Pickup trucks have the highest risk to others of all the light duty vehicles. Driver risk is higher for pickup trucks than for cars and for SUVs. Most compact pickup trucks have the same or higher risk to their drivers as full size pickup trucks. The pattern of driver risk in pickup trucks is highly regular with respect to size/capacity. Risk to other drivers increases with pickup truck capacity. The authors point out the body-on- frame SUVs and pickup truck share common design features that play a crucial role in its high risk to others in collisions.

Risk to driver in a roll-over crash (both as a first event or subsequent event) was highest for sports cars when numeric value of the measure was considered. However, risk from roll-over accounted for 34% of risk to drivers of sports cars. Although lower than the risk to driver for rollovers in sports cars, this risk was also high for SUVs in general, and accounted for 53% of the risk to driver. Risk to drivers from rollover for unibody SUVs was much lower, equal to about that of mid size cars. Among cars, the lowest risk to driver from rollovers was in import luxury cars, which accounted for 25% of the risk to driver, and for large cars, where it accounted for 16% of risk to the driver.

The authors note that driver behavior, or how, where, and when a vehicle is driven affects risk of death in a traffic crash, and examine the proportion of fatalities in each vehicle type that were young drivers (associated with risk-taking behaviors), elderly drivers (associated with fragility) and bad drivers (associated risk-taking behaviors such as drinking and driving, driving without a license, and reckless driving). They find that highest fractions of young male driver deaths are in the vehicle types with the highest risks, i.e., sports cars and compact pickup trucks. Highest proportions of older drivers are in large cars, which have a lower risk to drivers of the same vehicle but a higher risk to drivers of other vehicles. The fraction of bad driver deaths is highest for sports cars, compact pickups, and ½ ton pickups.

Vehicle type and driving environment were also explored. The authors argue that rural roads pose greater risk to drivers than other roads. The population density of the county in which the crash occurred was used as a proxy for road types. Their analysis by population density indicates that pick up crashes (¾ -ton and –ton pickups) are more likely to occur in the most rural (least dense) areas.

Results indicate that vehicles exhibit widely different levels of risk to drivers. They examine the driver fatality rates in the subject vehicle and also in the collision partner, thus including the concept of aggressivity. They find that mass and size correlate inversely with risk to drivers in cars. However, the correlation is not strong, and they argue that mass and size are only two of the elements of vehicle design, and that other features such as quality of design and manufacture and specific safety features are more important. The analyses, however, are restricted to vehicle types and make and models, and do not explicitly include or control for specific safety features.

The exploration of driver behavior, while not extensive, revealed patterns of association between vehicle types, driver variables, and risks. While noting that some of the risk to drivers and others is attributable to driver behavior, the authors argue that most of the range of risk is attributable to vehicle design. However, their analysis does not determine how much driver factors matter. Thus, the conclusion that risk to drivers is mostly attributable to design is speculative. One can look at the same data and argue that driver factors are most important. The answer probably lies somewhere in between, but there clearly are relationships between vehicle models and driver behavior, and these should not be marginalized in analyses of fatality risks.

Furthermore, it should be noted that the variables, risk to driver and risk to others, in this analysis are rates (fatalities per million registered vehicles). In general, least squares linear regression models without modification are not exposure-based and therefore not appropriate for the analysis of fatality risk. For the most part, inference drawn from these models tends to be weak because they do not account for differences in exposure measures in the denominators of the rates. Thus, the approach used is overly simple for the problem being investigated.

5.2 The Wenzel and Ross 2006 Paper

Increasing the Fuel Economy and Safety of New Light Duty Vehicles [22]
White Paper for the Flora Hewlett Foundation
T. Wenzel and M. Ross

The objective of the paper was to show that the substantial majority of casualties in motor vehicle crashes are unrelated to the masses of the vehicles involved. The paper first reviews the distribution of fatal crashes and fatalities, and shows that in 2004, collisions of two or more vehicles accounted for 40% of all crashes, 42% of driver fatalities, and 48% of occupant fatalities; single vehicle crashes (into objects) accounted for 34% of crashes, 33 % of driver fatalities, and 38% of occupant fatalities; and first event rollovers accounted for 9% of driver and 10% of occupant fatalities. They also noted that collisions between light duty trucks with cars and with other light duty trucks have been increasing. The authors review previous work on fatality rates including the 2003 NHTSA and the 2005 DRI reports noting that mass and size are correlated, and the correlation of these in future vehicles might be less. They point out that in the NHTSA analysis, net fatality from weight reduction increases in some crash types and decreases in others. They point out that DRI finds that a 100 lb reduction while holding track width and wheelbase constant would result in a net reduction in the number of fatalities, reducing track width or wheelbase would increase fatalities, and combining weight and size reductions would result in a small statistically insignificant increase in fatalities. They end the review by stating that both the NHTSA and DRI reports indicate that “size may be as important if not more important than mass in protecting drivers in many types of crashes; and that the analysis of size and mass using historical data is made difficult by their tendency to be correlated, at least in vehicles of current design”(page 18).

The authors then go on to introduce the Wenzel and Ross model which has been summarized above. At the conclusion of this report, the authors delve briefly into crash injury causation. The analysis lacks context, ignores decades of detailed injury biomechanics and restraint system research, and cites selectively a small fraction of current literature. They cite as “new” on-site crash investigation, an area of research that enjoyed widespread development in the early 1960s. The authors’ suggestion that the fact that more injuries are attributed to interior contact than to “restrained acceleration” is supportive of their hypothesis is without basis. Very briefly, the amount of force that the restraint system can apply to the occupants is regulated indirectly by limits on chest compression, chest acceleration, and other measures. Efforts to protect occupant’s chests from belt loads are in direct conflict with the desire to restrain forward motion of the head to reduce the likelihood of contact. This tradeoff is made more challenging by the increased vehicle accelerations that result from light weighting in collisions between mobile objects.

5.3 The Wenzel and Ross 2008 Paper

The Relationship between Vehicle Weight/Size and Safety [23]
To be published in Physics of Sustainable Energy
T. Wenzel and M. Ross
May 19, 2008 Draft Report

This paper is an update of the 2005 paper reviewed above. The risk to drivers, risk to others, and combined risk again served as the fatality rates of concern. The methods were the same as those in the 2005 paper.

Fatality risk as a function of mass is examined. There is a linear relationship with risk to driver decreasing with increasing mass, but again it is weak (R^2 of 0.17). They repeat the regressions of risk by mass segmented by manufacturers and separately by resale value at five years. The mass to risk categorized by manufacturers is stronger again, when just Japanese and German manufacturers are considered, the R^2 of the linear regression of curb weight to risk to drivers is 0.54 (again risk decreasing with curb weight). As in the 2005 paper, the authors show a better linear fit of resale price of 5-year old cars (MY 98 in 2003) to risk to drivers (R^2 of 0.82).

This analysis shows similar patterns of risk by vehicle type as in the 2005 report. Among cars, the highest combined risks are for subcompact cars (high risk) and sports cars. The lowest overall risk is in import luxury cars. Among light duty trucks, the highest risks are in the pickups, with the highest overall risk and risk to others for the 1-ton pickup, followed by the 4/3-ton, then by ½ –ton, then the compact pickup. The total risk in truck-based SUVs is similar to that of compact cars. Total risk in cross-over SUVs is very low, just slightly higher than that of import luxury cars. The risk in rollover crashes in compact cross over SUVs is more than one-half that of compact size truck-based SUVs, and the risk in rollover crashes in midsized crossover SUVs is less than one-quarter that of mid size truck-based SUVs.

To explore the effect of driver behavior on risk, the authors examined rollover crash data by vehicle type, age, gender, and bad driver rating (defined in the 2005 study). They found that the drivers of truck-based SUVs are not different in age, gender, and bad driver rating than low-risk subcompact cars. Therefore, the differences in high rollover risks to drivers and risk to others in SUVs and pickups are probably not caused by drivers, but are due to vehicle design. Pickups, however, are driven more on rural roads and the authors conclude some of the risk for drivers in pickup trucks is due to environment.

The same criticisms apply to this paper as to the previous ones. The division of risk into the two categories is very useful and does show the effects of vehicle design. However, the approach of simple linear regression models is not adequate for the analysis of rates.

5.4 The Wenzel 2009 Comment Paper

Comments on the Joint Proposed Rulemaking to Establish Light Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards [19]
Docket NHTSA-2009-0059 and Docket EPA-HQ-OAR-2009-0472
October 27, 2009
T. Wenzel

The statistical analyses of casualty risk presented in the October, 2009 document are from a draft of the following report, and are reviewed below.

5.5 The Wenzel 2010 Paper

Analysis of the Relationship between Vehicle Weight/Size and Safety
and Implications for Federal Fuel Economy Regulation [20]
Final Report prepared for the Office of Energy Efficiency and
Renewable Energy, US Department of Energy
LBNL-3143E
T. Wenzel

The analyses presented here examined the relationship between vehicle weight, size (wheelbase, track width, and their product footprint), on fatality and casualty risk for drivers of passenger cars and light trucks. Model types within the passenger car and light truck categories were considered in the analyses. Fatality risk was defined as the ratio of driver fatalities in crashes involving specific vehicle models to the number of vehicle registrations. Casualty risk was defined as the number of driver fatalities and serious injuries in crashes involving specific car models to the number of these car models involved in police reported crashes.

Statistical analyses consisted of a series of simple linear regressions. The first set of regressions examined the correlations between vehicle weight and each of the size variables. The second set examined the relationship between fatality risk, weight and size variables. The third set examined the relationship between casualty risk and weight and size variables. Each analysis in each set was conducted separately for all light duty vehicles (LDV, that is for the passenger cars and trucks together, for just cars, and for just light trucks.

Data and models

Data for Analysis 1 – curb weight and size

Data on curb weight, wheelbase, track width and foot print for 2005 and later (through mid 2008) car models was obtained from US DOE CAFÉ compliance data. The first set of analyses used data for 2005 model year vehicles. Models of footprint as a function of curb weight and of track width as a function of wheelbase were fitted separately for all LDV, cars, and light trucks.

Data for Analysis 2 – fatality rate, curb weight, and size

Driver fatality risk was defined as the number of driver deaths per million registered vehicles for model years 2003 to 2007. Driver fatalities were obtained from FARS, and the registrations were obtained from Polk data. Only vehicle models with at least 0.5 million registration years were included. Crown Victoria models were excluded because this model carries a much higher risk of fatality due to its frequent use as a police car. Models of fatality rates were estimated for drivers of the subject vehicle and also for drivers in the other vehicle in all crashes (including motorcycle and heavy truck). Models were fitted separately for all LDV, cars, and light trucks.

Models of fatality rates in frontal crashes, and left side crashes for drivers of subject LDV were also estimated.

Data for Analysis 3 – casualty rate, curb weight, and size

The third set of analyses consisted of linear regressions relating casualty risk to vehicle curb weight, footprint, and vehicle type. Casualty risk was defined as the ratio of driver fatalities and serious injuries (incapacitating) to all police-reported crashes involving specific vehicle models. Data from all police-reported crashes in five states (Florida, Illinois, Maryland, Missouri, and Pennsylvania) for the years 2000-2004 were used. Crashes involving young male drivers (under age 26) and involving older drivers (age 65 and older) as well as crashes in counties with very low and very high population densities were excluded from analyses. This exclusion was intended to restrict casualty risk to vehicle features by removing those crashes in which driver and location features are known to be important. Weight, wheelbase, and track width information was obtained from Motor Trend magazine from models that were discontinued before 2005.

Models of casualty risk were estimated for drivers of the subject vehicle and also for drivers in the other vehicle in all crashes (including motorcycle and heavy truck). The models included separate regressions for curb weight and footprint as independent variables, and models that used curb weight and footprint together as separate independent variables.

Results – Size and Weight

Table 10 shows the results of the regression models that examined the correlation between vehicle curb weight and footprint, and wheelbase and track width. Models were run with unweighted data, and data weighted by the sales of each model type in year 2005. There is little difference between the results of unweighted and weighted models.

Table 10 2005 Vehicle Weight and Size (Wenzel 2010)

Dependent Variable	Indpenedent Variable	Vehicle Type	Vehicle models(n)	Coefficient	R^2
Unweighted					
Foot print(sq ft)	Curb Weight(lb)	LDV	1079	0.007**	0.63
		Cars	653	0.006**	0.60
		Light Trucks	426	0.007**	0.45
Track width (inches)	Wheelbase(inches^2)	LDV	1079	0.189**	0.54
		Cars	653	0.181**	0.43
		Light Trucks	426	0.167**	0.40
Weighted by 2005 sales					
Foot print(sq ft)	Curb Weigh(lb)	LDV	1079	0.007**	0.62
		Cars	653	0.006**	0.69
		Light Trucks	426	0.008**	0.42
Track width (inches)	Wheelbase(inches^2)	LDV	1079	0.206**	0.61
		Cars	653	0.304**	0.62
		Light Trucks	426	0.153**	0.48

** significant at p<0.001

The regressions show that footprint and curb weight are highly correlated in the 2005 model year LDVs. The relationship is the weakest for light trucks, however it is still substantial with an R^2 of .42-.45. Trackwidth is also highly correlated with wheelbase. Again the relationship is weakest but still considerable for light trucks (R^2 of .40-.48).

Wenzel notes that the relationship between curb weight and footprint for LDV (R^2 =63) is not as strong as that reported in Kahane's 2003 report. However, the vehicle data are from different years, so there is no reason that the relationship should be the same. Indeed, one would expect the correlation coefficient to be different. The correlation between curb weight and size shown in Wenzel's data is high enough to raise concerns about multicollinearity in models using both weight and footprint as independent variables.

Results – Fatality Risk and Size

Table 11 and Table 12 summarize the results of fatality risk models between vehicle footprint and size variables on fatality risk. Only the unweighted results are shown as there was very little difference between the unweighted and weighted model results. The very small R^2 values indicate no relationship between footprint and fatality rates and for drivers of the subject vehicle. The significance of the parameters for LDV and light trucks and the higher R^2 values indicate an increase in fatality risk to drivers in the other vehicle with the increase in footprint. The lack of significance for this parameter in the model for cars, indicates that this relationship in the LDV model is driven by the contribution from light trucks.

Table 11 Fatality Risk and Vehicle Weight and Size (Wenzel 2010)

Dependent Variable	Indpenedent Variable	Vehicle Type	Vehicle models(n)	Coeff	R^2
Risk to drivers Fatals/10^6 reg yr	Footprint (sq ft)	LDV	108	-0.8	0.01
		Cars	56	-2.4	0.01
		Light Trucks	52	1.2*	0.05
Risk to drivers of other vehicle Fatals/10^6 reg yr	Footprint (sq ft)	LDV	108	2.0**	0.38
		Cars	56	0.5	0.02
		Light Trucks	52	1.8**	0.33
Risk to drivers in frontal crash Fatals/10^6 reg yr	Wheelbase (inches)	LDV	108	-0.2*	0.01
Risk to drivers in left side crash Fatals/10^6 reg yr	Track width (inches)	LDV	108	-1.2**	0.15
* p=0.01, ** p<0.001					

Table 12 Fatality Risk to Drivers of Other Vehicle (Wenzel 2010)

Dependent Variable	Independent Variable	Vehicle Type	Vehicle models(n)	Coeff	R^2
Risk to drivers Fatals/10^6 reg yr	Footprint (sq ft)	LDV	108	-0.5	0.01
		Cars	56	-3.7	0.01
		Light Trucks	52	1.5*	0.05
Risk to drivers of other vehicle Fatals/10^6 reg yr	Footprint (sq ft	LDV	108	1.9**	0.38
		Cars	56	0.7	0.02
		Light Trucks	52	1.7**	0.33
Risk to drivers in frontal crash Fatals/10^6 reg yr	Wheelbase(inches)	LDV	108	-0.1*	0.01
Risk to drivers in left side crash Fatals/10^6 reg yr	Trackwidth (inches)	LDV	108	-1.5**	0.15

* p=0.015, ** p<0.001

Results – Casualty Risk and Weight and Size

Table 13 shows the results for unweighted regression models of casualty risk, weight and size. Weighting by 2005 sales numbers had little effect on the results, so only the unweighted results are shown here. The casualty risk to drivers of subject vehicles decreases as the curb weight increases and also as the footprint increases. The positive signs on the coefficients in the models of casualty risk to drivers of other vehicles indicate that the casualty risk increases as the curb weight increases and also as the footprint increases. However, the R^2 values of these models are small, indicating that not much of the variation in casualty risk is explained by these variables.

Table 13 Casualty Risk to Drivers (Wenzel 2010)

Dependent Variable	Independent Variable	Vehicle Type	Vehicle models (n)	Coeff	R^2
Casualty risk to drivers Casualties/10^3 crashes	Curb weight (100 lb)	LDV	144	-5.5**	0.31
		Cars	81	-10.4**	0.36
		Light Trucks	63	-3.5^	0.14
	Footprint (sq ft	LDV	144	-7.1**	0.26
		Cars	81	-13.5**	0.37
		Light Trucks	63	-3.1	0.08
Casualty Risk to drivers in frontal crash Casualties/10^3 crashes	Wheelbase (inches)	LDV	90	-2.7**	0.13
Casualty Risk to drivers in left side crash Casualties/10^3 crashes	Trackwidth (inches)	LDV	30	-18.3**	0.34
Casualty risk to drivers of other	Curb weight	LDV	144	4.6**	0.30
		Cars	81	1.8	0.03

vehicle Casualties/10^3 crashes	(100 lb)	Light Trucks	63	4.6**	0.19
	Footprint (sq ft	LDV	144	5.7**	0.26
		Cars	81	1.5	0.01
		Light Trucks	63	5.2**	0.22
^ p=.025, * p=0.002, ** p<0.001					

Table 14 and Table 15 each summarize six linear regression models with casualty risk as the dependent variable and first curb weight and footprint separately, then both together as independent variables, then each with a set of categorical variables representing vehicle types (variable = 1 for the vehicle type, 0 otherwise). Results for casualty risk to drivers of subject vehicles are in Table 14 and results for drivers of other vehicles are in Table 15.

Table 14 Casualty Risk to Drivers of Subject Vehicle (Wenzel 2010)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Coef(p)	Coef(p)	Coef(p)	Coef(p)	Coef(p)	Coef(p)
Intercept	464 (<.001)	594 (<.001)	525 (<.001)	506 (<.001)	668 (<.001)	596 (<.001)
Curb weight	-5.5 (<.0001)		-4.1 (.001)	-7.1 (<.0001)		-4.6 (<.003)
Foot print		-7.1 (<.0001)	-2.4 (.145)		-8.7 (<.0001)	-3.8 (.049)
Sports car				102.4(<.0001)	88.7 (<.0001)	96.5 (<.0001)
Import luxury				-56.1 (.016)	-77.4 (.001)	-63 (.007)
Minivan				- 10.6 (.574)	-11.3 (.560)	-4.6 (.807)
SUV				59.2 (.001)	0.1 (.994)	41.9 (.028)
Crossover				- 39.2 (.026)	-72.7 (<.0001)	-50.3 (.006)
Pick up				61.3 (.002)	84.4 (<.0001)	77.9 (<.0001)
Model R^2	0.31	0.26	0.32	0.55	0.54	0.56

In Models 1 and 2 there is only one independent variable and each is significant in its model (curb weight in model 1 and footprint in Model 2). However, the R^2 in each model is very low, indicating that each variable does not contribute much to the variation in the casualty risk as defined here. Model 3 used both curb weight and footprint as independent variables. Only curb weight is significant in Model 3. This is not surprising given the high level of correlation between curb weight and footprint. The R^2 in these models is quite low.

The three remaining models in this set added categorical (dummy 1, 0 variables) variables for 6 vehicle types (sports cars, imported luxury vehicles, minivans, SUV, cross over SUVs, and pickup trucks). These were added to curb weight in Model 4, to foot print in Model 5, and to both curbweight and foot print in Model 6. The directions of the signs are in directions that would be expected, and most of the vehicle model variables are significant. The coefficients of the sport cars and pickup trucks are positive, indicating an increase in casualty risk to drivers of these vehicles, while the coefficients for cross over SUVs and luxury imports are negative, indicating a decrease in casualty risk for these vehicles. The inclusion of the vehicle model types greatly improves the model fit. The R^2 in these two models are .54 and 55, indicating that including model types greatly improves the explanation of the variation in casualty risk. Model 6 uses highly correlated variables, and should not be considered for interpretation.

Models of casualty risk to other drivers are summarized in Table 15. As before, Models 3 and 6 should be disregarded due to the multicollinearity brought about by using both curb weight and track width in the model. The other models indicate that casualty risk to drivers of the other vehicle increase with curb weight and with footprint, and including vehicle type in the model greatly improves the model fit. The model parameters for the pickup and sport car are significant and positive, meaning that these vehicles increase the casualty risk to drivers of other vehicles in a collision. Interestingly, the cross over SUV does not have a significant effect on the casualty risk to other drivers in a collision.

Table 15 Casualty Risk to Drivers of other Vehicle (Wenzel 2010)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Coef(p)	Coef(p)	Coef(p)	Coef(p)	Coef(p)	Coef(p)
Intercept	126 (<.001)	20 (.613)	65 (.124)	153 (<.0001)	103 (.024)	144 (.005)
Curb weight	4.1(<.0001)		2.7 (.008)	2.9 (<.0001)		2.6 (0.72)
Foot print		5.5 (<.0001)	2.4 (.088)		3.2 (.002)	4.2 (.820)
Sports car				50.7 (.008)	55.8 (.004)	51.4 (.008)
Import luxury				27.3 (.207)	36.3 (.095)	28.2 (.202)
Minivan				-3.5 (.842)	-0.4 (.982)	-4.2 (.816)
SUV				27.1 (.089)	-52.8 (<.001)	29.2 (.108)
Crossover				-0.5 (.977)	13.4 (.400)	0.7 (.966)
Pick up				97.6 (<.001)	92.1 (<.001)	5.8 (<.001)
Model R^2	0.25	0.23	0.27	0.43	0.41	0.43

Wenzel argues that there are many design features that affect the risk of death or injury in a collision, and weight and size are only two of these. While increases in curb weight and footprint decrease the risks to drivers of the vehicle, the correlations between the weight and risk and size and risk as defined in Wenzel are quite weak. Addition of vehicle types greatly increases the explanatory power of the models.

The analyses used are simple linear regressions. The removal of crashes of young male drivers, drivers over 65 years of age, and those from very sparsely and very densely populated counties removes the drivers with the highest crash rates, and the most fragile drivers, as well as locations with highest crash rates. The reason this is done is to try to minimize driver behavior and location effects on crash occurrence. However, it also removes a portion of drivers, whose probability of serious or fatal injuries in a crash are different than those of middle age drivers.

This work and the earlier work by Wenzel and Ross provide two contributions. First, joint consideration of risk-to-drivers and risk-to-others focuses attention on aggresivity as an important issue. Second, the plots of risk per registered vehicle are unlike presentations of data in other literature, and highlight the fact that vehicles, as driven, pose widely varying risks to occupants within the vehicle and in other vehicles.

The analyses offer some insight into vehicle design effects on risk to drivers and others. However, the analyses do not account for covariates that are likely to have strong effects on differences between vehicles. The conclusion is drawn that mass effects are not large because the method does not reveal them. But readily hypothesized driver factors could account for the trends, and the scatter around the trends that they observed. The major criticism of this work, however, is that the approach of simple linear regression models used is overly simplistic and inadequate for the problem. The risks in both the fatality and casualty models are rates, and in general, least square linear regressions without modifications are not appropriate models for rates. Inferences drawn on these models tend to be weak because they do not account for the differences in exposure measures in the denominators of the rates.

6. Review of the J.P. Research and Daimler Chrysler Corporation Papers

Vehicle Dynamics Evaluation of the Size/Weight Question

In this section we review the size/weight question with a focus on engineering principles. This is useful to supplement statistical analysis which suffers loss of resolution due to correlated regressors for size and weight (see below). On the other hand, the detailed mechanical processes involved in crashes are highly complex, so improved understanding is best gained from a combination of crash statistics, simple models and experimental crash results.

From the perspective of vehicle factors, those influencing fatality risk (given that a crash has occurred) are predominantly the following

- Masses and linear dimensions of the colliding vehicles.
- Other aspects of vehicle design: structural design, restraints, interior design, collapsible steering column, etc. and especially the stiffness and energy absorbing properties of the vehicle structures.
- Vehicle dynamic performance: the ability a typical driver has to control the vehicle in the pre-crash phase: being able to steer and brake effectively and without causing vehicle instability.

This list can be condensed using a few assumptions. First, we assume analysis is limited to comparable events, with comparable vehicles and drivers, and where only mass and size are changed. For the vehicle this means that structural design is largely similar within the classes of passenger car and light truck, and reflective of future design changes. If, for example, new materials are used in the future design of lighter weight cars, the analysis of historical data is not particularly informative. For the pre-crash conditions, even if the heavier vehicles are less responsive to immediate pre-crash braking and steering, our assumption is that the kinematics of the crashes are minimally affected.

In the following, we focus on frontal collisions involving passenger cars and light trucks, and review results from three especially relevant publications.

6.1 The J.P. Research 2003 Paper

Influence of Vehicle Size and Mass and Selected Driver Factors on Odds of Driver Fatality [11]
Association for the Advancement of Automotive Medicine, 47th Annual Proceedings
September, 2003
J. Padamanaban

This study was performed in two phases of analysis to estimate the contributions of size and mass parameters to fatality odds. The first phase is based on FARS data in which precisely one driver was killed, and fatality odds are based on differences between the two vehicles/drivers as to which one was killed. A comprehensive range of size metrics was initially defined, including linear measures (e.g. overall length, wheelbase, front overhang, etc.), area (e.g. length × width) and volume measures (e.g. length × width × height). Fifteen such measures were considered, a list that was expanded with 12 further measures of crash protection and structural compatibility.

Since this very large number of variables could not be varied in a systematic fashion, as in a factorial or fractional factorial design, the authors included an informal stage of factor prioritization: "results were then examined carefully to select the best statistical models based on statistical and engineering interpretation of findings". Explanatory power of the different factors was included, based on the Wald Chi-square test. Engineering considerations were presumably more subjective, but in any case the authors were able to reduce the set of parameters and assess the relative contributions of mass and size metrics. They concluded that mass ratio between the two vehicles accounted for 19% of the fatality odds in car-to-car crashes, and 27% for light truck-to-car crashes. By comparison, length and area measures contributed only a few percent. In the second class of crash, the fatality risk was also strongly increased if the striking vehicle was the light truck.

In Phase 2, an expanded set of crashes (not just fatal crashes) was analyzed using data from a much larger crash set; the baseline data was obtained from three states and extrapolated to a nationally representative set. In this case, fatality odds are defined relative to the overall crash population and are more easily interpreted. In this analysis the mass contribution is more consistent, at around 20% in both types of crash, and again size metrics are less important (around 7% associated with the front axle to windshield distance).

The results of the above paper were revisited by the same group (JP research) in the following publication:

6.2 The J.P. Research 2009 Paper

Updated Evaluation of Size and Mass Effects in Front-to-Front
Crashes Involving Light Vehicles [4]
SAE International Technical Paper Series, 2009-01-0375
V. Eyges and J. Padmanaban

This recent paper focuses on the difficulties of separating the contribution of various parameters to fatality risk, and especially the various size and weight parameters. In part, this was prompted by results from Van Auken and Zellner [DRI, 17] which were very different from those of Padmanaban.[11] Eyges and Padmanaban properly recognize the problem arising from "multicollinearity" among regressors, which result in quite different regression models depending on the precise approach taken. This is true for any form of regression where a pair of regressors is highly correlated; in such a case it is even possible to adjust one regressor to an arbitrary new value, and then fit the second to compensate, hardly affecting the goodness of fit.

Results presented in this paper (for example those shown in the appendix) are highly plausible, and the mass ratio term ranks very highly. The authors include a clear explanation of why fatality risk is expected to increase with increasing mass ratio, and use Monte-Carlo simulation to verify that the non-availability of impact speed should not cause major bias in the logistic regression. Overall their methodology seems appropriate and their conclusions are reasonable. However, their analysis does not rule out the possible validity of Van Auken. In the absences of actual calculation errors, the results in Tables 2 and 3 (of Eyges and Padmanaban, 2009) are all consistent with the source crash data, and from a purely statistical perspective it is not easy to decide that Eyges and Padmanaban are "right" and Van Auken is "wrong", or vice-versa. Mainly, the source data do not uniquely decide the correct interpretation, and other aspects need to be addressed.

The criticisms made in this 2009 paper of Van Auken's results are mainly that (i) statistically, the model is more complex than the data supports (ii) the trends are less plausible than those of Kahane and also the 2009 JP Research study. Here we focus on the second point and reinterpret the results. At the simplest level, one might argue that uniformly reducing mass across the fleet of passenger cars and light trucks would reduce the kinetic energy of impact, reducing intrusion and therefore fatality risk. In this case, the Van Auken assessment of a 6.7% reduction of fatalities in car-to-car frontal crashes for a 100 lb weight reduction may be considered reasonable. If on the other hand one considers the acceleration pulse to be of greatest relevance in matching to fatality risk, the car-to-car result would be expected to be close to zero, consistent with the results (Eyges and Padmanaban Table 2) of Kahane (1997) and the JP Research study. The "energy bias" explanation is not fully consistent; however, in Table 3 Van Auken's estimate for fatality rate in truck-to-car crashes is near zero (+0.6%) when there is a 100lb reduction in the weight of the light truck. Also, it is difficult to accept this excessive energy bias in the case of truck-to-car collisions, when reducing the weight of the car is deemed to have minimal effect on fatality rates, even though basic dynamics indicates that the magnitude of acceleration crash pulse experienced by the passenger car will be increased (irrespective of the assumption that size remains unchanged). The positive rate increases estimated by Kahane and JP Research are broadly more convincing.

Of course it may that there is a more complex interaction between mass and the larger dimensions of a light truck, so this cursory examination is not conclusive. To make the assessment more quantitative, it is worth reviewing in more detail a study on the mechanics of the crash itself. In the following paper, the authors use simple models and quantitative crash test data to evaluate the effects of independently changing size and mass.

6.3 The Daimler Chrysler 2003 Paper

Estimation of the Effects of Vehicle Size and Mass on Crash-Injury Outcome Through Parameterized Probability Manifolds [10]
SAE International Technical Paper Series, 2003-01-0905
G. Nusholtz, G. Rabbiolo, and Y. Shi

The focus here is again on frontal crashes, now restricted to a population of passenger cars only, as represented by a sample of 22 vehicles tested under the NCAP program.

Results are presented from FARS to demonstrate a well-tested relationship between fatality risk and mass ratio, $R = (\text{mass ratio})^{\alpha}$, which can be interpreted as $R \propto \Delta v^{\alpha}$ or $R \propto \overline{A}^{\alpha}$ where Δv is the change in vehicle velocity during the crash and $\overline{A}$ is the mean deceleration (these being equivalent if the duration of the impact is constant). This does not influence the conclusions above, since other factors – especially structural stiffness and crush space – may also be important when independent changes are made; while the mass ratio is known in the FARS data, the physical interpretation is somewhat open since Δv and $\overline{A}$ are not known in the crash data.

The mechanical interpretation is therefore extended based on a simple dynamic model. If the correct severity measure is given by $R \propto \Delta v^{\alpha}$, it is shown that mass ratio is the dominant vehicle variable – in fact, according to the model, the only influential vehicle parameter. The authors suggest, and indeed it seems correct, that the duration of the crash is also important, as this affects the mean magnitude of the crash pulse, $\overline{A}$. (The authors do not go into greater detail about injury mechanisms and criteria, and clearly the shape and peak of the acceleration pulse are important; however, when looking at major trends it seems reasonable to use a simple explanatory variable such as $\overline{A}$). In this case, simple analysis of momentum and energy changes are not sufficient to describe the risk dependency, and therefore the authors consider a simple model that includes a two-stage force vs. deflection characteristic for the impacting vehicles. Monte-Carlo simulation is conducted based on the assumed $R \propto \overline{A}^{\alpha}$ relationship, with constant obtained from another (NASS) database. Although limited in accuracy and fidelity, this physical model appears appropriate for estimating gross dependencies and certainly addresses the question of whether vehicle size can reasonably be the dominant vehicle factor for fatality risk. Based on the two-step model and the assumed risk model, it is found that changing the mean mass of the vehicle population (leaving variability unchanged) has a stronger influence on fatality risk than corresponding (feasible) changes in mean vehicle dimensions. If one accepts the methodology, there is an unequivocal conclusion that reducing vehicle mass while maintaining constant vehicle dimensions will increase fatality risk, and this conclusion is robust against realistic changes that may be made in the force vs. deflection characteristics of the impacting vehicles. This is in stark contrast to the prediction (Eyges and Padmanaban 2009, Table 2) of a 6.7% reduction in fatalities in car-to-car frontal crashes given a 100lb weight reduction in the fleet as a whole.

If the $R \propto \Delta v^{\alpha}$ is assumed instead, there is no change in fatality risk in this case (since mass ratios are unchanged), which is consistent with the results of the other authors in Table 2. It would also predict increases in risk when mass ratios increase (as in truck-to-car crashes when the passenger car weight is reduced) again in contrast to Van Auken's predictions.

7. Review of the Robertson Papers

7.1 The Robertson 2006 Paper

Blood and Oil: Vehicle Characteristics in Relation to Fatality Risk and Fuel Economy [12]
American Journal of Public Health, Vol 96, No 11
November, 2006
L. Robertson

It is difficult to provide a critical review of this paper. In general, it needs a lot of work. It took several readings to understand this paper, and still some of the methodology is not clear. The problem being solved in this work is very difficult, and the methods used are too simple to adequately cover the topic appropriately. It is understood that this paper passed a peer review process, but our recommendation would have been to review the paper again after a major revision. The biggest problem, other than the methodology, is that the paper needs to be documented better, with clearer explanations and concrete examples so that the reader understands exactly what was done.

Some of the claims in the paper appear to be overstated. One of the claims is that half the deaths involving passenger cars, vans, and SUVs could have been prevented if all vehicles had crashworthiness and stability equal to those of the top rated vehicles. Considering the complex nature of the events associated with fatal crash involvement, this is a very ambitious claim.

The methodology is outlined as best can be determined based on the information provided in the paper. FARS data were used to identify vehicles involved in fatal crashes. Exposure in the form of vehicle years was calculated using sales data provided by *Ward's Automotive Yearbook*. Disaggregate logistic regression was performed using the following predictor variables:

- Lateral distance needed to make a 180-degree turn (used as size)
- An average of four factors used as an index of crashworthiness (based on vehicle ratings by the Insurance Institute for Highway Safety)
- A categorical measure of stability (T/2H<1.2, else 1.2 where T is track width and H is the high center of gravity from the ground)
- Curb weight
- Indicators for vans and SUVs

Two separate logistic models were fit to estimate odds of death to drivers and all occupants. In one set of analyses, vehicles were limited to passenger cars, excluding sports cars. Standard least squares regression models were used in an attempt to identify issues with collinearity.

The following points are made about the methods and results:

- The paper likely overstates findings significantly. One of the assumptions of logistic regression is that the observations are independent. Since fatalities occur at the occupant level, fatalities in the same vehicle are not independent. For single-vehicle crashes, violating the independence assumption may not be too severe if there are not too many vehicles with multiple fatalities. However, in multiple-vehicle crashes, fatalities can occur not only in the "subject" vehicle, but also in the "other" vehicle. These outcomes are not independent and the logistic regression model is too simple for estimating the odds of fatality given the predictor variables. The effect is that estimates of standard errors of parameters in the model are too small, resulting in significant and likely overstated findings. Note that this paper does not fit single-vehicle and multiple-vehicle crash types separately. Both crash types are fit in the same model.

- Related to the point above is that treating each fatality as an independent outcome artificially increases the sample size. Likelihood-based estimation upon which the logistic regression is based will tend to identify significant results when the sample size is large, even when effects are small. Note that 7,263 driver deaths are reported with a total of 14,438 fatalities.

- Based on the complexity of the problem being investigated, the model is too simple to answer the questions being considered. The data are much too variable for the method used, and separate analyses, at least by crash type, should have been considered. A separate analysis was performed for passenger cars, excluding sports cars, but no provisions were made for single-vehicle versus multiple-vehicle crashes, or multiple-vehicle crashes involving light pickups or heavy trucks.

- Considering that vans and SUVs have very different operating characteristics than passenger cars, models for vans and SUVs should likely be fit separately and should not be included in models with passenger cars.

- No adjustment for driver behavior using surrogates such as age and gender was made.

- Using lateral distance needed to perform a 180-degree turn as a measure of size instead of wheelbase may be a useful alternative.

- The methodology section should include a better description of the exposure measure used. It appears that vehicle years were calculated based on published sales information. A reference is given, but the paper should explicitly describe the method and give a specific example of the structure of the database. Some preliminary descriptive statistics should be provided before model fitting other than total fatalities and total years of use.

- Intermediate calculations leading to the claims that fatality rates would have been reduced by 28 percent and fuel use reduced by 16 percent if vehicle weights had been reduced to the weight of vehicles with the lowest weight per size should be given. Intermediate results in tables would help the reader determine if these claims are valid.

7.2 The Robertson 2007 Paper

Prevention of Motor-Vehicle Deaths by Changing Vehicle Factors [13]
Injury Prevention, 13, p. 307-310
August, 2007
L. Robertson

There are many similarities between this paper and the one reviewed above. The data and methodology are basically the same; therefore, the differences will be highlighted. In the logistic regression model, curb weight is not fit. Based on high correlation with weight and horsepower, fuel economy is used as a surrogate. Another difference is that the effects of electronic stability control (ESC) on fatality risk is assessed by including a variable indicating whether ESC is standard or optional equipment for each make/model under investigation. In addition to a variable used to measure crashworthiness from front crash tests, a variable designed to measure crashworthiness from side impacts was also included in the model. Three separate models were fit to assess the effects of the predictors on risk to all road users, drivers only, and pedestrians and bicyclists.

The following points are made about the methods and results:

- The scope of the problem is too broad, while the logistic regression model is too simple. The model is fit to both single-vehicle and multiple-vehicle crashes. The data are too variable for such a simple model. Comments about the assumption of independence in logistic regression made above also apply here. Standard errors of parameter estimates are likely to small resulting in confidence intervals that are too tight.

- As in the 2006 paper, claims appear to be overstated. While many papers have been written about the positive effects of ESC, this paper concludes that ESC would reduce mortality by 42 percent if it were installed on all vehicles. An interesting exercise might first be to try to estimate the population of fatal crashes that would likely benefit from ESC and determine if 42 percent of fatalities is possible. Many crashes, even single-vehicle crashes, do not involve loss of control.

- Data recorded in the FARS database contain information about fatal involvements which tend to be high energy and high impact crashes. Many of these crashes are head-on, side impact, or single-vehicle ran-off-road type crashes. Separate analyses for these crash types, and others (rear-end) could be more powerful since they focus on certain prevalent fatal crash types.

- This paper would be greatly improved if it were written in a more clear and concise manner. The motivation for pursuing certain analyses should be given greater detail. For example, the methods used to obtain the results in Table 3 require more explanation.

- Fatalities occur at the person level, yet the databases are created at the vehicle level. That is, each record in the database is a vehicle. When total fatalities are being assessed (societal costs), it might be beneficial to develop a model that analyzes data at the person level. Such a model would require added complexity due to correlated outcomes between occupants in the same vehicle, and vehicles in the same crash, but improved inference would result especially with respect to estimates of standard errors. Our report to NHTSA discusses alternative models that can be fit to data at the person level.

8. Conclusions

In December 2007, Congress passed the Energy Independence and Security Act (EISA) that required NHTSA to set ”attribute-based” Corporate Average Fuel Economy (CAFE) standards, in which a manufacturer’s compliance obligation depends on the mix of vehicles they produce for sale. NHTSA selected the vehicle footprint (the measure of a vehicle’s wheelbase multiplied by its average track width) as the attribute upon which to base the CAFE standards for model year 2012-2016 passenger cars and light trucks. These standards are likely to result in weight reductions in new light duty vehicles. As part of its regulatory analysis, the government would like to estimate the effect of the new CAFE standards on safety in terms of crash injuries and fatalities. A number of fairly comprehensive statistical papers have been published analyzing associations between fatality/injury rates and vehicle weight, track width, and wheelbase. Many of the papers arrive at conclusions that are inconsistent.

This report is a review of papers analyzing associations between crash/fatality outcome and vehicle weight and size. The various studies are based on different data sources, model assumptions, and methodologies. The authors of these studies represent a mix of those in government, research institutes, and academia, and have a broad range of professional backgrounds and philosophies. The goal of this report is to provide an independent review of the papers and to critically assess the methods and conclusions presented. Technical reports and papers published by the following organizations and authors were reviewed:

- National Highway Traffic Safety Administration (NHTSA)
- Dynamic Research, Inc. (DRI)
- T. Wenzel and M. Ross
- J.P. Research
- Daimler Chrysler Corporation
- L. Robertson

The Executive Summary of this report provides a general overview of the conclusions in this review, so the main summary and findings can be found there. However, the more important points of the review will be reiterated with special attention given to the following topics:

- Use of logistic regression as an exposure-based risk model
- Issues associated with multicollinearity
- Use of induced exposure methods
- The crash types and vehicle types considered
- Analysis of the data at the person level rather than the vehicle level

Both NHTSA and DRI used logistic regression in one form or another in several reports to analyze fatality rates. Logistic regression is not one of the standard exposure-based risk models for analyzing rates. However, when rates are very small, as is the case when fatalities in the numerator are relatively rare and exposures in the denominators are very large, the model approximates the Poisson log-linear model for rates, which is a standard exposure-based risk model. However, in practice the Poisson model is generally too simple for use in observational studies.

Likelihood-based tests, derived from fitting logistic and Poisson models, tend to be significant even when results show small effects, as long as sample sizes are large enough. Construction of confidence intervals and tests of hypotheses depend on specification of a model that accommodates the variation in the data. The study under consideration is an observational one using various sources of data, and it could be argued that the logistic model is misspecified. In the presence of extra-variation, standard errors tend to be too small and significance can be overstated. A more robust model would at least adjust standard errors to account for the extra-variation often encountered in observational studies. In Section 3.2 and 3.4, alternative models and methods are described that could be used to account for the extra-variation that was likely present in the data analyzed.

As shown in several NHTSA and DRI reports, the variables curb weight, track width, and wheelbase tend to have strong positive correlations based on analysis of historical data. It is well-known that inclusion of highly correlated variables generally leads to multicollinearity, which can result in unstable estimation of parameters. If predictor variables are highly correlated and have a strong positive association with the response, those variables are potential surrogates for one another. When entered into separate models one at a time, they generally have strong associations in the same direction. However, if entered together in the same model, the potential exists for the magnitudes of the parameter estimates and associated standard errors to change significantly.

If the three variables are entered together in the same model, and one or more coefficients change sign, then making inference about the effects of certain predictors on the response variable can be misleading. For example, if the estimate for curb weight changes sign, making inference about changes in curb weight while holding track width and wheelbase constant could be regarded as overly ambitious. We seek the simplest model with the fewest parameters that explain the data well. Including highly correlated predictor variables in the same regression model leads to over-fitting and unstable estimation. Inference in the presence of multicollinearity should be judged with great concern.

Induced exposure vehicles are generally the non-culpable vehicles in two-vehicle crashes and were derived from various State data files. In the absence of a traditional exposure measure, such as vehicle miles traveled (VMT), induced exposure is a surrogate that represents the denominator of a rate. Admittedly, there are no other sources of exposure data available that are recorded at the level required to analyze fatality rates in the studies reviewed. In the NHTSA 2003 report, a novel approach was used whereby vehicle registration data and odometer readings were used to apportion vehicle miles traveled to each induced exposure crash. In the absence of viable alternatives, the approach seems logical. However, there is a concern that the method could introduce bias in certain situations. For example, non-culpable vehicles tend to have very different speed distributions than vehicles involved in fatal crashes. The authors of the studies seem to be aware of these and other differences and attempts were made to adjust for potential bias. The use of induced exposure that is limited to certain states is likely to be an issue for further investigation as long as other sources of exposure such as VMT remain unavailable.

In the NHTSA and DRI reports, six crash types were considered: principal rollover, hit fixed object, hit pedestrian/bicycle/motorcycle, car-to-heavy truck, car-to-car, and car-to-light truck. The three single-vehicle crash types seem to be well-specified. However, the three multiple-vehicle crash types seem to be too general. Many of the high-impact crashes in the FARS data are opposite direction or head-on crashes. Similarly, FARS data should support analysis of side impact and rear-end crashes. Possibly, focusing on additional multiple-vehicle crash types would lead to improved inference by reducing variability in the more broadly defined ones. Some of the other papers did not even distinguish and present separate analyses by crash type.

Due to design features, LTVs are often described as “niche” vehicles. It appears that estimation of effects for LTVs is more demanding than for passenger cars. A possible way to reduce variability and focus in on the effects of size and weight for LTVs could be to consider more vehicle types.

- Light pickups
- Heavier pickups
- Minivans (if possible)
- SUVs

An argument can possibly be made for analysis of fatality outcome at the person level. Casualties inside the same car and the same crash are more correlated than casualties involved in a different car or crash. The natural structure of crash data consists of correlated observations with nested crash-car-occupant levels. In this review, it was stated several times that the standard logistic regression model assumes independent observations.

A random effects logistic model was described in Section 3.2 that allows taking into account the differences in each crash and each car and permits modeling of correlations between occupants of the same car and between cars of the same crash. This random effects model could be fit to a database constructed at the person level and would provide adjusted standard errors relative to the ones produced by the ordinary logistic model. Since the goal of the analysis is to estimate the net change in fatalities to society as a whole, an analysis at the person level could be considered. These days, statistical software packages fit these models as a standard option, and could provide results for comparison with results presented in the reviewed reports.

9. References

1 Agresti, A. (2002). *Categorical Data Analysis 2nd ed.* Wiley.

2 Breslow, N., and Day, N.E. (1987). *The Design and Analysis of Cohort Studies* Vol. II. Lyon: IARC.

3 Dobson, A. (2001). *An Introduction to Generalized Linear Models 2nd ed.* Chapman and Hall.

4 Eyges, V. and Padmanaban, J. (2009) Updated Evaluation of Size and Mass Effects in Front-to-Front Crashes Involving Light Vehicles, SAE paper 2009-01-0375.

5 Kahane, C.J. (1997). Relationships between Vehicle Size and Fatality Risk in Model Year 1985-93 Passenger Cars and Light Trucks, DOT HS 808 570.

6 Kahane, C.J. (2003). Vehicle Weight, Fatality Risk and Crash Compatibility of Model Year 1991-99 Passenger Cars and Light Trucks, DOT HS 809 662

7 Kahane, C.J. (2004). Response to Docket Comments On NHTSA, Technical Report Vehicle Weight, Fatality Risk and Crash Compatibility of Model Year 1991 -99 Passenger Cars and Light Trucks [68 FR 661 531Docket No. NHTSA-2003-163 18.

8 Kahane, C.J. (2010). Relationships Between Fatality Risk, Mass, and Footprint in Model Year 1991-1999 and Other Passenger Cars and LTVs.

9 McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models 2nd ed.* Chapman and Hall.

10 Nusholtz G., .Rabbiolo, G., and.Shi, Y. (2003) Estimation of the Effects of Vehicle Size and Mass on Crash-Injury Outcome Through Parameterized Probability Manifolds, SAE Technical Paper Series 2003-01-0905.

11 Padmanaban, J. (2003).Influences of Vehicle Size and Mass and Selected Driver Factors on Odds of Driver Fatality, AAAM 2003 USCAR.

12 Robertson, L. (2006). Blood and Oil: Vehicle Characteristics in Relation to Fatality Risk and Fuel Economy, Commentary in American Journal of Public Health, Vol 96, No 11

13 Robertson, L. (2007). Prevention of Motor-Vehicle Deaths by Changing Vehicle Factors, Injury Prevention, 13, p. 307-310.

14 Van Auken, R.M., and Zellner, J.W. (2002). An Assessment of the Effects of Vehicle Weight on Fatality Risk in Model Year 1985-98 Passenger Cars and 1985-97 Light Trucks, DRI-TR-02-02.

15 Van Auken, R.M., and Zellner, J.W. (2003). A Further Assessment of the Effects of Vehicle Weight and Size Parameters on Fatality Risk in Model Year 1985-98 Passenger Cars and 1985-97 Light Trucks, DRI-TR-03-01.

16 Van Auken, R.M., and Zellner, J.W. (2004). A Review of the Results in the 1997 Kahane, 2002 DRI, 2003 DRI, and 2003 Kahane Reports on the Effects of Passenger Car and Light Truck Weight and Size on Fatality Risk, DRI-TR-04-02.

17 Van Auken, R.M., and Zellner, J.W. (2005). Supplemental Results on the Independent Effects of Curb Weight, Wheelbase, and Track on Fatality Risk in 1985-1998 Model Year Passenger Cars and 1985-1997 Model Year LTVs, DRI-TR-05-01.

18 Van Auken, R.M., and Zellner, J.W. (2009). DRI Comments on Safety Impacts of EPA-NHTSA Proposed Rule to Establish Light-Duty Vehicle Greenjouse Gas Emission Standards and Corporate Average Fuel Economy Standards Dated September 28 2009, DRI-TM-09-86.

19 Wenzel, T. (2009). Comments on the Joint Proposed Rulemaking to Establish Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards Docket No. NHTSA–2009–0059 and Docket No. EPA-HQ-OAR-2009-0472.

20 Wenzel, T. (2010). Analysis of Relationship between Vehicle Weight/Size and Safety and Implications for Federal Fuel Economy Regulation, Final Report prepared for the Office of Energy Efficiency and Renewable Energy, Department of Energy, LBNL-3143E.

21 Wenzel, T., and Ross, M. (2005). The Effects of Vehicle Model and Driver Behavior on Risk, *Accident Analysis and Prevention*, 37 (2005) 479–494.

22 Wenzel, T., and Ross, M. (2006). Increasing the Fuel Economy and Safety of New Light Duty Vehicles, White Paper for the William and Flora Hewlett Foundation.

23 Wenzel, T., and Ross, M. (2008). The Relationship between Vehicle Weight/Size and Safety, to be published in *Physics of Sustainable Energy* by the American Institute of Physics.

24 Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* 57, 348-368.

Appendix A: List of Papers Provided by NHTSA

<u>NHTSA</u>

The Effect of Decreases in Vehicle Weight on Injury Crash Rates
NHTSA
January 1997

Relationships between Vehicle Size and Fatality Risk in Model Year 1985-93 Passenger Cars and Light Trucks
Charles J. Kahane, Ph.D.
DOT HS 808 570 NHTSA Technical Report
January 1997

Vehicle Weight, Fatality Risk and Crash Compatibility of Model Year 1991-99 Passenger Cars and Light Trucks
Charles J. Kahane, Ph.D
October 2003
NHTSA Technical Report DOT HS 809 662

RESPONSE TO DOCKET COMMENTS ON NHTSA TECHNICAL REPORT
Vehicle Weight, Fatality Risk and Crash Compatibility
***of Model Year 1991 -99 Passenger Cars and Light Trucks* [68 FR 661 531**
Docket No. NHTSA-2003-163 18
Charles J. Kahane, Ph.D.,,National Highway Traffic Safety Administration
November 2004.

Relationships between Fatality Risk, Mass, and Footprint in Model Year 1991-1999 and Other Passenger Cars and LTVs
Charles J. Kahane, NCSA, NHTSA
March 24, 2010

Pages 464-542 of ***Final Regulatory Impact Analysis: Corporate Average Fuel Economy for MY 2012-MY 2016 Passenger Cars and Light Trucks***
Office of Regulatory Analysis and Evaluation National Center for Statistics and Analysis
National Highway Traffic Safety Administration, Washington, DC

<u>Dynamic Research</u>

An Assessment of the Effects of Vehicle Weight on fatality Risk in Model Year 1985-98 Passenger Cars and 1985-97 Light Trucks.
Volume I: Executive Summary
DRI-TR-02-02
R.M. Van Auken, J.W. Zellner, Dynamic Research
February 2002

A Further Assessment of the Effects of Vehicle Weight and Size parameters in Fatal Crash Risk in Model Year 1`985-1998 passenger Cars and 1985-1997 Light Trucks. Vol II, Technical Report DRI-TR-03-01.
R.M. Van Auken, J.W. Zellner, Dynamic Research
January 2003.

A Review of The Results in the 1997 Kahane, 2002 Dri, 2003 Dri, And 2003 Kahane Reports
on the Effects of Passenger Car and Light Truck Weight and Size on Fatality Risk
DRI-TR-04-02
R. M. Van Auken, J. W. Zellner, Dynamic Research Int.
March 2004

Supplemental Results on the Independent Effects of Curb Weight, Wheelbase, and Track on Fatality Risk In 1985-1998 Model Year Passenger Cars and 1985-1997 Model Year LTVs
Report DRI-TR-05-01
R.M. Van Auken, J.W. Zellner, Dynamic Research
May 20, 2005.

An Assessment of the Effects of Vehicle Weight and Size on Fatality Risk in 1985-1998 Model Year Passenger Cars and 1985-1997 Model Year Light Trucks and Vans
R. M. Van Auken, J. W. Zellner, Dynamic Research
SAE Paper 2005-01-1354

DRI Comments on Safety Impacts of EPA-NHTSA Proposed Rule to Establish Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards
R. M. Van Auken J. W. Zellner, Dynamic Research
September 28 2009
DRI-TM-09-86

<u>Wenzel</u>

The effects of vehicle model and driver behavior on risk
Tom Wenzel, Lawrence Berkeley National Laboratory
Marc Ross, Physics Department, University of Michigan
Accident Analysis and Prevention 37 (2005) 479–494
Accepted Aug 2004

Increasing the Fuel Economy and Safety of New Light-Duty Vehicles
Tom Wenzel, Lawrence Berkeley National Laboratory
Marc Ross, Physics Department, University of Michigan
September 18, 2006

White paper prepared for the William and Flora Hewlett Foundation's Workshop on Simultaneously Improving Vehicle Safety and Fuel Economy through Improvements in Vehicle Design and Materials

Comments on the Joint Proposed Rulemaking to Establish Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards
Docket No. NHTSA–2009–0059 and Docket No. EPA-HQ-OAR-2009-0472
Tom Wenzel, Lawrence Berkeley National Laboratory
October 27, 2009

Analysis of the Relationship Between Vehicle Weight/Size and Safety, and Implications for Federal Fuel Economy Regulation
Tom Wenzel, Lawrence Berkeley National Laboratory
March 2010
Final Report prepared for the Office of Energy Efficiency and Renewable Energy, US Department of Energy

The Relationship between Vehicle Weight/Size and Safety
Tom Wenzel and Marc Ross
To be published in *Physics of Sustainable Energy* by the American Institute of Physics
May 19, 2008 DRAFT – Please do not quote or circulate

Other papers

Blood and Oil: Vehicle Characteristics in Relation to Fatality Risk and Fuel Economy
Leon S. Robertson, PhD
Commentary, *American Journal of Public Health*, November 2006, Vol 96, No. 11, 1906-1909.

Prevention of motor-vehicle deaths by changing vehicle factors
Leon S Robertson
Injury Prevention 2007;13:307-310. doi: 10.1 136/ip.2W7.01 6204

Critique and Discussion of Robertson's Two Reports **(Kahane)**

April 15, 2010

Response to Kahane critique (Robertson)
July 9 2010

Influences of Vehicle Size and Mass and Selected Driver Factors on Odds of Driver Fatality
Jeya Padmanaban, JP Research
AAAM 2003 USCAR

Updated Evaluation of Size and Mass Effects in Front-to-Front Crashes Involving Light Vehicles
Vitaly Eyges and Jeya Padmanaban, JP Research, Inc.

SAE paper 2009-01-0375

<u>**Nusholtz**</u>

Estimation of the Effects of Vehicle Size andMass on Crash-Injury Outcome Through Parameterized Probability Manifolds
G. S. Nusholtz, G. Rabbiolo and Y. Shi, DaimlerChrysler Corporation
SAE TECHNICAL PAPER SERIES 2003-01-0905
March 2003.

Appendix B: Statement of Work

SECTION C - DESCRIPTION, SPECIFICATION, WORK STATEMENT

Independent Review Statistical Analyses of Relationship between Vehicles Curb Weight, Track Width, Wheelbase and Fatality Rates

C.1 BACKGROUND

The proposed Corporate Average Fuel Economy (CAFE) standards for MYs 2012 to 2016 are likely to result in weight reductions in new cars and Light Truck Vehicles (LTVs), possibly ranging from 100 to 600 pounds, depending on the size and type of vehicle. In the past, CAFE standards were set as universal or flat standards, with a single mpg number being applicable to every manufacturer – for example, the standard for passenger cars from 1990 through 2010 has been 27.5 mpg. Flat standards encourage manufacturers to comply, in part, by building smaller and lighter vehicles that achieve better fuel economy to “average out” more popular larger and heavier vehicles that achieve worse fuel economy. Down-weighting, whether by reducing footprint or otherwise, tends to be one of the most cost-effective means by which a manufacturer can meet CAFE standards.

With the passage of the Energy Independence and Security Act (EISA) in December 2007, Congress required that NHTSA set “attribute-based” CAFE standards, such that a manufacturer’s compliance obligation depends on the mix of vehicles they produce for sale. NHTSA chose “footprint” (the measure of a vehicle’s wheelbase multiplied by its average track width) as the attribute upon which to base CAFE standards for MYs 2012-2016 passenger cars and light trucks. The footprint-based passenger car and light truck standards are defined by constrained linear function target curves, with more stringent mpg targets for smaller-footprint vehicles and less stringent targets for larger-footprint vehicles. Manufacturers need not meet the target mpg for every single vehicle, but their overall fleet average mpg must meet or exceed the average mpg required by all the targets to which their vehicles are subject.

C.2 SCOPE OF WORK

NHTSA chose footprint primarily for safety reasons and to try to avoid “gaming” of the curves by manufacturers. Compared to incrementally reducing vehicle mass, it is comparatively difficult for a manufacturer to incrementally increase vehicle footprint in order to obtain a less stringent target. Additionally, because targets become more stringent as footprint decreases, NHTSA believes it less likely that manufacturers will respond to the MYs 2012-2016 standards through down-weighting accompanied by reductions in footprint.

As part of its regulatory analysis, the government desires to estimate the anticipated effect on crash fatalities and injuries of the weight reductions. One approach is to use the results of past statistical analyses of relationships between fatality rates and weight or size attributes such as curb weight, track width, and wheelbase.

A number of fairly comprehensive statistical analyses have been published or currently exist in draft, including reports by Van Auken and Zellner, Wenzel and Ross, and Kahane. These reports have different results: some associate a significant fatality increase with weight reductions, while others a decrease. They are also based on different databases and statistical methods. In some studies, curb weight is the only weight-size attribute (and reductions in curb weight implicitly include accompanying reductions in other attributes such as track width); in others, curb weight, track width, and wheelbase are separate independent variables. However, all these studies are based on existing or previous vehicles (where less weight typically meant smaller footprint), whereas in the future, weight reductions may be achieved while maintaining footprint, by techniques such as materials substitution or engine downsizing.

C.3 SPECIFIC TASKS

NHTSA desires an independent review of recent and updated statistical analyses of relationships between vehicles' curb weight, track width, wheelbase, and fatality rates. The contractor shall review the validity of the studies in modeling the data upon which they are based, clearly explain their methodology, exploratory data analysis and their potential utility in predicting the possible effects on fatalities and injuries of weight reductions for future vehicles.

The Contractor shall provide acquisition support services as described below.

C.3.1 TASKS

The contractor shall review the following published and draft statistical studies of relationships between vehicle weight-size and fatality or injury rates. National Highway Traffic Safety Administration (NHTSA) will provide a paper or electronic copy of each study:

Van Auken, R.M., and J.W. Zellner. 2002. An assessment of the effects of vehicle weight on fatality risk in model year 1985-98 passenger cars and 1985-97 light trucks. DRI-TR02-02. Dynamic Research Inc., Torrance, California.

Van Auken, R.M., J.W. Zellner, J.P. Boughton, and J.M. Brubacher. 2003. A further assessment of the effects of vehicle weight and size parameters on fatality risk in model year 1985-98 passenger cars and 1985-97 light trucks. DRI-TR03-01. Dynamic Research Inc., Torrance, California.

Van Auken and Zellner, 2004. A Review of the Results in the 1997 Kahane, 2002 DRI, 2003 DRI, and 2003 Kahane Reports on the Effects of Passenger Car and Light Truck Weight and Size on Fatality Risk. DRI-TR-04-02. Dynamic Research Inc., Torrance, California.

Van Auken, R.M., and J.W. Zelkner, 2004/2005. An Assessment of the Effects of Vehicle Weight and Size on Fatality Risk in 1985 to 1998 Model Year Passenger Cars and 1985 to 1997 Model Year Light Trucks and Vans. SAE 2005-01-1354.

Van Auken, R.M., and J.W. Zellner. 2005. Supplemental results on the independent effects of curb weight, wheelbase, and track on fatality risk in 1985-1998 model year passenger cars and 1986-97 model year LTVs. DRI-TR05-01. Dynamic Research Inc., Torrance, California.

Wenzel, T.P., and M. Ross. 2005. The effects of vehicle model and driver behavior on risk. *Accident Analysis and Prevention* 37: 479-494.

Wenzel, T.P., and M. Ross. 2006. Increasing the Fuel Economy and Safety of New Light-Duty Vehicles. Paper prepared for the William and Flora Hewlett Foundation's Workshop on Simultaneously Improving Vehicle Safety and Fuel Economy through Improvements in Vehicle Design and Materials. Supported by U.S. DOE Contract No. DE-AC03-76SF00098.

Wenzel, T. and Ross, R. 2008. The Relationship between Vehicle Weight/Size and Safety. AIP Conference Proceedings -- September 12, 2008 -- Volume 1044, pp. 251-265

Padmanaban, J. 2004. Influences of Vehicle Size and Mass on Odds of Driver Fatality. Association for the Advancement of Automotive Medicine.

Eyges, V and Padmanaban, J. 2009. Updated Evaluation of Size and Mass Effects in Front-to-Front Crashes Involving Light Vehicles. SAE Technical Paper Series, 2009-01-0375. SAE International, March. Warrendale, PA: SAE.

Nusholtz, G.S., G. Rabbiolo, and Y. Shi. 2003. Estimation of the Effects of Vehicle Size and Mass on Crash-Injury Outcome Through Parameterized Probability Manifolds. **SAE Technical Papers, Document Number:** 2003-01-0905.

Kahane, C.J. 1997. Relationships between vehicle size and fatality risk in model year 1985-93 passenger cars and light trucks. NHTSA DOT HS 809 662. U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C.

Kahane, C.J. 2003. Vehicle weight, fatality risk and crash compatibility of model year 1991-99 passenger cars and light trucks. NHTSA DOT HS 809 662. U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C. http://www-nrd.nhtsa.dot.gov/Pubs/809662.PDF

Kahane, C.J. 2004. Response to docket comments on NHTSA technical report "Vehicle weight, fatality risk and crash compatibility of model year 1991-99 passenger cars and light trucks". Submission to docket no. NHTSA-2003-16318-16. U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, D.C.

NHTSA. 2010. Additional analyses of 1991-99 passenger cars and more recent vehicles. **Draft**.

NHTSA. 1997. The Effect of Decreases in Vehicle Weight on Crash Injury Risk.

DRI comments on safety impacts of EPA-NHTSA proposed rule to establish light-duty vehicle greenhouse gas emission standards and corporate average fuel economy standards, Dated September 28, 2009, DRI-TM-09-86, Docket # NHTSA-2009-0059-0112.1, 11/30/09

Wenzel, T, comments on the joint proposed rulemaking to establish light-duty vehicle greenhouse gas emission standards and corporate average fuel economy standards, Dated October 27, 2009, Docket # NHTSA-2009-0059-0060.1, 11/18/09

Leon S. Robertson, Prevention of motor-vehicle deaths by changing vehicle factors, Injury Prevention 2007, 13:307-310. doi:1136/ip.2007.016204

Leon S. Robertson, Blood and Oil: Vehicle Characteristics in Relation to Fatality Risk and Fuel Economy, American Journal of Public Health/November 2006, Vol. 96, No. 11, Pages 1906-1909

OPTION TO EXERCISE: TO REVIEW ANY NEW REPORTS AND/OR UPDATED PUBLISHED AND DRAFT STATISTICAL STUDIES OF RELATIONSHIPS BETWEEN VEHICLES WEIGHT-SIZE AND FATALITY OR INJURY RATES.

C.3.2 PROCUREMENT PLANNING AND DOCUMENTATION

The contractor shall review the indicated documents and provide a detailed evaluation on the data sources, analysis methods, assumptions and the reports' ability to support its conclusions. The report evaluations should compare the data sources used for each study and any influence upon the resulting conclusions. This evaluation should also consider and contrast the statistical evaluation methods, their suitability to the individual study, and the overall effect on the accuracy of the results safety estimates. All underlying assumptions or future projections for material or cost performance should be identified for each report to assist in interpreting the differing safety forecasts. The contractor shall utilize the detailed review to evaluate the relative suitability of the various estimates and their predictive value for future vehicle safety.