UMTRI-94-8

Analytical Review and Expansion of Certain Statistical Models Relating Rollover Risk to Precrash Factors

Hans C. Joksch

October 1994

FINAL REPORT

The University of Michigan
Transportation Research Institute
Ann Arbor, Michigan 48109-2150

Disclaimer

The research reported herein was conducted under general research funds provided by the American Automobile Manufacturers Association. The opinions, findings, and conclusions expressed in this publication are not necessarily those of the AAMA.

Technical Report Documentation Page

1. Report No. UMTRI-94-8	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle ANALYTICAL REVIEW AND EXPANSION OF CERTAIN STATISTICAL MODELS RELATING ROLLOVER RISK TO PRE-CRASH FACTORS		5. Report Date October, 1994
		6. Performing Organization Code
7. Author(s) Hans C. Joksch		8. Performing Organization Report No. UMTRI-94-8
9. Performing Organization Name and Address The University of Michigan Transportation Research Institute 2901 Baxter Road, Ann Arbor, Michigan 48109-2150		10. Work Unit No. 94134A
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address American Automobile Manufacturers Association 7430 Second Avenue, Suite 300 Detroit, Michigan 48202		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The ratio of rollover to single vehicle accidents was studied, using 1986 to 1988 Michigan data on single-vehicle accidents. Logistic regression models developed by NHTSA were replicated, and more complex models were examined. However, this analysis had a limited scope, and did not examine the conceptual validity of the approach, model fit, or model sensitivity.

Separate models for urban and for rural accidents were developed. Not only is the rollover probability in rural environments much higher, but also all model coefficients differ. Thus, environment interacts with all other factors. When adding vehicle type, drive-wheel configuration, the presence of ABS and the single vehicle accident rate to the models, the coefficients of the static stability measures decreased.

More complex models were also developed, adding the wheelbase and several quadratic and interaction terms. Wheelbase had, over its full range, about the same effect as stability over its full range. In urban environments, the effect of stability appeared too be slightly greater, in rural environments that of wheelbase was greater. Thus, the relative importance of static stability measures and wheelbase differ in rural as compared to urban environments.

17. Key Words Rollover Tilt-table ratio Critical sliding velocity Logistic regression	18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	34	

Executive Summary

NHTSA developed statistical models that express the probability of rollover in a single vehicle accident as a function of several vehicle, driver, and environmental variables. The objective of this work was to replicate the NHTSA models, and to explore expanded models to assess the roles of the original variables and additional variables. The analysis uses 1986-88 Michigan data on single vehicle accidents obtained from NHTSA. Statistical models including the same independent variables NHTSA used were fitted to the data. NHTSA reports only the coefficients of three variables for two of its models. The coefficients of the replicated models agreed well with the reported values.

The statistically most important, and also most influential, variable was the urban/rural environment. The rollover probability is much higher in a rural than in an urban environment. Because interactions with other factors were suspected, separate models for urban and rural accidents were developed. All coefficients differed between the environments, sometimes very much. Thus, models not distinguishing the environment, or including the corresponding extensive interaction terms, do not adequately reflect reality.

In all models, stability measures Critical Sliding Velocity (CSV) or Tilt Table Ratio (TTR) had very significant coefficients. Extended models included as additional variables vehicle type, drive-wheel configuration, the presence of ABS, and the single vehicle accident rate per registered vehicle. The effect of these additional factors was to decrease the coefficients of the stability measures.

More complex models included, in addition, the wheelbase, quadratic terms of the stability measure and the wheelbase, and interactions between any pair of the stability measure, wheelbase, and ABS. These models showed that wheelbase had, overall, about the same mathematical effect as stability on rollover risk. In urban environments, stability appeared to have a stronger influence than wheelbase; in rural environments wheelbase appeared to have a stronger effect than stability. ABS had either no discernible effect, or a beneficial effect, approximately comparable to that of a 15-inch increase in wheelbase.

The analyses described here followed the modeling approach taken by NHTSA. We did not examine the conceptual validity of the approach or the suitability of the logistic model used. Detailed model evaluations to identify outliers, or the sensitivity of coefficients to a few observations, were also not carried out. In addition, one must make the usual reservations for analyses based on police reported accident data, in terms of possible bias in accident reporting as well as completeness and reliability of the reported information.

Nevertheless, one can conclude that the more complex models provide a much more accurate representation of the data analyzed. The more complex models provide strong evidence not only that the stability factors studied have different mathematical effects in different environments (rural/urban), but that wheelbase also is an important explanatory factor that needs

to be included in the model. The analysis described here indicates that the relative importance of stability factors and wheelbase differs in rural as compared to urban environments.

Contents

List of Figures

1 Introduction

In the context of rulemaking concerning rollover in crashes, NHTSA prepared two papers, an addendum [1] to a previous Technical Analysis Paper [3], and one estimating potential reductions in fatalities and injuries resulting from assumed changes in vehicle static stability [2]. Both papers use statistical models, relating the probability of rollover in a single vehicle crash to a number of vehicle, environment, and driver factors.

NHTSA gave only brief descriptions of the models and presented very limited quantitative information on them. It is not possible to assess and interpret the models on this basis. Therefore, models similar to NHTSA's had to be recreated, based on the information in the two papers, other data available, and data files provided by NHTSA to UMTRI.

NHTSA specifically states that it does not rely on the statistical models to establish a causal link between reliable stability and rollover, which is based on physical relationships.

In addition to recreating the models, more extensive and complex models were developed to assess whether additional factors played a role, and whether NHTSA's models adequately reflected the role of the factors included. This preliminary analysis deals only with the mathematical aspects of NHTSA's, and similar models. The analyses are preliminary, because they follow the traditional course of fitting models of assumed mathematical structure to the data. It was not tested whether the assumed mathematical form was the best possible, nor whether certain vehicle models exerted a disproportionate influence on the results. Other factors for which no variables were in the database could, of course, not be examined.

2 The Structure of the Statistical Models

The statistical models are logistic regression models, relating a dependent variable y, the probability of rollover in a single vehicle crash, to a number of independent variables x_1, x_2, ... x_n, by an expression

$$y=1/(1+\exp(L)),$$

where

$$L=a + b_1x_1 + b_2x_2 + \dots b_nx_n$$

This model implies a certain form of the relationship between y and the x_i, which may or may not reflect the real form of the relationship. If not, it may be possible to improve the fit by adding squares or higher powers of the variables, and/or products of variables, which reflect interactions.

The coefficients a and b_i are shown in the various tables. A negative sign indicates that if the variable increases, the rollover probability decreases. The standard error of the coefficient indicates how precisely the coefficient can be estimated within the context of the statistical model; it does not reflect any error with regard to a "real" value it might have. The Wald chi-square is a simple function of the coefficient and its standard error, the P a function of chi-square. A larger chi-square and smaller P indicates that the variable can be estimated with greater statistical precision and is more "powerful" in a statistical sense. It does not mean that it has a stronger influence on y.

For instance, in Table 3a the coefficient -0.7655 of UV shows that being a utility vehicle or not has almost twice the effect as having front wheel drive or not; FWD has a coefficient of -0.4978. The chi-square for FWD, 24.6, however, is larger than that for UV, 19.0.

3 The Basic Models

The coefficients of the models corresponding to those in reference [1] are shown in Tables A-1a and A-1b, referring to the "long" file. NHTSA shows (Tables 9 and 11 in [1]) the values for the three statistically "most powerful" variables. They agree well with those in Tables A-1a and A-1b. Thus, the recreation of NHTSA's models was successful. Figures 1a and 1b show the modeled relationships between rollover risk and a stability measure, holding all other variables constant at their mean values to control for their influence (light lines labeled "all").

The large and very significant coefficient that the variable RURAL has in both models, and the findings of previous work raises the suspicions that RURAL might interact with other variables. Therefore interactions of RURAL with all other factors should be tentatively included in the model. This amounts to fitting two separate models, urban and rural.

Tables 1a and 1b show the coefficients. All differed between urban and rural environments, most by relatively large amounts, and most very significantly. Some factors that were statistically significant in one environment were not so in the other. Therefore, one single model for both environments, with no interaction terms, can not quantitatively reflect any causal relationship that might exist between rollover and the independent variables.

Figures 1a and 1b illustrate the differences between urban and rural environments. The light lines labeled "urban" and "rural" show the modeled relationships between rollover risk and stability measures, keeping all other variables constant at their mean values.

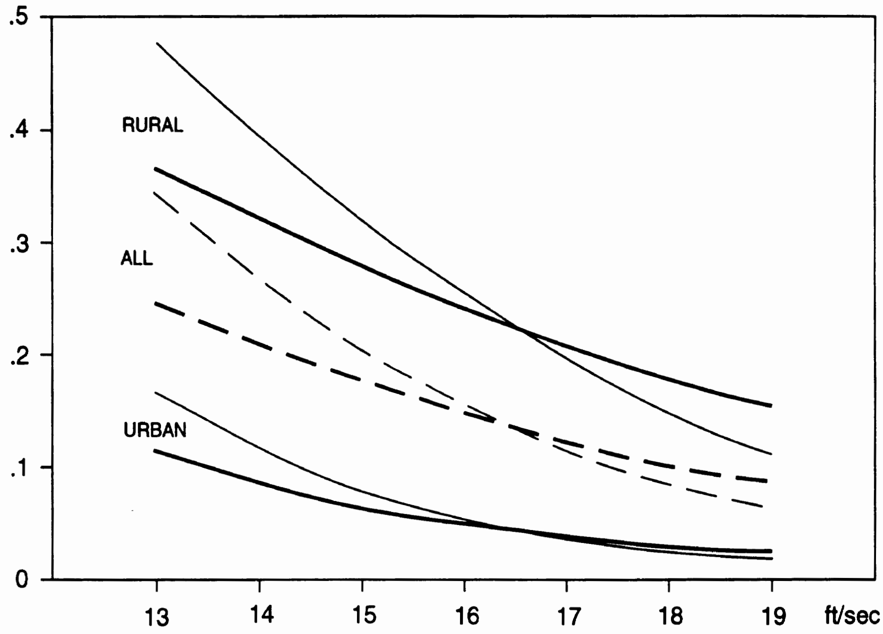

Figure 1a. Modeled rollover risk versus CSV, for mean values of all other variables. Light lines represent the basic models, heavy lines the extended models.

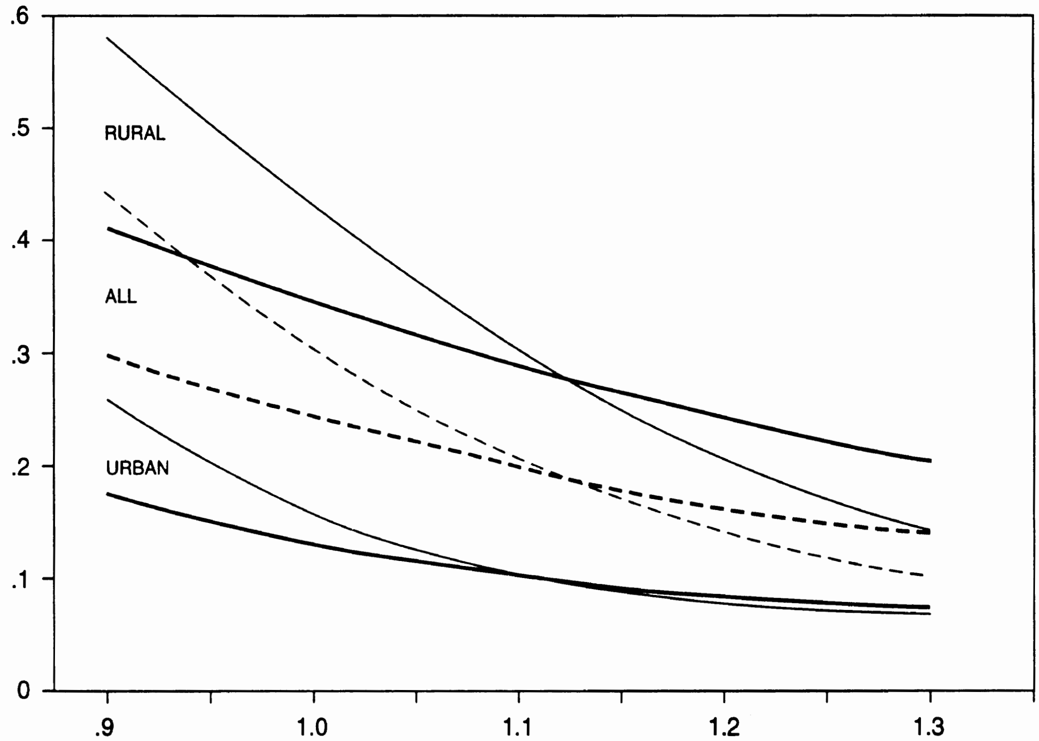

Figure 1b. Modeled rollover risk versus TTR, for mean values of all other variables. Light lines represent the basic models, heavy lines the extended models.

4 Extended Models

In reference [2], NHTSA used models including more variables, which were described only in general terms: "Additional variables were added to the statistical model to improve its predictive power for the determination of sensitivity of rollover risk to change in vehicle stability. Variables representing anti-lock braking equipment, vehicle age, single vehicle accident involvement rate (SVA/registered vehicle make/model) and a vehicle class variable which combines the vehicle type, relative size, and drive configuration were added." Since vehicle age is already mentioned in the description of the basic model, it does not need to be added.

The vehicle class variable is not sufficiently defined to reconstruct it. In the captions of figures, 20 class variables are mentioned. Therefore, only a simpler vehicle classification could be used. Variables were introduced for utility vehicles, pickup trucks, and vans, thus quantifying differences in rollover risk relative to passenger cars. Similarly, variables for all-wheel drive and front-wheel drive quantify differences relative to rear-wheel drive.

Models with these variables were developed, for all accidents and separately for urban and rural accidents. The coefficients for the latter are given in Tables 2a and 2b. The modeled relationships between rollover risk and a stability measure are shown for rural, urban, and all accidents in Figs 1a and 1b by heavy curves.

Comparing the models represented by the light and the heavy lines, it is obvious that the apparent effect of stability on rollover risk is much weaker when the additional variables are included in the model. A comparison of the coefficients in Tables 1a and 1b with those in Tables 2a and 2b shows that the statistically most significant coefficients have changed only little. Exceptions are stability, AGE, and MALE. The latter two are probably reduced because of the addition of SVA/RV, which is related to driver age and sex. The coefficient of stability is changed because some of the added vehicle characteristics are correlated with stability.

5 More Complex Models

The extended models included more variables then the simple models. Still, they do not include all available variables that may be related to rollover. For instance, the data file contains, as one of the contributing circumstances, information on skidding. Since skidding may plausibly lead to rollover, it was added to the independent variables.

Previous work had found a relationship between rollover and wheelbase. Though a direct causal relationship between wheelbase and rollover is not obvious, it may have indirect effects, e. g., by influencing the risk of losing control. Therefore, it was added to the independent variables.

Logistic regression specifies a certain mathematical relationship between the independent and the dependent variables. It is used for modeling probabilities because the dependent variable can vary only between zero and one, as necessary for probabilities. There is no reason to believe that any physical relationship between a causal factor and rollover probability can be adequately described by a simple logistic function using only linear terms in the variables. Adding quadratic terms makes it more flexible to represent a wider range of physical relationships. Therefore, quadratic terms in the stability measure, and in wheelbase, were added to the models. Also, there is no reason to assume that physical effects are additive in the exponent of the logistic function. Therefore, as a second approximation, interaction terms between the stability measure and wheelbase, between the stability measure and ABS, and between wheelbase and ABS were added.

These models have a large number of independent variables, some of which were selected because they might have an effect, not because an effect was established. Therefore, one would like to exclude terms that are not necessary to represent the data. We decided to exclude terms that might, with a high probability (over 5 percent) be due to chance. One approach is a backward selection: starting with a model using all variables, variables with coefficients not significantly different from zero are dropped stepwise, until no nonsignificant coefficients are left. Empirically, this procedure tends very often to find the "best" model representing the data using the given variables. Therefore, this procedure was used. As a check, another common procedure, the forward stepwise selection, was also used. Here, variables are added in order of their significance, and dropped when they become nonsignificant. Though frequently used, this procedure often does not lead to the best model using the available data.

For urban environments, the forward stepwise and the backward selection resulted in the same model. Thus, we can be fairly certain that these are the models "best" describing the data, using the available variables. For rural environments, there were some differences. For both stability measures, the forward stepwise procedure included driver AGE, but as the least or second-least significant term, the backward elimination procedure did not include it; we did not include AGE in our final model (but UNDER25 remained in the model).

The rural models using CSV as stability measure included CSV squared by both selection procedures, but only the stepwise selection included CSV itself as the least significant variable. Therefore, it was not included in our final model.

The coefficients of our complex models are shown in Tables 3a and 3b. The relationship between the stability measures and modeled rollover probability, by wheelbase and ABS, are shown versus the stability measures in Figures 2a and 2b, and versus wheelbase in Figures 3a and 3b. All other variables are fixed at the population means.

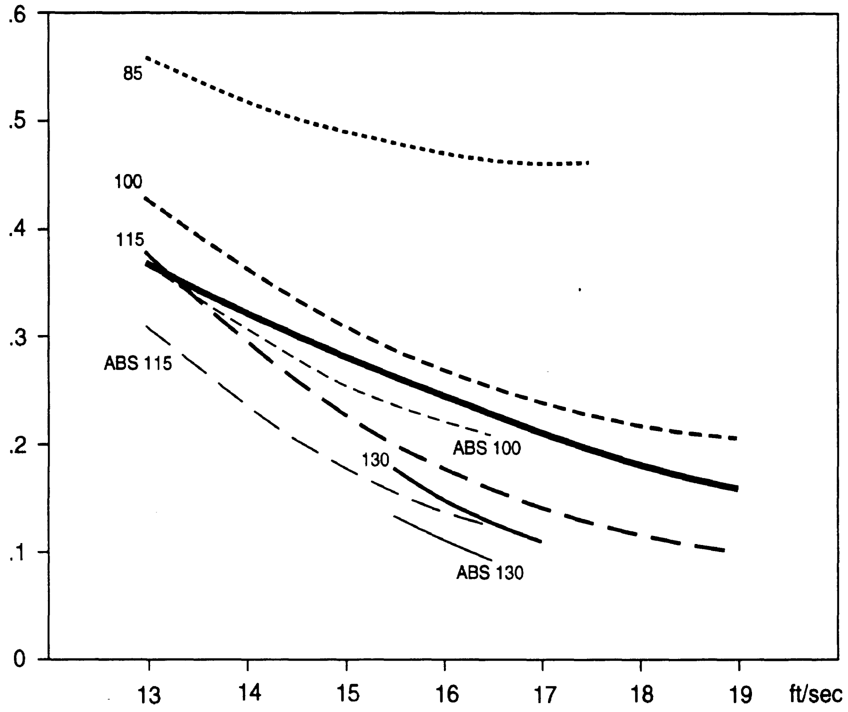

Figure 2a. Modeled rollover risk in rural environment versus CSV, for selected values of wheelbase, and presence or absence of ABS, for mean values of all other variables. Curves for ABS are lighter. The heavy curve represents the extended model.

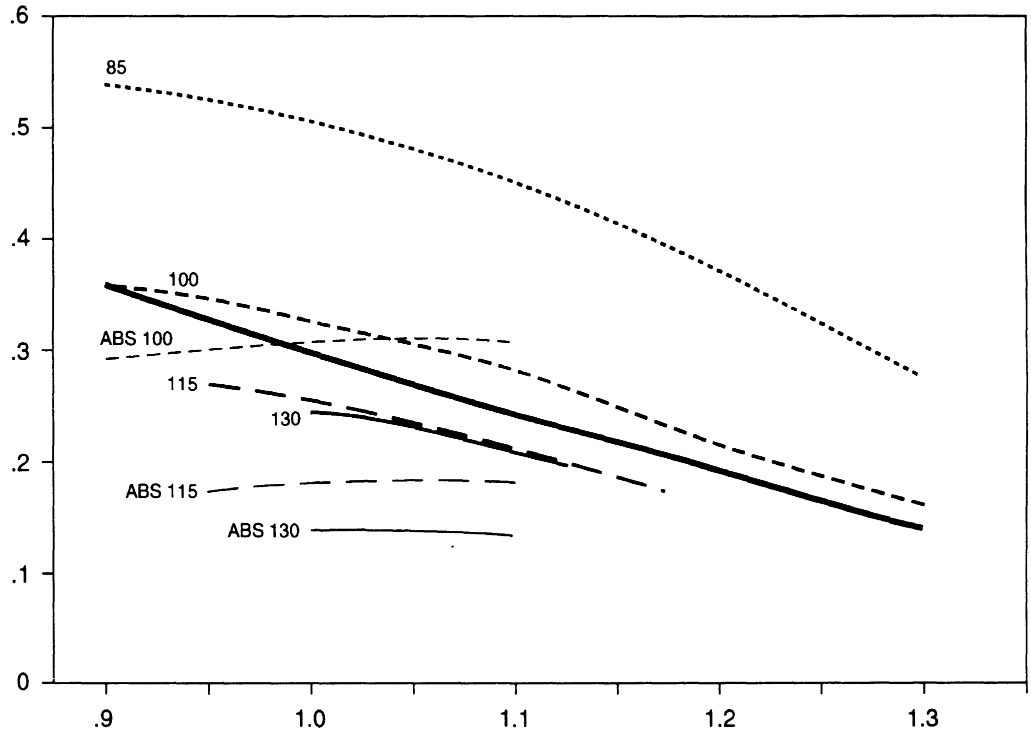

Figure 2b. Modeled rollover risk in rural environment versus TTR, for selected values of wheelbase, and presence or absence of ABS, for mean values of all other variables. Curves for ABS are lighter. The heavy curve represents the extended model.

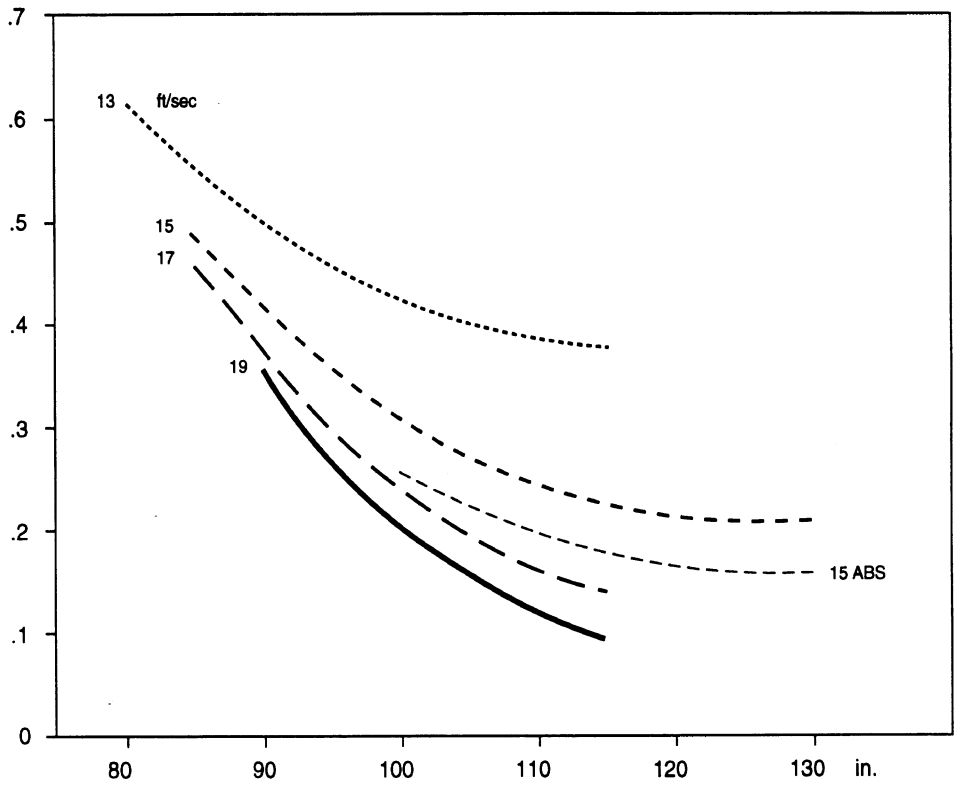

Figure 3a. Modeled rollover risk in rural environment versus wheelbase, for selected values of critical sliding velocity, and presence or absence of ABS, for mean values of all other variables. Curves for ABS are lighter.

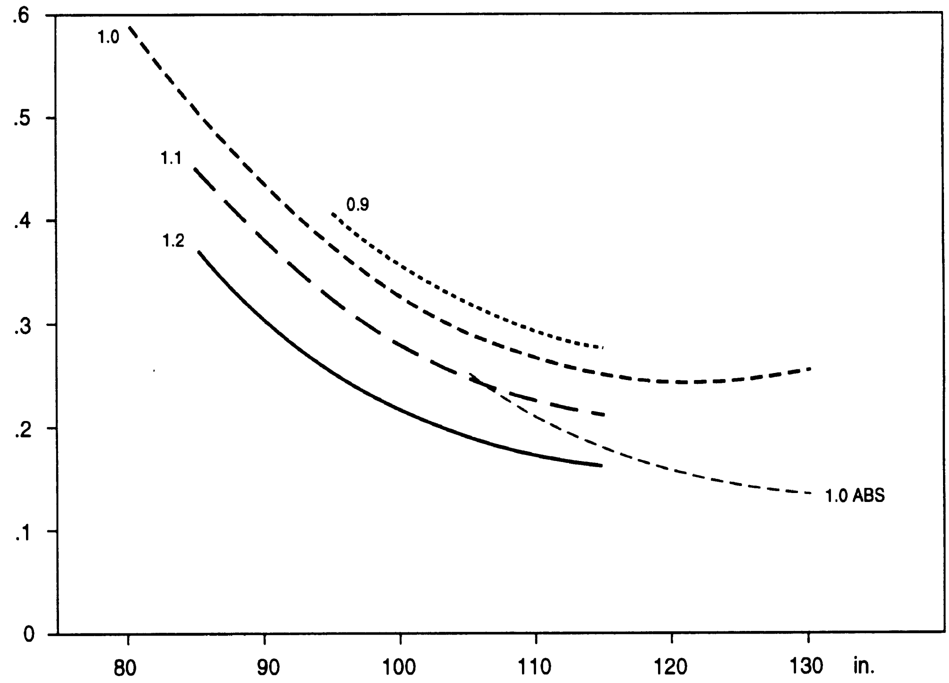

Figure 3b. Modeled rollover risk in rural environment versus wheelbases for selected values of TTR, and presence or absence of ABS, for mean values of all other variables. Curves for ABS are lighter.

The model for rural accidents, using CSV as stability measure (Figures 2a and 3a), shows a declining relationship of rollover probability with CSV. It also shows a strong dependence of rollover probability on wheelbase. Indeed, the variation of rollover risk over the range of wheelbases is larger than over the range of CSV values. The effect of wheelbase is nonlinear: the step from 85 inches to 100 inches results in a greater reduction of rollover risk than that from 100 inches to 115 inches, which in turn results in a greater reduction than a change from 115 inches to 130 inches. The curve reflecting the "extended" model (section 4) is fairly similar and close to that for a vehicle with 100-inch wheelbase. ABS reduces the rollover probabilities for each wheelbase by a nearly constant amount, the amount declining with wheelbase.

Figures 2b and 3b reflect the rural model with TTR as stability measure. Again, a strong effect of wheelbase is present, greater than that of TTR. The effect of wheelbase is even more nonlinear than with CSV as stability measure: there is no difference in rollover risk for vehicles with 115-inch and with 130-inch wheelbases. The effect of ABS is puzzling: for vehicles with ABS, there is practically no relationship between rollover risk and TTR.

Figures 4a and 5a reflect the model for urban accidents using CSV as stability measure. There, relationships between rollover risk and stability measures are highly nonlinear, much more so then in the "extended" model. An effect of wheelbase is clear, but it is much weaker than that of stability. Curiously, ABS does not seem to have any effect.

Figures 4b and 5b reflect the model for urban accidents with TTR as stability measure. Again, there are clear relationships between rollover risk and TTR, and with wheelbase. In this case, the total magnitudes of the effects are comparable. The curve representing the "extended" model fits well into this picture, corresponding to a vehicle with about 105-inch to 108-inch wheelbase. The effect of ABS is simple: it is equivalent to adding 15 inches to the wheelbase.

Though these models show qualitatively similar effects—with the exception of ABS—there are large quantitative differences. Some of these differences might be real, reflecting effects of the different environments, or that CSV and TTR measure somewhat different vehicle characteristics. However, some of the differences may be due to limitations of the database. As discussed above, the number of data points representing physically different vehicles is small. Therefore, the experience of a few or even one single vehicle could have a strong influence on the statistical models. A thorough sensitivity analysis would be needed to determine this.

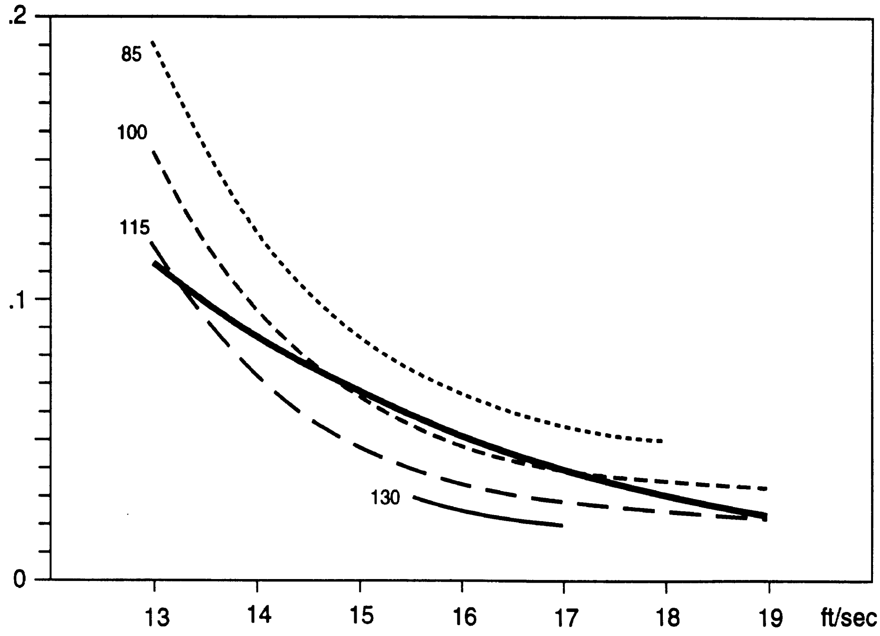

Figure 4a. Modeled rollover risk in urban environment versus CSV, for selected values of wheelbase (this model does not show an effect of ABS), for mean values of all other variables. The heavy curve represents the extended model (Figure 2a, heavy curve).

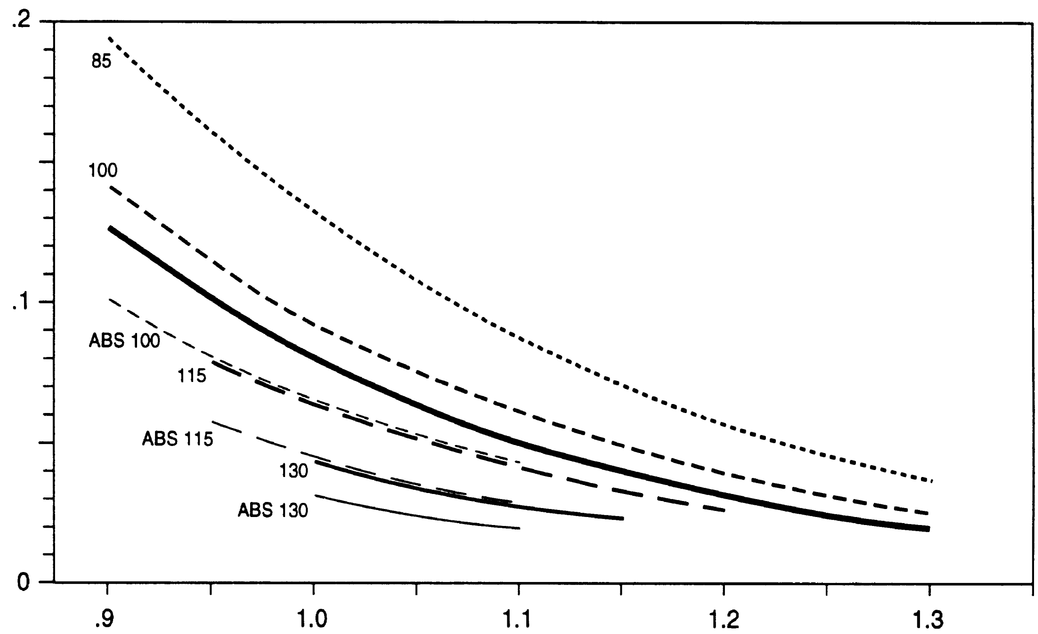

Figure 4b. Modeled rollover risk in urban environment versus TTR, for selected values of wheelbase, and presence or absence of ABS, for mean values of all other variables. Curves for ABS are lighter. The heavy curve represents the extended model (Figure 2b, heavy curve).

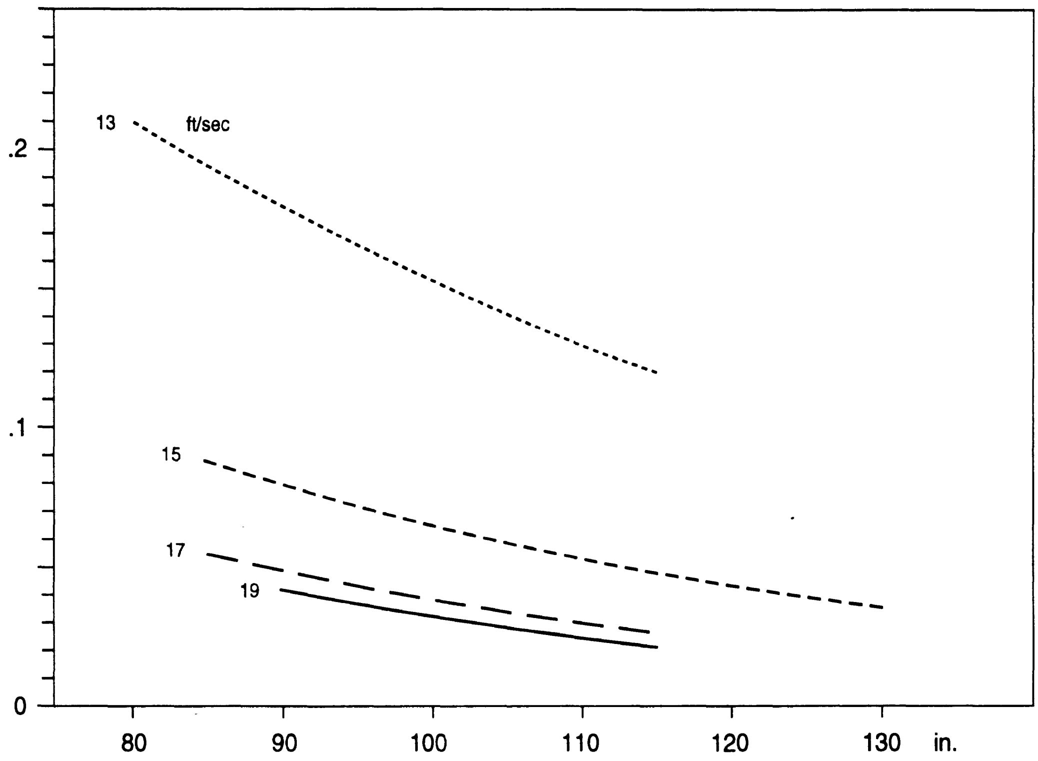

Figure 5a. Modeled rollover risk in urban environment versus wheelbase, for selected values of CSV (this model does not show an effect of ABS), for mean values of all other variables.

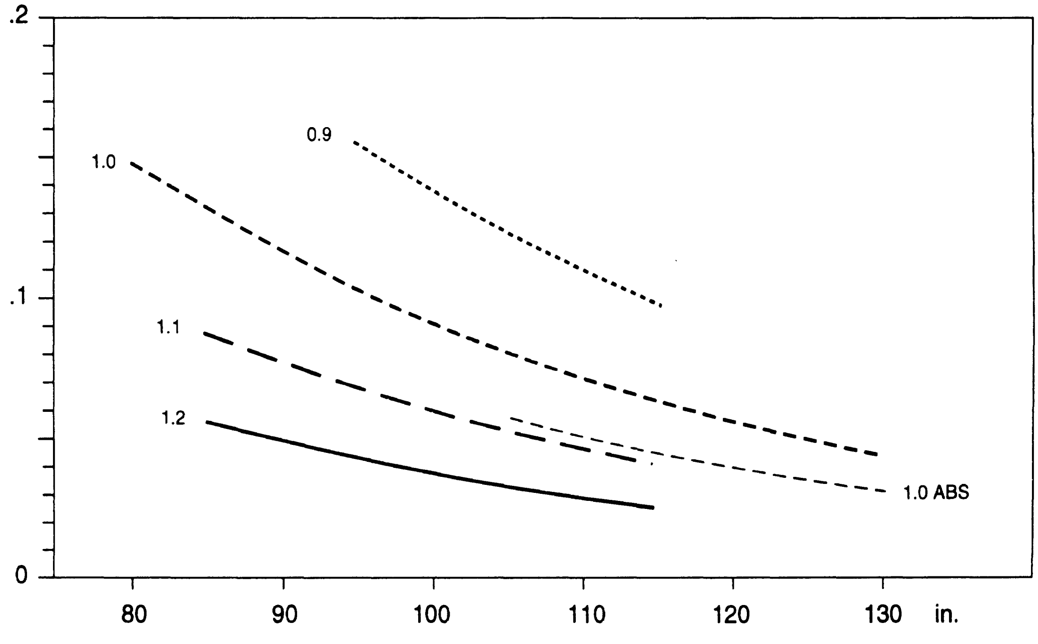

Figure 5b. Modeled rollover risk in urban environment versus wheelbase, for selected values of TTR, and presence or absence of ABS, for mean values of all other variables. Curves for ABS are lighter.

6 Some Evidence on the Effect of Speed

Table 7 in reference [2] allows one to calculate the ratio of occupants exposed in single vehicle rollover accidents to occupants exposed in single vehicle accidents, for four speed limit ranges $\leq 25^1$, 30-35, 40-50, and 55-65 mph. If vehicle occupancy is not related to speed limit, these ratios represent rollover risks. Though speed limits and actual travel speeds differ, one can expect some relationship between speed limit and travel speed (if only because of the fairly common practice to set speed limits at the 85th percentile of actual travel speeds). Figure 6 shows the ratio of occupants in rollovers to occupants in all single vehicle crashes versus the speed limit, together with a regression line through the points. The relationship is practically perfect. That the rollover risk increases with speed is plausible. That it triples from the lowest to the highest speed range is remarkable.

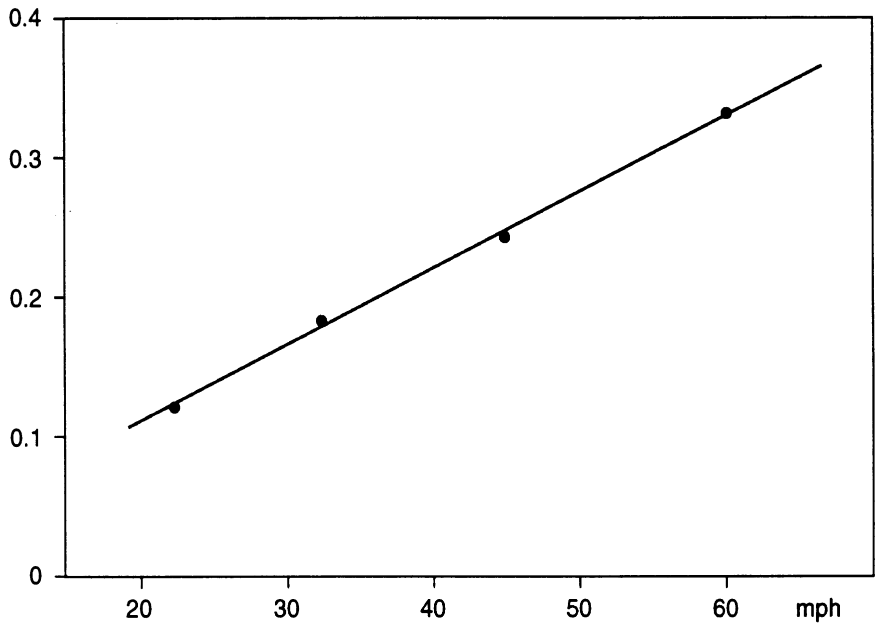

Figure 6. Ratio of occupants in single vehicle rollover crashes to occupants in single vehicle crashes, versus road speed limit. Based on data from reference [2]. The line represents a linear regression.

The strong relationship between speed range and rollover risk suggests that a comprehensive model of rollover risk should include an indicator of speed. To some extent this might have been achieved by using the urban/rural variable. To check this, we did the following.

A logistic regression was fitted to the four proportions with the midpoint of the speed limit ranges (for the lowest range, 22.5 was used, because very rarely are speed limits less than 20 mph). The proportions were used, and not the counts shown in the table, because the latter are

[1]The table indicates "<25," but from the context it appears to be <=25."

obtained by expanding from relatively few actual NASS cases, using factors reflecting NASS's complex sampling plan. The resulting function is

$$L = -2.68 + .039*v$$

This model represented the data fairly well. A difference in travel speeds of 25 mph corresponds to a change of L by 0.85, one of 30 mph by 1.02, and one of 35 mph by 1.19. In the other models, the difference between rural and urban environments is expressed by coefficients ranging from 1.64 to 1.68. Thus, between 50 and 70 percent of the urban/rural difference might be explained by plausible differences in speed. This suggest that statistical models should treat effects of speed and environment separately.

7 Discussion

This study had a very limited scope. It studied statistical models of rollover probability developed by NHTSA, compared them, and developed more refined models. However, it did not study how well these models represented the data from which they were derived. Neither did it study whether the contributions of the various factors included in the models were mathematically adequately reflected by the models. It was not studied whether the models were strongly influenced by the experience of a few vehicle models, or possibly vehicle types. Also, the relatively small number of actually different vehicle models limited the analysis. Despite the limited scope, some interesting observations were made.

NHTSA developed two types of models. The simpler one contained a number of driver and environmental factors, and only two vehicle factors (vehicle age and a stability measure). They showed a large mathematical effect of stability on the rollover probability. The extended models included, in addition, the single vehicle accident rate per registered vehicle, as an indicator of environmental and driver effects, which are not reflected in the other driver and environmental variables, and vehicle type, drive configuration, and ABS. We found that addition of these variables changed the coefficients of the variables for driver age and sex, presumably because of the correlation between the accident rate and driver age and sex. We also found that the coefficients of stability were considerable reduced, by between 30 and over 50 percent, varying between measures and models, presumably because the additional vehicle factors contributed to the rollover risk.

The statistically most important, and also most influential, variable was the urban/rural environment. The rollover probability is much higher in a rural than in an urban environment. Because interactions with other factors were suspected, separate models for urban and rural accidents were developed. All coefficients differed between the environments, sometimes very much. Thus, models not distinguishing the environment or including the corresponding, extensive interaction terms do not adequately reflect reality.

Wheelbase is a factor that could influence the rollover probability. Therefore, we included it in complex models. Also, there is no reason to believe that logistic models with only linear terms can adequately represent a physical relationship between rollover probability and the independent factors. Therefore, quadratic and interaction terms were added for stability, wheelbase, and ABS. These models showed that some nonlinear terms were indeed statistically significant, some even highly so. Wheelbase showed a strong relationship with rollover probability in all models. In rural environments, this relationship with wheelbase was even stronger than that with stability; in urban environments they were about equal or the stability relationship was stronger. ABS sometimes showed a reduction of the rollover risk; because of the small number of vehicle models with ABS, and its concentration among certain vehicle types, its effect could not be well determined.

Despite the limited scope of the analyses, one can safely conclude that realistic statistical models for rollover probability must separate urban and rural accidents, include wheelbase, should probably include ABS, and that nonlinear and interaction terms of critical variables should be allowed to enter the models.

References

1. *Technical Assessment Paper: Relationship Between Rollover and Vehicle Factors. Addendum.* National Highway Traffic Safety Administration, Office of Vehicle Safety Standards, Crash Avoidance Division, 1994.

2. *Potential Reductions in Fatalities and Injuries in Single Vehicle Rollover Crashes as a Result of a Minimum Rollover Stability Standard.* National Highway Traffic Safety Administration, Office of Vehicle Safety Standards, Crash Avoidance Division, no date.

3. *Technical Assessment Paper: Relationship Between Rollover and Vehicle Factors.* National Highway Traffic Safety Administration, July 1991.

TABLES

Table 1a. Comparing coefficients of urban and rural models, short file. Stability measure CSV.

	urban model				rural model			
Variable	Param. Est.	Std. Error	Wald Chi-Sq.	P	Param. Est.	Std. Error	Wald Chi-Sq.	P
INTERCPT	-4.8937	0.3011	264.1	0.0001	-4.3061	0.1232	1221.0	0.0001
ALC_DRUG	-0.3480	0.0652	28.4	0.0001	-0.2178	0.0280	60.5	0.0001
DRIV_ERR	-0.3695	0.0863	18.3	0.0001	-0.0038	0.0382	0.0	0.9215
AGE	0.0151	0.0032	21.8	0.0001	0.0052	0.0012	17.1	0.0001
UNDER25	-0.0230	0.0759	0.0	0.7623	-0.1254	0.0311	16.2	0.0001
MALE	-0.1651	0.0614	7.2	0.0072	0.1244	0.0232	28.6	0.0001
BAD_SURF	0.0210	0.0715	0.0	0.7691	-0.0468	0.0273	2.9	0.0858
CURVED	-0.8774	0.0677	167.7	0.0001	-0.2293	0.0274	70.2	0.0001
OFF_ROAD	0.9872	0.0696	201.3	0.0001	0.0421	0.0424	0.9	0.3205
BADWEATH	-0.0612	0.0798	0.5	0.4434	0.1156	0.0278	17.3	0.0001
VEH_AGE	-0.0025	0.0120	0.0	0.8325	-0.0286	0.0049	33.7	0.0001
CSV	0.4222	0.0172	604.4	0.0001	0.3333	0.0068	2368.2	0.0

Table 1b. Comparing coefficients of urban and rural models, short file. Stability measure TTR.

	urban model				rural model			
Variable	Param. Est.	Std. Error	Wald Chi-Sq.	P	Param. Est.	Std. Error	Wald Chi-Sq.	P
INTERCPT	-6.8190	0.3682	342.9	0.0001	-5.7438	0.1530	1409.1	0.0
ALC_DRUG	-0.3917	0.0652	36.0	0.0001	-0.2396	0.0279	73.5	0.0001
DRIV_ERR	-0.3762	0.0864	18.9	0.0001	-0.0026	0.0382	0.0	0.9465
AGE	0.0185	0.0032	32.2	0.0001	0.0074	0.0012	35.5	0.0001
UNDER25	-0.0405	0.0766	0.2	0.5972	-0.1374	0.0312	19.4	0.0001
MALE	-0.2071	0.0612	11.4	0.0007	0.0817	0.0231	12.5	0.0004
BAD_SURF	0.0014	0.0715	0.0	0.9846	-0.0672	0.0272	6.1	0.0134
CURVED	-0.8781	0.0678	167.8	0.0001	-0.2247	0.0273	67.7	0.0001
OFF_ROAD	1.0070	0.0696	209.0	0.0001	0.0332	0.0423	0.6	0.4328
BADWEATH	-0.0695	0.0799	0.7	0.3844	0.1163	0.0278	17.5	0.0001
VEH_AGE	-0.0028	0.0119	0.0	0.8151	-0.0232	0.0049	22.6	0.0001
TTR	7.8574	0.3181	610.1	0.0001	6.1288	0.1305	2203.9	0.0

Table 2a. Coefficients of extended models for urban and rural environments. Stability measure CSV.

	urban model				rural model			
Variable	Param. Est.	Std. Error	Wald Chi-Sq.	P	Param. Est.	Std. Error	Wald Chi-Sq.	P
INTERCPT	-1.6525	0.6505	6.4	0.0111	-0.7662	0.2772	7.6	0.0057
ALC_DRUG	-0.3682	0.0657	31.3	0.0001	-0.2394	0.0282	72.0	0.0001
DRIV_ERR	-0.3743	0.0867	18.6	0.0001	-0.0126	0.0384	0.1	0.7427
AGE	0.0127	0.0032	15.2	0.0001	0.0034	0.0013	7.4	0.0064
UNDER25	-0.0195	0.0765	0.0	0.7985	-0.1297	0.0316	16.8	0.0001
MALE	-0.2378	0.0627	14.3	0.0001	0.0817	0.0240	11.6	0.0006
BAD_SURF	0.0437	0.0721	0.3	0.5447	-0.0448	0.0275	2.6	0.1032
CURVED	-0.8784	0.0682	166.0	0.0001	-0.2228	0.0275	65.5	0.0001
OFF_ROAD	0.9472	0.0702	181.9	0.0001	0.0286	0.0426	0.4	0.5023
BADWEATH	-0.0584	0.0804	0.5	0.4677	0.1178	0.0280	17.6	0.0001
VEH_AGE	-0.0018	0.0124	0.0	0.8856	-0.0324	0.0050	41.2	0.0001
CSV	0.2965	0.0337	77.3	0.0001	0.1932	0.0141	187.1	0.0001
AWD	0.1079	0.1354	0.6	0.4255	-0.1325	0.0470	7.9	0.0048
FWD	-0.4978	0.1002	24.6	0.0001	-0.6111	0.0418	213.8	0.0001
UV	-0.7655	0.1753	19.0	0.0001	-0.5937	0.0674	77.5	0.0001
PU	0.0543	0.1233	0.1	0.6596	-0.3133	0.0484	41.8	0.0001
VN	-0.6446	0.1452	19.7	0.0001	-0.7925	0.0635	155.6	0.0001
SVARATE	-52.2754	7.4523	49.2	0.0001	-48.6357	3.0648	251.8	0.0001

Table 2b. Coefficients of extended models for urban and rural environments. Stability measure TTR.

	urban model				rural model			
Variable	Param. Est.	Std. Error	Wald Chi-Sq.	P	Param. Est.	Std. Error	Wald Chi-Sq.	P
INTERCPT	-2.4655	0.7012	12.3	0.0004	-0.7789	0.3021	6.6	0.0099
ALC_DRUG	-0.3812	0.0658	33.5	0.0001	-0.2457	0.0282	75.7	0.0001
DRIV_ERR	-0.3748	0.0868	18.6	0.0001	-0.0136	0.0384	0.1	0.7242
AGE	0.0141	0.0032	18.7	0.0001	0.0041	0.0013	10.7	0.0011
UNDER25	-0.0180	0.0767	0.0	0.8147	-0.1317	0.0316	17.4	0.0001
MALE	-0.2403	0.0627	14.6	0.0001	0.0804	0.0240	11.2	0.0008
BAD_SURF	0.0405	0.0723	0.3	0.5756	-0.0501	0.0275	3.3	0.0685
CURVED	-0.8789	0.0683	165.6	0.0001	-0.2212	0.0275	64.6	0.0001
OFF_ROAD	0.9564	0.0704	184.7	0.0001	0.0269	0.0426	0.3	0.5281
BADWEATH	-0.0693	0.0805	0.7	0.3892	0.1154	0.0280	16.9	0.0001
VEH_AGE	-0.0013	0.0123	0.0	0.9186	-0.0305	0.0050	36.8	0.0001
TTR	5.1197	0.5484	87.1	0.0001	2.9218	0.2336	156.4	0.0001
AWD	0.1657	0.1357	1.4	0.2220	-0.1146	0.0481	5.6	0.0171
FWD	-0.3842	0.0999	14.7	0.0001	-0.5535	0.0418	175.3	0.0001
UV	-0.7099	0.1760	16.2	0.0001	-0.6144	0.0675	82.8	0.0001
PU	0.1142	0.1255	0.8	0.3630	-0.3291	0.0486	45.8	0.0001
VN	-0.3920	0.1562	6.2	0.0121	-0.6975	0.0684	103.9	0.0001
SVARATE	-67.1234	6.6735	101.1	0.0001	-59.4815	2.6687	496.7	0.0001

Table 3a. Coefficients of complex models for urban and rural environments. Stability measure CSV. A dash indicates that the variable was not selected.

	urban model				rural model			
Variable	Param. Est.	Std. Error	Wald Chi-Sq.	P	Param. Est.	Std. Error	Wald Chi-Sq.	P
INTERCPT	-14.1900	2.7582	26.4	0.0001	-5.4168	1.0757	25.3	0.0001
ALC_DRUG	-0.4057	0.0654	38.5	0.0001	-0.2429	0.0276	77.5	0.0001
DRIV_ERR	-0.4158	0.0868	22.9	0.0001	-			
UNDER25	-				-0.1816	0.0225	64.9	0.0001
AGE	0.0130	0.0024	29.2	0.0001	-			
MALE	-0.2460	0.0628	15.3	0.0001	0.0857	0.0239	12.8	0.0003
CURVED	-0.8509	0.0682	155.7	0.0001	-0.2235	0.0275	66.2	0.0001
BADWEATH	-				0.0977	0.0240	16.5	0.0001
VEH_AGE	-				-0.0142	0.0053	7.1	0.0076
OFF_ROAD	0.9595	0.0701	187.5	0.0001	-			
CSV	1.7109	0.3372	25.7	0.0001	-			
CSVSQ	-0.0492	0.0102	23.4	0.0001	-0.0183	0.0022	66.1	0.0001
WB	-				0.0691	0.0195	12.6	0.0004
WBSQ	-				-0.0007	0.0001	78.3	0.0001
WBCSV	0.0015	0.0004	13.9	0.0002	0.0076	0.0007	129.2	0.0001
WBABS	-				0.0026	0.0005	31.4	0.0001
AWD	0.3195	0.1452	4.8	0.0278	-			
FWD	-0.4945	0.1025	23.2	0.0001	-0.5259	0.0426	152.5	0.0001
UV	-0.9826	0.1785	30.2	0.0001	-0.9616	0.0604	253.5	0.0001
PU	-0.5614	0.1928	8.4	0.0036	-1.1065	0.0733	228.0	0.0001
VN	-0.9921	0.1649	36.2	0.0001	-1.0549	0.0718	216.0	0.0001
SVARATE	-24.8160	8.6358	8.2	0.0041	-			
SKIDDING	-0.3686	0.0808	20.7	0.0001	-0.0876	0.0293	8.9	0.0028

Table 3b. Coefficients of complex models for urban and rural environments. Stability measure TTR. A dash indicates that the variable was not selected

	urban model				rural model			
Variable	Param. Est.	Std. Error	Wald Chi-Sq.	P	Param. Est.	Std. Error	Wald Chi-Sq.	P
INTERCPT	-4.9345	0.8659	32.4	0.0001	-6.9101	1.8233	14.3	0.0002
ALC_DRUG	-0.4106	0.0652	39.6	0.0001	-0.2493	0.0276	81.5	0.0001
DRIV_ERR	-0.4162	0.0867	23.0	0.0001	-			
AGE	0.0132	0.0024	30.4	0.0001	-			
UNDER25	-				-0.1851	0.0225	67.3	0.0001
MALE	-0.2455	0.0628	15.3	0.0001	0.0870	0.0240	13.2	0.0003
CURVED	-0.8517	0.0681	156.3	0.0001	-0.2233	0.0275	65.9	0.0001
BADWEATH	-				0.0959	0.0240	15.9	0.0001
OFF_ROAD	0.9675	0.0701	190.7	0.0001	-			
TTR	4.6982	0.5260	79.7	0.0001	-8.2173	2.8267	8.4	0.0036
TTRSQ	-				5.0124	1.3056	14.7	0.0001
WB	0.0268	0.0060	19.8	0.0001	0.2052	0.0175	137.9	0.0001
WBSQ	-				-0.0008	0.0001	113.9	0.0001
ABS	0.3806	0.1478	6.6	0.0100	-			
TTRABS	-				-2.1966	0.7634	8.2	0.0040
WBABS	-				0.0229	0.0069	10.9	0.0009
AWD	-				-0.2472	0.0526	22.0	0.0001
FWD	-0.3140	0.1019	9.4	0.0021	-0.4983	0.0439	128.7	0.0001
UV	-0.7009	0.1423	24.2	0.0001	-0.5499	0.0714	59.3	0.0001
PU	-0.5272	0.1896	7.7	0.0054	-0.7884	0.0804	96.2	0.0001
VN	-0.7432	0.1712	18.8	0.0001	-0.9098	0.0767	140.6	0.0001
SVARATE	-43.8880	8.1149	29.2	0.0001	-45.5611	3.5519	164.5	0.0001
SKIDDING	-0.3803	0.0807	22.1	0.0001	-0.0913	0.0293	9.7	0.0018

Appendix: Analysis of the Data Base

A-1 The Data Files

In reference [1], NHTSA used Michigan single vehicle accident data from the years 1986-90. Our analysis is based on a data set CARDA.MI86TO90 obtained from NHTSA in 1993. This file contains 118,544 single vehicle accidents from the year 1986-90.

Vehicle parameters CSV, TTR, wheelbase, and ABS presence were obtained for 181 make/series from various NHTSA docket submissions and data files. These 181 make/series, however, represent only 85 different vehicle parameter combinations (109 if vehicles with and without ABS are distinguished); the remaining 96 represent corporate twins.

From the file MI86TO90, the vehicles for which the parameters were known were selected, and cases where one or several of the variables used in the models were missing were excluded. The remaining "long" file contained 89,579 cases.

The statistical models used in reference [2] contained also the variable "single vehicle accidents per registered vehicle" (SVA/REG). This variable is not in the data file. Thus, it had to be created from other sources. A NHTSA program MIPGM.SAS contains registration counts for 1986, 1987, and 1988 for a number of vehicle make/series. These registration data were combined with 1986-88 single vehicle accident data (file OVERTURN.MI86TO88 also obtained from NHTSA) to calculate SVA/REG values.

Vehicle parameters, and SVA/REG values were available for 93 makes/series (of which, however, only 55 represented different parameter combinations, or 61 if ABS is considered). Selecting these make/series, and excluding cases with missing relevant variables resulted in the "short" file with 75,333 cases. While going from the long to the short file does not reduce the number of accidents very much, it reduces the number of different vehicle configurations with ABS dramatically: from 32 to 10. Thus, estimates of ABS effects based on the short file are very uncertain.

Figures A-1a and A-1b show the combinations of certain vehicle parameters for the make/series in the short file. There is no apparent correlation between stability measures and wheelbase. Therefore, statistical models should be able to separate effects of stability and of wheelbase. Points representing vehicles with ABS, however, cluster around a line. That means that the effects of ABS may not be reliably separable from those of wheelbase and stability.

That the data base contains a limited number of vehicle parameter combinations limits the power of the statistical model to separate the effects of various vehicle factors. Table A1 shows that this holds also for vehicle classes. Most utility vehicles have all-wheel drive, in addition only relatively few pickup trucks have all-wheel drive. Thus, any specific effects of utility vehicles be they due to physical characteristics, or to use patterns are difficult to separate from those of all-wheel drive which again might be use related, in addition to its effect on vehicle handling. Most passenger cars in terms of parameter combinations in the data base had front-wheel drive, and only one van, thus making it difficult to separate any effects of front-wheel drive from passenger car effects. With one exception, ABS was available only for some utility vehicles, and some pickup trucks, thus limiting estimates of its effectiveness to these vehicle classes.

Table A-1. Combinations of vehicle type, drive configuration, and ABS, in the short data file. The number of different vehicle parameter combinations CSV, TTR, and wheelbase is shown. The first figure is for vehicles without, the second for vehicles with ABS.

	AWD	FWD	RWD
Utility vehicle	13,3	0	1,1
Pickup truck	2,2	0	7,3
Van	0	1	5,1
Passenger	0	14	8

Furthermore, utility vehicles and pickup trucks tend to have lower stability parameters than passenger cars. This will interact with the other vehicle factors discussed.

To determine to what extent the effects of the three vehicle class parameters, two driver parameters, ABS, the single vehicle accident rate, and a stability measure can be separated on the basis of essentially only 61 data points would require a very thorough analysis. With 20 vehicle class variables as used by NHTSA in reference [2], this would be even more important.

A-2 Variables

NHTSA's models used the following independent variables: alcohol or drug use, evidence of driver error, driver age and sex, rural or urban road, slippery road conditions, accidents in curves, accidents off the road shoulder, driver under age 25, bad weather, and vehicle age. The models in reference [2] use, in addition, the single vehicle accident rate per registered vehicle, presence of ABS, and variables distinguishing 20 vehicle classes. The latter combined vehicle type, relative size, and driver configuration. The definition of these classes were not given. Therefore, only a simpler classification could be used. Categorical variables were introduced for utility vehicles, pickup trucks, and vans, thus quantifying differences with passenger cars. Categorical variables are also introduced for all-wheel drive, and front-wheel drive, thus quantifying differences versus rear-wheel drive. Though not directly comparable, these variables should to some extent capture the effects of the variables reflecting NHTSA's twenty vehicle classes.

The data files contain some information which is potentially relevant for rollover, but was not included in NHTSA's models. One is wheelbase. While wheelbase has no obvious relationship with rollover, it is possibly related to loss of control, which in turn can lead to rollover. Also, some previous work indicated relationships between wheelbase and rollover. Therefore, we included it in some models.

Also of potential relevance is skidding which implies at least some loss of control. "Skidding" is one of the categories of a variable "contributing circumstance." There is also a category "loss of control." It is, however, not relevant in this context because it is specified as being "due to shifting load, wind or vacuum;" also, it is extremely rare.

The variable "contributing circumstance" is also the source of NHTSA's variable "evidence of driver error," and one of the sources of the variable "alcohol or drug use." If codes for "reckless or careless driving...," or "ill, fatigued, inattention;" or "failed to comply with license restriction" were given, NHTSA considered this as indication of driver error. Considering that another variable hazardous action provides codes for actions which most users would consider driver errors, the choice of the name is unfortunate.

"Contributing circumstance," however, is a "soft" variable. It is not coded by the officer investigating the accident at the scene. but by coders in a central office, who interpret the officer's sketch and description of the accident. If a certain circumstance is coded, it provides some evidence for its presence, but if it is not coded one can not necessarily take this as evidence for its absence. Another complication is that all 10 listed circumstances compete for one code. Thus if "reckless or careless driving" is coded, "skidding" or "under the influence of alcohol or drugs" can not be coded, etc. Thus, all findings regarding driver error and skidding should be interpreted with great suspicion. This does not hold for alcohol or drug use, because there is also another variable providing this information.

The variable from which rollover is derived also indicates collisions with a parked vehicle among the non-rollover accidents. Of the single vehicle accidents in the original file, 18% involve collisions with parked vehicles. The events leading to collisions with vehicles parked on the road may have more similarities with those leading to some collisions between vehicles than with accidents involving running off the road, or striking a roadside object. Also, parked vehicles might be more common in certain environments than in others. Then, inclusion of collisions with parked vehicles could influence the results.

Therefore, some analyses were performed excluding collisions with parked vehicles. Compared with models based on all single vehicle accidents, the coefficients of ALC_DRUG, DRIV_ERR, AGE, BAD_SURF, CURVED, and OFF-ROAD were very different. Those of TTR, and of CSV, however, differed only by less than one standard error. Therefore, we decided not to exclude collisions with parked vehicles from the data base, to keep our models comparable with NHTSA's.

A-3 Comparing Models Based on the Two Data Files

NHTSA's models in reference [1] are based on the "long" data file, and we could develop similar models using the same file.

Models from reference [2] must be based on the short file because they require SVA/REG. Comparisons between these models could be questionable, because one would not know whether differences are due to differences in the variables included in the models, or differences between the data files. Could one develop models similar to those in reference [1] and [2] using the same "short" file and still maintain compatibility?

To address this question, the model of reference [1] was developed once from the short, once from the long file. Table A-2a shows the coefficients of the model if CSV is used as stability measure. Comparing the parameter estimates with the three shown in Table 11 of reference [1] shows good agreement. Table A-2b shows the corresponding information for the model using TTR as stability measure. Again, the three parameter estimates agree well with NHTSA's.

Comparing the coefficients of the models based on the short, and on the long file shows mostly good agreement, the differences being less than the standard error. In a few cases, they are slightly larger than one standard error, and in only one case (vehicle age, when CSV is used on stability measure) does the difference approach two standard errors. The differences in the critical variables CSV and TTR were small. Thus, we believe that we can safely perform all analyses with the short file to make the results compatible.

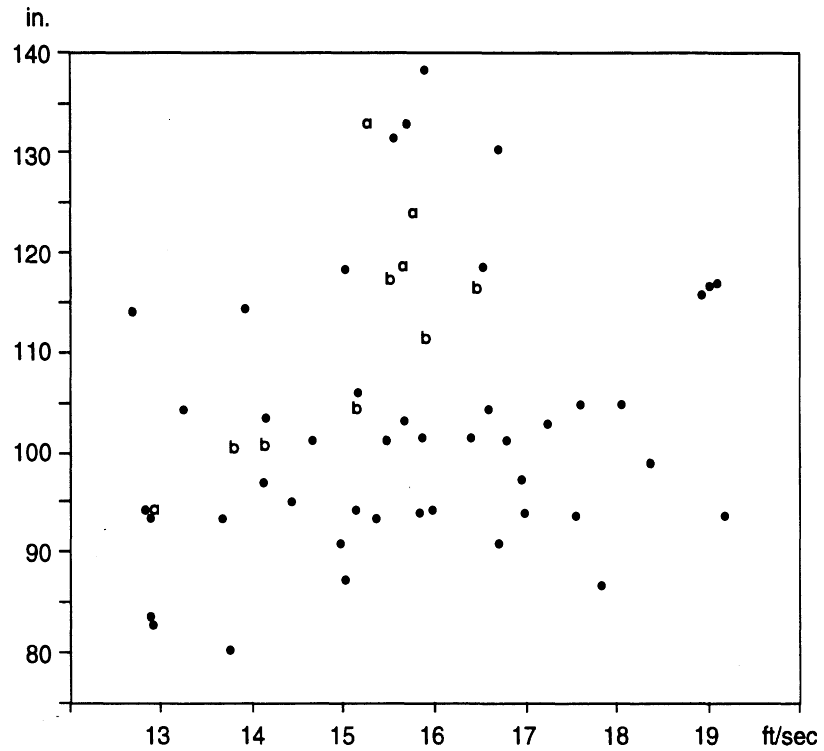

Figure A-1a. CSV wheelbase combinations appearing in the "short" data file. "a" indicates vehicles with ABS, "b" indicates vehicles with and without ABS.

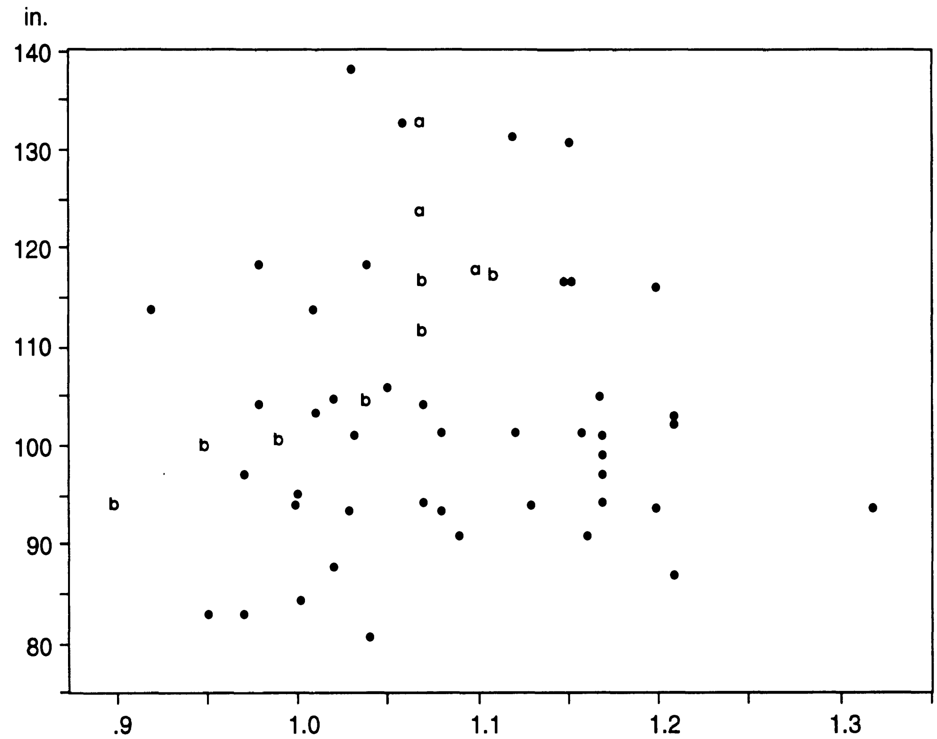

Figure A-1b. TTR and wheelbase combinations appearing in the "short" data file. "a" indicates vehicles with ABS, "b" indicates vehicles with and without ABS.

Table A-2a. Comparing coefficients of models based on long and short file. Stability measure CSV.

	long file				short file			
Variable	Param. Est.	Std. Error	Wald Chi-Sq.	P	Param. Est.	Std. Error	Wald Chi-Sq.	P
INTERCPT	-2.9564	0.1084	743.2	0.0001	-3.0762	0.1166	696.3	0.0001
ALC_DRUG	-0.2668	0.0238	125.7	0.0001	-0.2418	0.0258	87.9	0.0001
DRIV_ERR	-0.0807	0.0325	6.1	0.0132	-0.0615	0.0352	3.0	0.0806
AGE	0.0067	0.0011	39.7	0.0001	0.0071	0.0012	37.3	0.0001
UNDER25	-0.1100	0.0266	17.1	0.0001	-0.1068	0.0288	13.7	0.0002
MALE	0.0882	0.0200	19.4	0.0001	0.0873	0.0217	16.2	0.0001
RURAL	-1.6517	0.0265	3897.2	0.0	-1.6441	0.0288	3255.7	0.0
BAD_SURF	-0.0557	0.0235	5.6	0.0178	-0.0611	0.0254	5.8	0.0160
CURVED	-0.3226	0.0236	186.1	0.0001	-0.3203	0.0256	156.5	0.0001
OFF_ROAD	0.2636	0.0337	61.0	0.0001	0.2723	0.0366	55.3	0.0001
BADWEATH	0.0750	0.0243	9.5	0.0020	0.0937	0.0263	12.7	0.0004
VEH_AGE	-0.0167	0.0042	15.7	0.0001	-0.0249	0.0046	29.9	0.0001
CSV	0.3383	0.0059	3293.6	0.0	0.3456	0.0064	2942.8	0.0

Table A-2b. Comparing coefficients of models based on long and short file. Stability measure TTR.

	long file				short file			
Variable	Param. Est.	Std. Error	Wald Chi-Sq.	P	Param. Est.	Std. Error	Wald Chi-Sq.	P
INTERCPT	-4.4321	0.1345	1085.3	0.0001	-4.5687	0.1436	1012.4	0.0001
ALC_DRUG	-0.2918	0.0238	150.8	0.0001	-0.2664	0.0258	106.9	0.0001
DRIV_ERR	-0.0811	0.0325	6.2	0.0126	-0.0608	0.0351	2.9	0.0835
AGE	0.0092	0.0011	75.4	0.0001	0.0094	0.0012	66.7	0.0001
UNDER25	-0.1273	0.0267	22.7	0.0001	-0.1199	0.0289	17.2	0.0001
MALE	0.0542	0.0199	7.4	0.0065	0.0447	0.0215	4.3	0.0378
RURAL	-1.6800	0.0264	4039.8	0.0	-1.6617	0.0288	3332.2	0.0
BAD_SURF	-0.0796	0.0234	11.5	0.0007	-0.0819	0.0253	10.4	0.0012
CURVED	-0.3188	0.0236	182.7	0.0001	-0.3160	0.0255	153.0	0.0001
OFF_ROAD	0.2592	0.0337	59.2	0.0001	0.2671	0.0365	53.4	0.0001
BADWEATH	0.0774	0.0243	10.1	0.0014	0.0939	0.0262	12.7	0.0003
VEH_AGE	-0.0173	0.0042	17.1	0.0001	-0.0201	0.0045	19.8	0.0001
TTR	6.2993	0.1139	3060.2	0.0	6.3704	0.1211	2765.4	0.0