

STATISTICAL ESTIMATION  
OF  
ROLLOVER RISK

Peter Mengert  
Santo Salvatore  
Robert DiSario  
Robert Walter

Review Copy  
March 10, 1988

Prepared for  
National Highway Traffic Safety Administration  
Office of Crash  
Avoidance Research

## TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
LIST OF TABLES	iii
LIST OF FIGURES	iv
EXECUTIVE SUMMARY	v
1.0 BACKGROUND	1
2.0 TECHNICAL APPROACH	3
2.1 CARDfile ACCIDENT DATABASE	3
2.2 DATA STRUCTURING	8
2.3 COMPUTER TECHNIQUE	10
3.0 METHODS	12
4.0 RESULTS	17
4.1 LOGISTIC REGRESSION	17
4.2 GRAPHICAL RESULTS	24
4.3 RELATIVE STRENGTH OF MODEL COEFFICIENTS WHEN APPLIED TO DATA	34
5.0 CONCLUSIONS	42
REFERENCES	46
APPENDIX A SOME INITIAL VARIABLE SELECTION	A-1
APPENDIX B EXAMINATION OF THE LOGISTIC MODEL	B-1
APPENDIX C FURTHER EXPLORATIONS OF THE URBAN - RURAL VARIABLE	C-1

## LIST OF TABLES

- TABLE E-1. VEHICLE DATA
- TABLE E-2. VARIABLE DEFINITIONS
- TABLE E-3. LOGISTIC REGRESSION MODELS FOR ROLLOVER PROBABILITY  
CONDITIONAL ON SINGLE VEHICLE ACCIDENT
- TABLE 1. SUMMARY OF CARDfile STATE CRASH EXPERIENCE (1983-1985)
- TABLE 2. CARDfile DATA ELEMENTS
- TABLE 3. VEHICLE DATA
- TABLE 4. VARIABLE DEFINITIONS
- TABLE 5. LOGISTIC REGRESSION MODELS FOR ROLLOVER PROBABILITY  
CONDITIONAL ON SINGLE VEHICLE ACCIDENT
- TABLE 6. MODEL DESCRIPTION
- TABLE 7. MAKE/MODELS CONSIDERED IN ROLLOVER STUDY
- TABLE 8. NOMINAL VALUES FOR VARIABLES AND SPECIFICATIONS FOR CASES  
ONE, TWO, AND THREE
- TABLE 9. DISTRIBUTION OF p VALUES: CASE 1
- TABLE 10. DISTRIBUTION OF p VALUES: CASE 2
- TABLE 11. DISTRIBUTION OF p VALUES: CASE 3
- TABLE 12. MEAN PROBABILITIES BY MAKE/MODEL FOR CASES ONE, TWO AND  
THREE
- TABLE A1 SIXTEEN VARIABLE MODEL
- TABLE C1 LOGISTIC REGRESSION MODELS FOR URBAN / RURAL ANALYSIS

## LIST OF FIGURES

- FIGURE 1. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR ELEVEN-FACTOR BASIC MODEL (11F)
- FIGURE 2. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR SEVEN-FACTOR BASIC MODEL (7F)
- FIGURE 3. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR MODEL 11F-SF-WB
- FIGURE 4. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR MODEL 11F-SF
- FIGURE 5. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR MODEL 11F-WB
- FIGURE 6. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR MODEL SF ONLY
- FIGURE B1. SCATTERGRAM OF 7 P\_LINEAR VERSUS 7 P\_CUBIC
- FIGURE C1. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR MODEL 6F
- FIGURE C2. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR MODEL 6F-RURAL
- FIGURE C3. ACTUAL VERSUS PREDICTED ROLLOVER RATES FOR MODEL 6F-SF

## EXECUTIVE SUMMARY

### PROBLEM

Studies have indicated that vehicles with a high center of gravity are more likely to roll over when involved in single vehicle accidents (reference 1 and 2). The popularity of the utility vehicle has prompted both Congressional and NHTSA action to further examine this issue. Utility vehicles, with their high centers of gravity and increased road clearance, are designed for both on- and off-road use. Most of the studies of vehicle rollover use accident data to compare utility vehicles with passenger vehicles. These studies attempt to construct mathematical models using the accident data that predict a vehicle's rollover potential during a single vehicle accident\* based upon vehicle properties and accident variables. These models are usually developed with linear regression techniques. The most recent studies by Robertson and Kelly and by Harwin and Brewer, using linear regression techniques, developed models that indicated that vehicle factors are the most important indicators of rollover potential in a single vehicle accident. Specifically, the stability factor, (SF) which is defined as the ratio of one-half the tread width to the center of gravity height, was most highly correlated with rollover rates in single vehicle accidents. The addition of other accident factors relating to the accident environment or the driver did not significantly improve the ability of the model to predict rollovers. During NHTSA review of these results, it was suggested that improvements could be made in these analyses by the use of logistic regression techniques. This paper reports on the results of that analysis.

---

\*This is usually an estimate of the fraction of single vehicle accidents which result in rollover. All the analyses in this report are based on single vehicle accidents and rollovers as a subset of these accidents.

## **APPROACH**

The Transportation Systems Center (TSC) performed logistic regression analyses using the accident data previously analyzed by Harwin and Brewer. The accident data was derived from the CARDfile database which contains the police accident reports from six states - Texas (TX), Maryland (MD), Washington (WA), Indiana (IN), Michigan (MI), and Pennsylvania (PA). From over 2 million accidents in TX, MD, and WA, (1983 to 1985) 39,956 single vehicle accidents (SVAs) were analyzed. This analysis used 40 different make/models of utility vehicles, and domestic and imported passenger cars with known stability and other vehicle factors (see Table E-1; note that Tables E-1, E-2, and E-3 are identical to Tables 3, 4, and 5 respectively in the body of this report). In the data-set, the stability factor varied from 1.01 to 1.57. The data contained 4910 rollovers (RO). The ratio of ROs to SVAs varied by make/model from 0.021 to 0.489. Using both the Statistical Analysis System (SAS) and the Biomedical Data Package (BMDP) at the National Institutes of Health computer facility, we performed logistic regression analyses with single vehicle accidents which involved the selected vehicles and their related accident variables. Mathematical models were developed that contained both vehicle factors (stability factor and wheelbase) and accident variables relating to the driver, the vehicle, and the environment. The complete list of CARDfile variables and those used in the analysis can be found in Table E-2. During a preliminary analysis, we identified those variables that were most highly correlated with rollover. These variables were then used in various combinations to predict the actual rollover experience during a single vehicle accident.

## **RESULTS**

The main results of the logistic regression analyses are given in Table E-3. These results show that at the accident level the variables which are very useful in

Table E-1 Vehicle Data

Vehicle No.	Make/Model	Year	Wheel B. (inches)	Cg. Ht. (inches)	Tread W. /2 in.	S.F.	Data Contents Roll Ov. Acc.	Sgl.Veh. Acc.	Ratio RO/SVA
1	JEEP CJ-5	: 1972-75	83.5	26.5	26.8	1.01	88	180	.489
2	JEEP CJ-7	: 1963-85	93.5	26.3	27.9	1.06	89	248	.359
3	JEEP CHEROKEE	: 1975-83	108.7	26.6	30.7	1.15	17	64	.266
4	FORD BRONCO	: 1973-83	104.7	27.9	30.0	1.08	216	710	.304
5	CHEVY BLAZER S-10	: 1983	100.5	27.2	29.6	1.09	77	246	.313
6	CHEVY BLAZER	: 1982	106.5	28.1	32.6	1.16	30	89	.337
7	TOYOTA LANDCRUISER	: ALL	98.0	27.7	29.1	1.05	75	197	.381
8	INTER. H. SCOUT	: <1979	100.0	27.7	29.3	1.06	86	257	.335
9	CAD. DEVILLE/BROUGHAM	: 1981 - 84	121.4	21.7	30.6	1.41	5	161	.031
10	CHEVY CITATION	: 1980 - 84	104.9	21.0	28.9	1.38	132	1197	.110
11	OLDS. OMEGA	: 1980 - 81	104.9	21.0	28.9	1.38	21	182	.115
12	BUICK SKYLARK	: 1980 - 84	104.9	21.0	28.9	1.38	31	328	.095
13	PONTIAC PHOENIX	: 1980 - 84	104.9	21.0	28.9	1.38	43	267	.161
14	CHEVY CHEVETTE	: 1979	97.3	19.8	27.0	1.36	46	273	.168
15	CHEVY CORVETTE	: 1973	98.0	18.2	28.6	1.57	3	40	.075
16	CHEVY CAMARO	: ALL	108.0	18.7	29.4	1.57	353	7156	.049
17	PONTIAC FIREBIRD	: ALL	108.0	18.7	29.4	1.57	192	3585	.054
18	CHEVY MALIBU	: 1978 - 81	108.0	21.7	30.4	1.40	82	1171	.070
19	OLDS. CUTLASS	: 1978 - 81	108.0	21.7	30.4	1.40	130	2272	.057
20	CHEVY MONTE CARLO	: 1978 - 81	108.0	21.7	30.4	1.40	108	1659	.065
21	BUICK CENTURY/REGAL	: 1978 - 81	108.0	21.7	30.4	1.40	74	1401	.053
22	PONTIAC LEMANS	: 1978 - 81	108.0	21.7	30.4	1.40	20	355	.056
23	CHRYSLER CORDOBA	: 1977 - 81	114.7	20.3	29.9	1.47	21	509	.041
24	DODGE MIRANDA	: 1977 - 81	114.7	20.3	29.9	1.47	1	48	.021
25	DODGE DIPLOMAT	: 1977 - 81	112.7	20.8	29.9	1.44	13	187	.070
26	CHRYSLER LEBARON	: 1977 - 81	112.7	20.8	29.9	1.44	22	377	.058
27	FORD MUSTANG	: 1979 - 81	100.4	20.0	28.3	1.42	209	1869	.112
28	MERCURY CAPRI	: 1979 - 81	100.4	20.0	28.3	1.42	58	587	.099
29	FORD LTD	: 1979 - 81	114.0	21.2	28.4	1.34	148	1461	.101
30	MERCURY MARQUIS	: 1979 - 81	114.0	21.2	28.4	1.34	11	204	.054
31	AMC CONCORD	: 1980	108.0	19.6	26.7	1.36	36	549	.066
32	AUDI 4000	: ALL	99.8	20.4	26.9	1.32	19	120	.158
33	DATSUN 2, ZX	: ALL	91.3	19.4	27.2	1.40	283	2060	.137
34	DATSUN B210	: ALL	92.1	20.3	24.2	1.19	551	2433	.226
35	RENAULT LE CAR	: ALL	95	20.9	24.5	1.17	27	113	.239
36	HONDA CIVIC	: <1983	94.5	20.7	26.3	1.27	335	1654	.203
37	TOYOTA COROLLA	: <1979	93.3	20.7	25.5	1.23	307	1388	.221
38	VW BEETLE	: <1980	94.5	22.5	26.6	1.18	588	2404	.245
39	VW RABBIT	: ALL	94.5	21.1	27	1.28	288	1686	.171
40	MAZDA 6LC	: <1980	91.1	20.5	24.6	1.20	75	269	.279
Totals							4910	39956	.123

Table E-2  
Variable Definitions

MODEL VARIABLE	CARDFILE FILENAME	CARDFILE VARIABLE	VARIABLE VALUES	FREQUENCY	CATEGORY	CARDFILE SUBCATEGORIES INCLUDED DICHOTOMIZED MODEL VARIABLES
ALCAD	DRIVER	ALC-DRUG	-1	31886	NO USE	NO INDICATION, MISS, UNK.
	DRIVER	RESTRAIN	1	8070	USE	ALCOHOL, DRUGS
BELT	DRIVER	RESTRAIN	-1	32838	NO BELT	NOT USED, NOT EQUIP, MISS, UNK
			1	7118	BELT	USED
CLIMATE	ACCIDENT	WEATHER	-1	32147	CLEAR/CLOUD	CLEAR, CLOUDY, MISS, UNK
			1	7809	OTHER	RAIN, SNOW/ICE, OTHER
CURVE	ACCIDENT	ROAD-ALIG	-1	29224	STRAIGHT	STRAIGHT, MISS, UNK
			1	10732	CURVED	CURVED
DURBAN	ACCIDENT	LAND-USE	-1	22280	MISSING	MISSING, UNK
			1	17676	URB/RUR	URBAN, RURAL
HERR	DRIVER	DRIVER-ERR	-1	10463	NOERROR	NONE, MISS, UNK,
			1	29493	ERROR	SPEED, SIGN/SIGNAL, PASSING, ASLEEP, ETC.
PROFILE	ACCIDENT	ROAD-PRO	-1	32968	LEVEL	LEVEL, MISS, UNK
			1	6988	GRADE	GRADE
ROADLOC	ACCIDENT	IMPLOC	-1	8624	ON ROAD	ON ROADWAY, MISS, UNK
			1	31332	OFF ROAD	ON SHOULDER, OFF ROADWAY
RURAL	ACCIDENT	LAND-USE	-1	32412	URBAN	URBAN, MISS, UNK
			1	7544	RURAL	RURAL
SEXY	DRIVER	SEX	-1	13065	FEMALE	FEMALE
			1	26891	MALE	MALE, MISS, UNK
STABLE	VEHICLE	PRESTAB	-1	34557	STABLE	TRACKING, NOT APPLICABLE, MISSING, UNK
			1	5399	NONSTABLE	SKIDDING, SPINNING, JACKKNIFING
STEER	VEHICLE	AVOID	-1	37142	NO AVOID	NO AVOIDANCE, MISS, UNK,
			1	2814	AVOID	AVOID VEHICLE, PEDESTRIAN, ETC
SURF	ACCIDENT	ROAD-SUR	-1	27848	DRY	DRY, MISS, UNK
			1	12108	ICY	WET, SNOW/ICE, OTHER
YOUTH	DRIVER	AGE	-1	18206	OLD	25 AND OVER
			1	21750	YOUNG	LESS THAN 25
SF		RANGE	MIN/MAX	MEAN	SD	
		.56	1.01	1.38	.147	
			1.57			
HB		37.9	83.5	102.81	7.3	
			121.4			

Table E-3. Logistic Regression Models for Rollover Probability Conditional on Single Vehicle Accident

MODEL FACTOR	SF ONLY	7 F	7F REDUCED	11 F	11F-SF	11F-SF-WB	11F-WB
1 SF	-4.903800 (45.41)	-3.965900 (28.70)	-4.483400 (21.20)	-4.090000 (29.37)	***** *****	***** *****	-4.934200 (44.52)
2 WB	***** *****	-.029990 (10.83)	-.030302 (8.437)	-.028107 (10.06)	-.080502 (36.47)	***** *****	***** *****
3 RURAL	***** *****	.456340 (19.00)	.449650 (13.22)	.456570 (18.94)	.475520 (20.03)	.481750 (20.68)	.452140 (18.77)
4 DURBAN	***** *****	-.393490 (17.13)	-.443740 (13.44)	-.397440 (17.03)	-.392100 (17.02)	-.293670 (13.08)	-.369840 (15.99)
5 CURVE	***** *****	.26478 (15.34)	.25394 (10.45)	.26075 (15.00)	.242860 (14.20)	.254560 (15.20)	.26693 (15.37)
6 HERR	***** *****	.404720 (18.01)	.407270 (13.51)	.392950 (17.26)	.375640 (16.65)	.374080 (16.79)	.397090 (17.45)
7 STABLE	***** *****	.215000 (9.59)	.265380 (8.675)	.287710 (12.07)	.286430 (12.25)	.297370 (13.01)	.289080 (12.14)
8 YOUTH	***** *****	***** *****	***** *****	.066667 (4.02)	.018644 (1.146)	.026348 (1.652)	.085073 (5.15)
9 ALCAD	***** *****	***** *****	***** *****	.093045 (4.71)	.079714 (4.11)	.070307 (3.71)	.094859 (4.812)
10 BELT	***** *****	***** *****	***** *****	.091940 (4.53)	.084718 (4.20)	.109700 (5.56)	.101900 (4.983)
11 SURF	***** *****	***** *****	***** *****	-.170520 (8.32)	-.148740 (7.85)	-.142570 (7.68)	-.173010 (8.95)
CONSTANT	4.620500 (32.56)	6.598100 (29.36)	7.318400 (22.37)	6.664100 (29.47)	6.488000 (29.21)	-1.6256 (49.9)	4.9588 (33.41)
r-squared (Make/ Model Based)	.907	.9448	.9467	.9444	.5272	.6812	.9296
LIS (See Text)	1109	1832	*****	1907	1478	781	1856

*See Table E-3*

predicting the probability of rollover in single vehicle accidents are SF, wheelbase, land use (rural/urban), and driver error. However, the primary importance of the stability factor (SF) is seen as the result of several observations:

1. Leaving SF and wheelbase (WB) out of a large 11-factor model lowered the likelihood ratio (as measured by the LIS explained in Section 3) more than leaving out all variables except SF (compare LIS of 781 for the former case with 1109 for the latter case).
2. Leaving SF out of the 11-factor model lowered the LIS much more than leaving out WB (LIS= 1478 vs. LIS=1856).
3. Although SF and WB are collinear ( $r = 0.64$ ) and tend to proxy for each other in predicting rollover probability, there is evidence in the coefficients that the major predictive power is in SF. This is because the coefficient of SF shrinks by only 17% upon the introduction of WB (from -4.934 to -4.090) while the coefficient of WB shrinks by almost a factor of 3 on the introduction of SF (from - 0.0805 to 0.0281).
4. The coefficient of SF does not shrink on the introduction of all nonvehicle variables. It changes only by a trivial amount: from -4.904 to -4.934.

A regression analysis that excluded Texas accidents (56% of the cases) indicated that the importance of the land-use variable (rural/urban) was underestimated in the previous models as Texas had no land-use variable (see Appendix C). This result indicated that land-use may be as important as SF in predicting rollover at the accident level. However, of more importance in the evaluation of SF as a predictor of rollover is the fact that the coefficient of SF changed little when land use was added to or taken out of the model. Moreover, the predictive capability of land-use is greatly reduced at the make/model level and will not affect the conclusions given below.

where is this shown?

When attention shifts to the performance of the models with predicted and actual rollover rates aggregated to the make/model level, the primary importance of stability factor is accentuated.

1. The vehicle make/model  $r^2$  of the model with SF only is far higher than that for the model with all other factors (compare 0.907 to 0.5272).
2. Several plots discussed in the body of the text show that any model containing stability factor predicts rollover rate at least fairly well and any model which does not contain SF predicts rollover rate very poorly.
3. The larger models containing both WB and SF lead to exceptionally accurate predicted rollover rates.

When the distributions of predicted probabilities based on actual and nominal data are observed, there is confirmation of the importance of SF in predicting rollover rate. There is also evidence that with regard to the influence on predicted rollover rates, the nonvehicle variables are remarkably well balanced over make/models.

## 1.0 BACKGROUND

Recent studies by Robertson and Kelly<sup>1</sup> and Harwin and Brewer<sup>2</sup> using statistical regression analysis, indicated that a vehicle's propensity to roll over is directly related to a "stability factor." The stability factor is defined as the ratio of one-half the track or tread width to the center of gravity height.

Based partly on the Robertson-Kelly study, Congressman T. Wirth (D-Co) petitioned NHTSA to establish a rule, based on the stability factor, to limit a vehicle's rollover potential (Congressman Wirth proposed a stability factor of 1.2 as being the minimally acceptable level). He also requested that NHTSA further study this issue, open a defect investigation, and warn the public of this potential problem. The major parts of this petition were denied, based in part on the limitations of the Robertson-Kelly study as well as the need of more evidence of the connection between rollover and stability factor and the need to study the role of other vehicle parameters in this question. The Robertson-Kelly limitations included the use of 14 make/models which tended to cluster the data and the use of the Fatal Accident Reporting System (FARS) data which made the results applicable to fatal accidents only. Harwin and Brewer improved on the Robertson and Kelly study by using forty make/models and approximately 40,000 single vehicle accidents, including but not limited to fatals, from the CARDfile database. Their results indicated a strong relationship between the stability factor and rollover accidents. An internal NHTSA review agreed that this study was a significant improvement over the previous study, but suggested that the number of observations was insufficient for the number of predictors that were tested. It was also suggested that a logistic regression be performed where each single vehicle accident would be treated as an observation rather than the vehicle make/model as the observation. The dependent variable

would be rollover. Logistic regression lends itself well to analysis when using a dichotomous dependent variable such as rollover/nonrollover.

NHTSA requested that TSC assist them in enhancing the Harwin-Brewer study by performing the logistic regression. This report details the results of an analysis of the relationship of the stability factor to rollover propensity using logistic regression analysis at the individual accident level.

## 2.0 TECHNICAL APPROACH

The approach that TSC used was to restructure the Harwin-Brewer (HB) CARDfile data on single vehicle accidents (SVA) so that a logit analysis could be performed at the accident level. The HB database that contained all SVAs, including rollovers, from the states of Maryland and Texas for 1984 and 1985 and Washington for 1983, 1984, and 1985 was used. Other predictors were also used in addition to the stability factor. These included those available from CARDfile relating to the driver, the vehicle, and the accident together with other variables relating to the vehicle geometry.

### 2.1 CARDfile ACCIDENT DATABASE

The Crash Avoidance Research Database (CARDfile) was developed by NHTSA to define problem areas and support research in crash avoidance. The police accident reports from the states of Texas, Maryland, Washington, Pennsylvania, Indiana, and Michigan are assembled into a common format in a Statistical Analysis System (SAS) structure. CARDfile had approximately 4 million accidents from these states for 1983 through 1985 that were available for analysis (Table 1). (CARDfile for 1986 is now available and will be used in future analyses.) The CARDfile database is subdivided into three subfiles relating to the accident, the driver, and the vehicle. The data elements in each of these files is shown in Table 2. Another study (Ref. 3) has indicated that CARDfile is representative of both national demographics and the accident experience. For a more detailed description of CARDfile, the reader is referred to the Harwin-Brewer study and to Reference 3.

**TABLE 1. SUMMARY OF CARDFILE STATE CRASH EXPERIENCE (1983-1985)**

<b>STATE</b>	<b>NO. CRASHES</b>	<b>NO. VEHICLES</b>
Indiana	480,399	854,571
Maryland	384,450	717,284
Michigan	1,023,366	1,724,288
Pennsylvania	414,210	694,854
Texas	1,341,415	2,326,103
Washington	338,307	617,093
<b>Totals</b>	<b>3,982,117</b>	<b>6,934,193</b>