U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

DOT HS 808 061
NHTSA Technical Report

January 1994

# Correlation of NCAP Performance with Fatality Risk in Actual Head-On Collisions

This document is available to the public from the National Technical Information Service, Springfield, Virginia 22161.

Technical Report Documentation Page

| | | |
|---|---|---|
| 1. Report No. DOT HS 808 061 | 2. Government Accession No. | 3. Recipient's Catalog No. |
| 4. Title and Subtitle CORRELATION OF NCAP PERFORMANCE WITH FATALITY RISK IN ACTUAL HEAD-ON COLLISIONS | | 5. Report Date January 1994 |
| | | 6. Performing Organization Code |
| 7. Author(s) Charles J. Kahane, Ph.D. | | 8. Performing Organization Report No. |
| 9. Performing Organization Name and Address Evaluation Division, Office of Strategic Planning and Evaluation National Highway Traffic Safety Administration Washington, D.C. 20590 | | 10. Work Unit No. (TRAIS) |
| | | 11. Contract or Grant No. |
| 12. Sponsoring Agency Name and Address Department of Transportation National Highway Traffic Safety Administration Washington, D.C. 20590 | | 13. Type of Report and Period Covered NHTSA Technical Report |
| | | 14. Sponsoring Agency Code |
| 15. Supplementary Notes | | |

16. Abstract


The New Car Assessment Program (NCAP) has gauged the performance of vehicles in frontal impact tests since model year 1979. In response to Congressional direction, the National Highway Traffic Safety Administration studied the relationship between vehicle test scores in NCAP and the fatality risk in crashes of vehicles on the road. This study is based on head-on collisions, where the effect of crashworthiness can be separated from the effects of extraneous factors that influence fatality rates (drivers, roadways, mileage). Collisions between two 1979-91 passenger cars in which both drivers were wearing safety belts were selected from the Fatal Accident Reporting System. There were 396 collisions (792 cars) in which both cars were identical with or very similar to vehicles which had been tested in NCAP. In the analyses, adjustments were made for the relative weights of the cars, and for the age and sex of the drivers - factors which substantially affect fatality risk.

There are statistically significant correlations between NCAP scores for head injury, chest acceleration and femur loading and the actual fatality risk of belted drivers. A composite NCAP score, based on the test results for all three body regions, has excellent correlation with fatality risk: in a head-on collision between a car with good composite score and a car of equal weight with poor score, the driver of the car with the better NCAP score has, on average, a 20 to 25 percent lower risk of fatal injury. Slightly smaller, but still significant fatality reductions are obtained even when the NCAP scores for just one body region (just HIC, or chest g's, or femur load) are used to partition the fleet into "good" and "poor" performance groups. The borderline between good and poor NCAP scores that optimizes the differences in actual fatality risk is close to the criteria specified in Federal Motor Vehicle Safety Standard (FMVSS) 208 for each of the three body regions.

Cars built from 1979 through 1982 had, on the average, the poorest NCAP scores. Test performance improved substantially in 1983 through 1986 models, and continued to improve in 1987 through 1991 cars. In parallel, fatality risk for belted drivers in actual head-on collisions decreased by 20 to 25 percent in model years 1979-91, with the largest decreases just after 1982. The 35 mph test speed for NCAP is 5 mph higher than the test speed for FMVSS 208. By now, most passenger cars meet the FMVSS 208 criteria at the NCAP test speed. The study shows that achievement of this enhanced level of test performance has been accompanied by a significant reduction in actual fatality risk. However, being a statistical study, it does not address what portion of the fatality reduction was directly "caused" by NCAP. Also, these results do not guarantee that any individual make-model with low NCAP scores will necessarily have lower fatality risk than another make-model with higher NCAP scores.


| | |
|---|---|
| 17. Key Words NCAP; New Car Assessment Program; crashworthiness; FARS; head-on collision; accident analysis; statistical analysis; frontal crash; fatality risk index; HIC; chest g's; femur load; safety belts; Polk; | 18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161 |

| | | | |
|---|---|---|---|
| 19. Security Classif. (of this report) Unclassified | 20. Security Classif. (of this page) Unclassified | 21. No. of Pages 164 | 22. Price |

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

## SUMMARY

The Appropriations Act for Fiscal Year 1992 directs "NHTSA to provide a study to the House and Senate Committees on Appropriations comparing the results of NCAP data from previous model years to determine the validity of these tests in predicting actual on-the-road injuries and fatalities over the lifetime of the models." In December 1993, the agency responded with a Report to the Congress that compared NCAP results and real-world crash experience, based on various analyses of accident data files. One set of analyses demonstrated a statistically significant correlation between NCAP performance and the fatality risk of <u>belted</u> drivers in actual <u>head-on collisions</u>. This technical report provides a more detailed exposition of the data sources, analytic approach and statistical findings in the analysis of head-on collisions.

NHTSA's goal was to see if cars with poor NCAP scores had more belted-driver fatalities than would be expected, given the weights of the cars, and the age and sex of the drivers involved in the crashes. Without adjustment for vehicle weight, driver age and sex, the large diversity of fatality rates in accident data mainly reflects the types of people who drive the cars, not the actual crashworthiness of the cars. For example, "high-performance" cars popular with young male drivers have an exceptionally high frequency of fatal crashes - because they are driven in an unsafe manner - even though they may be just as crashworthy as other models. NHTSA's analysis objective was to isolate the actual <u>crashworthiness</u> differences between cars, removing differences attributable to the way the cars are driven, the ages of the occupants, etc., and then to correlate NCAP performance with crashworthiness on the highway.

Analysis overview

Since NCAP is a frontal impact test involving dummies protected by safety belts, the agency limited the accident data to frontal crashes involving belted occupants. However, NHTSA did not consider all types of frontal crashes, but further limited the data to head-on collisions between two passenger cars, each with a belted driver, which resulted in a fatality to one or to both of the drivers. A head-on collision is a special type of highway crash ideally suited for studying frontal crashworthiness differences between two cars. Both cars are in essentially the same frontal collision. It doesn't matter if one of them had a "safe" driver and the other, an "unsafe" driver; at the moment they collide head-on, how safely they were driving before the crash is nearly irrelevant to what happens in the crash. Which driver dies and which survives depends primarily on the intrinsic relative crashworthiness of the two cars, their relative weights, and the age and sex (vulnerability to injury) of the two drivers.

If car 1 and car 2 weigh exactly the same, and both drivers are the same age and sex, the likelihood of a driver fatality in a head-on collision would be expected to be equal in car 1 and car 2. If car 1 and car 2 have different weights, etc., it is still possible to calibrate formulas predicting the expected fatality risk for each driver in a head-on collision between the two cars, as a function of each vehicle's weight and each driver's age and sex. The formulas measure the relative vulnerability to fatal injury of the two drivers, given that their cars had a head-on collision. The risk is greater in the lighter car than the heavier car, and a female or older driver is more vulnerable to injury than a male or younger driver. For example, given 100 fatal head-on collisions between 3000-pound-cars driven by belted, 20-year-old males and 2500 pound cars driven by belted, 50-year-old females, these formulas predict 10.8 times as many deaths among the older females in the lighter cars as among the young males in the heavier cars.

Cars with <u>average</u> crashworthiness capabilities will experience an actual number of fatalities very close to what is predicted by these formulas, which are calibrated from the collision experience of production vehicles. If a group of cars, however, consistently experiences more fatalities than expected in their head-on collisions, then the empirical evidence suggests that this group of cars is less crashworthy than the average car of similar mass. The gist of the analyses is to see if groups of cars with poor NCAP scores have significantly more belted-driver fatalities per 100 actual head-on collisions than expected (and there are several ways to define a "poor" score). The analyses measure the reduction in fatality risk, in actual head-on collisions, for a car with good NCAP scores relative to a car with poor NCAP scores. They measure the overall reduction in fatality risk, for belted drivers in head-on collisions, since model year 1979, when NCAP testing began, until 1991, the latest model year for which substantial accident data were available as of mid 1993.

The analyses require a data file of actual head-on collisions, with both drivers belted, resulting in a fatality to at least one of the drivers, indicating, for both cars, the curb weight, the driver's age and sex, and the HIC, chest g's and femur loads that were recorded for the driver dummy when that car was tested in NCAP. NHTSA's Fatal Accident Reporting System (FARS), complete through mid-1992, provided the basic accident data for the study. The FARS data were supplemented with accurate curb weights, derived from R. L. Polk's files and NHTSA compliance tests. Insufficient NCAP and FARS data were available to include light trucks, vans or sport utility vehicles in the analyses. Thus, the study is limited to collisions between two 1979-91 passenger cars.

NHTSA staff reviewed the cars involved in head-on collisions on FARS and identified, where possible, the NCAP test car that came closest to matching the FARS case. They found 396 head-on collisions, involving 792 cars, in which both drivers were belted and both cars match up acceptably with an NCAP case: (1) The make-models on FARS and NCAP are identical or true "corporate cousins" (e.g., Dodge Omni and Plymouth Horizon). (2) The model years on FARS and NCAP are identical, or the FARS model year is later than the NCAP model year, but that model was basically unchanged during the intervening years. The FARS cases were supplemented with the matching NCAP test results for each car. The sample is large enough for a statistical analysis of NCAP scores and fatality risk.

FARS data do not single out those head-on collisions that closely resemble an NCAP test: perfectly aligned collisions of two nearly identical cars, with minimal offset, a closing speed close to 70 mph, and both drivers 50th-percentile males. In addition, FARS cases may involve injury to the neck or abdomen: the potential for injury to these body regions is not specifically measured in NCAP. It is inappropriate to expect perfect correlation between NCAP test results and actual fatality risk in the full range of head-on collisions represented in the FARS sample. Moreover, if there is any significant correlation between the two, it suggests that the NCAP scores say something about actual crashworthiness in a range of crashes that goes far beyond the specific type tested in NCAP.

Correlation of NCAP scores and fatality risk

The goal of the analysis is to test if cars with poor scores on the NCAP test have higher fatality risk for belted drivers, in actual head-on collisions, than cars with good or acceptable scores. There are many ways to define "poor" and "good" scores and measure the difference in fatality risk. All of the methods tried out by NHTSA staff demonstrate a statistically significant relationship between NCAP scores and actual fatality risk, as shown in the accompanying table.

A straightforward way to delineate "poor" and "good" scores is to partition the cars based on their NCAP score for a single body region - chest g's, HIC or femur load - and to consider only a subset of the 392 head-on crashes where one car has a score in the "poor" range and the other car has a score in a good or acceptable range. This subset should contain approximately 120 crashes, which is equivalent to defining the worst 20 percent of cars as "poor" performers and the remaining 80 percent as good or acceptable. Do the cars with the poor NCAP scores have significantly more driver fatalities than expected?

When chest g's are used to partition the cars into acceptable and poor performance groups, the cars with high chest g's almost always have significantly more fatalities than the cars with acceptable chest g's. For example, there are 125 actual head-on collisions (both drivers belted) in which one of models had more than 56 chest g's for the driver when it was tested in NCAP, and the other had 56 g's or less. In the 125 cars with chest g's > 56, 80 drivers died, whereas only 68.2 fatalities were expected, based on car weight, driver age and sex. In the 125 cars with chest g's $\leq$ 56, there were 74 actual and 77.6 expected driver fatalities. That is a statistically significant fatality reduction of

COLLISIONS OF CARS WITH "GOOD" NCAP SCORES INTO CARS WITH "POOR" NCAP SCORES
(N of crashes approximately 120 in each analysis)

| "Good" NCAP Performance | "Poor" NCAP Performance | Performance in Actual Crashes: N of Crashes | Performance in Actual Crashes: Fatality Reduction for Good Car (%) |
|---|---|---|---|
| Chest g's ≤ 56 | Chest g's > 56 | 125 | 19* |
| HIC ≤ 1000 | HIC > 1200 | 113 | 14* |
| L Femur ≤ 1600 AND<br>R Femur ≤ 1600 AND<br>L+R Femur ≤ 2600 | L Femur > 1600 OR<br>R Femur > 1600 OR<br>L+R Femur > 2600 | 132 | 20** |
| HIC ≤ 1100 AND<br>Chest g's ≤ 60 | HIC > 1300 OR<br>Chest g's > 60 | 125 | 19* |
| Chest g's ≤ 56 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 134 | 22** |
| HIC ≤ 900 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | HIC > 1300 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 121 | 19* |
| HIC ≤ 900 AND<br>Chest g's ≤ 56 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 118 | 21** |
| NCAPINJ ≤ .6 | NCAPINJ > .6 | 117 | 26** |

*Statistically significant at the .05 level
**Statistically significant at the .01 level

$$1 - [(74/80) / (77.6/68.2)] = 19 \text{ percent}$$

for the cars with the lower chest g's. The relationship between chest g's on the NCAP test and fatality risk over the range of head-on collisions experienced on the highway, although statistically significant, is not perfect. Merely having the lower NCAP score of the two cars in the collision does not guarantee survival, even if the two cars are of the same weight and the drivers of the same age and sex. Yet, on the average, in collisions between cars with ≤ 56 chest g's on NCAP and cars with > 56 chest g's, the driver of the car with the better NCAP score had 19 percent less fatality risk than the driver of the car with the poorer NCAP score, after controlling for weight, age and sex.

Fifty-six chest g's are just one possible boundary value between "good" and "poor" performance. The fatality reduction for "good" performers can be magnified by using a higher boundary value or by replacing a single boundary value with a gap, putting some distance between the "good" and the "poor" groups. For example, in collisions of cars with chest g's ≤ 60 into cars with chest g's > 60 (the pass-fail criterion in FMVSS 208), the fatality reduction in the "good" performers is 24 percent. However, there are only 92 crashes meeting those criteria. Many other boundary values between low and high chest g's will also produce statistically significant fatality reductions for the group with low chest g's, but the boundary value of 56 maximizes the fatality reduction for an accident sample close to 120 crashes.

The Head Injury Criterion (HIC) can be used to partition the cars into two performance groups. In 113 head-on collisions between a car with HIC ≤ 1000 on the NCAP test and a car with HIC > 1200, the fatality risk was a statistically significant 14 percent lower in the cars with HIC ≤ 1000. The femur loads measured on the NCAP tests can also, by themselves, differentiate safer from less safe cars. The "good" performers are defined to be the cars with ≤ 1600 pounds on each leg, and the sum of the two loads ≤ 2600 pounds. The "poor" performers are those with > 1600 pounds on either leg, or a sum > 2600 pounds. In 132 head-on collisions, the fatality reduction for the "good" NCAP femur load performers was a statistically significant 20 percent.

One reason that chest g's, HIC and femur load all "work" by themselves is that the three NCAP test measurements are not independent observations on isolated body regions. Cars with intuitively excellent safety design tend to have low scores on all parameters, while cars with crashworthiness problems tend to have high scores on one or more parameters, but it is not always predictable which one. Still, the reasons for the significant correlation between NCAP femur load and actual fatality risk are not completely understood at this time, since injuries to the lower extremities, by themselves, are generally not fatal.

Any two NCAP parameters, working together, can do an even more reliable job than any single parameter. In 125 actual head-on collisions between cars with driver HIC ≤ 1100 and chest g's ≤ 60 on the NCAP test and cars with either HIC > 1300 or chest g's > 60, the fatality risk was a statistically significant 19 percent lower in the cars with low HIC and chest g's. The accompanying table shows how chest g's and femur load, or HIC and femur load can be used to partition the cars, with statistically significant 19-22 percent fatality reductions for the "good" performers, in samples of 121-134 crashes.

NCAP scores for all three body regions, with an independent "pass-fail" criterion on each score, work about as well as scores for any two body regions. "Good" performance could be defined as HIC $\leq$ 900 and chest g's $\leq$ 56 and femur load $\leq$ 1400 on each leg and $\leq$ 2400, total, while HIC > 1300 or chest g's > 60 or femur load > 1700 on either leg or > 2700, total defines "poor" performance. The fatality risk in 118 actual head-on collisions between a good and a poor NCAP performer is a statistically significant 21 percent lower for the drivers of the cars with good NCAP scores, after controlling for vehicle weight, driver age and sex. These criteria can be varied by a moderate amount and the fatality reduction for the "good" performers will still be statistically significant, as long as the HIC cutoff is reasonably close to or slightly above the FMVSS 208 value of 1000, the chest g cutoff is not far from the FMVSS 208 value of 60 g's, and the femur load cutoff ranges from about 1400 pounds up to the FMVSS 208 value of 2250 pounds.

A highly efficient way to use the NCAP scores for the three body regions, however, is to combine them into a single composite score, wherein excellent performance on two body regions might compensate for moderately poor performance on the third. The composite score could be some type of weighted or unweighted average of the scores for the various body regions. For example, a weighted average measure of NCAP performance, NCAPINJ, was derived by a two-step process. First, the actual NCAP results for the driver dummy were transformed to logistic injury probabilities, HEADINJ, CHESTINJ, LFEMURINJ and RFEMURINJ, each ranging from 0 to 1. The weighted average

NCAPINJ = .21 HEADINJ + 2.7 CHESTINJ + 1.5 (LFEMURINJ + RFEMURINJ)

has the empirically strongest relationship with fatality risk for belted drivers in the specific data set of actual head-on collisions described above (396 collisions, 792 cars). The accident data include 117 head-on collisions of a car with NCAPINJ $\leq$ 0.6 into a car with NCAPINJ > 0.6. Fatality risk is a statistically significant 26 percent lower in the cars with NCAPINJ $\leq$ 0.6. Since NCAPINJ is a weighted sum of NCAP scores for all of the body regions, the cars with NCAPINJ $\leq$ 0.6 have, on the average, substantially lower HIC, chest g's and femur loads than cars with NCAPINJ > 0.6.

The purpose of defining NCAPINJ was to illustrate the strength of the overall relationship between NCAP performance and fatality risk. However, NCAPINJ is not a "magic bullet" or "ideal" way to combine the NCAP scores, resulting in far higher correlations than other methods. Many other weighted averages, or even an unweighted sum of the logistic injury probabilities, work almost as well for differentiating the safer from the less safe cars on the principal accident data set. On a more restricted alternative accident data set of 310 collisions and 620 cars, where the FARS vehicles are also required to have the same number of doors as their matching NCAP test vehicles, NCAPINJ is not the optimum weighted average (although it comes close to the optimum), and it is only slightly more correlated with fatality risk than an unweighted sum of the logistic injury probabilities. Moreover, on this alternative data set, HIC and femur load have about equally strong correlation with fatality risk.

Improvements in actual crashworthiness and NCAP performance during 1979-91

The performance of passenger cars on the NCAP test has greatly improved since the program was initiated in 1979. That was demonstrated in NHTSA's 1992-93 reports to the Congress and several other studies, which cite specific improvements in vehicle structures and occupant protection systems resulting in better NCAP performance. Has the historical trend of better performance on the NCAP test been matched by a reduction in the actual fatality risk of belted drivers in head-on collisions?

In general, it is not easy to compare the crashworthiness of cars of different model years. Fatality rates per 100 million vehicle miles have been declining for a long time. In any given year, the fatality rate per 100 million miles or per 100 crashes is lower for new cars than for old cars. Both trends create the impression that "cars are getting safer all the time," but, in fact, the declines in fatality rates to a large extent reflect changes in driving behavior, roadway environments, demographics or accident-reporting practices. A head-on collision between cars of two different model years, however, reveals their relative crashworthiness. Both cars are in essentially the same frontal collision, on the same road, in the same year, on the same accident report. The behavior of each driver, prior to the impact, has little effect on who dies during the impact. After adjustment for differences in car weight, driver age and sex, the model year with more survivors is more crashworthy.

There have been 241 actual head-on collisions between a model year 1979-82 car and a model year 1983-91 car, in which both drivers were belted. These collisions allow a comparison of cars built during the first four years of NCAP to subsequent cars, where manufacturers have had time to build in safety improvements. In the 241 older cars, 146 drivers died, whereas only 126.6 fatalities were expected, based on car weight, driver age and sex. In the newer cars, there were 132 actual and 147.1 expected driver fatalities. For the 1983-91 cars, that is a statistically significant fatality reduction of

$$1 - [(132/146) / (147.1/126.6)] = 22 \text{ percent}$$

A more generalized analysis, which allows a larger sample size of 1189 crashes, applies to head-on collisions in which the "case" vehicle of interest is a 1979-91 car that matches up with an NCAP test, whose driver wore belts, but the "other" vehicle in the crash can be any 1976-91 passenger car with a belted driver. For any subset of crashes, a fatality risk index can be computed for the "case" vehicles, based on the ratio of actual to expected fatalities in the case and other vehicles. The lower the risk index, the more crashworthy the car (100 = average). The actual fatality risk indices can be compared in three model-year groups, 1979-82, 1983-86 and 1987-91. So can the NCAP test performance, as measured by a composite score such as NCAPINJ, or by the average values of the actual NCAP parameters for the three body regions:

| | Model Years | | |
|---|---|---|---|
| | **1979-82** | **1983-86** | **1987-91** |
| Fatality risk index in actual head-on collisions | 119 | 95 | 91 |
| Average value of NCAPINJ | .59 | .40 | .37 |
| Percent of cars with NCAPINJ > 0.6 | 49 | 14 | 9 |
| Average HIC | 1052 | 915 | 827 |
| Average chest g's | 54.9 | 46.8 | 46.5 |
| Average left femur load | 928 | 883 | 1002 |
| Average right femur load | 1079 | 784 | 1018 |

The trends in the actual fatality risk and the average value of NCAPINJ are almost identical. The risk index decreased by a statistically significant 20 percent from 1979-82 to 1983-86, and by another 4 percent from then until 1987-91 (nonsignificant). In all, the actual fatality risk for belted drivers in head-on collisions decreased by a statistically significant 24 percent from model years 1979-82 to 1987-91. A composite NCAP score, such as NCAPINJ, nicely portrays the improvement in NCAP performance over time. Parallel to the reduction in the fatality risk index, NCAPINJ greatly improved from an average of 0.59 in model years 1979-82 to 0.40 in 1983-86, with an additional, modest improvement to 0.37 in 1987-91. If NCAPINJ = 0.6 is defined as the limit of "acceptable" NCAP performance, the passenger car fleet has truly progressed since the inception of NCAP: initially, 49 percent of the cars had NCAPINJ > 0.6, but that decreased to 14 percent in 1983-86 and 9 percent in 1987-91. Average HIC and chest g's declined substantially during the NCAP era; average femur loads stayed about the same, but well below the 2250 pounds permitted in FMVSS 208.

## Principal findings, conclusions and caveats

- o There is a statistically significant correlation between the performance of passenger cars on the NCAP test and the fatality risk of belted drivers in actual head-on collisions. Since many head-on collisions differ substantially from NCAP test conditions, this suggests NCAP scores are correlated with actual crashworthiness in a wide range of crashes.

- o In a head-on collision between a car with "acceptable" NCAP performance and a car of equal mass with "poor" performance, the driver of the "good" car has, on the average, about 15-25 percent lower fatality risk.

- o A highly effective way to differentiate "good" from "poor" NCAP performance is by a single, composite NCAP score, such as a weighted combination of the scores for the three body regions. However, even the NCAP score for any single body region can be used to partition the fleet so that the cars with "good" scores have significantly lower fatality risk than the cars with "poor" scores. The borderline between "good" and "poor" NCAP scores that optimizes the differences in actual fatality risk is close to the FMVSS 208 criteria for each of the three body regions.

- NCAP scores have improved steadily since the inception of the program in 1979, with the greatest improvement in the early years. By now, most passenger cars meet the FMVSS 208 criteria in the 35 mph NCAP test. This achievement has been paralleled by a 20-25 percent reduction of fatality risk for belted drivers in actual head-on collisions in model years 1979-91, with the largest decreases during the early 1980's.

- This is a statistical study and it is not appropriate for conclusions about cause and effect. It shows that passenger cars became significantly safer in head-on collisions during 1979-91, as NCAP scores improved. It does not prove that the NCAP program was the stimulus for each of the vehicle modifications that saved lives during 1979-91. (For example, the automatic protection requirement of FMVSS 208 was another important stimulus.)

- The correlation between NCAP scores and actual fatality risk is statistically significant, but it is far from perfect. On the whole, cars with poor NCAP scores have higher-than-average fatality risk in head-on collisions, but there is no guarantee that every specific make-model with poor NCAP scores necessarily has higher fatality risk than the average car. Conversely, there is no guarantee that a specific model with average or even excellent scores necessarily has average or lower-than-average fatality risk in head-on collisions.

- The data show that cars with poor NCAP scores (e.g., above the FMVSS 208 criteria) have significantly elevated fatality risk in head-on collisions, but they do not show a significant difference between the fatality risk of cars with exceptionally good NCAP performance and those with merely average performance.

# CHAPTER 1

## INTRODUCTION AND ANALYSIS OVERVIEW

The Appropriations Act for Fiscal Year 1992 directs "NHTSA to provide a study to the House and Senate Committees on Appropriations comparing the results of New Car Assessment Program (NCAP) data from previous model years to determine the validity of these tests in predicting actual on-the-road injuries and fatalities over the lifetime of the models" [5], p. 35. In February 1992, the agency responded to the directive with a plan to compare NCAP results and real-world crash experience, based on various analyses of accident data files [23]. A Report to Congress, presenting the highlights of the analyses, was completed in December 1993 [24]. One analytic approach, described in Section 3 of the Report to Congress, addressed the correlation between NCAP performance and the fatality risk of <u>belted</u> drivers in actual <u>head-on collisions</u>. This technical report provides a more detailed exposition of the data sources, analytic approach and statistical findings in the analysis of head-on collisions.

The New Car Assessment Program was developed in response to Title II of The Motor Vehicle Information and Cost Savings Act of 1973 (MVICS) [20], which authorized NHTSA to develop consumer information on the <u>crashworthiness</u> of passenger vehicles. Since 1979, NCAP has been a program of <u>frontal</u> impact tests at 35 mph into a barrier, with <u>belted</u> dummies at the driver and right-front seat positions. The 35 mph impact speed is 5 mph faster than the test speed in NHTSA's Federal Motor Vehicle Safety Standards for occupant protection in frontal crashes (FMVSS 204, 208, 212, 219 and 301), and it produces a velocity change close to the average in actual fatal frontal impacts. Measurements on the dummies are used to calculate the Head Injury Criterion (HIC), chest g forces (3 millisecond peak) and left and right femur loads (peak axial loads at knee). HIC measures the cumulative impact force on the head during the crash. An average of 30 passenger cars and light trucks are tested each year, including make/models that are new or significantly redesigned in that model year.

### 1.1 NCAP performance vs. crashworthiness on the highway

FMVSS 208 requires all passenger cars to have HIC $\leq$ 1000, chest g's $\leq$ 60 and femur load $\leq$ 2250 pounds on a 30 mph test. NCAP is not a regulatory program and does not set pass-fail levels of performance for its 35 mph test. Nevertheless, it could be argued that the level of frontal occupant protection guaranteed by the basic FMVSS at 30 mph has largely been extended to 35 mph since NCAP started in 1979. In model year 1979, fewer than 25 percent of cars met the FMVSS 208 criteria at 35 mph. In subsequent years, NCAP results were widely disseminated to consumers, manufacturers and insurers. By 1986-91, over 60 percent of passenger cars met the FMVSS 208 criteria at 35 mph. While statistics do not prove that the NCAP program was solely responsible for the improvement in test results (e.g., automatic occupant protection installed in response to FMVSS 208 was another obvious factor), the trend is certainly in the right direction and it appears to fulfill one promise of a consumer information program: the manufacturers significantly enhanced safety performance as measured by the publicized test protocol. They are now "designing their vehicles to 35 mph."

While there is overwhelming evidence that vehicle performance on the NCAP test has improved since the inception of the program, that evidence, by itself, does not prove that cars have become safer in actual crashes on the highway. The ultimate goal of all safety programs, including consumer information programs such as NCAP is the reduction of deaths and injuries on the highway. There is a desire for evidence that cars with poor NCAP scores are less safe in actual crashes than cars with acceptable scores, and, more generally, that cars have become safer in actual crashes since the beginning of NCAP. Researchers have eagerly explored the correlation between NCAP performance and fatality risk in actual crashes since the initial years of NCAP [4]. There are two reasons why their efforts have had little success in past years. NCAP is a test program involving belted dummies, and, until very recently, there simply have not been enough fatal or serious-injury accident data involving belted occupants for a meaningful comparison with NCAP results. NCAP describes differences in the crashworthiness of vehicles on identical 35 mph tests, whereas in accident data it is quite difficult to isolate the effects of crashworthiness (the ability of a vehicle to protect its occupants from death or injury, given that a crash has occurred) from other factors that affect fatality rates of cars: the types of people who drive the cars, and the environments where they are driven.

Thanks to the steady increase in belt use after 1984, as more and more States enacted belt use laws, enough accident data involving belted occupants had accumulated, by 1993, for meaningful statistical analyses. But it is still necessary to find a method which isolates the crashworthiness differences between cars and filters out the differences attributable to the way the cars are driven. The method used in this report is to analyze fatal head-on collisions between two passenger cars.

## 1.2 The difficulty of isolating crashworthiness effects

Before any discussion of the unique advantages of head-on collisions as a data source, it helps to review the foibles of conventional measures of fatality risk, such as the occupant fatality rate per million vehicle years. It is well known that "high-performance" cars popular with young male drivers have a higher frequency of fatal crashes than family sedans, and it is generally suspected that the difference is primarily due to the way the cars are driven, not crashworthiness. But a look at some actual fatality rates helps clarify the extent to which differences in drivers and exposure influences the variation in fatality rates.

For example, Table 1-1 displays the actual rate of fatalities per million vehicle registration years for model year 1985-87 cars in calendar years 1986-88 (data compiled by the Insurance Institute for Highway Safety [28]). The actual rates range from 60 in the Volvo 740/760 to 520 in the Corvette - almost a 9:1 ratio. The 15th percentile of actual risk is 120 (Pontiac Grand Am) and the 85th percentile is 310 (Dodge Daytona) - still a 3:1 variation in risk across the middle 70 percentiles. It is intuitively obvious that 9:1 differences between cars are due primarily to the types of people who drive them, rather than real variation in crashworthiness. It is most unlikely than one make-model is intrinsically 9 times as dangerous as another. The make-models in Table 1-1 with the lowest fatality rates are primarily luxury and family cars. The models with the highest rates are "performance" cars and small economy cars. Even within a specific make-model, station wagons have lower fatality rates than four-door sedans, while two-door coupes have higher rates.

These differences are somewhat diminished, but still persistent, even <u>after</u> "adjusting" the rates for key variables such as car weight, driver age and sex. The Insurance Institute attempted to control for at least some of the driver differences by computing, for each make/model, a <u>predicted</u> fatality rate

TABLE 1-1

FATALITY RISK INDICES BASED ON FATALITIES
PER MILLION REGISTERED VEHICLE YEARS
(Model year 1985-87 cars in calendar years 1986-88)

| | Fatality Rate | | Fatality | | Fatality Rate | | Fatality |
|---|---|---|---|---|---|---|---|
| | Actual | Predictd | Risk Index | | Actual | Predictd | Risk Index |
| Volvo 740/760 4dr | 60 | 140 | 43 | Ford Escort 4dr | 180 | 270 | 67 |
| Ford Taurus Wagon | 70 | 150 | 47 | Ford Tempo 4dr | 180 | 180 | 100 |
| Lincoln Town Car | 80 | 120 | 67 | Buick LeSabre | 180 | 140 | 129 |
| VW Jetta 4dr | 110 | 250 | 44 | Olds Calais 2dr | 190 | 190 | 100 |
| Chev Cavalier Wagon | 110 | 200 | 55 | Ford Tempo 2dr | 200 | 260 | 77 |
| Toyota Cressida | 110 | 190 | 58 | VW Golf 4dr | 200 | 250 | 80 |
| Audi 5000 | 110 | 170 | 65 | Nissan Maxima | 200 | 250 | 80 |
| Olds Ciera Wagon | 110 | 150 | 73 | Chev Nova 4dr | 200 | 210 | 95 |
| Caddy DeVille 2dr | 110 | 140 | 79 | Buick Regal 2dr | 200 | 190 | 105 |
| Caddy DeVille 4dr | 110 | 120 | 92 | Subaru 4dr | 200 | 180 | 111 |
| Ford Escort Wagon | 120 | 220 | 55 | Pont Grand Am 2dr | 210 | 280 | 75 |
| Volvo 240 | 120 | 190 | 63 | Honda Civic 2dr | 230 | 280 | 82 |
| Pont Grand Am 4dr | 120 | 190 | 63 | Ford T-Bird | 230 | 250 | 92 |
| Olds Ciera 2dr | 120 | 180 | 67 | Dodge Omni 4dr | 230 | 210 | 110 |
| Pont Grand Prix | 120 | 170 | 71 | Chev Cavalier 4dr | 230 | 190 | 121 |
| Buick Century 4dr | 120 | 160 | 75 | Mercury Cougar | 240 | 220 | 109 |
| Mercury Gr Marquis | 120 | 150 | 80 | Chev Celebrity 2dr | 240 | 150 | 160 |
| Mercury Sable | 130 | 200 | 65 | Toyota Corolla 2dr | 250 | 380 | 66 |
| Pontiac 6000 | 130 | 170 | 76 | Nissan 200SX | 250 | 330 | 87 |
| Chev Celebrity Wagon | 130 | 170 | 76 | Pont Sunbird 4dr | 250 | 180 | 139 |
| Olds Ciera 4dr | 130 | 150 | 87 | BMW 300 2dr | 260 | 340 | 76 |
| Buick Electra | 130 | 140 | 93 | Hyundai Excel 4dr | 260 | 260 | 100 |
| Ford Taurus | 140 | 200 | 70 | Plym Reliant 4dr | 260 | 160 | 163 |
| Olds Calais 4dr | 140 | 190 | 74 | Chev Cavalier 2dr | 270 | 260 | 104 |
| Honda Accord 2dr | 140 | 180 | 78 | Pont Sunbird 2dr | 280 | 240 | 117 |
| Subaru Wagon | 140 | 170 | 82 | Plym Horizon 4dr | 280 | 210 | 133 |
| Chev Caprice Wagon | 140 | 170 | 82 | Chev Monte Carlo | 280 | 210 | 133 |
| Ford Crown Vic | 140 | 160 | 88 | Dodge Aries 4dr | 290 | 190 | 153 |
| Nissan Sentra 2dr | 150 | 430 | 35 | Ford Escort 2dr | 300 | 290 | 103 |
| Honda Prelude | 150 | 310 | 48 | Dodge Daytona | 310 | 320 | 97 |
| Buick Somerset 2dr | 150 | 220 | 68 | Chev Spectrum 2dr | 320 | 250 | 128 |
| Mazda 626 | 150 | 200 | 75 | Chev Chevette 2dr | 340 | 250 | 136 |
| Honda Accord 4dr | 150 | 170 | 88 | Pontiac Fiero | 360 | 380 | 95 |
| Olds 98 | 150 | 150 | 100 | Plym Turismo | 360 | 260 | 138 |
| Olds Delta 88 | 150 | 130 | 115 | Pontiac Firebird | 380 | 310 | 123 |
| Chrys 5th Avenue | 150 | 120 | 125 | Honda CRX | 390 | 530 | 74 |
| Toyota Celica | 160 | 280 | 57 | Chev Sprint | 410 | 290 | 141 |
| Toyota Corolla 4dr | 160 | 230 | 70 | Chev Chevette 4dr | 410 | 190 | 216 |
| Mercury Topaz 4dr | 160 | 200 | 80 | Nissan 300ZX | 420 | 420 | 100 |
| Chrys New Yorker | 160 | 160 | 100 | Ford Mustang | 440 | 370 | 119 |
| Chev Caprice 4dr | 160 | 140 | 114 | Dodge Charger | 450 | 330 | 136 |
| Honda Civic 4dr | 170 | 260 | 65 | Chev Camaro | 490 | 380 | 129 |
| Chev Celebrity 4dr | 170 | 160 | 106 | Chev Corvette | 520 | 360 | 144 |

**Actual fatality rate** = actual fatalities per million registration years (source: IIHS [28])

**Predicted fatality rate** "attempts to take into account the age and sex of drivers involved and the car size [28]."

**Fatality risk index** = 100 * Actual/Predicted which "takes into account the age and sex of drivers involved and the car size" to the extent that they affect annual mileage, collision propensity and vulnerability to fatal injury. The predicted rates are shown next to the actual fatality rates in Table 1-1. The adjusted fatality risk index, equal to the ratio of the adjusted to the predicted fatality rate (and multiplying by 100) was computed for each make-model and is shown in the right columns of Table 1-1. Cars with an index below 100 have lower fatality rates than expected; indices above 100 denote "higher than expected" fatality rates. The index ranges from 35 in the Nissan Sentra 2 door to 216 in the Chevrolet Chevette 4 door - almost a 6:1 ratio. The 15th percentile of the risk index is 67 (Lincoln Town Car) and the 85th percentile is 129 (Chevrolet Camaro) - nearly a 2:1 variation in the risk index across its middle 70 percentiles.

This fatality risk index filters out some of the worst disparities in the actual fatality rates, but still does not isolate crashworthiness differences. The differences between the best and the worst cars still seem larger than what could likely be ascribed to variations in crashworthiness. The index shows some large differences between cars that ought to be about equally crashworthy. For example, the Celebrity wagon has an index of 76, the Celebrity 4-door sedan's index is 106 and the 2-door model's index is 160. It is true that 4-door cars have a safety advantage over 2-door cars in certain types of crashes, but not that large an advantage. Clearly, the types of people who drive station wagons are much less likely to have serious accidents than the drivers of the same age and sex who drive 2-door coupes. Similarly, Chevrolets and Fords consistently have higher risk indices than "corporate cousin" vehicles sharing essentially identical components. For example, the Chevrolet Monte Carlo has an index of 133, while the Pontiac Grand Prix has an index of 71; the Ford Tempo 4-

door has an index of 100 while the Mercury Topaz 4-door has an index of 80. Differences like these are much too large to ascribe to crashworthiness and probably reflect socioeconomic or geographic differences of the drivers. Almost all the imported cars have risk indices below 100, often far below 100. The advantage for imported cars, however, may be due to the clientele, not the vehicle: the Chevrolet Nova is essentially the same car as the Toyota Corolla 4-door, but the Nova has an index of 95 while the Corolla's index is 70. In summary, simple fatality rates per million car years, even if they are adjusted for driver age and sex, are poor measures of crashworthiness because 30-year-old males who drive Volvos have completely different driving patterns and far lower accident proneness than 30-year-old males who drive Corvettes.

## 1.3 Analysis overview

The objective of isolating the actual crashworthiness differences between cars is better attained by studying head-on collisions between two passenger cars, each with a belted driver, which resulted in a fatality to one or to both of the drivers. A head-on collision is a special type of highway crash ideally suited for studying crashworthiness differences between two cars; it comes close to a controlled laboratory test. Both cars are in essentially the same frontal collision. The outcome is a verdict on the intrinsic relative crashworthiness of the two cars and the intrinsic relative vulnerability to injury of the two drivers. Events that happened before the moment of impact - unsafe driving acts, crash avoidance capabilities - are mostly irrelevant in deciding which driver survives the crash and which dies. Records of every fatal head-on collision since 1975 may be found in NHTSA's Fatal Accident Reporting System (FARS).

To set the scene, consider a head-on collision between a Volvo 740 and a Corvette. Both cars weigh just over 3000 pounds. Both drivers, in this collision, are 30-year-old males. At the moment these two cars hit head-on, it becomes irrelevant that Corvettes have 9 times as high an overall fatality rate per million car years as Volvos, as shown in Table 1-1. It is irrelevant that one of the drivers (guess who) had an unblemished record and used his car only to commute between his office and his home at a prudent 5 mph below the speed limit, while the other was weaving at high speed down the wrong side of the road in a drunk and drugged haze, and had a long record of accidents and violations. The commendable past history of the one driver will not protect him in the head-on collision. If there are any survivors, the likelihood is that the driver of the intrinsically more crashworthy car will be the one to survive.

The preceding example of a head-on collision was a special case in that both cars had identical weights and both drivers were 30-year-old males. Neither car had an inherent advantage. In the absence of specific knowledge about the intrinsic crashworthiness of the two cars, each driver would be expected to have the same fatality risk in the crash. In most head-on collisions, the two cars have different weights, and their drivers are not necessarily the same age. Still, it is possible to predict the <u>expected</u> fatality risk for each driver in a head-on collision between these two cars, as a function of the weights of the two cars and the age and sex of each driver. Weight, age and sex are important because the lighter car experiences a greater velocity change than the heavier car, and an older/female driver is more vulnerable to injury than a younger/male driver. The fact that age and sex are also correlated with crash-proneness is irrelevant here, because the attempt is to predict the relative fatality risk of each driver, <u>given</u> that the two cars have already collided, head-on.

The expected fatality risk of each driver is calibrated from the accident data by a logistic regression. Regression coefficients vary slightly, depending on the calibration data set, but the following pair of regression formulas is typical for head-on collisions in which both drivers are belted. The expected fatality risk for driver 1 is

$$\frac{\exp[.616 - 5.427(\log W_1 - \log W_2) + .0531(A_1-A_2) + .34(F_1-F_2)]}{1 + \exp[.616 - 5.427(\log W_1 - \log W_2) + .0531(A_1-A_2) + .34(F_1-F_2)]}$$

where $W_1$ is the curb weight of car 1, $A_1$ is the age of driver 1 and $F_1$ is 1 if driver 1 is female, 0 if male. The expected fatality risk for driver 2 is

$$\frac{\exp[.616 + 5.427(\log W_1 - \log W_2) - .0531(A_1-A_2) - .34(F_1-F_2)]}{1 + \exp[.616 + 5.427(\log W_1 - \log W_2) - .0531(A_1-A_2) - .34(F_1-F_2)]}$$

These formulas, as stated above, measure the relative vulnerability to fatal injury of the two drivers, given that their cars had a head-on collision not the propensity of cars to get involved in head-on collisions. For example, given 100 fatal head-on collisions between 3000 pound cars driven by belted, 20-year-old males and 2500 pound cars driven by belted, 50-year-old females, the formulas predict 9 deaths among the young males in the heavier cars and 97 deaths among the older females in the lighter cars (for a total of 106 fatalities in the 100 collisions, since some of them resulted in fatalities to both drivers).

Cars with average crashworthiness capabilities will experience an actual number of fatalities very close to what is predicted by these formulas, which are calibrated from the collision experience of production vehicles. If a particular group of cars, however, consistently experiences more driver fatalities than expected in their head-on collisions, then it has to be concluded, based on the empirical evidence, that this group of cars is less crashworthy than the average car of similar mass. In the preceding example, if the 3000 pound cars with the young male drivers had 8, 9 or 10 deaths in the 100 crashes, they are doing about as well as expected, but if they had 30 deaths, they are less crashworthy than the 2500 pound cars with the older female drivers.

More generally, given a set of head-on collisions between one group of cars A and another group of cars B, it is possible to compare the crashworthiness of the two groups. The cars in group A are less crashworthy in head-on collisions than the cars in group B if the actual number of driver fatalities in group A is higher than the expected number of fatalities in the collisions, given the weight, driver ages, etc. in groups A and B. The actual fatalities and expected probabilities of fatality are summed for over all the crashes for groups A and B, as follows:

Head-On Collisions between Groups A and B

| | Car Group A | Car Group B |
|---|---|---|
| Actual fatalities | 100 | 60 |
| Expected fatalities | 91.8 | 68.2 |

To the extent that the cars in group A (in this example) are, on the average, slightly smaller than those in group B, more fatalities are expected in A than in B. If the actual and expected fatalities had been equal, groups A and B would have been judged equally crashworthy. In fact, group A performed slightly worse than expected. There were more fatalities than expected in A and fewer than expected in B. The increase in fatality risk for A relative to B is

$$[\ (100/91.8)\ /\ (60/68.2)\ ] - 1 = 23.8 \text{ percent}$$

Conversely, the fatality reduction for B relative to A is

1 - [ (60/68.2) / (100/91.8) ] = 19.2 percent

In the central analyses of this report, group A is a set of passenger cars with "poor" NCAP scores and group B is a set of cars with "acceptable" NCAP scores. In the actual head-on collisions between group A cars and group B cars, do the cars with poor NCAP scores have significantly more driver fatalities than would be expected? The analyses will measure the reduction in fatality risk, in actual head-on collisions, for a car with good NCAP scores relative to a car with poor NCAP scores. There are several methods to define "poor" and "acceptable" NCAP performance - e.g., based on a single NCAP parameter (chest g's, HIC or femur load), or based on a composite of these parameters. How big is the fatality reduction, for the "acceptable" vs. the "poor" cars, by each method? Other analyses will measure the overall reduction in fatality risk, for belted drivers in head-on collisions, since model year 1979, when NCAP testing began, until 1991, the latest model year for which substantial accident data were available as of mid 1993.

Two studies utilized the special advantages of head-on collisions for isolating crashworthiness differences between cars. Zador, Jones and Ginsburg analyzed the relative fatality risk of the two drivers in a fatal head-on collision and, even with the quite limited data on belted drivers available in the 1975-83 FARS, found some significant correlations between NCAP scores and fatality risk [30]. NHTSA's 1988 Evaluation of Occupant Protection in Frontal Interior Impact also studied head-on collisions, but concentrated on the unrestrained driver [19], pp. 111-140. After significant gains during the late 1960's, little net improvement in frontal crashworthiness was found for the unrestrained driver during model years 1970-84. Since these two studies were published, there has been a vast increase in FARS cases involving belted drivers, permitting detailed analyses of the crashworthiness of passenger cars in head-on collisions for belted drivers.

### 1.4 Some preliminary caveats

While head-on collisions, as reported in the Fatal Accident Reporting System, have many advantages for correlational analyses with NCAP results, it must be pointed out that many of these head-on collisions do not come close to resembling an NCAP test. FARS data can be used to distinguish head-on collisions from other crashes, but they currently do not identify many important details about the collisions, such as the impact speeds, the exact alignment of the vehicles, the height and weight of the drivers, or the specific body region with fatal lesions. All NCAP tests are 35 mph impacts straight ahead into a flat barrier, with contact over the entire front of the car, which is regarded equivalent to a perfectly aligned head-on collision of two identical cars, each travelling 35 mph. The driver dummies simulate 50th percentile males. NCAP test results are limited to three body regions (head, chest, femur).

The FARS sample, on the other hand, includes the full range of closing speeds that may occur on the highway, and the cars, although hitting front-to-front, may be aligned at an angle, and with small or substantial offset. The drivers may be any height or weight, and may have adjusted the seat forward or backward as they wish. Many fatal lesions are in the neck or abdomen: body regions not specifically tested in NCAP. It is not possible, with FARS, to single out those head-on collisions that come really close to an NCAP test. As a consequence, it is inappropriate to expect perfect correlation between NCAP test results and actual fatality risk in the FARS sample. Moreover, if there is <u>any</u> significant correlation between the two, it suggests that the NCAP scores say something about actual crashworthiness in a range of crashes that goes far beyond the specific type tested in NCAP.

None of the analyses of this report are conducted <u>at the make-model level</u>. There are not nearly enough head-on crashes with belted drivers to compute a fatality risk index by make-model and to compare this index with NCAP scores, by make-model. Thus, the analyses will indicate whether, <u>on the whole</u>, cars with poor NCAP scores have higher-than-average fatality risk in head-on collisions, but they will not identify any <u>specific</u> make-model (with or without poor NCAP scores) as being significantly less safe than the average car.

Since the accident data in this study are limited to head-on collisions between two cars with belted drivers, the correlations found here do not necessarily extend to other types of frontal impacts, such as collisions with fixed objects or trucks, let alone side impacts, rear impacts, rollovers or crashes involving drivers who do not wear safety belts.

# CHAPTER 2

## AN ACCIDENT DATA FILE WITH NCAP INFORMATION

Most of the analyses of this report examine head-on collisions between two passenger cars with belted drivers. Two groups of cars are selected, based on their NCAP scores. The collisions between cars of the two groups are examined, and the actual number of driver fatalities in each group is compared to the expected number, given each car's curb weight and each driver's age and sex. Thus, the type of data needed for the analysis, ideally, would be a file of actual head-on collisions, with both drivers belted, resulting in a fatality to at least one of the drivers, indicating the age and sex of each driver, the curb weight of each vehicle, and the HIC, chest g's and femur load that were recorded for the driver dummy when that vehicle was tested in NCAP.

NHTSA's Fatal Accident Reporting System (FARS) contains a record of every fatal crash in the United States since 1975. FARS data identify what crashes were head-on collisions; indicate the age, sex and belt use of each driver; and identify the vehicles by their Vehicle Identification Numbers (VIN). However, FARS data, themselves, do not include an accurate measure of curb weight or any information about NCAP results for the vehicles involved in the crash. Accurate curb weights are indispensable, because the relative fatality risk for two vehicles in a head-on collision is so sensitive to the relative weight (as evidenced by the coefficient of 5.427 in the formulas of Section 1.3). This chapter describes how the VIN and other vehicle codes are used to link FARS with other data files - the R. L. Polk National Vehicle Population Profile and the NHTSA file of NCAP test results - so that accurate curb weights and NCAP scores can be appended to the accident data.

## 2.1 Initial FARS data reduction

At the time of this study, FARS data were available through mid 1992 [10], [11], [12]. The type of crash of specific interest for this analysis is a head-on collision between two passenger cars resulting in a fatality to at least one of the drivers. Although light trucks have been tested in NCAP since model year 1983, collisions between a car and a light truck, or between two light trucks were not included in the study, because of problems in obtaining accurate weight information on trucks, and also because the samples of crashes involving belted truck drivers were insufficient.

The model years included in the study should, at least, range from 1979, the first year of NCAP testing, through 1991, the last year before air bags became the predominant type of occupant protection. In the initial data reduction, cars of model years 1976-78 were also included, because the designs from those model years sometimes carried over into the NCAP era. Cars of model year 1975 and earlier were excluded because they usually had different belt systems from later models (ignition interlock or separate lap and shoulder belts).

From 1975 through mid 1992, FARS contains 1,006 records of head-on collisions involving two passenger cars of model years 1976-91, fatal to at least one driver, in which both drivers wore safety belts (2,012 cars). A 2-vehicle file is designed, with one record for each collision, containing information on vehicle no. 1 and its driver and on vehicle no. 2 and its driver. A "head-on" collision has to be a crash involving exactly two vehicles (VE_FORMS = 2); both vehicles have to be passenger cars (BODY_TYP 1-9); both have to have frontal damage (IMPACT2 = 11, 12 or 1); the "most harmful event" for each vehicle has to be a collision with another motor vehicle, in transport or in "other roadway" (codes 12-13; prior to 1979, this variable was not defined on FARS, so it is not used as a filter); for both drivers, FARS must record their age (and it must be in the range of 14-98) and sex. A driver is "belted" if either a manual or an automatic belt was used, according to FARS (MAN_REST = 1, 2 or 8 or AUT_REST = 1 in 1975-90; REST_USE = 1, 2 or 8 in 1991-92). Cars with air bags are included in the study only if drivers wore their belts.

There are questions about the accuracy and completeness of FARS belt-use data, which are mostly based on information in police reports. Officers are usually not present at the scene at the moment of the crash and must rely on statements by survivors and witnesses, physical evidence and judgment. Belt use is coded "unknown" for 18 percent of the drivers and is not necessarily accurate in the remaining cases. The greatest concern is in States with buckle-up laws, where belt use may be overreported by survivors to escape penalties. Based on 1983-92 trends in reported belt use among FARS fatalities vs. actual belt use observed on the road, NHTSA believes that the belt use of the fatally injured occupants, at least, is quite accurately reported in FARS. In many cases, these fatally injured occupants may not have been moved between the time of the crash and arrival of police, allowing easy identification of belt use. While the belt use of survivors may not be as accurately reported as for fatalities, at least there is no reason to suspect that reported belt use is in any way confounded with a vehicle's NCAP performance.

Before FARS data can be linked to the Polk or NCAP files by make-model, it is obviously necessary to have accurate make-model information on FARS. The make-model codes on the basic FARS file, which are manually entered and not decoded from the VIN, are not suitable for the analysis. Many cars are miscoded, especially where model names are easily confused (e.g., Cutlass, Cutlass Ciera, Cutlass Calais, Cutlass Supreme). Also, the model is often coded "unknown" when there is a valid, decodable VIN. The basic FARS file contains a 3-digit VINA_MOD code, which is obtained from the VIN, accurate, and suitable for linking FARS to Polk data, which contain a similar code with the name SERS_ABR. The VINA_MOD code, however, is not well-suited for linking FARS with other data files, such as NCAP results, which do not have a corresponding code.

A program was written to decode VINs and define make-models, using approximately the same 4 digit numeric scheme as in the basic FARS, superseding the values in the original FARS data. Cases in which one or both of the vehicle records has a blank or nonvalid VIN are deleted. To prevent excessive deletions, however, one set of "minor" errors in the VIN is permitted: if a field which must have a numeric code has alphabetic O the program "corrects" it to numeric 0; likewise I to 1, Z to 2, S to 5, G to 6 and B to 8 - and vice versa if look-alike numeric codes appear in an alphabetic field. The model year decoded from the VIN supersedes the model year code on FARS. The exclusion of cases with unknown or nonvalid VINs reduces the file to 934 collisions (1,868 cars).

The make-model code, by itself, is not sufficient for linking FARS to the file of NCAP test results. The same make-model code may be used for two quite different cars (e.g., 1979 and 1991 Honda Civic), sometimes even in the same year (e.g., 1988 Buick LeSabre H-body sedan or a B-body station wagon). Conversely, the same or quite similar cars can have different model codes (e.g., Dodge Colt and Plymouth Colt). As will be seen, nearly identical make-models in the accident data will sometimes be linked to the same NCAP test. Based on the

VIN, passenger cars of the 1976-91 era were classified into about 300 car groups with shared body platforms - e.g., all GM N-body cars. A 4-digit code for the car group supplements the make-model code. When a car gets a major redesign, a new car group is defined - e.g. Toyota Celica in 1976, 1978 and 1987. "Shared body platform" generally means the same wheelbase, track width and drive system (front-wheel or rear-wheel). Not all cars in a car group are nearly identical "corporate cousins." Sometimes, they may vary by several hundred pounds in weight or have easily visible differences in structure or interior layout (e.g., 1983 Cadillac Seville and Eldorado). These differences will be discussed further in Section 2.4.

2.2 Curb weight data from Polk files

The single most important safety factor in a head-on collision is the relative weight of the two cars. As stated above, a 1 percent weight advantage for one of the cars translates into more than a 5 percent reduction of expected fatality risk for that car. By the same token, a 1 percent error in the weights of a group of cars can throw off their expected fatality risk by 5 percent. Vehicle weights should be as accurate as possible and biases must be avoided.

The weight variable in the FARS data, VIN_WGT, is not usable for several reasons. It lists the "shipping weight" (unoccupied car without fuel or other fluids) of some cars in some years and, arbitrarily, the "curb weight" (unoccupied car with fuel and other fluids) at other times, especially after 1981: about a 100 pound discrepancy [19], p. 118. It is defined at the make-model level (which, itself, is inaccurately reported in the basic FARS) and does not take into account the extra weight of optional engines, station wagon bodies, etc. Curb weights from Automotive News Almanacs [2] should also be avoided in this study (although, at least, they do not have severe year-to-year biases). In general, the Almanac lists only one or two curb weights for a specific make-model in a given year and does not indicate exactly which engine and level of decor (e.g., L, GL, or LX) this weight applies to.

The most accurate listings of curb weights are the official Automobile Specifications supplied by the manufacturers through the American Automobile Manufacturers Association (then called the Motor Vehicle Manufacturers Association, or MVMA). The books list the baseline curb weight of every make-model and level of decor plus the incremental weight of each optional engine and other equipment. The vast amount of data in these hard-copy files has been encoded in the tapes of R. L. Polk's National Vehicle Population Profile [21], which lists a curb weight for each combination of make, model year, series (model and decor level, expressed in the 3-digit SERS_ABR code), body style, engine code and, possibly, fuel code. These Polk weights are highly accurate for a car which contains no equipment beyond that which is standard in a particular make-model and subseries. (The Polk file, however, does not include curb weights for light trucks. Although curb-weight information may be available from other sources, its utility is uncertain, because a truck's weight may be substantially augmented by cargo.) A program was written to define variables on FARS that mimic those on the Polk files and merge the two files. Make and model year are already on FARS. So is the 3 digit series code (called VINA_MOD on FARS and, if missing, obtained manually by analyzing the VIN). The body style, engine and fuel codes are derived from the VIN. After the initial computer merge, and after a manual search through the MVMA specification books in those cases where the Polk file had missing weights, it was possible to identify a curb weight for both vehicles in 926 head-on collisions (1852 vehicles).

## 2.3 Adjusting Polk weights based on actual weight measurements

Although the Polk weights are detailed and based on authoritative sources, they are still not the weights of actual cars on the road. Before the Polk weights are accepted at face value, it is wise to compare them to measured weights of some actual cars. Moreover, it is likely that the actual cars would be heavier, because most cars contain at least some optional equipment such as air conditioning, radios, etc. Those items are not included in the Polk weight unless they are standard equipment on a particular make-model and subseries.

NHTSA's data bank of compliance tests for new cars is a reliable source of actual curb weights. Since 1968, NHTSA has performed hundreds of compliance tests each year, under contract at test laboratories, checking selected new vehicles or safety equipment to see if they meet certain Federal Motor Vehicle Safety Standards (FMVSS) or other regulations. A subset of the compliance tests involve FMVSS where the measurement of curb weight is an integral part of the test. For example, it is essential to know the curb weight when testing Roof Crush Resistance (FMVSS 216), because 150 percent of the curb weight (or 5000 pounds, whichever is less) has to be applied to the roof structure. The FMVSS and other NHTSA regulations whose compliance tests always include a measurement of curb weight [3] are FMVSS 105 (hydraulic brake systems), 110 (tire selection and rims), 204 (steering control rearward displacement), 208 (occupant crash protection), 212 (windshield mounting), 214 (side door strength), FMVSS 215/Part 581 (bumpers), 216 (roof crush resistance), 219 (windshield zone intrusion), 301 (fuel system integrity) and Part 575 (consumer information regulations). In addition, the curb weight is sometimes measured and included in test reports for FMVSS 124 (accelerator control systems) and 207 (seating systems), even though it is not an essential part of the compliance test. Twenty contractors have worked on compliance tests which include weighing the cars.

NHTSA does not test every make-model for every FMVSS every year, but operates according to a sampling plan. In general, new FMVSS and new or redesigned make-models are tested intensively, while existing FMVSS and carryover make-models are spot-checked on a cyclical basis. On the average, 100 curb weights are measured each year, but not for 100 different make models; 2-4 cars of the same make-model may be tested one year, especially if this is a new or redesigned model. Moreover, a single car may be tested for several different FMVSS by one or more test laboratories. For example, a nondestructive test (FMVSS 105) may be followed by a crash test which produces data on several FMVSS (204, 212, 219 and 301). The car may be weighed several times or it may be weighed just once and the same weight entered in more than one test report.

An important feature of the compliance tests is that they are performed on "real" cars. Contractors go to nearby retail dealerships and buy cars off the lot, in all likelihood equipped with the types of options consumers usually want for that make-model (automatic transmission, air conditioning, radios, fancy decor, popular engines, etc.). Sometimes the contractor gets a car more "loaded" than usual and sometimes more "stripped," but it averages out to the typical car of that type.

In all, as of July 1990, NHTSA compliance test reports furnish 2006 curb weight data points for passenger cars of model years 1968-89. The curb weight data were manually retrieved from compliance test reports and encoded along with the VINs of the cars, the number of the FMVSS being tested, the name of the contractor, etc. The VIN decode program that was developed for FARS data (see above) was also used on the compliance test file to define the make-model and the car group. VINs were further decoded to make the compliance test file compatible for merging with Polk's National Vehicle Population Profile - by defining the 3-digit series code (VINA_MOD), body style, engine code and, possibly, fuel code. The merged file contains 1966 records of passenger cars with an "actual" curb weight measured by the contractor and a "normative" or "prescriptive" curb weight from the Polk file (40 cases in the compliance test file were lost due to errors in the VIN or because they were rare cars for which Polk has no corresponding record). The 1966 records comprise 1840 distinct weight measurements (126 cases are entries of a previous weight measurement into a 2nd or 3rd compliance test report), 1563 distinct vehicles (277 cases are 2nd or 3rd weighings of the same vehicle) and 1192 distinct combinations of the merge variables (make-model, MY, series code, body style, engine - as stated above, NHTSA often tests more than one car of a particular type).

The comparison of actual vs. Polk weights would be simple if only the "actual" weights themselves were completely accurate or, at worst, imprecise only to the extent of tolerances allowed in scales. In fact, a few of the weights are inaccurate, as evidenced, for example, by discrepancies as high as 350 pounds in two weighings of the same car by different contractors and 100 pounds in two weighings of the same car by the same contractor on different FMVSS. A case-by-case review was conducted to eliminate records in which the measured weight was suspected of inaccuracy. The review took into account the FMVSS being tested; the contractor; the size of the discrepancy between the measured and the Polk weight; and, when a car was weighed more than once, the discrepancy between the various "actual" weights. The 126 records which were merely entries of a previous weight measurement into a 2nd or 3rd compliance test report were also deleted, since they provide no new information.

Some FMVSS were associated with fairly evident biases in the curb weight measurements. For example, with FMVSS 124 and 207, where measurement of curb weight is not really needed for performing the test, the weights were biased upward and had to be discarded in most cases. In general, cases where the measured weight was more than 10 percent above or 5 percent below the Polk weight were discarded unless they demonstrated a problem with the Polk weight (in which case the Polk weight was corrected, based on backup sources). For FMVSS-contractor combinations where the weight seemed to be biased in a particular direction, discrepancies of more than 8 percent above or 3 percent below Polk weight were not tolerated. If the same car was weighed twice and there was more than 3 percent discrepancy between the weights, the less plausible measurement was discarded. In all, 114 records were eliminated in the case-by-case review, leaving a file of 1726 distinct weight measurements.

The 1726 weight measurements were aggregated into 61 make-model groups. Most of the 61 groups had 10 or more weight measurements. The simple arithmetic percentage average of the excess of "actual" weight over Polk weight was calculated for each group and shown in Table 2-1. On the average, actual curb weights are 2 to 3 percent higher than those on the Polk files - i.e., 70 to 105 pounds for a 3500 pound car, which seems about right for optional equipment included in the typical car. The average excess ranged from 0.4 in Mazdas to 5.3 percent in GM X-body cars. In general, domestic cars of the 1970's, which were usually sold with automatic transmission and air conditioning as <u>optional</u> equipment, had the highest excess of actual weight over Polk weight.

TABLE 2-1

EXCESS OF ACTUAL CURB WEIGHTS OVER POLK WEIGHTS BY MAKE-MODEL GROUP

(actual weights from NHTSA compliance test reports)

| **Make-Model Group** | **N of Test Reports** | **Avg. Excess of Actual over Polk Weight (%)** |
|---|---|---|
| AMC older models | 23 | 1.95 |
| AMC newer models | 47 | 2.70 |
| Chrysler Dart/Valiant | 22 | 2.53 |
| Chrysler Belvedere/Coronet | 15 | 4.03 |
| Chrysler old fullsized | 47 | 3.65 |
| Chrysler Cordoba/Charger | 21 | 3.58 |
| Chrysler Aspen/Volare/later RWD | 46 | 3.34 |
| Chrysler Omni/Horizon | 17 | 2.23 |
| Chrysler K car derivs. | 46 | 2.51 |
| Ford LTD till 78 | 23 | 3.01 |
| Ford old luxury cars | 24 | 2.36 |
| Ford Maverick | 15 | 4.53 |
| Ford Pinto | 17 | 1.51 |
| Ford Torino | 26 | 3.86 |
| Ford Mustang II | 5 | 3.48 |
| Ford Granada | 19 | 2.99 |
| Ford Fairmont | 23 | 2.59 |
| Ford Mustang 79- | 16 | 3.16 |
| Ford new fullsized 79- | 17 | 1.75 |
| Ford Escort | 19 | 1.58 |
| Ford new midsized RWD | 17 | 1.75 |
| Ford new midsized FWD | 16 | 3.55 |
| GM Corvette till 82 | 8 | 2.36 |
| GM Nova RWD | 32 | 2.23 |
| GM Camaro till 81 | 18 | 4.09 |
| GM low-priced fullsized till 76 | 12 | 2.58 |
| GM med-priced fullsized till 76 | 15 | 3.55 |
| GM luxury till 76 | 39 | 4.04 |
| GM Vega | 14 | 4.10 |
| GM midsized 116" wb 73-77 | 11 | 4.09 |
| GM midsized 112" wb 73-77 | 22 | 3.27 |
| GM Monte Carlo 73-77 | 10 | 4.74 |
| GM Monza 75-80 | 22 | 2.76 |

TABLE 2-1 (continued)

EXCESS OF ACTUAL CURB WEIGHTS OVER POLK WEIGHTS BY MAKE-MODEL GROUP

(actual weights from NHTSA compliance test reports)

| **Make-Model Group** | **N of Test Reports** | **Avg. Excess of Actual over Polk Weight (%)** |
|---|---|---|
| GM Chevette | 29 | 3.22 |
| GM downsized big cars RWD 77- | 29 | 3.35 |
| GM downsized luxury RWD 77- | 10 | 2.23 |
| GM downsized intermeds RWD 78- | 22 | 2.66 |
| GM Mt Carlo/Supreme G 78-89 | 20 | 2.90 |
| GM X cars | 21 | 5.33 |
| GM J cars | 9 | 4.36 |
| GM Camaro/Corvette 82- | 10 | 2.38 |
| GM midsized FWD 82- | 30 | 3.25 |
| GM big/luxury FWD 79- | 26 | 0.85 |
| | | |
| VW rear engine | 17 | 0.93 |
| VW front engine | 72 | 1.17 |
| European sports cars | 31 | 2.22 |
| European luxury cars | 71 | 1.61 |
| European economy cars | 89 | 2.54 |
| | | |
| Nissan midsized RWD till 81 & sports cars RWD till 79 | 18 | 4.09 |
| Nissan economy RWD | 8 | 1.72 |
| Nissan sports cars RWD 79- | 19 | 1.80 |
| Nissan FWD | 45 | 1.54 |
| Hondas of the 70's | 13 | 1.88 |
| Hondas of the 80's | 42 | 1.60 |
| Mazda | 41 | 0.37 |
| Subaru | 33 | 0.83 |
| Toyota Corolla RWD | 17 | 3.24 |
| Toyota Celica/RWD | 36 | 1.40 |
| Toyota FWD | 32 | 2.31 |
| Mitsubishi | 76 | 1.58 |
| | | |
| All other cars | 58 | 1.27 |

Finally, each of the 1,852 Polk weights on the head-on collision file were adjusted upwards by the percentage shown in Table 2-1, depending on the make-model group to which the car belonged. The type of occupant protection system at the driver's position was decoded from each car's VIN, based on programs developed in NHTSA's evaluation of occupant protection [6]; 84 percent of the cars on the file had manual belts only, 3 percent had an air bag plus manual belts and 16 percent had some type of automatic belt.

## 2.4 A file of NCAP test results

An average of 30 passenger vehicles are tested each year. They are crashed into a rigid barrier at a target speed of 35 mph, which is 5 mph faster than the speed for compliance with Standards 204, 208, 212, 219 and 301. There are correctly restrained, instrumented 50th percentile male "Part 572" dummies at the driver and right front passenger seat locations.

NHTSA maintains a data file containing information about each NCAP test conducted since the program began with model year 1979. The information on the data file matches the listing of test results in several NHTSA publications [13], [14], [15], [25], [26]. The variables on the file include the make and model (written in plain English, not in a numerical code), the model year, the vehicle's body style and type of occupant protection (depicted by 2-digit codes) and the NCAP scores for the driver dummy: the Head Injury Criterion (HIC), chest g's (3 millisecond peak) and left and right femur loads (peak axial loads at knee). Some of the NCAP scores are missing in a few cases when there were operational problems with parts of the test instrumentation. The file includes 305 NCAP tests of passenger cars of model years 1979-91.

The original NCAP test file was modified to facilitate linkage with the accident data. The plain-English make and model descriptions were replaced by the pair of 4-digit numeric codes indicating the make-model and car-group, as defined in Section 2.1. The codes for body style and type of occupant protection were converted to the numeric codes defined on the accident file.

Although NCAP scores most accurately characterize the performance of the specific car that was tested, they may also apply, with some accuracy, to cars of the next several model years. The customary procedure in NCAP is to test a make-model with high sales-volume in the first model year of its existence, or as soon as possible thereafter. A given make-model is not retested until it is significantly redesigned. For example, models are retested after a change in the body platform or vehicle structure, a shift in the type of occupant protection (e.g., from manual belts only to air bags plus belts), or when manufacturers inform NHTSA that they have modified safety-related interior components in a way that might significantly change test results. Thus, an NCAP test result may be considered valid for subsequent model years up to the next significant redesign [26].

NHTSA staff reviewed the 1979-91 NCAP tests and determined the "end year" for each test: the last year before the car was discontinued or redesigned. NCAP test results are considered valid from the model year of the test vehicle to the "end year." Results of 305 NCAP tests are listed in [22], ordered by car-group, make-model and model year, indicating the test number, type of occupant protection, body style, *HIC*, *chest g's* and *femur load*; and the "end year."

## 2.5 Matching NCAP tests with FARS cases

The ideal matching of NCAP tests with the accident data would be a simple merge by make-model, model-year, type of occupant protection and body style. In other words, given a specific car involved in an actual head-on collision, if a basically identical car was tested in NCAP, the driver's HIC, chest g's and femur load are transcribed to the accident file. The problem is that thousands of different cars (make-model-model year-body style combinations) were sold during 1979-91, but only 305 cars were tested in NCAP. Many of the cars on FARS do not match up exactly with an NCAP case; there are only 12 head-on collisions in which both drivers were belted and both vehicles match up exactly with an NCAP case.

However, as noted above, NCAP test results are considered valid for several subsequent model years, until a car is redesigned [26]. Moreover, when two or more make-models, produced by the same manufacturer, not only share a body platform, but also have nearly identical interior and exterior components (e.g., Dodge Omni and Plymouth Horizon), a test for one of these models is considered valid for its "corporate cousins" [25]. These two extensions in the reach of NCAP results make it possible to match an NCAP test to a lot more accident cases. Perhaps there are yet other situations where NCAP test results could be accepted for somewhat dissimilar crash-involved cars, further extending the size of the accident sample that can be matched with NCAP. For example, the results for a 2-door car might be acceptable for a 4-door car of the same make-model, and vice-versa.

NHTSA staff reviewed each of the cars in the head-on collision file and identified the NCAP-tested car, if any, which most closely resembled it, based on four affinity factors. Each of the four affinity factors has several quality levels ranging from best (complete agreement of the NCAP car and the FARS car) to worst. The affinity factors and their quality levels are the following:

Body platform and make-model

4 (best) FARS and NCAP cars have the same make-model, body and chassis.

3 FARS and NCAP cars are "true corporate cousins" (identical body and chassis, as evidenced by equal wheelbase, weight and exterior dimensions). Example: Dodge Aries and Plymouth Reliant. Different nameplates suggest, at most, slight differences of interior components.

2 FARS and NCAP cars are built on the body platform, but are not true corporate cousins. Above the chassis, the cars are not the same, as evidenced by unequal weights, exterior dimensions, or appearance. Example: FWD Buick LeSabre and FWD Buick Electra (or Olds 98).

1 (worst) FARS and NCAP cars are built on different chassis, as evidenced by unequal wheelbase, but one chassis is basically a "stretch" version of the other, and the overall designs are similar. Example: RWD Olds 98 and Cadillac Fleetwood Brougham.

General model year range

A (best) The model year of the FARS car is within the range of applicable model years for the NCAP test - i.e., no earlier than the model year of the NCAP test vehicle and no later than the "end year."

B The model year of the FARS car precedes the model year of the "matching" NCAP test.

C (worst) The model year of the FARS car is later than the "end year" for the matching NCAP test.

Specific model year

Best The FARS and NCAP cars are the identical model year.

The FARS and NCAP model years are not identical, but differ by N years.

Body style

| | |
|---|---|
| Best | The FARS and NCAP cars have the same number of doors and exactly the same body style (sedan/coupe, hatchback, station wagon, convertible). |
| | The FARS and NCAP cars have the same number of doors and almost the same body style (one is a hatchback and the other is a sedan/coupe). |
| | The FARS and NCAP cars have the same number of doors, but different body styles (station wagon vs. sedan, convertible vs. coupe). |
| Worst | One is a 2-door and the other is a 4-door. |

The "best" NCAP match was identified case-by-case, based on staff discussions, rather than by an automated procedure. When there is no perfect match, but two or more choices among imperfect matches, the best choice depends on the specific make-model involved. For example, if the FARS case is a 4-door car, and the two NCAP tests for the same make-model are a 2-door car of the same model year and a 4-door car of a different model year, the best choice depends on whether, for this particular make-model, the difference between the 2-door and the 4-door version exceed the change in the 4-door version over time. In all cases, though, the FARS and NCAP cars had to have the same type of occupant protection.

Reference [9] lists every car on the FARS file (model year 1976-91 cars involved in head-on collisions where both drivers were belted) and, next to it, the NCAP test vehicle, if any, which was judged to be the best match. It exhibits, side by side, the make-model, model year and body style of the FARS and NCAP cars, illustrating how well (or poorly) they match. The MATCHLVL data field, a number followed by a letter, indicates the quality of the match according to the first two criteria: body platform/make-model and general model year range. For example, the first car on the FARS file, when it is ordered by car group and make-model, is a 1980 AMC Spirit 2-door hatchback with manual belts. The only NCAP test vehicle that comes close is a 1981 AMC Spirit 2-door hatchback with manual belts. These two cars match exactly on the car group, make-model, body style and type of occupant protection. However, the FARS model year (1980) is less than the NCAP model year (1981) and precedes the time span from the NCAP model year to the "end year" for that NCAP test (1981-83). Thus, the MATCHLVL is rated 4B: 4 because the make-model and car group match exactly, B because the FARS model year precedes the NCAP year. It should be noted that quite a few cars on FARS, such as the 1978 AMC Pacer, do not closely resemble <u>any</u> car tested in NCAP, and do not have an NCAP match.

MATCHLVL 3A and 4A may be considered especially important in the analyses. Here, the FARS and NCAP vehicles are of the same make-model or true corporate cousins, and the FARS model year is within the "valid" range of the NCAP test. As noted above, NHTSA has not asserted that NCAP test results can be extended to cars which only match an NCAP test at a lower level than that [25], [26]. The head-on collisions in which both cars match an NCAP test at the 3A or 4A level and, possibly, also match on number or doors, then, would seem to be the most natural data sets to look for correlations between NCAP scores and fatality risk. However, lower levels of matching, such as 2A or 4B, are not excluded from the data set at this time; in Chapter 3, these cases will be empirically tested for correlation between NCAP and fatality risk.

## 2.6 <u>Creation of the analysis file</u>

In all, there are 739 head-on collisions, involving 1,478 model year 1976-91 passenger cars, in which both drivers were belted and <u>both</u> cars match up <u>at any level</u> with an NCAP case. From the previous FARS file of 926 head-on collisions (1,852 vehicles), about 20 percent of the cases cannot be used in most of the analyses, because one or both of the crash-involved vehicles do not match up with any NCAP case. The following variables are defined for each vehicle and driver on the file:

- VIN
- Model year
- Car group (4 digit code derived from VIN)
- Make-model (4 digit code derived from VIN)
- Fatality outcome: "1" if the driver died; "0" if the driver survived
- Polk weight
- Curb weight (Polk weight escalated by correction factor)
- Driver age (has to be 14-98)
- Driver sex (has to be known)
- Body style (convertible, 2 dr coupe/sedan, 2 dr hatchback, 2 dr hardtop, 4 dr sedan, 4 dr hatchback, 4 dr hardtop, station wagon)
- Type of occupant protection (manual belts only, air bag plus manual belt, motorized belt, automatic 3-point belt, nonmotorized 2-point belt)
- Test number of the matching NCAP test car
- Model year of the matching NCAP test car
- Car group of the matching NCAP test car
- Make-model of the matching NCAP test car
- Body style of the matching NCAP test car
- "End year" for the matching NCAP test car
- Match level for the FARS-NCAP match
- HIC for the driver in the matching NCAP test car
- Chest g's for the driver in the matching NCAP test car
- Left femur load for the driver in the matching NCAP test car
- Right femur load for the driver in the matching NCAP test car

It should be noted that the HIC, chest g's and femur load numbers written on this file are those recorded on the driver dummy in the NCAP test vehicle during a 35 mph barrier crash and <u>not</u> those actually experienced by the driver of the crash-involved vehicle on FARS, which are, of course, unknown.

On the analysis file, the two vehicles in the collision are referred to as the "case" vehicle and the "other" vehicle, rather than vehicles "1" and "2." Each record in the original FARS file is written <u>twice</u> onto the analysis file: first with the original vehicle number 1 as the "case" vehicle and vehicle number 2 as the "other" vehicle; then with the original vehicle number 2 as the "case" vehicle and vehicle number 1 as the "other" vehicle. The concept here is that a head-on collision is essentially a symmetrical event; while FARS may call one of the vehicles "No. 1" and the other "No. 2" (arbitrarily, or based on pre-crash events that are no longer relevant to the analysis) it could just as well have reversed the order. Both vehicles have participated as "case" vehicles in a head-on collision. Thus the analysis file contains 1,852 records of head-on collisions, comprising 1,852 different vehicles (each of which appears twice on the file, once as the "case" vehicle and once as the "other" vehicle). The analyses will primarily deal with the subset of 1,478 collision records in which both the "case" and the "other" vehicle match up with an NCAP test.

# CHAPTER 3

## CORRELATION OF FATALITY RISK WITH INDIVIDUAL NCAP PARAMETERS

The initial investigation of the relationship between NCAP test performance and fatality risk in actual head-on collisions is based on regression analyses of fatality risk by HIC, chest g's, femur load, vehicle weight, driver age and sex. Specifically, since there are two cars in a head-on collision, the fatality risk for the driver of the case vehicle is modeled as a function of the relative HIC scores for the two vehicles on the NCAP test, the relative chest g's, etc. Does fatality risk increase significantly with higher NCAP scores?

### 3.1 Analysis objective

In Chapter 2, a file of head-on collisions was created, including 1,478 records in which both vehicles could be "matched" to an NCAP test. These 1,478 cases are the raw material for the regression analyses, because they contain all the necessary variables (weight, age, sex and NCAP scores for both cars). The quality of the matches, however, varied in the 1,478 cases. Sometimes, the crash-involved car and the "matching" NCAP car had nothing more in common than a similar body platform, while at other times they were essentially identical vehicles.

The principal task of this chapter is to identify a subset of the 1,478 cases which best indicates the relationship between NCAP parameters and fatality risk. That involves a trade-off between sample size and the quality of the matches. The full data set has the largest sample size, but the poor quality of some of the matches could obscure the relationship: the NCAP scores assigned to some of the crash-involved vehicles may simply be inappropriate for those cars. On the other hand, demanding too close a match between the FARS and the NCAP cars could reduce the sample size to the point where significant correlation is unlikely. The approach of this chapter is to try out the regression analysis on various subsets of the data file, defined by how closely the FARS and NCAP cases match, and to find a subset which yields excellent correlations and is also intuitively reasonable. Based on NHTSA statements on the applicability of NCAP tests, the best results might be expected when the FARS and NCAP vehicles are of the same make-model or true corporate cousins, the FARS model year is within the "valid" range of the NCAP test, and, possibly, the FARS and NCAP vehicles have the same number of doors [25], [26]. However, regressions will also be performed on a variety of other subsets of the data file.

### 3.2 Regression analysis procedure

The method for calibrating fatality risk as a function of relative NCAP scores, weight, age and sex is logistic regression on disaggregate data, using maximum likelihood principles [18]. Logistic regression uses a large number of individual observations of success (case driver survival) or failure (driver fatality) given different actual combinations of the independent variables to predict the driver's probability of fatality under any hypothetical combination of the independent variables. Specifically, the model generates an equation which expresses the log-odds of a fatality as a linear function of the independent variables.

However, the scores for HIC, chest g's and femur load, as actually measured on the NCAP tests, are not well-suited as independent variables in a regression analysis. Their distributions are skewed in one direction - e.g., there are a few tests with extremely high HIC (above 2000). The extreme values of HIC would have excessive weight in any regression analysis and that would seriously distort the calibration of fatality risk as a function of HIC. For the regression analyses to work, each actual injury criterion needs to be transformed to variables with a normal distribution - or, at least, to a variable with a symmetric distribution that has a wide peak and narrow tails. Several procedures exist for normalizing variables; one of these was especially appropriate for the present regression analyses. The actual NCAP results for the driver dummy were transformed to logistic injury probability functions for each body region:

HEADINJ = 1 / [1 + exp(5.02 - .00351 HIC)]
CHESTINJ = 1 / [1 + exp(5.55 - .06930 chest g's)]
LFEMURINJ = 1 / [1 + exp(7.59 - .00294 left femur load)]
RFEMURINJ = 1 / [1 + exp(7.59 - .00294 right femur load)]

These functions were developed by General Motors and others, based on empirical testing with human surrogates and dummies [13], [27], [29]. They measure the probability of life-threatening or fatal head and chest injury (4-6 on the Abbreviated Injury Scale [1]) and severe leg injury (AIS $\geq$ 3), as a function of HIC, chest g's and femur load. The logistic injury probabilities correspond to actual NCAP scores as follows:

| Logistic Injury Probability | HIC | Chest g's | Femur Load |
|---|---|---|---|
| .001 | | | 232 |
| .01 | 121 | | 1019 |
| .02 | 321 | 24 | 1258 |
| .05 | 591 | 38 | 1580 |
| .10 | 804 | 48 | 1834 |
| .20 | 1035 | 60 | 2110 |
| .30 | 1189 | 68 | 2293 |
| .40 | 1315 | 74 | 2444 |
| .50 | 1430 | 80 | 2582 |
| .60 | 1546 | 86 | 2720 |
| .70 | 1672 | 92 | 2870 |
| .80 | 1825 | 100 | 3053 |
| .90 | 2056 | 112 | |
| .95 | 2269 | | |
| .99 | 2739 | | |
| .999 | 3398 | | |

The original NCAP scores have been transformed into measures of relative injury risk that can readily be used in regression analyses. The transformed variables for the different body regions can be tested for correlation, added to one another, and combined into weighted or unweighted averages. The logistic injury probabilities compress high values of the original scores into a narrow band and eliminate the skew to the high side present in the original scores - e.g., all HIC over 2056 are compressed into a range from .90 to 1. The low values of the original scores (e.g., HIC below 800, chest g's below 48) are also compressed into a narrow band. The mid-ranges of the original scores, which are critical for differentiating between acceptable and poor safety performance, occupy a wide middle band (.1 to .9) of the logistic injury probability distribution. The logistic transformation acts like a lens that magnifies differences in the middle of the range, but diminishes them at the low and high ends. The resulting distributions, as desired, have short tails and wide peaks.

As explained in Section 2.6, each record on the analysis file contains information on the two vehicles in a head-on collision, and their drivers: the "case" vehicle and the "other" vehicle. This information is now supplemented with logistic injury probabilities derived from the NCAP tests. CHIC, CCG, CLFEM and CRFEM are the logistic injury probability scores for HIC, chest g's, left and right femur load for the driver of the <u>case</u> vehicle. OHIC, OCG, OLFEM and ORFEM are the corresponding scores for the driver of the <u>other</u> vehicle.

In the initial regression model, each of the 1,478 head-on collision records in which both vehicles match up with NCAP tests becomes a data point. The dependent variable is the actual outcome of the crash for the driver of the case vehicle, equaling 1 for a fatality and 0 for a survivor. There are 6 independent variables W, A, S, DELHIC, DELCG and DELFEM, all of which are calculated for the case vehicle relative to the other vehicle, as follows:

W is the difference of the natural logs of the curb weight of the case vehicle and the other vehicle (NHTSA's Evaluation of Frontal Interior Impact [19], pp. 138-140, showed exceptionally good fit when the weight variable is expressed in this form).

A is the simple arithmetic difference of the ages of the two drivers, the case driver's age minus the other driver's age (with 14 or 15 year old drivers counted as 16 year olds). Evans [8] showed exceptionally good fit when the age variable is expressed in this form.

S is 0 if both drivers were males or both were females; -1 if the driver of the case vehicle was male and the other, female; and +1 if the driver of the case vehicle was female and the other, male

DELHIC = CHIC - OHIC

DELCG = CCG - OCG

DELFEM = (CLFEM + CRFEM) - (OLFEM + ORFEM) if all 4 of these are known,
= 2 CLFEM - (OLFEM + ORFEM) if CRFEM is unknown, the others known,
= 2 CRFEM - (OLFEM + ORFEM) if CLFEM is unknown, the others known,
etc.

W should have negative correlation with the dependent variable (the heavier the case vehicle the lower the fatality risk for its driver). A and S should have positive correlation with the dependent variable (older drivers and female drivers are more vulnerable to fatal injury). DELHIC, DELCG and DELFEM should also have positive correlation with the dependent variable: high HIC in the case vehicle would be associated with high fatality risk in the case vehicle. High HIC in the other vehicle would be associated with high fatality risk in the other vehicle and, since most head-on crashes kill only one of the drivers, low fatality risk in the case vehicle.

### 3.3 The initial regression - including all NCAP matches

There are 1478 data points in the full data set of head-on collisions where both drivers were belted and both cars could be matched to an NCAP test at any level of match quality. After excluding 86 data points where DELHIC, DELCG or DELFEM could not be defined because of missing data in the NCAP tests, there are 1392 cases available for the initial regression. The regression coefficients and their associated statistical significance levels are:

Initial Regression - All NCAP Matches

| | Reg. Coeff. | Chi Square | Stat. Sig.? | Partial Corr. |
|---|---|---|---|---|
| INTERCEPT | .5555 | 51.45 | RR | |
| W (car weight) | -5.5296 | 281.72 | RR | -.384 |
| A (driver age) | .0532 | 242.35 | RR | .356 |
| S (driver sex) | .3400 | 9.06 | RR | .061 |
| DELHIC | .0225 | .01 | | |
| DELCG | 1.8400 | 6.62 | R | .049 |
| DELFEM | .3975 | .77 | | |

In the preceding table, a Chi-square ($\chi^2$) statistic is calculated for each regression coefficient, to see if the variable makes a statistically significant contribution to fatality risk. If $\chi^2$ is greater than 6.64 and the coefficient has the "right" sign (as discussed above), the variable has a highly significant association with fatality risk (two-sided alpha less than .01), even after controlling for the other variables, as indicated by an "RR" in the statistical significance column. If $\chi^2$ is between 3.84 and 6.64 and the coefficient has the "right" sign the variable has a significant association with fatality risk (two-sided alpha between .05 and .01), as indicated by an "R." If the regression coefficient is nonsignificant, the statistical significance column is left blank. The partial correlation coefficient measures the direction and relative strength of the contribution of a variable to the prediction of fatality risk (if $\chi^2$ is less than 2, this coefficient is set to zero).

In this initial regression, knowledge of NCAP chest g's significantly enhanced the ability to predict whether the driver of the case vehicle was killed in a head-on collision ($\chi^2 = 6.62$, $p < .05$). If all other factors such as car weight, driver age, etc. are equal, the driver of the car with the lower NCAP chest g's has a significantly better chance of survival than the other driver. Of course, chest g's are far less important than relative car weight and driver age as predictors of fatality risk in head-on collisions, as demonstrated by the much larger $\chi^2$ and partial correlation coefficients for W and A. But chest g's have almost as much influence on relative fatality risk as the sex of the drivers (S). Knowledge of HIC and femur load add little more to the ability to predict fatality risk in this initial analysis, as evidenced by $\chi^2$ of 0.01 and 0.77, although, at least, both coefficients have the right sign (positive).

A possible reason that the HIC and femur load variables do not contribute much to the prediction of fatality risk is that both of them are intercorrelated with chest g's, and, as a result, some of the information potentially conveyed by HIC and femur load is already contained in the chest g variable. The logistic injury probability CHESTINJ has correlation coefficients of .281 with HEADINJ, .162 with RFEMURINJ and .062 with LFEMURINJ, all statistically significant, suggesting at least a partial overlap in the scores. Although, ordinarily, these are not damaging levels of collinearity for independent variables, it will be seen in Chapter 5 that HIC and femur load, when analyzed <u>separately</u> from chest g's, have stronger relationships with fatality risk than those revealed in the regression approach of this chapter.

This initial data set, however, obscures the relationship between <u>any</u> NCAP scores and actual fatality risk primarily because of the poor quality of many of the FARS-NCAP vehicle "matches." The HIC, femur load and chest g's, in some cases, are derived from NCAP tests of cars that are fairly distinct from the FARS vehicle and might be inaccurate for that vehicle. The regression needs to be rerun with subsets of crashes in which both cars match up more closely with NCAP tests. The regression coefficients can be expected to increase while the sample sizes decrease. If the right kind of subsets are selected, the gains in the regression coefficients will overshadow the loss of sample size, and the $\chi^2$ will increase.

### 3.4 Regressions on data sets with closer FARS-NCAP matches

The closeness of the FARS-NCAP match was described in Section 2.5 by a quality rating on each of four affinity factors. The affinity factors and their quality levels were body platform/make-model (4 = same make-model, 3 = true corporate cousin, 2 = same platform only, 1 = similar platform); general model year range (A = FARS MY within NCAP MY-END MY range, B = FARS MY precedes NCAP MY, C = FARS MY after END MY); specific model year (FARS and NCAP MY identical, MY's off by 1, off by 2, ...); and body style (identical body style, similar body style, identical N of doors only, different N of doors). The affinity factors and their quality levels are a basis for defining subsets of the head-on collision file. A minimum acceptable match-quality level is specified for each affinity factor, and the regression is run for the subset of crashes in which both cars meet or exceed the match-quality levels. For example, NHTSA has never claimed that NCAP tests are valid for cars that are less than true corporate cousins or for cars outside the range of model years from the NCAP test vehicle year up to the "end year" [25], [26]. That is equivalent to demanding at least levels 3 and A, and excluding levels 1, 2, B and C, although not setting any minimum requirement on the specific model year. It would define a subset of crashes in which both cars match an NCAP test at either level 3A or 4A. Surely, that has to be one of the candidate subsets for the regression analysis. Another important subset is the one where both FARS cars match an NCAP test at level 3A or 4A and on the number of doors.

Although the 3A/4A subset may seem, intuitively, the "right" data set, it is worthwhile first to analyze some intermediate subsets, larger than the 3A/4A group, but smaller than the full data set. The first subset to be analyzed consists of crashes where both cars must match an NCAP test at level 3 or better (identical make/model or true corporate cousin). No demands are made on the general model year range, the specific model year or the body style. This subset eliminates some of the least satisfactory FARS-NCAP matches: the crashes where one or both vehicles only matched NCAP at level 2 (same chassis, different body) or level 1 (similar chassis). Since there were relatively few level 1 and 2 matches (see [9]), the sample size only decreased from 1392 to 1110 cases. The regression coefficients and their statistical significance levels are:

Level 3 and 4 Matches - Identical Make-Model or Corporate Cousin

| | Reg. Coeff. | Chi Square | Stat. Sig.? | Partial Corr. |
|---|---|---|---|---|
| INTERCEPT | .6234 | 47.51 | RR | |
| W (car weight) | -5.9405 | 218.11 | RR | -.378 |
| A (driver age) | .0590 | 205.49 | RR | .366 |
| S (driver sex) | .3065 | 5.60 | R | .049 |
| DELHIC | .2204 | .55 | | |
| DELCG | 2.1339 | 7.06 | RR | .058 |
| DELFEM | .8514 | 2.04 | | .005 |

The results are a definite improvement on the initial regression. The coefficient for chest g's increased from 1.84 to 2.13 and the $\chi^2$ increased from

6.62 to 7.06, despite the reduction in sample size. The contribution of chest g's is now significant at the .01 level and its $\chi^2$ is greater than the $\chi^2$ for gender. The coefficient for HIC increased from .02 to .22 and for femur load from .40 to .85; their $\chi^2$ increased substantially, although not to the level of statistical significance.

Table 3-1 summarizes the main findings of the two preceding regressions, plus the others that will be discussed in this chapter. It shows the sample size for each subset, the regression coefficients for head, chest and femur injury, and the $\chi^2$ for those variables.

The next subset to be analyzed takes one more step upwards on the first affinity factor. It consists of crashes where both cars must match an NCAP test at level 4 - i.e. have the same make-model as the NCAP test. As before, no demands are made on the general model year range, the specific model year or the body style. The third line of Table 3-1 shows that eliminating "corporate cousins" does not improve the results. The sample size drops severely, from 1110 to 612. Although the regression coefficient for chest g's increased from 2.13 to 2.37, its $\chi^2$ fell from 7.06 to 4.93, because of the reduced sample size, dropping it out of the .01 significance level. The coefficient and $\chi^2$ for femur load increased, but the coefficient for HIC dropped out of the positive range to a value close to zero.

The initial data set contained a moderate number of level B and C matches, where the FARS model year preceded the NCAP model year or came after the "end year" specified for the NCAP test. These are rather questionable matches and good candidates for deletion. The 4th line of Table 3-1 shows results for

## TABLE 3-1

## EFFECT OF NCAP-FARS MATCH QUALITY ON LOGISTIC REGRESSIONS FOR NCAP VARIABLES

| Both Cars Match with NCAP Tests at the Following Level: | N of Cars | Betas (Regression Coefficients) | | | Chi-Squares | | |
|---|---|---|---|---|---|---|---|
| | | Head Injury | Chest Injury | Femur Injury | Head Injury | Chest Injury | Femur Injury |
| At any level | 1392 | .02 | 1.84 | .40 | .01 | 6.62* | .77 |
| Level 3 or 4: identical make/model or "corporate cousin" | 1110 | .22 | 2.13 | .85 | .55 | 7.06** | 2.04 |
| Level 4: identical make/model | 612 | -.01 | 2.37 | 1.42 | .00 | 4.93* | 2.83 |
| Level A: NCAP MY $\le$ FARS MY $\le$ ENDYR | 872 | -.07 | 2.26 | 1.53 | .06 | 6.35* | 5.98* |
| 3A or 4A: same make/model or corporate cousin, NCAP MY $\le$ FARS MY $\le$ ENDYR | 740 | .21 | 2.70 | 1.41 | .41 | 7.70** | 3.94* |
| 4A: same make/model, NCAP MY $\le$ FARS MY $\le$ ENDYR | 402 | .14 | 2.43 | 3.67 | .09 | 3.28 | 6.62* |
| 3A or 4A, and (FARS MY - NCAP MY) $\le$ 2 | 416 | -.04 | 2.70 | 2.30 | .01 | 3.79 | 2.31 |
| 3A or 4A, and (FARS MY - NCAP MY) $\le$ 1 | 252 | .60 | 3.98 | 3.44 | 1.40 | 3.41 | 2.21 |
| 3A or 4A, and FARS MY = NCAP MY | 78 | .93 | 2.92 | 2.59 | .33 | .39 | .13 |
| 3A or 4A, and N of doors match | 588 | .42 | 2.52 | .73 | 1.20 | 5.62* | .98 |
| 4A, and N of doors match | 306 | .23 | 1.89 | 2.90 | .18 | 1.74 | 3.50 |
| 3A or 4A, and body style matches exactly | 452 | .07 | 2.72 | .67 | .03 | 4.94* | .44 |

*statistically significant, alpha < .05

**statistically significant, alpha < .01

the subset of level A matches, where both cars in the crash must have their model year in the range from the NCAP model year to the "end" year; no demands are made on the other affinity factors (e.g., level 1 and 2 matches are included here). The sample size is reduced to 872. Compared to the initial regression, the results for chest g's and, especially, femur load are improved. Both of these coefficients are statistically significant at the .05 level. But the coefficient for HIC drops out of the positive range.

Since limiting the data to levels 3 and 4 helped, and restricting to level A also helped, good results can be expected if both limitations are applied at the same time - i.e., both cars in the crash have to have be the same make model or a true corporate cousin of their NCAP match, and the FARS model years have to be in the "valid" range for the NCAP tests. The specific model year or the body styles are not required to match. The sample size is 740 vehicles, which is slightly more than half the initial data set (actually, 792 vehicles, but 52 had to be excluded from this regression because DELHIC, DELCG or DELFEM were missing). The regression coefficients (summarized in the 5th, bold line of Table 3-1) are the following:

Level 3A and 4A Matches

| | Reg. Coeff. | Chi Square | Stat. Sig.? | Partial Corr. |
|---|---|---|---|---|
| INTERCEPT | .6883 | 39.19 | RR | |
| W (car weight) | -5.7355 | 142.47 | RR | -.374 |
| A (driver age) | .0579 | 134.54 | RR | .364 |
| S (driver sex) | .3665 | 5.33 | R | .058 |
| DELHIC | .2116 | .41 | | |
| DELCG | 2.7004 | 7.70 | RR | .075 |
| DELFEM | 1.4109 | 3.94 | R | .044 |

They are the most satisfactory results of any of the regressions, and this subset will be used for most of the analyses in the remainder of this report. The coefficient for chest g's is 2.7, and its $\chi^2$ is 7.70, which is significant at the .01 level and is the highest $\chi^2$ found in any of the regressions summarized in Table 3-1. Femur load has a coefficient of 1.41, consistent with the two preceding regressions in Table 3-1 and significant at the .05 level. Head injury has a positive coefficient, although not statistically significant.

A comparison of the $\chi^2$ here vs. the initial regression on the full data set shows a higher $\chi^2$ for each of the NCAP variables in the 3A/4A regression than in the full data set, despite a reduction in the sample size from 1392 to 740. That suggests there is little or no "correlation" between NCAP and FARS when the FARS vehicles match NCAP at affinity levels <u>less than</u> 3A. The overall correlations found in the larger subsets merely reflects the 3A/4A cases within those subsets. Indeed, when the same regression is performed on the 652 cases (i.e., 1392 - 740) where one or both cars do <u>not</u> match an NCAP test at level 3A or 4A, the coefficients for the NCAP variables are all nonsignificant: -0.22 for head injury ($\chi^2$ = 0.22), 1.23 for chest injury ($\chi^2$ = 1.22) and -0.36 for femur injury ($\chi^2$ = 0.66).

A further subsetting of the data, from level 3A/4A to exclusively level 4A matches does not improve the results. The sample size is reduced to 402 cases. When sample sizes drop much below 500, the set of head-on collisions becomes too small to include a representative mix of cars, and anomalous results can be expected when the regression model seizes on certain properties of the vehicles in the sample and "attributes" the results to HIC, chest g's or femur load. The regression coefficients (summarized in the 6th line of Table 3-1) are the following:

Level 4A Matches

| | Reg. Coeff. | Chi Square | Stat. Sig.? | Partial Corr. |
|---|---|---|---|---|
| INTERCEPT | .5846 | 15.38 | RR | |
| W (car weight) | -5.7218 | 77.50 | RR | -.371 |
| A (driver age) | .0630 | 72.27 | RR | .358 |
| S (driver sex) | .4360 | 4.06 | R | .061 |
| DELHIC | .1433 | .09 | | |
| DELCG | 2.4291 | 3.28 | | .048 |
| DELFEM | 3.6708 | 6.62 | R | .092 |

The regression coefficient for chest g's drops to 2.43, which is not statistically significant at that sample size. The coefficient for femur load climbs to an unexpectedly high 3.67, and is significant at the .05 level. The femur load coefficient is higher than in any of the other regressions in Table 3-1 and seems out of line with the general trend in that table.

So far, the analyses have shown that level 3A and 4A matches between FARS and NCAP are satisfactory, while level 1, 2, B and C matches should not be used. None of the subsets demanded the FARS and NCAP vehicles to match on specific model year or body style. In the remaining regressions, FARS and NCAP will always have to match up at the 3A or 4A level, and the effect of further restricting the data to close matches on model year or body style will be considered.

Given that the FARS model year is within the valid range of model years for the NCAP test (level A), there is no advantage to further limits on the model year. If it is demanded that the FARS and NCAP model years can differ by no more than 2 (7th line of Table 3-1), the sample size is reduced from 740 to 416, and <u>all</u> $\chi^2$ drop out of the significant range. The coefficient for HIC drops out of the positive range. Further limiting the FARS and NCAP model years to be within one, or to be identical, cuts sample sizes to unsatisfactory levels and does not produce statistically significant coefficients.

When NHTSA staff matched the FARS and the NCAP cases, they placed high priority on matching the vehicles by body style. They generally preferred to match a corporate cousin or an NCAP test several years old, but with the same body style, than an NCAP test of the same make-model and model year, but with a different body style. The great majority of the 740 level 3A and 4A matches also had identical body styles or at least the same number of doors. The results were slightly modified, but not necessarily improved by limiting the analyses to subsets of crashes in which both vehicles matched NCAP test vehicles on N of doors and/or exact body style.

There were 588 crashes in which both cars had the same number of doors as their NCAP test matches. The regression coefficients (summarized in the 10th line of Table 3-1) are the following:

FARS/NCAP Match at Level 3A and 4A <u>and</u> Same N of Doors

| | Reg. Coeff. | Chi Square | Stat. Sig.? | Partial Corr. |
|---|---|---|---|---|
| INTERCEPT | .6600 | 29.17 | RR | |
| W (car weight) | -5.6001 | 109.31 | RR | -.367 |
| A (driver age) | .0545 | 103.41 | RR | .356 |
| S (driver sex) | .4066 | 5.04 | R | .062 |
| | | | | |
| DELHIC | .4218 | 1.20 | | |
| DELCG | 2.5179 | 5.62 | R | .067 |
| DELFEM | .7316 | .98 | | |

The regression coefficients show some changes from the 3A/4A analysis. The head-injury coefficient increased from .21 to .42, which is the highest level it reached in any of the analysis of Table 3-1 that are based on more than 500 cases; nevertheless, it did not reach statistical significance ($\chi^2 = 1.20$). The chest-injury coefficient stayed about the same (2.52 vs. 2.70), although its statistical significance dropped from the .01 level to the .05 level, in part due to the reduction in sample size. The coefficient for femur load, which was significant in the 3A/4A analysis, dropped out of the significant range here, having a $\chi^2$ slightly lower than the head-injury coefficient.

The preceding analysis suggests a possibility that the higher correlation coefficient for head injury is a result of requiring FARS and NCAP to match on N of doors. If so, an even stronger requirement - viz. that FARS and NCAP cases have the exactly the same body style (4-door sedan, station wagon, etc.) - could be expected to maintain or even further increase the head-injury coefficient. There are 452 crashes where both vehicles matched an NCAP test at levels 3A or 4A <u>and</u> exact body style: only a small reduction from the 588 cases in the preceding analysis. However, the last line of data in Table 3-1 shows that, for this subset, the head-injury coefficient dropped back close to zero. Given the sample sizes on which the various regressions are based, the subtle variations in the regression coefficients are quite probably due to chance.

## 3.5 <u>Summary</u>

The main purpose of this chapter was to identify a large set of head-on collisions in which both cars match up close enough with NCAP test vehicles that the scores for the NCAP vehicles accurately depict the 35 mph barrier performance of the actual crash-involved vehicles. The empirical evidence is that a "level 3A or 4A" match between the FARS and NCAP vehicles is close enough - i.e., the FARS and NCAP vehicles are of the same make-model or true corporate cousins, and the FARS model year is within the "valid" range of the NCAP test. Anything less than a level 3A/4A match is insufficient. An alternative, somewhat smaller data set that produces good, but slightly different correlations consists of FARS cases in which both cars match an NCAP test at level 3A or 4A <u>and</u> have the same number of doors as the NCAP test vehicle. The empirical findings are consistent with NHTSA's earlier claims that an NCAP test result can be extended to a car's corporate cousins, and for subsequent model years until the car is redesigned [25], [26].

While the methods of this chapter are not a preferred way to test for correlation between NCAP performance and actual fatality risk, they still showed that significant correlations exist. In the analyses summarized in Table 3-1, NCAP chest g's had a statistically significant regression coefficient in <u>every</u> regression with a sample size greater than 500, while the coefficient for femur load reached significance in three analyses. (The methods in subsequent chapters will confirm these correlations, and also show significant correlation with HIC under certain conditions). At the same time, the statistics in Table 3-1 indicate that the regression coefficients can vary quite a bit in response to moderate changes in the calibration data set. The range of Chi-Squares for chest g's overlaps with the range for femur load, which, in turn, overlaps with the range for HIC. In other words, while the accident data set is sufficient to indicate an overall significant relationship between NCAP scores and actual fatality risk, there are not enough data indicate the exact relative importance of the three NCAP body regions in predicting fatality risk.

# CHAPTER 4

# A COMPOSITE NCAP SCORE: CORRELATION WITH FATALITY RISK

HIC, chest g's and femur load each provide some information about a vehicle's safety performance on an NCAP test. An appropriate <u>weighted average</u> of the scores for the three body regions could provide more information about a car's overall safety performance than any score for a single body region, and have greater correlation with actual fatality risk than any single NCAP score. The objectives of this chapter are to identify a composite NCAP score, NCAPINJ, that has maximum correlation with the fatality risk of belted drivers in the principal calibration data set of actual head-on collisions, and to measure the extent of that correlation. NCAPINJ is a specific weighted average of head, chest and femur scores. However, sensitivity tests in this chapter will show that NCAPINJ is not the only composite score that has excellent correlation with fatality risk; other weighted averages, and even an unweighted sum of logistic injury probabilities, also correlate well with actual risk. NCAPINJ may not be the optimum composite score on another calibration data set. Thus, the purpose of defining NCAPINJ is not to find a unique "magic bullet" that is the best and only way to express the NCAP results, but to show that existing NCAP test scores for the three body regions, when combined by some reasonable scheme, have highly significant correlation with actual fatality risk in head-on collisions.

## 4.1 A composite measure of NCAP performance

The regression analyses of Chapter 3 supply most of the framework for generating a composite NCAP measure that has excellent correlation with fatality risk. Here are some of the relevant analytic tools developed in Chapter 3. The actual NCAP results for the driver dummy were transformed to logistic injury probability functions, ranging from 0 to 1:

$$\text{HEADINJ} = 1 / [1 + \exp(5.02 - .00351\ \text{HIC})]$$
$$\text{CHESTINJ} = 1 / [1 + \exp(5.55 - .06930\ \text{chest g's})]$$
$$\text{LFEMURINJ} = 1 / [1 + \exp(7.59 - .00294\ \text{left femur load})]$$
$$\text{RFEMURINJ} = 1 / [1 + \exp(7.59 - .00294\ \text{right femur load})]$$

NCAP performance for the "case" vehicle relative to the "other" vehicle in a head-on collision was defined in terms of these functions:

$$\text{DELHIC} = \text{HEADINJ}_{\text{CASE}} - \text{HEADINJ}_{\text{OTHER}}$$
$$\text{DELCG} = \text{CHESTINJ}_{\text{CASE}} - \text{CHESTINJ}_{\text{OTHER}}$$
$$\text{DELFEM} = (\text{LFEMURINJ}_{\text{CASE}} + \text{RFEMURINJ}_{\text{CASE}}) - (\text{LFEMURINJ}_{\text{OTHER}} + \text{RFEMURINJ}_{\text{OTHER}})$$

The most appropriate data set of head-on collisions for studying correlation with NCAP was found to be the crashes in which both cars matched up with an NCAP test vehicle at levels "3A" or "4A": the model year on FARS is within the range of model years considered valid on the NCAP test, and the make-models on FARS and NCAP are identical or true corporate cousins. That data set includes 396 head-on collisions (792 vehicles); however, DELHIC, DELCG, or DELFEM are undefined, due to missing data on NCAP tests, in 26 collisions, leaving a sample of 370 collisions (740 vehicles) for the initial regression analysis. The logistic regression model that best predicts the fatality risk of the driver of the case vehicle, in those 740 cases, has the following regression coefficients and chi-squares ($\chi^2$):

| | Regression Coefficient | Chi-Square |
|---|---|---|
| INTERCEPT | .69 | 39.19 |
| W (car weight) | -5.74 | 142.47 |
| A (driver age) | .0579 | 134.54 |
| S (driver sex) | .367 | 5.33 |
| DELHIC | .21 | .41 |
| DELCG | 2.70 | 7.70 |
| DELFEM | 1.41 | 3.94 |

The initial goal is to find a single variable DELNCAP, which would replace DELHIC, DELCG and DELFEM in the preceding regression and get a high $\chi^2$. If the choice of DELNCAP is limited to linear combinations of DELHIC, DELCG and DELFEM, i.e.,

$$\text{DELNCAP} = \text{W1 DELHIC} + \text{W2 DELCG} + \text{W3 DELFEM}$$

then the preceding regression coefficients .21, 2.70 and 1.41, if substituted for W1, W2 and W3, generate the DELNCAP that maximizes $\chi^2$ in those 740 cases. With minor modifications, that will become the composite measure of NCAP performance.

The regression in Chapter 3 was limited to the 740 cases in which HIC, chest g's and femur load on at least one leg were known for both vehicles (i.e., they were successfully measured in the NCAP test that is matched with the crash-involved vehicle). It is desired to expand the analysis to include cases where the NCAP results are partially missing, to include as many of the 792 level 3A and 4A matches as possible. Just as DELNCAP, as defined above, was a linear combination of the relative scores for two vehicles, it is possible to define a composite logistic injury score for the driver of one vehicle if the NCAP results are all known:

$$\text{NCAPINJ} = \text{W1 HEADINJ} + \text{W2 CHESTINJ} + \text{W3 (LFEMURINJ + RFEMURINJ)}$$

where W1, W2 and W3 are constants which remain to be determined. LFEMURINJ and RFEMURINJ have similar means and distributions, and one can be used as a surrogate for the other, if it is unknown - e.g., if only LFEMURINJ is unknown,

$$\text{NCAPINJ} = \text{W1 HEADINJ} + \text{W2 CHESTINJ} + 2\ \text{W3 RFEMURINJ}$$

The situation is more complicated if HEADINJ or CHESTINJ are unknown or if both LFEMURINJ and RFEMURINJ are unknown, because these variables have different means and make different contributions to NCAPINJ. In these 792 cases, the average of HEADINJ is .196, average CHESTINJ is .123, and average (LFEMURINJ + RFEMURINJ) is .057. If HEADINJ is unknown, define an inflation factor

$$MISSHIC = (.196\ W1 + .123\ W2 + .057\ W3) / (.123\ W2 + .057\ W3)$$

and inflate the NCAPINJ based on the other two body regions by this factor:

$$NCAPINJ = MISSHIC\ [W2\ CHESTINJ + W3\ (LFEMURINJ + RFEMURINJ)]$$

or, if LFEMURINJ is also unknown,

$$NCAPINJ = MISSHIC\ [W2\ CHESTINJ + 2\ W3\ RFEMURINJ]$$

Similar inflation factors are applied if just chest g's or if both femur loads are unknown. That expands the analysis to 784 cases in which NCAP scores were known on at least two body regions for each vehicle. Only 8 cases had to be deleted because NCAP results were unknown on more than one body region. There are 756 cases in which NCAP scores are known for the chest (which has the highest of the three relative weights) and at least one other body region.

Starting with the values of .21, 2.7 and 1.41 for W1, W2 and W3, and defining DELNCAP = $NCAPINJ_{CASE} - NCAPINJ_{OTHER}$, regressions are run in which the dependent variable is fatality risk in the "case" vehicle, and the independent variables are DELNCAP plus W, A and S (as defined in Chapter 3). These regressions are performed for the full set of 784 cases (NCAP scores known for at least two body regions) and its subset of 756 cases where chest g's are known for both vehicles. Two series of regressions are run with alternative values for W1 and W3, relative to W2. In the first series, a sort of fine tuning to maximize $\chi^2$, examines the effects of slight variations from the starting values of W1, W2 and W3. The $\chi^2$ for DELNCAP were as follows:

| Relative Weights | | | Chi-Squares | |
|---|---|---|---|---|
| Head Injury | Chest Injury | Femur Injury | 784 Cases | 756 Cases |
| .21 | 2.7 | 1.41 | 16.08 | 13.93 |
| .1 | 2.7 | 1.41 | 16.25 | 13.75 |
| .3 | 2.7 | 1.41 | 15.84 | 13.94 |
| .21 | 2.7 | 1.3 | 16.03 | 13.96 |
| **.21** | **2.7** | **1.5** | **16.09** | **13.89** |
| .21 | 2.7 | 1.6 | 16.06 | 13.81 |

All of the regressions produce $\chi^2$ close to the first one, indicating a plateau rather than a peak of optimum correlation. While none of the regressions has the maximum $\chi^2$ for <u>both</u> the 784 and the 756 cases,

**NCAPINJ = .21 HEADINJ + 2.7 CHESTINJ + 1.5 (LFEMURINJ + RFEMURINJ)**

can be considered the best of the composite injury scores. The $\chi^2$ for the 784 cases reaches a local maximum of 16.09, while the $\chi^2$ of 13.89 for the 756 cases is still close to the maximum. Although the model with W1 = .1 for head injury has a higher $\chi^2$ for the 784 cases, the $\chi^2$ for the 756 cases is the lowest of the group.

The second series of regressions examines the width of the plateau of near-optimum correlation. It compares the $\chi^2$ for the NCAPINJ with the optimum weights (.21, 2.7 and 1.5), the $\chi^2$ for an unweighted injury function

INJ = HEADINJ + CHESTINJ + LFEMURINJ + RFEMURINJ

and for three intermediate injury functions, proceeding by harmonic steps from the optimum weights to the unweighted function:

| Relative Weights | | | Chi-Squares | |
|---|---|---|---|---|
| Head Injury | Chest Injury | Femur Injury | 784 Cases | 756 Cases |
| 0.21 | 2.70 | 1.50 | 16.09 | 13.89 |
| 0.31 | 2.10 | 1.36 | 15.58 | 13.67 |
| 0.46 | 1.64 | 1.22 | 14.19 | 12.71 |
| 0.68 | 1.28 | 1.11 | 11.40 | 10.42 |
| 1 | 1 | 1 | 7.86 | 7.34 |

All $\chi^2$ are statistically significant at the .01 level, and in the second and third line the $\chi^2$ are still close to the optimum values even though the relative weights are quite different from their optimum values. This second series of regressions shows that the correlation of fatality risk with a composite NCAP score is relatively insensitive to the exact choice of the weight factors and that <u>any</u> reasonable combination of head, chest and femur injury scores will correlate well with fatality risk.

The optimum score,

$$NCAPINJ = .21\ HEADINJ + 2.7\ CHESTINJ + 1.5\ (LFEMURINJ + RFEMURINJ)$$

can range from 0 to 5.91. It is a relative measure of overall NCAP performance (the higher the NCAPINJ, the worse the performance), but specific values of NCAPINJ, such as 0.5 or 1.0, do not correspond to any intuitive, absolute level of injury.

NCAPINJ, at first glance, seems to give a very low weight to head injury and a surprisingly high weight to femur injury, as indicated by W1 = .21 and W3 = 1.5. However, the Wi's, by themselves, do not indicate the relative weights of the body regions. As noted above, the average value of HEADINJ is

.196, average CHESTINJ is .123, and average (LFEMURINJ + RFEMURINJ) is .057. Thus, the relative contribution of HEADINJ to NCAPINJ is .21 x .196 = .041; the contribution of CHESTINJ is 2.7 x .123 = .332; and the contribution of femur injury is 1.5 x .057 = .086. To the extent that HIC, chest g's and femur load are intercorrelated, their "relative contributions" to the composite score need not reflect the actual relative importance of head, chest and lower-body injuries in crashes - i.e., CHESTINJ may be making such a large contribution to NCAPINJ because it incorporates information of the probability of head and femur injury, in addition to chest injury. The composite score is a mathematical method of combining NCAP information to get the best correlation with fatality risk in the current data set of head-on collisions. (NCAPINJ is optimized for the current data file of fatal head-on collisions, but when additional accident data become available, the relative weights for the three body regions might change).

The complete results of the regression model, with independent variables W, A, S and DELNCAP = $NCAPINJ_{CASE}$ - $NCAPINJ_{OTHER}$, where NCAPINJ = .21 HEADINJ + 2.7 CHESTINJ + 1.5 (LFEMURINJ + RFEMURINJ), for the 784 head-on collision cases in which both cars matched up with an NCAP test vehicle at levels "3A" or "4A" and DELNCAP could be calculated, are as follows:

DELNCAP = .21 DELHIC + 2.7 DELCG + 1.5 DELFEM

| | Reg. Coeff. | Chi Square | Stat. Sig.? | Partial Corr. |
|---|---|---|---|---|
| INTERCEPT | .6345 | 37.42 | RR | |
| W (car weight) | -5.3305 | 146.49 | RR | -.368 |
| A (driver age) | .0558 | 143.87 | RR | .365 |
| S (driver sex) | .4200 | 7.91 | RR | .075 |
| DELNCAP | 1.0665 | 16.09 | RR | .115 |

The $\chi^2$ of 16.09 for DELNCAP indicates a strongly significant correlation ($p < .0001$) between the composite NCAP score and fatality risk. It exceeds the sum of the $\chi^2$ for head, chest and femur injury when they were entered as separate variables in the regression, because cases that had to be excluded from that regression due to missing NCAP data can now be included. However, the $\chi^2$ of the regression coefficient for DELNCAP is just one of several ways to measure the extent or strength of the relationship between NCAPINJ and fatality risk in actual head-on collisions. The remainder of this chapter presents other methods to gauge the relationship, and to measure the actual fatality reduction for a good NCAPINJ relative to a poor score.

## 4.2 RELEXP: actual safety performance relative to expectations

DELNCAP = $\text{NCAPINJ}_{\text{CASE}}$ - $\text{NCAPINJ}_{\text{OTHER}}$ is a measure of the relative NCAP performance of two vehicles that became involved in a head-on collision. If the case vehicle had better performance on the NCAP test than the other vehicle, DELNCAP is negative. Another variable, RELEXP, will now be defined for each head-on collision, measuring the relative actual performance of the two vehicles in that collision. RELEXP will be negative when the driver of the case vehicle did better than expected (e.g., survived) and the driver of the other car did worse than expected, given the weights of the two cars and the age and sex of each driver. RELEXP can be tested for correlation with DELNCAP; more generally, the average value of RELEXP can be computed for various groups of crashes (e.g., collisions of good NCAP performers with poor NCAP performers).

The first step in computing RELEXP is a regression on the file of 784 head-on collision cases defined in Section 4.1 (both cars match an NCAP test vehicle at levels 3A or 4A, and DELNCAP could be calculated), but without any

NCAP variable. In other words, the dependent variable is the outcome for the driver of the case vehicle (fatality = 1, survival = 0) and the independent variables are only W, A and S - relative vehicle weight, driver age and sex:

Without NCAP Information

| | Reg. Coeff. | Chi Square | Stat. Sig.? | Partial Corr. |
|---|---|---|---|---|
| INTERCEPT | .616 | 36.57 | RR | |
| W (car weight) | -5.427 | 154.15 | RR | -.378 |
| A (driver age) | .0531 | 142.35 | RR | .363 |
| S (driver sex) | .34 | 5.39 | R | .056 |

The intercept and coefficients for W, A and S are similar to those obtained in the preceding regression with DELNCAP. This logistic regression model can be used to predict the <u>expected</u> fatality risk for each driver in a head-on collision, in the absence of NCAP information. The expected fatality risk $E_{case}$ for the driver of the case vehicle is

$$\frac{\exp[.616 - 5.427(\log W_{case} - \log W_{other}) + .0531(A_{case}-A_{other}) + .34(F_{case}-F_{other})]}{1 + \exp[.616 - 5.427(\log W_{case} - \log W_{other}) + .0531(A_{case}-A_{other}) + .34(F_{case}-F_{other})]}$$

where $W_{case}$ is the curb weight of the case vehicle, $A_{case}$ is the age of the driver of the case vehicle and $F_{case}$ is 1 if the driver of the case vehicle is female, 0 if the driver is male. The expected fatality risk $E_{other}$ for the driver of the other vehicle is

$$\frac{\exp[.616 + 5.427(\log W_{case} - \log W_{other}) - .0531(A_{case}-A_{other}) - .34(F_{case}-F_{other})]}{1 + \exp[.616 + 5.427(\log W_{case} - \log W_{other}) - .0531(A_{case}-A_{other}) - .34(F_{case}-F_{other})]}$$

These formulas measure the relative vulnerability to fatal injury of the two drivers, <u>given</u> that their cars had a head-on collision. The risk is greater in the lighter car than the heavier car, and the older/female driver is more vulnerable to injury than the younger/male driver. The formulas do not address the propensity of cars to get involved in head-on collisions as a function of driver age, sex, etc. For example, if the case vehicle is a 2500-pound car driven by a belted, 50-year-old female and the other vehicle is a 3000-pound car driven by belted, 20-year-old male, $E_{case}$ = .97 and $E_{other}$ = .09 (for a total of 1.06 fatalities expected in the collision, since there is a 6 percent chance that both drivers died).

If $A_{case}$ and $A_{other}$ are the actual outcome of the collision for the driver of each car (fatality = 1, survival = 0),

$$\textbf{RELEXP} = (A_{case} - E_{case}) - (A_{other} - E_{other})$$

measures actual performance "relative to expectations." It can range from -2 to +2. The more negative it is, the better the actual performance of the case vehicle relative to expectations. For example, if the 50-year-old female in the 2500-pound case car actually survived, while the 20-year-old male in the 3000-pound other car died, RELEXP = -1.88 (much better than expectations). If she died and he survived, RELEXP = +0.12 (about what would be expected). If both drivers died in the crash, RELEXP = -0.88 (not a good outcome for either driver, of course, but the case vehicle performed better than expected, relative to the other vehicle). Note that RELEXP is measured for a two-car crash, not for a vehicle. It does not measure the absolute safety of a vehicle, just the performance of the case vehicle relative to the other vehicle.

The population standard deviation of RELEXP was computed for the full set of 784 crashes and for many subsets of these crashes. In every case, the standard deviation was very close to 0.64. That makes it easy to test if the average value of RELEXP is significantly less than zero for a specific group of crashes (i.e., the case vehicles were significantly safer than the other cars), or if the difference in average RELEXP for two groups of crashes is statistically significant.

## 4.3 Correlation of DELNCAP and RELEXP

DELNCAP, a measure of the relative NCAP performance of two vehicles that became involved in a head-on collision, and RELEXP, a measure of their relative actual performance are defined for each of 392 collisions on the file. DELNCAP and RELEXP are both close enough to a normal distribution that their correlation can be tested by the conventional Pearson method. (In the analysis file, there are 784 collision records, but there are only 392 actual collisions, since each crash is listed twice. Reversing the "case" and the "other" vehicle merely changes the sign of both DELNCAP and RELEXP, so the second listing of each crash provides no new information for the analysis, and using N = 784 would spuriously inflate significance levels.)

DELNCAP and RELEXP have a correlation coefficient of .166, which has strong statistical significance ($p = .001$, $N = 392$). In other words, the higher the composite NCAP score for car 1 relative to car 2, the higher the fatality risk for driver 1 relative to driver 2, after adjusting for car weight, driver age and sex.

The correlation coefficient and its significance level both say a lot about the relationship between NCAP performance and actual fatality risk in all types of head-on collisions on the highway. On the one hand, the correlation of .166 is far from perfect: the driver of the car with the lower NCAP score will not always be the survivor in any type of head-on collision with a car having a higher score, even if both cars have the same weight. Furthermore, needless to say, which driver dies in a head-on collision has a lot more to do with relative vehicle weight and driver age than with NCAP scores. On the other hand, the significance level of .001 suggests beyond doubt that there is <u>some</u> correlation between NCAP and actual fatality risk in head-on collisions: <u>on the average</u>, cars with acceptable NCAP scores have lower fatality risk, across the range of head-on collisions, than cars of the same weight with high scores.

This analysis approach also makes it possible to test if NCAP information for a single body region, in the absence of information about the other two body regions, is correlated with fatality risk. The approach is different from Chapter 3, in which information for all three body regions was simultaneously entered in a regression, and the relative contribution of each NCAP score to fatality risk was estimated. DELCG, DELFEM and DELHIC are the measures of relative NCAP performance, based on logistic injury probability functions (see Section 3.2). DELCG has a statistically significant correlation with RELEXP ($r = .136$, $p = .008$, $N = 378$). In other words, there is a significant correlation between chest g's, by itself, and fatality risk. DELFEM has a positive correlation with RELEXP, but not quite statistically significant ($r = .094$, $p = .065$, $N = 387$); DELHIC also has a nonsignificant positive correlation ($r = .050$, $p = .321$, $N = 389$).

## 4.4 Fatality reduction for the car with lower NCAPINJ

The accident data file contains 784 head-on collision records for which $DELNCAP = NCAPINJ_{CASE} - NCAPINJ_{OTHER}$ (relative composite NCAP performance) is known. The records can be ranked by DELNCAP and listed in order, from the case with DELNCAP = -1.98 (largest differential in favor of the case vehicle) to the case with DELNCAP = +1.98. The 16 records in the middle, with DELNCAP = 0, are deleted from the list: both cars in the collision matched up with the same NCAP test, so NCAP gives no information to favor one car or the other. The last 384 records, with DELNCAP > 0, are also deleted: they are merely the same crashes as the first 384 records, with the "case" and "other" vehicles reversed. That leaves a file of 384 distinct head-on collisions, comprising 768 distinct vehicles, in which the NCAP performance of the case vehicle is always better than the performance of the other vehicle. DELNCAP ranges from -1.98 for the first record to -0.000243 for the last (a very small advantage for the case vehicle).

The objective of this analysis is to compute the reduction in fatality risk for the 384 "better" NCAP performers relative to the 384 "poorer" NCAP performers, and to test if the reduction is statistically significant. Moreover, if the analysis is limited to the first half/quarter/tenth of the file, where the NCAP performance advantage of the case vehicle successively increases, does the fatality reduction for the case vehicle also escalate?

As explained in Section 4.2, each collision has an outcome $A_{case}$ for the driver of the case vehicle (fatality = 1, survival = 0) and $A_{othr}$ for the driver of the other car. The expected outcomes $E_{case}$ and $E_{othr}$ (expected probabilities of fatality) are based on the relative vehicle weights and the age and sex of the drivers. The actual and expected fatalities are summed over all the crashes included in the analysis: sum($A_{case}$) and sum($A_{othr}$) are the actual numbers of driver fatalities in the case and the other vehicles; sum($E_{case}$) and sum($E_{othr}$) are the numbers of driver fatalities that would be expected in the case vehicles and the other vehicles, given the relative weight, age and sex in each crash. The fatality reduction for the case vehicles, relative to the other vehicles, is

$$\text{Fatality Reduction} = 1 - ([\text{sum}(A_{case})/\text{sum}(A_{other})] \,/\, [\text{sum}(E_{case})/\text{sum}(E_{other})])$$

Specifically, in the 384 collisions with DELNCAP < 0, there were 202 actual driver fatalities in the case vehicles and 246 fatalities in the other vehicles. Since the vehicle weights, driver age and sex distributions are similar, on the average, in the better and poorer NCAP performers, the expected numbers of fatalities are about the same: 220.7 in the case vehicles and 227.7 in the other vehicles. That is a relative fatality reduction of

$$1 - [(202/246) \,/\, (220.7/227.7)] = 15.3 \text{ percent}$$

for the better NCAP performers. Conversely, the poorer performers had 18 percent higher fatality risk.

The test for statistical significance of the fatality reduction is based on the variable RELEXP, which is computed for each collision record, as explained in Section 4.2. RELEXP measures actual performance of the case vehicle relative to expectations. The average value of RELEXP is

$$\text{mean RELEXP} = [(202 - 220.7) - (246 - 227.7)] \,/\, 384 = -.096$$

For this population of crashes, as for most others, the standard deviation of RELEXP is very close to .64. With a sample size of 384, the t statistic for RELEXP is 2.96, which is significant at the .01 level. In other words, the better NCAP performers had significantly fewer fatalities than expected, relative to the poorer performers.

Table 4-1 shows the fatality reductions and other statistics for all crashes in which DELNCAP is less than zero or is more negative than some specified amount (i.e., the case vehicles did better on NCAP than the other

TABLE 4-1

FATALITY REDUCTION FOR THE DRIVER OF THE CAR WITH THE LOWER COMPOSITE NCAP SCORE

| NCAPINJ Differential between the "Case" Car and the "Other" Car | Comparison of Real-World Performance | | | | |
|---|---|---|---|---|---|
| | N of Crashes | % Fat Red for Better Car | Mean Relexp | Sum Relexp | T-Test for Relexp |
| DELNCAP < 0 | 384 | 15.3 | -.096 | -37.00 | 2.96 |
| DELNCAP < -0.18 | 203 | 15.8 | -.102 | -20.67 | 2.46 |
| DELNCAP < -0.42 | 95 | 24.5 | -.171 | -16.25 | 2.84 |
| DELNCAP < -0.635 | 38 | 40.1 | -.253 | - 9.61 | 2.82 |

| NCAPINJ Differential between the "Case" Car and the "Other" Car | Mean NCAP Scores and Model Year ("Case" vs. "Other" Car) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | HIC | | Chest G's | | L Femur | | R Femur | | Model Year | |
| | Case | Other | Case | Other | Case | Other | Case | Other | Case | Other |
| DELNCAP < 0 | 815 | 1007 | 43.9 | 53.5 | 858 | 1025 | 843 | 1033 | 85.6 | 85.0 |
| DELNCAP < -0.18 | 820 | 1082 | 43.0 | 57.7 | 873 | 1141 | 831 | 1103 | 85.8 | 84.6 |
| DELNCAP < -0.42 | 835 | 1095 | 43.5 | 62.7 | 891 | 1213 | 854 | 1081 | 85.7 | 83.3 |
| DELNCAP < -0.635 | 832 | 1072 | 43.7 | 61.8 | 883 | 1738 | 903 | 1225 | 85.6 | 83.0 |

vehicles, by some specified amount). The first line in the upper half of Table 4-1 displays statistics for the analysis described above: the crashes with DELNCAP < 0. It shows the sample size, the fatality reduction for the case vehicles, mean RELEXP, the t test and one other statistic: the sum of RELEXP. In this case, the sum is -37 = 384 x -.096. Intuitively, this sum describes the total amount of "information" provided by the NCAP results for identifying safer cars; the more negative the sum, the better. The bigger the actual safety difference between the good and poor NCAP performers, and the more cars involved in the analysis, the more negative the sum.

The first line in the lower half of Table 4-1 compares the actual NCAP performance of the case vehicles and the other vehicles. The case vehicles (the better NCAP performers), appropriately, had lower HIC, chest g's and femur loads, <u>on the average</u>, than the poor performers. Average HIC was 815 in the case vehicles, 1007 in the other cars. Chest g's averaged 10 less in the case vehicles; femur loads averaged 167 pounds less on the left side and 190 pounds less on the right side. These are <u>average</u> differences; they do not mean that HIC, chest g's and femur loads are lower for the case vehicle in each individual crash. Since NCAPINJ is a weighted sum of HIC, chest g's and femur load, the case vehicle might have the higher HIC in some crashes, but DELNCAP would still be negative if the case vehicle has much lower chest g's. The average model year is slightly more recent in the case vehicles (85.6) than the other vehicles (85.0).

In the preceding analysis, the case vehicle was only required to have better NCAP performance than the other vehicle; even an infinitesimal difference was sufficient. The fatality reduction for good NCAP performance is even greater when a "gap" is placed between good and poor performance and DELNCAP is required to be more negative than some specified amount. The results are shown in the 2nd, 3rd and 4th lines of each section of Table 4-1. Approximately half the crashes (203) on the original file have DELNCAP < -0.18. In that group, the better NCAP performers have 15.8 percent lower fatality risk than the poor performers; the reduction is statistically significant (t for RELEXP is 2.46, p < .05). The case vehicles enjoy an even larger advantage in HIC, chest g's and femur load than in the preceding analysis. The gap in average model year also grows (85.8 vs. 84.6), reflecting the superior NCAP performance of more recent cars.

The sample size is again halved when DELNCAP < -0.42. In those 95 crashes, the case vehicles had 24.5 percent lower fatality risk than the poor NCAP performers, which is a statistically significant reduction (t for RELEXP is 2.84, p < .01). Although the fatality reduction is greater than in the preceding analyses, the sum of RELEXP is smaller: here, NCAP identifies a small number of cars that are quite unsafe; previously, NCAP identified a much larger number of cars that were slightly less safe than average. The lower half of Table 4-1 shows that chest g's are much higher in the other cars (62.7) than in the case cars (43.5). The gap in HIC and femur load, on the other hand, is about the same as in the preceding analysis. The difference in average model year continues to escalate.

Finally, when DELNCAP < -.635, the accident file is reduced to 38 crashes, one-tenth of the original number. In these crashes with a large contrast in NCAP performance, the good cars had 40 percent lower fatality risk than the poor performers. The reduction is statistically significant (t for

RELEXP is 2.82, $p < .01$). The "other" cars have much higher chest g's and femur loads than the case vehicles. The difference in fatality risk between these best and worst NCAP performers is almost as great as the difference between a belted occupant and an unrestrained occupant.

## 4.5 NCAP performance of cars that did better than expected in crashes

How about crashes with astonishing outcomes - where an older driver in a smaller car walks away and the younger driver in the larger car dies? Did the car with the unanticipated good performance have lower NCAP scores? This analysis is the converse of the preceding one, comparing the NCAP scores of two cars in a head-on collision, when the driver of the case vehicle, in the actual crash, did better than expected (as evidenced by negative RELEXP). The 784 head-on collision records are ranked by RELEXP and listed in order, from the case with RELEXP = -1.89 (largest differential in favor of the case vehicle) to the case with DELNCAP = +1.89. The last 392 records, with RELEXP > 0, are deleted: they are merely the same crashes as the first 392 records, with the "case" and "other" vehicles reversed. That leaves a file of 392 distinct head-on collisions, comprising 784 distinct vehicles, in which the actual outcome, relative to expectations, is always better for the case vehicle than the other vehicle. RELEXP ranges from -1.89 for the first record to -0.004 for the last. The crashes with RELEXP more negative than -1 are the ones in which the case car driver survived even though the expected fatality risk was lower for the other driver. A negative RELEXP that is close to zero does not signify an astonishing outcome: it means that the driver with a heavy advantage did, in fact, survive; or that two cars were almost evenly matched, and both drivers died.

Table 4-2 compares the NCAP performance of the "case" and the "other"

TABLE 4-2

NCAP PERFORMANCE OF CARS WHOSE DRIVERS FARED BETTER THAN EXPECTED IN ACTUAL HEAD-ON COLLISIONS

| RELEXP Differential between the "Case" Car and the "Other" Car | N of Crashes | Mean DELNCAP | T-Test for DELNCAP | HIC Case | HIC Other | HIC T-Test | Chest G's Case | Chest G's Other | Chest G's T-Test |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean NCAP Scores ("Case" vs. "Other" Car) | | | | | |
| RELEXP < 0 | 392 | -.063 | 3.17 | 899 | 924 | 1.24 | 47.7 | 49.7 | 3.14 |
| RELEXP < -0.5 | 152 | -.132 | 4.08 | 903 | 942 | 1.05 | 46.6 | 50.3 | 3.47 |
| RELEXP < -0.75 | 88 | -.117 | 3.17 | 896 | 907 | .32 | 47.1 | 49.7 | 2.02 |
| RELEXP < -1.0 | 49 | -.128 | 3.11 | 885 | 895 | .02 | 46.2 | 48.1 | 1.32 |



Mean NCAP Scores and Model Year ("Case" vs. "Other" Car)

| RELEXP Differential between the "Case" Car and the "Other" Car | L Femur Case | L Femur Other | R Femur Case | R Femur Other | T-Test for Equal Femur Inj | Model Year Case | Model Year Other | Model Year T-Test |
|---|---|---|---|---|---|---|---|---|
| RELEXP < 0 | 935 | 940 | 944 | 932 | 1.11 | 85.5 | 85.2 | 1.42 |
| RELEXP < -0.5 | 879 | 953 | 858 | 941 | 2.05 | 86.1 | 85.0 | 3.66 |
| RELEXP < -0.75 | 884 | 959 | 829 | 1007 | 2.50 | 86.4 | 85.3 | 2.62 |
| RELEXP < -1.0 | 847 | 1000 | 791 | 1053 | 2.53 | 86.2 | 85.0 | 2.02 |

vehicles, in the crashes where RELEXP is less than zero or is more negative than some specified amount. The first line in the upper half of Table 4-2 analyzes all 392 cases in which RELEXP < 0. The average value of DELNCAP is -.063, i.e., the cars with better actual performance also had better NCAP performance, as evidenced by NCAPINJ being .063 lower. These are <u>average</u> differences; they do not mean that NCAPINJ is lower for the case vehicle in each individual crash. However, the average difference in NCAPINJ performance is statistically significant at the .01 level: a test of the hypothesis that DELNCAP = 0 yields a t value of 3.17.

The case vehicles had lower average HIC (899) than the other vehicles (924). Those are arithmetic averages of the original HIC scores. A significance test for the difference in HIC can be based on the variable DELHIC, the difference in the logistic injury probabilities. The t value for the hypothesis that DELHIC = 0 is 1.24, so the difference is not significant. But the difference in average chest g's, 47.7 vs. 49.7 is significant at the .01 level (t for DELCG is 3.14, p < .01). The first line in the lower section of Table 4-2 shows that the case and other vehicles had about the same average femur loads and model years.

There are 152 crashes in which the actual performance of the case vehicle was a fair amount better than expected, as evidenced by RELEXP < -0.5. NCAP performance of the case and the other vehicle is compared in the second line of Table 4-2. The average difference of NCAPINJ is -.132, which is more than double the value in the preceding analysis. The cars with better actual performance had lower NCAPINJ, by a highly significant amount (t for DELNCAP is 4.08, p < .0001). Moreover, the cars with better actual performance had better

NCAP results on <u>every</u> body region. HIC averaged 903 for the case vehicles and 942 for the other vehicles; chest g's averaged 46.6 for the case vehicles and 50.3 for the other vehicles; femur loads averaged about 80 pounds lower on both legs. The difference in chest g's is significant at the .01 level, and for femur load, at the .05 level. The cars with better actual performance had a significantly more recent model year (86.1 vs. 85.0, $t = 3.66$, $p < .001$).

RELEXP was more negative than -0.75 in about one-fourth of the crashes. The difference in NCAPINJ (-.117) is about the same as in the preceding analysis, and it is statistically significant (t for DELNCAP is 3.17, $p < .01$). However, the difference in HIC and chest g's decreased from the preceding cases, while the gap in femur load increased. Finally, in the 49 crashes with really surprising outcomes, as evidenced by RELEXP < -1.0, NCAPINJ is once again significantly lower in the cars with better actual performance (DELNCAP = -.128, $t = 3.11$, $p < .01$). HIC is about the same in both cars; chest g's are lower in the case cars, on the average, by 2. Femur loads are significantly lower in the case cars, by an average of 150 pounds on the left side and 260 pounds on the right side. The average difference in model year remains close to one year.

The four analyses in Table 4-2 are strong evidence that cars with better-than-expected actual performance, as evidenced by negative RELEXP, had significantly better NCAP performance than the cars they hit. That, by itself, is not really a new finding, since a highly significant correlation between RELEXP and DELNCAP was already shown in Section 4.3. This analysis, however, reveals some traits of the relationship.

First, the average differences in NCAP performance, although statistically significant, are not vast in absolute terms. Average HIC differed by 40 or less, chest g's by 2-4 and femur load by no more than 250 pounds. Who survives and who dies in a specific head-on collision, depends on many factors besides vehicle performance as measured in an NCAP test; it depends a lot on the personal vulnerability to injury of the individual occupants and the unique circumstances that may be present in that crash. Nevertheless, on the average, the cars with better actual performance had lower NCAP scores.

Second, the composite performance measure NCAPINJ had a stronger relationship with actual performance than did any of the NCAP scores for individual body regions. In all four analyses, the difference in NCAPINJ was significant at the .01 level, with t-values always over 3. Although HIC, chest g's and femur loads were consistently lower, on the average, in the safer case vehicles, the differences were not always statistically significant, and only rarely at the .01 level (chest g's in the first two analyses).

## 4.6 Fatality reduction for the car with lower head or chest injury risk

NHTSA's December 1993 Report to Congress contains an analysis similar to the approach in Section 4.4, but with the collision records ranked by head or chest injury risk, rather than NCAPINJ [24], p. 72. The analysis is called "Case A" in the Report to Congress. The accident data file contains 740 head-on collision records in which NCAP HIC and chest g's are known for both drivers. The **maximum head/chest injury** for a driver is the greater of the logistic injury probability functions, HEADINJ and CHESTINJ (see Section 4.1):

$$\text{MAXHCINJ} = \max(\text{HEADINJ}, \text{CHESTINJ})$$

The performance for the case vehicle relative to the other vehicle is

$$\text{DELMAXHC} = \text{MAXHCINJ}_{\text{CASE}} - \text{MAXHCINJ}_{\text{OTHER}}$$

The records were ranked by DELMAXHC, starting with the crash having the largest differential in favor of the case vehicle. The 12 records in the middle, with DELMAXHC = 0, and the last 384 records, with DELMAXHC > 0, are deleted, as in Section 4.4. That leaves a file of 364 distinct head-on collisions, comprising 728 distinct vehicles, in which the NCAP head/chest injury performance of the case vehicle is always better than the performance of the other vehicle.

In the 364 collisions with DELMAXHC < 0, the average NCAP scores for the "case" and "other" vehicles, and the actual and expected fatality counts were as follows:

| | **Case Vehicle** | **Other Vehicle** |
|---|---|---|
| Average HIC | 721 | 1111 |
| Chest g's | 45.0 | 52.7 |
| Left femur load | 1012 | 895 |
| Right femur load | 1002 | 902 |
| Actual fatalities | 199 | 228 |
| Expected fatalities | 207.8 | 217.4 |
| Fatality reduction (%) | 8.7 | |
| Mean RELEXP | - .053 | |
| T-test for RELEXP | 1.62 | |

Chest g's and, especially, HIC are lower, on the average, in the "case" vehicles. Femur loads, which are not factored into the calculation of MAXHCINJ, are actually slightly higher in the case vehicles. The reduction in actual fatality risk, adjusted for vehicle weight, driver age and sex, is

$$1 - [(199/228) / (207.8/217.4)] = 8.7 \text{ percent}$$

in the case vehicles, and it is not statistically significant, although it comes close to significance (t = 1.62).

4.7 Sensitivity test: NCAPINJ on a different calibration data set

NCAPINJ was the specific weighted average of head, chest and femur scores that had maximum correlation with fatality risk on the basic calibration data set of FARS cases in which both cars match with an NCAP test at levels 3A or 4A. In Section 4.1, regression analyses with that data set showed excellent correlations of NCAP scores and fatality risk even when the relative weights for the three body regions diverged substantially from .21, 2.7 and 1.5, the weights in NCAPINJ. As an additional sensitivity test, these regressions can also be run on a subset of the FARS cases: where each FARS car not only matches up with an NCAP car at level 3A or 4A, but also must have the **same number of doors** as the NCAP car. In Section 3.4, it was shown that the regression analysis with DELHIC, DELCG and DELFEM as separate variables assigned them regression coefficients of .42, 2.52 and .73. Thus, the optimum composite measure of injury for this data set would have relative weights for the three body regions close to these values - i.e., a slightly higher weight for HIC and a lower weight for femur load than in the full data set.

The new calibration data set contains 620 cases in which NCAP scores are known on at least two body regions for each vehicle, and 598 cases in which the scores are known for the chest and at least one other body region. The second series of regressions in Section 3.1, which compared the $\chi^2$ for the NCAPINJ with the original optimum weights (.21, 2.7 and 1.5), for an unweighted injury function, and for three intermediate injury functions, is rerun for the new calibration data set. A regression is also run for a composite injury function based on the new optimum weights (.42, 2.52 and .73):

| Relative Weights | | | Chi-Squares | |
|---|---|---|---|---|
| Head Injury | Chest Injury | Femur Injury | 620 Cases | 598 Cases |
| 0.21 | 2.70 | 1.50 | 10.80 | 8.43 |
| 0.31 | 2.10 | 1.36 | 10.73 | 8.52 |
| 0.46 | 1.64 | 1.22 | 10.41 | 8.47 |
| 0.68 | 1.28 | 1.11 | 9.31 | 7.78 |
| 1 | 1 | 1 | 7.49 | 6.46 |
| 0.42 | 2.52 | 0.73 | 11.24 | 9.61 |

All $\chi^2$ are statistically significant at the .01 level (except the regression on 598 cases with the unweighted injury function, which is significant at the .05 level). The last regression produces the highest $\chi^2$. The original NCAPINJ is not the optimal injury function on this calibration data set, although its $\chi^2$ is just slightly less than the maximum values. The drop-off in $\chi^2$, as the regressions proceed from the original NCAPINJ to the unweighted injury function, are less precipitous than on the original calibration data set. Here, they drop by about 30 percent; there, the descent was closer to 50 percent.

The sensitivity tests confirm that <u>any</u> reasonable combination of head, chest and femur injury scores, not just NCAPINJ, will correlate well with fatality risk. While the FARS sample is adequate to show that chest g's need to be given a substantial weight in any composite score, the accident sample is not really large enough to determine exactly the relative importance of the head and the femur injury scores.

# CHAPTER 5

## COLLISIONS BETWEEN A "GOOD" AND A "POOR" NCAP PERFORMER

Probably, the simplest way to estimate the fatality reduction associated with good NCAP scores is to partition the cars based on a specific NCAP score - e.g., HIC, chest g's, femur load or a composite score such as NCAPINJ - and to consider only the subset of head-on collisions in which the case vehicle has a score in the "good" range and the other vehicle has a score in the "poor" range. Do the cars with the poor NCAP scores have significantly more driver fatalities than expected, after control for the curb weight, driver age and sex?

Most of the analyses in this chapter are based on the data set of FARS head-on collisions between two passenger cars in which both cars matched up with an NCAP test vehicle at levels "3A" or "4A": the model year on FARS is within the range of model years considered valid for the NCAP test, and the make-models on FARS and NCAP are identical or true corporate cousins. That data set includes 396 head-on collisions (792 vehicles). As a sensitivity test, some of the analyses are repeated, in Section 5.9, on the subset of head-on collisions in which both vehicles not only match an NCAP test vehicle at levels 3A or 4A, but also have the same number of doors as the NCAP test vehicle (310 collisions, 620 cars).

### 5.1 Cars with low NCAPINJ hit cars with high NCAPINJ

A composite measure of NCAP performance, NCAPINJ was defined in Section 4.1, as a weighted average of logistic injury probability functions for the head, chest and femurs. The weights were chosen to maximize the correlation of NCAPINJ with fatality risk in the principal calibration data set (crashes where both vehicles match an NCAP test at the '3A or 4A level') but are not necessarily optimal for other accident data sets (e.g., crashes where both vehicles match an NCAP test at the '3A or 4A level and N of doors'). In Section 4.3, 384 head-on collisions were identified in which the case vehicle had a lower (better) NCAPINJ score than the other vehicle. The fatality risk was a statistically significant 15 percent lower in the case vehicles. In that analysis, the NCAPINJ for the case vehicle did not have to be below any specified level, nor did the NCAPINJ for the other vehicle have to be above some specified level: it was only required that the case vehicle did better than the other vehicle. Thus, the set of 384 collisions includes some where both vehicles did quite well (in absolute terms) and some where both did poorly.

The approach of this chapter is generate subsets of the 384 collisions in which all the case vehicles had "good" NCAP performance: better than some specified level A. All the other vehicles had "poor" performance: higher than another specified level B, where B $\geq$ A. By eliminating the cases where both cars did well, or both did poorly, there should be an even larger differentiation of fatality risk between the case and the other vehicles.

Table 5-1 presents the results of nine analyses for the special case where B = A; i.e., there is a single boundary between "good" and "poor" performance. All cars with NCAPINJ lower than the boundary are "good" and all above it are "poor." The nine analyses use boundary values of 0.2, 0.3, ... , 1.0, respectively. In every analysis, the fatality risk is significantly lower in the good NCAP performers than in the poor performers, as evidenced by t-test results greater than 1.65 ($p < .05$). The fatality reduction for a good NCAPINJ

TABLE 5-1

COLLISIONS OF CARS WITH LOW NCAPINJ INTO CARS WITH HIGH NCAPINJ:
EFFECT OF MOVING THE BOUNDARY BETWEEN "LOW" AND "HIGH" NCAPINJ

| Definition of "Low" NCAPINJ | Definition of "High" NCAPINJ | N of Crashes | Comparison of Real-World Performance: % Fat Red for Low NCAPINJ | Mean Relexp | Sum Relexp | T-Test for Relexp |
|---|---|---|---|---|---|---|
| NCAPINJ ≤ .2 | NCAPINJ > .2 | 115 | 19.1 | -.112 | -12.91 | 1.83 |
| NCAPINJ ≤ .3 | NCAPINJ > .3 | 186 | 11.6 | -.072 | -13.47 | 1.68 |
| NCAPINJ ≤ .4 | NCAPINJ > .4 | 186 | 17.0 | -.108 | -20.17 | 2.56 |
| NCAPINJ ≤ .5 | NCAPINJ > .5 | 147 | 17.3 | -.114 | -16.83 | 2.44 |
| **NCAPINJ ≤ .6** | **NCAPINJ > .6** | **117** | **26.4** | **-.181** | **-21.13** | **3.22** |
| NCAPINJ ≤ .7 | NCAPINJ > .7 | 108 | 27.2 | -.189 | -20.46 | 3.34 |
| NCAPINJ ≤ .8 | NCAPINJ > .8 | 80 | 28.7 | -.201 | -16.10 | 3.13 |
| NCAPINJ ≤ .9 | NCAPINJ > .9 | 55 | 31.9 | -.216 | -11.88 | 2.73 |
| NCAPINJ ≤ 1.0 | NCAPINJ > 1.0 | 27 | 41.2 | -.279 | - 7.53 | 2.84 |

Mean NCAP Scores and Model Year ("Low" vs. "High" NCAPINJ Car)

| Definition of "Low" NCAPINJ | Definition of "High" NCAPINJ | HIC | | Chest G's | | L Femur | | R Femur | | Model Year | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Low | High | Low | High | Low | High | Low | High | Low | High |
| NCAPINJ ≤ .2 | NCAPINJ > .2 | 647 | 952 | 36.0 | 50.3 | 721 | 941 | 694 | 993 | 85.9 | 85.4 |
| NCAPINJ ≤ .3 | NCAPINJ > .3 | 709 | 1064 | 40.4 | 53.5 | 825 | 1044 | 749 | 1078 | 85.9 | 85.4 |
| NCAPINJ ≤ .4 | NCAPINJ > .4 | 801 | 1070 | 42.1 | 55.7 | 816 | 1121 | 775 | 1136 | 85.8 | 85.0 |
| NCAPINJ ≤ .5 | NCAPINJ > .5 | 852 | 1103 | 43.9 | 59.8 | 908 | 1119 | 901 | 1031 | 85.8 | 84.3 |
| **NCAPINJ ≤ .6** | **NCAPINJ > .6** | **898** | **1106** | **45.0** | **62.6** | **878** | **1161** | **846** | **1117** | **85.8** | **83.1** |
| NCAPINJ ≤ .7 | NCAPINJ > .7 | 922 | 1103 | 45.9 | 62.7 | 866 | 1226 | 865 | 1164 | 85.5 | 83.2 |
| NCAPINJ ≤ .8 | NCAPINJ > .8 | 942 | 1090 | 47.5 | 64.1 | 940 | 1179 | 953 | 1058 | 85.0 | 83.8 |
| NCAPINJ ≤ .9 | NCAPINJ > .9 | 962 | 1069 | 47.8 | 63.8 | 947 | 1487 | 951 | 1294 | 84.9 | 83.2 |
| NCAPINJ ≤ 1.0 | NCAPINJ > 1.0 | 1026 | 908 | 49.0 | 62.4 | 979 | 1979 | 974 | 1200 | 85.0 | 84.0 |

score, relative to a poor score, ranges from 12 percent when 0.3 is the boundary value to 41 percent when 1.0 is the boundary value. In general, the higher the boundary value, the greater the fatality reduction for good vs. poor performance. However, the analyses with high boundary values have sharply reduced sample sizes, because there are few cars on the file which had really high NCAPINJ results. These last analyses only provide information about a fraction of the cars on the file; they don't say much about the "typical" car on the road.

The ideal analysis should combine a large fatality reduction and a large sample size. The variable "sum RELEXP," which is the product of mean RELEXP and sample size, intuitively describes the total "information" provided by an analysis. Sum RELEXP reaches a maximum of 21.13 when "good" NCAP performance is defined as NCAPINJ $\leq$ 0.6 and "poor" NCAP performance is defined as NCAPINJ > 0.6. There are 117 head-on collisions of a "good" performer with a "poor" performer, in which both drivers are belted. In the 117 cars with NCAPINJ > 0.6, 77 drivers died, whereas only 65.5 fatalities were expected, based on car weight, driver age and sex. In the 117 cars with NCAPINJ $\leq$ 0.6, there were 62 actual and 71.6 expected driver fatalities. (The good performers weighed almost the same as the poor performers - 2868 vs. 2869 pounds, on the average, but the drivers of the low-NCAPINJ cars were older than the drivers of the high-NCAPINJ cars - 44.7 vs. 40.5 years, on the average; thus, the expected fatalities are slightly higher in the cars with low NCAPINJ.) The fatality risk is

$$1 - [(62/77) \ / \ (71.6/65.5)] = 26 \text{ percent}$$

lower in the good performers than in the poor performers, after controlling for vehicle weight, driver age and sex (t for RELEXP is 3.22, $p < .001$).

Of course, even in these accident samples tailored to highlight the safety benefits associated with good NCAP scores, the relationship between the NCAPINJ and fatality risk over the range of head-on collisions experienced on the highway is not perfect. Merely having the lower composite NCAP score of the two cars in the collision does not guarantee survival, even if the two cars are of the same weight and the drivers of the same age and sex. Yet, <u>on the average</u>, in collisions between cars with NCAPINJ $\leq$ 0.6 and cars with NCAPINJ > 0.6, the driver of the car with the better composite NCAP score had 26 percent less fatality risk than the driver of the car with the poorer NCAP score, even after controlling for weight, age and sex.

The sample size of 117 is about a third of the 384 cases considered in Section 4.3. Although it seems small, it is close to the "ideal" sample size for the analyses of this chapter, whose technique is to exclude the crashes between two "good" cars or two "poor" cars, and include only the crashes between a "good" and a "poor" car. If exactly half the cars had "poor" NCAP performance, that would eliminate exactly half the crashes, leaving a sample of 190. Intuitively, though, substantially less than half the cars can be called really "poor" performers on NCAP. If the "poor" performers are the worst 20 percent or so, and the "acceptable" performers are the best 70-80 percent (with perhaps 10 percent in a borderline area), the file of 384 crashes can be expected to contain about 110-130 collisions between an "acceptable" and a "poor" performer. In general, the objective in this chapter is to find boundary values between "acceptable" and "poor" performance that <u>maximize the fatality reduction for "acceptable" relative to "poor" performance while maintaining a sample size close to the target of 120 crashes</u>.

The lower half of Table 5-1 compares the average scores of "good" and "poor" NCAPINJ performers on the individual NCAP body regions. Since NCAPINJ is a weighted sum of injury probabilities for all of the body regions, the cars with NCAPINJ ≤ 0.6 have, on the average, lower HIC than the cars with NCAPINJ > 0.6 (898 vs. 1106), also lower chest g's (45 vs. 62.6), and lower femur loads (878 vs. 1161 on the left; 846 vs. 1117 on the right). Similar patterns are found in the other analyses, except when the boundary value is 1.0, where the sample is quite small. As the boundary values rise, so do HIC, chest g's and femur loads, for both the "acceptable" and the "poor" groups. As NCAPINJ rises above 0.7, though, HIC and chest g's tend to level off, while femur loads continue to escalate. Reflecting the trend of improvement in NCAP results during 1979-91, the average model year for the "good" performers ranges from 0.4 to 2.7 years more recent than for the "poor" performers.

Table 5-2 shows what happens when a "gap" or borderline area is placed between "low" and "high" NCAPINJ. More and more crashes drop out of the sample, as one or both cars have NCAPINJ in the borderline region. Table 5-2 starts with a single boundary of NCAPINJ = 0.6 (the "best" analysis in Table 5-1) and expands, by 0.1 at a time, the distance between the lower and upper boundary values. For example, in the second analysis of Table 5-2, the "good" cars are the ones with NCAPINJ ≤ 0.5 and the "poor" ones have NCAPINJ > 0.6. As the sample size drops from 117 to 22, the fatality reduction for good performance rises from 26 to 57 percent. Although the analyses with the larger gaps have impressive fatality reductions and high statistical significance (t values as high as 4.11), they don't really mean as much as the analysis without a gap, as evidenced by steadily declining sum RELEXP. In the lower half of Table 5-2, average scores for the individual body regions indicate that the analyses with big gaps compare really good all-around NCAP performers to really poor all-around

TABLE 5-2

COLLISIONS OF CARS WITH LOW NCAPINJ INTO CARS WITH HIGH NCAPINJ:
EFFECT OF PLACING A GAP BETWEEN "LOW" AND "HIGH" NCAPINJ

| | | Comparison of Real-World Performance | | | | |
|---|---|---|---|---|---|---|
| Definition of "Low" NCAPINJ | Definition of "High" NCAPINJ | N of Crashes | % Fat Red for Low NCAPINJ | Mean Relexp | Sum Relexp | T-Test for Relexp |
| NCAPINJ ≤ .6 | NCAPINJ > .6 | 117 | 26.4 | -.181 | -21.13 | 3.22 |
| NCAPINJ ≤ .5 | NCAPINJ > .6 | 102 | 24.1 | -.165 | -16.78 | 2.99 |
| NCAPINJ ≤ .5 | NCAPINJ > .7 | 92 | 26.1 | -.181 | -16.66 | 3.28 |
| NCAPINJ ≤ .4 | NCAPINJ > .7 | 71 | 29.6 | -.208 | -14.78 | 3.28 |
| NCAPINJ ≤ .4 | NCAPINJ > .8 | 45 | 38.2 | -.276 | -12.40 | 4.11 |
| NCAPINJ ≤ .3 | NCAPINJ > .8 | 34 | 42.9 | -.282 | - 9.60 | 3.44 |
| NCAPINJ ≤ .3 | NCAPINJ > .9 | 22 | 57.0 | -.370 | - 8.15 | 3.72 |

| | | Mean NCAP Scores and Model Year ("Low" vs. "High" NCAPINJ Car) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Definition of | Definition of | HIC | | Chest G's | | L Femur | | R Femur | | Model Year | |
| "Low" NCAPINJ | "High" NCAPINJ | Low | High | Low | High | Low | High | Low | High | Low | High |
| NCAPINJ ≤ .6 | NCAPINJ > .6 | 898 | 1106 | 45.0 | 62.6 | 878 | 1161 | 846 | 1117 | 85.8 | 83.1 |
| NCAPINJ ≤ .5 | NCAPINJ > .6 | 846 | 1093 | 43.6 | 62.3 | 891 | 1198 | 873 | 1111 | 85.7 | 83.2 |
| NCAPINJ ≤ .5 | NCAPINJ > .7 | 864 | 1093 | 44.1 | 62.4 | 880 | 1247 | 887 | 1117 | 85.7 | 83.2 |
| NCAPINJ ≤ .4 | NCAPINJ > .7 | 857 | 1102 | 42.3 | 62.8 | 783 | 1195 | 771 | 1187 | 85.5 | 83.0 |
| NCAPINJ ≤ .4 | NCAPINJ > .8 | 850 | 1086 | 42.8 | 64.3 | 786 | 1166 | 793 | 1079 | 85.1 | 82.9 |
| NCAPINJ ≤ .3 | NCAPINJ > .8 | 748 | 1064 | 41.8 | 63.6 | 777 | 1152 | 727 | 1098 | 85.4 | 82.9 |
| NCAPINJ ≤ .3 | NCAPINJ > .9 | 733 | 1207 | 41.6 | 62.8 | 813 | 1399 | 746 | 1327 | 85.1 | 82.6 |

NCAP performers. Crashes of that type are rare, but the advantage is strongly with the good NCAP performer.

5.2 Cars with low NCAP chest g's hit cars with high chest g's

The composite score NCAPINJ is quite efficient for partitioning the cars into a "safer" and a "less-safe" group, as evidenced by the 26 percent lower fatality risk for cars with NCAPINJ $\leq$ 0.6 in 117 crashes where they hit cars with NCAPINJ > 0.6. NCAPINJ is a weighted sum of NCAP scores for three body regions. Do the NCAP scores for any single body region have comparable efficiency for identifying differences in actual safety performance?

Chest g's, which had significant correlation with actual fatality risk throughout Chapters 3 and 4 and are the largest component of NCAPINJ, are a reliable single parameter for partitioning the cars into safer and less-safe groups. Table 5-3 describes 14 analyses, each using a different single boundary between "good" and "poor" chest g's. The boundary ranges from 42 to 68 g's. For example, 60 chest g's, the maximum value allowed by Federal Motor Vehicle Safety Standard (FMVSS) 208, is used as the boundary in one of the analyses. There are 92 actual head-on collisions (both drivers belted) in which one of models had > 60 chest g's for the driver when it was tested in NCAP, and the other had $\leq$ 60 g's. The fatality risk is a statistically significant 24 percent lower in the cars with $\leq$ 60 g's than in the cars with > 60 g's (t for RELEXP is 2.74, $p < .01$). However, 60 g's, the pass-fail value in FMVSS 208 is just one possible boundary. All of the other cutoff points from 42 to 68 g's, except 48 and 50, produce statistically significant differences, as evidenced by t values greater than 1.65. The boundary value that yields a sample size closest to the target of 120 crashes is 56 g's: the fatality reduction for the "good" cars is 19

TABLE 5-3

COLLISIONS OF CARS WITH "GOOD" NCAP CHEST G SCORES INTO CARS WITH "POOR" CHEST G's
EFFECT OF MOVING THE BOUNDARY BETWEEN "GOOD" AND "POOR" NCAP CHEST G's

Comparison of Real-World Performance

| Definition of a "Good" Car | Definition of a "Poor" Car | N of Crashes | % Fat Red for Good Car | Mean Relexp | Sum Relexp | T-Test for Relexp |
|---|---|---|---|---|---|---|
| Chest g's ≤ 42 | Chest g's > 42 | 162 | 15.8 | -.089 | -14.49 | 1.85 |
| Chest g's ≤ 44 | Chest g's > 44 | 172 | 16.2 | -.099 | -17.10 | 2.26 |
| Chest g's ≤ 46 | Chest g's > 46 | 183 | 16.9 | -.107 | -19.58 | 2.51 |
| Chest g's ≤ 48 | Chest g's > 48 | 182 | 10.9 | -.068 | -12.32 | 1.56 |
| Chest g's ≤ 50 | Chest g's > 50 | 166 | 7.9 | -.050 | - 8.22 | 1.09 |
| Chest g's ≤ 52 | Chest g's > 52 | 145 | 15.0 | -.098 | -14.22 | 2.02 |
| Chest g's ≤ 54 | Chest g's > 54 | 130 | 17.0 | -.112 | -14.56 | 2.20 |
| **Chest g's ≤ 56** | **Chest g's > 56** | **125** | **18.7** | **-.123** | **-15.42** | **2.32** |
| Chest g's ≤ 58 | Chest g's > 58 | 94 | 22.0 | -.150 | -14.12 | 2.49 |
| Chest g's ≤ 60 | Chest g's > 60 | 92 | 24.2 | -.166 | -15.28 | 2.74 |
| Chest g's ≤ 62 | Chest g's > 62 | 66 | 24.9 | -.172 | -11.33 | 2.50 |
| Chest g's ≤ 64 | Chest g's > 64 | 49 | 29.7 | -.209 | -10.23 | 2.60 |
| Chest g's ≤ 66 | Chest g's > 66 | 48 | 29.6 | -.210 | -10.06 | 2.56 |
| Chest g's ≤ 68 | Chest g's > 68 | 43 | 25.9 | -.189 | - 8.13 | 2.06 |

Mean NCAP Scores and Model Year ("Good" vs. "Poor" Car)

| Definition of a "Good" Car | Definition of a "Poor" Car | HIC | | Chest G's | | L Femur | | R Femur | | Model Year | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Good | Poor | Good | Poor | Good | Poor | Good | Poor | Good | Poor |
| Chest g's ≤ 42 | Chest g's > 42 | 752 | 1021 | 38.5 | 54.1 | 947 | 969 | 848 | 978 | 86.1 | 85.0 |
| Chest g's ≤ 44 | Chest g's > 44 | 750 | 1039 | 39.2 | 55.0 | 964 | 955 | 851 | 1014 | 86.3 | 84.8 |
| Chest g's ≤ 46 | Chest g's > 46 | 782 | 1022 | 39.8 | 55.4 | 1003 | 943 | 891 | 1018 | 86.1 | 84.7 |
| Chest g's ≤ 48 | Chest g's > 48 | 818 | 1044 | 41.4 | 57.7 | 977 | 936 | 944 | 948 | 85.8 | 84.7 |
| Chest g's ≤ 50 | Chest g's > 50 | 829 | 1069 | 42.7 | 59.5 | 968 | 933 | 949 | 937 | 85.9 | 84.2 |
| Chest g's ≤ 52 | Chest g's > 52 | 855 | 1094 | 43.0 | 61.3 | 959 | 925 | 942 | 963 | 86.1 | 83.9 |
| Chest g's ≤ 54 | Chest g's > 54 | 902 | 1091 | 44.1 | 62.8 | 923 | 958 | 928 | 946 | 85.7 | 83.7 |
| **Chest g's ≤ 56** | **Chest g's > 56** | **914** | **1088** | **44.6** | **63.4** | **949** | **964** | **945** | **930** | **85.8** | **83.5** |
| Chest g's ≤ 58 | Chest g's > 58 | 896 | 1135 | 45.7 | 65.8 | 978 | 876 | 970 | 817 | 85.4 | 83.2 |
| Chest g's ≤ 60 | Chest g's > 60 | 896 | 1145 | 46.0 | 66.1 | 970 | 864 | 953 | 839 | 85.3 | 83.2 |
| Chest g's ≤ 62 | Chest g's > 62 | 932 | 1063 | 47.5 | 68.2 | 1008 | 826 | 997 | 696 | 85.0 | 83.8 |
| Chest g's ≤ 64 | Chest g's > 64 | 995 | 893 | 48.3 | 70.1 | 1016 | 831 | 1001 | 674 | 85.1 | 83.3 |
| Chest g's ≤ 66 | Chest g's > 66 | 993 | 873 | 48.1 | 70.2 | 1010 | 831 | 1002 | 667 | 85.2 | 83.4 |
| Chest g's ≤ 68 | Chest g's > 68 | 1023 | 844 | 49.1 | 70.6 | 1062 | 707 | 1010 | 671 | 85.1 | 83.7 |

percent (t for RELEXP is 2.32, $p < .05$).

The lower half of Table 5-3 compares the average NCAP performance of the cars with "good" and "poor" chest g's. Needless to say, there is a large difference in the average chest g's. The difference is about 16 g's in the first four analyses, and gradually gets larger as the boundary value increases, eventually reaching 22 g's. However, the "good" cars are also better, in many cases, on the other body regions. As noted in Section 3.3, chest g's have a strong correlation coefficient of .281 with HIC, and weaker correlations of .162 with right femur load and .062 with left femur load. Thus, the cars with the higher chest g's also tend to have higher HIC, by about 200-250, on the average, in all of the analyses with boundary values up to 62 g's. The femur loads are about the same, or slightly higher in the high-chest g cars, up to the analysis with a boundary of 56 chest g's. But, above those boundary values, the pattern reverses. The small groups of cars with very high chest g's tend to compensate for it with lower HIC and femur loads than their counterparts with low chest g's. Reflecting the trend of improvement in NCAP results during 1979-91, the average model year for the "good" performers ranges from 1.1 to 2.3 years more recent than for the "poor" performers.

On the whole, chest g's are not as efficient as NCAPINJ for discriminating safer and less-safe cars, as evidenced by a comparison of Tables 5-1 and 5-3. For accident samples of comparable size, the fatality reduction for low NCAPINJ is consistently greater than the reduction for low chest g's - e.g., 26 vs. 19 percent at the target sample size of 120 crashes. Sum RELEXP exceeds 20 three times in Table 5-1 and never in Table 5-3. The t value is always significant and goes above 3 in three NCAPINJ analyses; but in the chest g analyses it is usually close to 2 and is nonsignificant in two cases. In almost every analysis of Table 5-1, the cars with higher NCAPINJ had, on the average, higher NCAP scores on all three body regions; in Table 5-3, high chest g's were not always accompanied by high HIC and often coincided with low femur loads.

Table 5-4 shows that the fatality reduction for good vs. poor chest g's can be magnified by placing a gap between the "good" and the "poor" groups. The results parallel earlier findings for NCAPINJ (Table 5-2). When the gap is 8 g's or more, the fatality reductions for the "good" cars approach 40 percent, with high statistical significance (t values of 4 or more). However, the samples of crashes are quite limited; these analyses really don't say much about the relationship between chest g's and fatality risk in the overall vehicle fleet.

5.3 <u>Cars with low NCAP HIC hit cars with high HIC</u>

Table 5-5 presents 30 analyses that partition cars into "good" and "poor" groups based on the Head Injury Criterion (HIC). Some of them use a single boundary between "good" and "poor" HIC, while others have two boundary values and gap. A HIC of 1000 is the maximum amount permitted by FMVSS 208; boundary values ranging from 800 to 1600 are considered in Table 5-5. The analyses are ordered by the lower boundary value and by the size of the gap.

Table 5-5 shows that HIC is moderately reliable, by itself, for identifying differences in actual safety performance. In 29 of the 30 analyses, the cars with low HIC have lower fatality risk than the high-HIC cars they collided with, and the reduction is statistically significant in 5 of the analyses ($t > 1.65$, $p < .05$). The comparison that maximizes fatality reduction with a sample size close to 120 defines HIC $\leq$ 1000 as a "good" car and HIC > 1200

TABLE 5-4

COLLISIONS OF CARS WITH "GOOD" NCAP CHEST G SCORES INTO CARS WITH "POOR" CHEST G's
EFFECT OF PLACING A GAP BETWEEN "GOOD" AND "POOR" NCAP CHEST G's

**Comparison of Real-World Performance**

| Definition of a "Good" Car | Definition of a "Poor" Car | N of Crashes | % Fat Red for Good Car | Mean Relexp | Sum Relexp | T-Test for Relexp |
|---|---|---|---|---|---|---|
| Chest g's ≤ 60 | Chest g's > 60 | 92 | 24.2 | -.166 | -15.28 | 2.74 |
| Chest g's ≤ 58 | Chest g's > 60 | 89 | 22.8 | -.156 | -13.92 | 2.51 |
| Chest g's ≤ 58 | Chest g's > 62 | 60 | 26.3 | -.184 | -11.03 | 2.50 |
| Chest g's ≤ 56 | Chest g's > 62 | 56 | 32.9 | -.242 | -13.58 | 3.70 |
| Chest g's ≤ 56 | Chest g's > 64 | 40 | 37.9 | -.297 | -11.88 | 4.11 |
| Chest g's ≤ 54 | Chest g's > 64 | 39 | 40.7 | -.325 | -12.67 | 4.75 |
| Chest g's ≤ 54 | Chest g's > 66 | 39 | 40.7 | -.325 | -12.67 | 4.75 |

**Mean NCAP Scores and Model Year ("Good" vs. "Poor" Car)**

| Definition of a "Good" Car | Definition of a "Poor" Car | HIC | | Chest G's | | L Femur | | R Femur | | Model Year | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Good | Poor | Good | Poor | Good | Poor | Good | Poor | Good | Poor |
| Chest g's ≤ 60 | Chest g's > 60 | 896 | 1145 | 46.0 | 66.1 | 970 | 864 | 953 | 839 | 85.3 | 83.2 |
| Chest g's ≤ 58 | Chest g's > 60 | 895 | 1147 | 45.6 | 66.2 | 977 | 882 | 967 | 834 | 83.4 | 83.3 |
| Chest g's ≤ 58 | Chest g's > 62 | 913 | 1088 | 46.1 | 68.4 | 1016 | 839 | 1009 | 631 | 85.4 | 83.7 |
| Chest g's ≤ 56 | Chest g's > 62 | 905 | 1108 | 45.3 | 68.3 | 994 | 856 | 981 | 621 | 85.6 | 83.6 |
| Chest g's ≤ 56 | Chest g's > 64 | 979 | 907 | 45.7 | 70.2 | 967 | 801 | 950 | 579 | 85.6 | 82.8 |
| Chest g's ≤ 54 | Chest g's > 64 | 982 | 914 | 45.4 | 70.2 | 970 | 822 | 944 | 594 | 85.6 | 82.8 |
| Chest g's ≤ 54 | Chest g's > 66 | 982 | 914 | 45.4 | 70.2 | 970 | 822 | 944 | 594 | 85.6 | 82.8 |

TABLE 5-5

COLLISIONS OF CARS WITH "GOOD" NCAP HIC SCORES INTO CARS WITH "POOR" HIC

| | | | Comparison of Real-World Performance | | | |
|---|---|---|---|---|---|---|
| Definition of a "Good" Car | Definition of a "Poor" Car | N of Crashes | % Fat Red for Good Car | Mean Relexp | Sum Relexp | T-Test for Relexp |
| HIC ≤ 800 | HIC > 800 | 182 | 1.3 | -.006 | - 1.14 | .13 |
| HIC ≤ 800 | HIC > 900 | 133 | 4.2 | -.024 | - 3.26 | .44 |
| HIC ≤ 800 | HIC > 1000 | 85 | 2.6 | -.016 | - 1.33 | .22 |
| HIC ≤ 800 | HIC > 1100 | 73 | none | +.007 | + .52 | n.a. |
| HIC ≤ 800 | HIC > 1200 | 59 | 14.1 | -.090 | - 5.29 | 1.21 |
| HIC ≤ 900 | HIC > 900 | 190 | 10.7 | -.066 | -12.53 | 1.39 |
| HIC ≤ 900 | HIC > 1000 | 127 | 10.7 | -.067 | - 8.47 | 1.23 |
| HIC ≤ 900 | HIC > 1100 | 109 | 8.0 | -.049 | - 5.34 | .86 |
| HIC ≤ 900 | HIC > 1200 | 92 | 19.0 | -.126 | -11.55 | 2.13 |
| HIC ≤ 900 | HIC > 1300 | 68 | 26.5 | -.181 | -12.31 | 2.61 |
| HIC ≤ 1000 | HIC > 1000 | 155 | 10.4 | -.064 | - 9.99 | 1.30 |
| HIC ≤ 1000 | HIC > 1100 | 133 | 7.8 | -.048 | - 6.36 | .91 |
| HIC ≤ 1000 | HIC > 1200 | 113 | 14.2 | -.090 | -10.22 | 1.68 |
| HIC ≤ 1000 | HIC > 1300 | 81 | 20.2 | -.130 | -10.56 | 2.03 |
| HIC ≤ 1000 | HIC > 1400 | 58 | 15.8 | -.102 | - 5.91 | 1.25 |
| HIC ≤ 1100 | HIC > 1100 | 139 | 6.2 | -.037 | - 5.15 | .72 |
| HIC ≤ 1100 | HIC > 1200 | 118 | 13.1 | -.082 | - 9.69 | 1.54 |
| HIC ≤ 1100 | HIC > 1300 | 85 | 17.1 | -.107 | - 9.13 | 1.70 |
| HIC ≤ 1100 | HIC > 1400 | 62 | 11.6 | -.072 | - 4.49 | .92 |
| HIC ≤ 1100 | HIC > 1500 | 48 | 13.2 | -.084 | - 4.01 | .92 |
| HIC ≤ 1200 | HIC > 1200 | 120 | 10.8 | -.067 | - 8.03 | 1.25 |
| HIC ≤ 1200 | HIC > 1300 | 87 | 14.0 | -.086 | - 7.48 | 1.35 |
| HIC ≤ 1200 | HIC > 1400 | 64 | 7.3 | -.044 | - 2.83 | .56 |
| HIC ≤ 1200 | HIC > 1500 | 50 | 7.7 | -.047 | - 2.35 | .52 |
| HIC ≤ 1300 | HIC > 1300 | 91 | 12.0 | -.071 | - 6.50 | 1.12 |
| HIC ≤ 1300 | HIC > 1400 | 66 | 8.5 | -.051 | - 3.34 | .66 |
| HIC ≤ 1300 | HIC > 1500 | 52 | 9.1 | -.055 | - 2.87 | .63 |
| HIC ≤ 1300 | HIC > 1600 | 44 | 9.1 | -.056 | - 2.48 | .64 |
| HIC ≤ 1400 | HIC > 1400 | 68 | 8.5 | -.051 | - 3.44 | .68 |
| HIC ≤ 1500 | HIC > 1500 | 54 | 9.1 | -.055 | - 2.96 | .65 |

TABLE 5-5 (Continued)

COLLISIONS OF CARS WITH "GOOD" NCAP HIC SCORES INTO CARS WITH "POOR" HIC

| Definition of a "Good" Car | Definition of a "Poor" Car | Mean NCAP Scores and Model Year ("Good" vs. "Poor" Car): HIC | | Chest G's | | L Femur | | R Femur | | Model Year | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Good | Poor | Good | Poor | Good | Poor | Good | Poor | Good | Poor |
| HIC ≤ 800 | HIC > 800 | 620 | 1159 | 44.8 | 51.4 | 997 | 899 | 974 | 885 | 86.4 | 84.4 |
| HIC ≤ 800 | HIC > 900 | 618 | 1272 | 45.7 | 52.8 | 946 | 901 | 934 | 888 | 86.5 | 84.1 |
| HIC ≤ 800 | HIC > 1000 | 605 | 1449 | 44.8 | 52.6 | 971 | 837 | 992 | 737 | 86.6 | 84.7 |
| HIC ≤ 800 | HIC > 1100 | 599 | 1512 | 44.4 | 51.5 | 993 | 838 | 1016 | 739 | 86.6 | 84.9 |
| HIC ≤ 800 | HIC > 1200 | 590 | 1598 | 44.2 | 54.3 | 988 | 862 | 1009 | 794 | 86.4 | 84.8 |
| HIC ≤ 900 | HIC > 900 | 691 | 1299 | 46.5 | 53.3 | 921 | 918 | 915 | 894 | 86.1 | 84.1 |
| HIC ≤ 900 | HIC > 1000 | 690 | 1468 | 45.8 | 53.5 | 919 | 880 | 939 | 787 | 86.1 | 84.5 |
| HIC ≤ 900 | HIC > 1100 | 686 | 1537 | 45.7 | 52.7 | 942 | 871 | 963 | 772 | 86.1 | 84.6 |
| HIC ≤ 900 | HIC > 1200 | 688 | 1608 | 45.8 | 54.7 | 932 | 895 | 956 | 812 | 86.0 | 84.4 |
| HIC ≤ 900 | HIC > 1300 | 703 | 1741 | 46.3 | 54.5 | 926 | 914 | 946 | 762 | 85.9 | 84.6 |
| HIC ≤ 1000 | HIC > 1000 | 738 | 1476 | 47.2 | 53.5 | 939 | 852 | 968 | 789 | 85.5 | 84.5 |
| HIC ≤ 1000 | HIC > 1100 | 735 | 1546 | 47.0 | 52.9 | 958 | 840 | 988 | 765 | 85.5 | 84.5 |
| HIC ≤ 1000 | HIC > 1200 | 739 | 1616 | 47.3 | 54.8 | 964 | 861 | 995 | 807 | 85.3 | 84.3 |
| HIC ≤ 1000 | HIC > 1300 | 746 | 1767 | 47.4 | 54.2 | 958 | 899 | 983 | 761 | 85.3 | 84.8 |
| HIC ≤ 1000 | HIC > 1400 | 747 | 1936 | 48.1 | 51.2 | 1000 | 978 | 1035 | 678 | 85.5 | 85.4 |
| HIC ≤ 1100 | HIC > 1100 | 749 | 1556 | 47.5 | 53.0 | 965 | 829 | 986 | 761 | 85.4 | 84.5 |
| HIC ≤ 1100 | HIC > 1200 | 752 | 1629 | 47.8 | 54.9 | 970 | 851 | 993 | 804 | 85.2 | 84.3 |
| HIC ≤ 1100 | HIC > 1300 | 761 | 1782 | 48.2 | 54.4 | 963 | 881 | 986 | 762 | 85.2 | 84.7 |
| HIC ≤ 1100 | HIC > 1400 | 767 | 1946 | 49.1 | 51.7 | 1005 | 953 | 1035 | 684 | 85.3 | 85.3 |
| HIC ≤ 1100 | HIC > 1500 | 781 | 2084 | 49.3 | 50.4 | 989 | 927 | 1011 | 695 | 85.0 | 85.0 |
| HIC ≤ 1200 | HIC > 1200 | 759 | 1634 | 47.6 | 54.7 | 967 | 853 | 980 | 800 | 85.2 | 84.2 |
| HIC ≤ 1200 | HIC > 1300 | 769 | 1786 | 48.0 | 54.3 | 959 | 883 | 969 | 757 | 85.2 | 84.7 |
| HIC ≤ 1200 | HIC > 1400 | 778 | 1947 | 48.9 | 51.5 | 998 | 953 | 1010 | 680 | 85.3 | 85.2 |
| HIC ≤ 1200 | HIC > 1500 | 795 | 2080 | 49.0 | 50.3 | 981 | 928 | 980 | 690 | 84.9 | 84.9 |
| HIC ≤ 1300 | HIC > 1300 | 790 | 1777 | 48.4 | 54.3 | 945 | 872 | 971 | 769 | 85.1 | 84.7 |
| HIC ≤ 1300 | HIC > 1400 | 792 | 1943 | 49.1 | 51.6 | 985 | 936 | 1000 | 681 | 85.2 | 85.3 |
| HIC ≤ 1300 | HIC > 1500 | 812 | 2069 | 49.3 | 50.4 | 966 | 907 | 969 | 691 | 84.9 | 85.1 |
| HIC ≤ 1300 | HIC > 1600 | 811 | 2164 | 49.6 | 50.2 | 959 | 856 | 972 | 695 | 84.6 | 85.0 |
| HIC ≤ 1400 | HIC > 1400 | 808 | 1938 | 49.6 | 51.5 | 983 | 940 | 1008 | 679 | 85.1 | 85.3 |
| HIC ≤ 1500 | HIC > 1500 | 832 | 2058 | 49.9 | 50.3 | 963 | 913 | 979 | 687 | 84.8 | 85.1 |

as a "poor" car: the fatality reduction for the good cars is a statistically significant 14 percent (t for RELEXP is 1.68, p < .05). The other analyses that show statistically significant differences, with smaller samples, also have boundary values for HIC close to 1000 or just above it, and they have a modest gap: 900/1200, 900/1300, 1000/1300 and 1100/1300. Boundary values above 1300 did not produce large fatality reductions (unlike the situation with NCAPINJ and chest g's, where high boundary values produced large fatality reductions, although with small sample sizes.)

The second page of Table 5-5, which describes the average NCAP scores for the low-HIC and high-HIC groups, explains some of the trends in the fatality reductions. HIC is significantly correlated with chest g's. In <u>every</u> case, the low-HIC group has average chest g's under 50 and the high-HIC group has over 50 g's. However, the divergence between the low-HIC and the high-HIC group varies from 0.4 to 8.9. The average chest g's for the high-HIC group varies from 50.2 to 54.9. The analyses with a statistically significant or near-significant fatality reduction almost all have average chest g's over 54 in the high-HIC group, and vice-versa. These high-HIC groups contain a rich selection of high-chest g cases, and have elevated fatality risk. When the boundary value for HIC goes above 1300, the divergence in chest g's is diminished, and so is the difference in fatality risk. Table 5-5 also shows that the driver dummies with high HIC had, on the average, slightly or even appreciably lower femur loads than the dummies with low HIC.

## 5.4 <u>Cars with low NCAP femur loads hit cars with high femur loads</u>

In the NCAP test, femur loads are measured separately on the dummy's left and right legs (2250 pounds on either leg is the critical value on FMVSS

208). The definition of a "car" with high [low] femur load has to take into account the results for both legs. The approach used here is to say a car is a "good" performer if the left femur load $\leq$ A and the right-side measurement is also $\leq$ A and the sum of the two measurements $\leq$ another number B $\leq$ 2A. A car is a "poor" performer if the left femur load > C or the right femur load > C or the sum of the loads > D, where A $\leq$ C and B $\leq$ D $\leq$ 2C. Table 5-6 presents six analyses in which there is no gap between the lower and upper boundary values (A = C and B = D) and the critical value for the sum of the loads is 1000 above the load on either leg. For example, in the first analysis of Table 5-6, "good" cars must have femur loads $\leq$ 1300 pounds on each leg and $\leq$ 2300, total; performance is "poor" if femur load exceeds 1300 pounds on either leg or 2300 total.

Table 5-6 shows that femur load is rather reliable for differentiating safer from less safe cars. Over a range of boundary values from 1300 to 1800 pounds on one leg (and 2300 to 2800 pounds, total), the "good" performers consistently have a fatality reduction that is statistically significant at the .05 level. The significance never reaches the .01 level, as with chest g's and NCAPINJ, but it never falls below the .05 level, as with HIC. The comparison that maximizes sum RELEXP with a sample size close to 120 defines femur load $\leq$ 1600 on each leg (and 2600 total) as a "good" car and femur load > 1600 on either leg (or 2600 total) as a "poor" car: the fatality reduction for the good cars is a statistically significant 20 percent (t for RELEXP is 2.36, $p < .05$). The fatality reduction remains close to 20 percent for boundary values in the 1400-1800 pound range (for one leg; 2400-2800 pounds for both legs). Analyses were also tried with various gaps between "good" and "poor" performance; the gaps merely reduced sample size without appreciably escalating the fatality reduction for "good" performers. The second page of Table 5-6 shows that cars with low

TABLE 5-6

COLLISIONS OF CARS WITH "GOOD" NCAP FEMUR LOAD SCORES INTO CARS WITH "POOR" FEMUR LOADS:
EFFECT OF MOVING THE BOUNDARY BETWEEN "GOOD" AND "POOR" NCAP FEMUR LOADS

| | | Comparison of Real-World Performance | | | | |
|---|---|---|---|---|---|---|
| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **N of Crashes** | **% Fat Red for Good Car** | **Mean Relexp** | **Sum Relexp** | **T-Test for Relexp** |
| L Femur ≤ 1300 AND<br>R Femur ≤ 1300 AND<br>L+R Femur ≤ 2300 | L Femur > 1300 OR<br>R Femur > 1300 OR<br>L+R Femur > 2300 | 164 | 13.5 | -.084 | -13.81 | 1.73 |
| L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | L Femur > 1400 OR<br>R Femur > 1400 OR<br>L+R Femur > 2400 | 157 | 17.7 | -.113 | -17.74 | 2.24 |
| L Femur ≤ 1500 AND<br>R Femur ≤ 1500 AND<br>L+R Femur ≤ 2500 | L Femur > 1500 OR<br>R Femur > 1500 OR<br>L+R Femur > 2500 | 142 | 18.0 | -.116 | -16.53 | 2.18 |
| **L Femur ≤ 1600 AND<br>R Femur ≤ 1600 AND<br>L+R Femur ≤ 2600** | **L Femur > 1600 OR<br>R Femur > 1600 OR<br>L+R Femur > 2600** | **132** | **20.1** | **-.131** | **-17.30** | **2.36** |
| L Femur ≤ 1700 AND<br>R Femur ≤ 1700 AND<br>L+R Femur ≤ 2700 | L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 128 | 20.2 | -.132 | -16.85 | 2.30 |
| L Femur ≤ 1800 AND<br>R Femur ≤ 1800 AND<br>L+R Femur ≤ 2800 | L Femur > 1800 OR<br>R Femur > 1800 OR<br>L+R Femur > 2800 | 123 | 19.0 | -.123 | -15.07 | 2.08 |

TABLE 5-6 (Continued)

COLLISIONS OF CARS WITH "GOOD" NCAP FEMUR LOAD SCORES INTO CARS WITH "POOR" FEMUR LOADS:
EFFECT OF MOVING THE BOUNDARY BETWEEN "GOOD" AND "POOR" NCAP FEMUR LOADS

| | | Mean NCAP Scores and Model Year ("Good" vs. "Poor" Car) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **HIC** | | **Chest G's** | | **L Femur** | | **R Femur** | | **Model Year** | |
| | | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** |
| LFem ≤ 1300 AND RFem ≤ 1300 AND L+R Fem ≤ 2300 | LFem > 1300 OR RFem > 1300 OR L+R Fem > 2300 | 942 | 877 | 48.9 | 50.0 | 713 | 1442 | 709 | 1486 | 85.1 | 85.6 |
| LFem ≤ 1400 AND RFem ≤ 1400 AND L+R Fem ≤ 2400 | LFem > 1400 OR RFem > 1400 OR L+R Fem > 2400 | 913 | 881 | 48.6 | 49.8 | 746 | 1480 | 738 | 1545 | 85.4 | 86.0 |
| LFem ≤ 1500 AND RFem ≤ 1500 AND L+R Fem ≤ 2500 | LFem > 1500 OR RFem > 1500 OR L+R Fem > 2500 | 906 | 889 | 48.3 | 50.7 | 763 | 1505 | 747 | 1619 | 85.4 | 85.7 |
| **LFem ≤ 1600 AND RFem ≤ 1600 AND L+R Fem ≤ 2600** | **LFem > 1600 OR RFem > 1600 OR L+R Fem > 2600** | **913** | **898** | **48.4** | **51.0** | **759** | **1531** | **734** | **1659** | **85.5** | **85.6** |
| LFem ≤ 1700 AND RFem ≤ 1700 AND L+R Fem ≤ 2700 | LFem > 1700 OR RFem > 1700 OR L+R Fem > 2700 | 920 | 896 | 48.3 | 51.1 | 763 | 1542 | 740 | 1665 | 85.5 | 85.6 |
| LFem ≤ 1800 AND RFem ≤ 1800 AND L+R Fem ≤ 2800 | LFem > 1800 OR RFem > 1800 OR L+R Fem > 2800 | 918 | 880 | 48.6 | 51.4 | 777 | 1540 | 752 | 1703 | 85.4 | 85.5 |

femur loads also have, on the average, slightly lower chest g's than the cars with high femur loads; the difference in chest g's ranges from 1.1 to 2.8. There is little difference in HIC. The modest reduction in chest g's that accompanies low femur load may be a contributing factor in the fatality reduction, but probably not an important one.

In the preceding analyses, chest g's were usually more efficient than femur load for discriminating the actual safety performance of cars; femur load, in turn, was more usually reliable than HIC, although there was some overlapping in the results. The findings, which are consistent with the correlation analyses of Chapters 3 and 4, raise two interesting, related questions. Why are chest g's especially efficient? Given that femur injuries are rarely fatal, why does femur load correlate at all with fatality risk in actual crashes? The answer appears to be that the three NCAP test measurements are not independent observations on isolated body regions. There is not just a statistical correlation but, probably, also an intuitive overlap between the scores. Cars with intuitively excellent safety design tend to have low scores on all parameters. Cars with crashworthiness problems tend to have high scores on one or more parameters, but it is not always predictable which one. Thus, high femur load could reflect a more general problem affecting injury risk to other body regions in some crashes. Chest g's have two special advantages. Since the chest is the body region "in the middle," chest g's are correlated with both HIC and femur load; a poor score on chest g's often reflects poor scores on the other parameters. The measurement of chest g's tends to be less sensitive than the other parameters to moderate changes in the test conditions. That will make chest g's work especially well with the accident data used here, which, of necessity, include vehicles that do not exactly match the NCAP test vehicle, occupants of various heights and weights, and all types of head-on collisions, not just those that closely resemble an NCAP test.

## 5.5 Partitions of the fleet based on two NCAP parameters

Any single NCAP parameter, as shown above, can do an adequate job of partitioning the cars into a safer and a less safe group. Two parameters for two separate body regions, working together, can often do an even more reliable job. Table 5-7 examines the relative fatality risk in cars with good HIC and chest g scores when they hit cars with poor HIC or chest g scores. The approach used here is to say a car is a "good" performer if HIC $\leq$ A <u>and</u> chest g's $\leq$ B. A car is a "poor" performer if HIC > C <u>or</u> chest g's > D, where A $\leq$ C and B $\leq$ D. In every analysis of Table 5-7, the fatality risk is lower for the good performers. The reduction is statistically significant in every analysis which uses boundary values for HIC close to 1000, with a modest gap, and boundary values for chest g's close to 60, without a gap or with a modest gap. Some of the analyses show fatality reductions well over 20 percent, with samples of 80-90 crashes. However, in the comparison that maximizes fatality reduction with a sample size close to 120, the "good" performers are defined as the cars with driver HIC $\leq$ 1100 <u>and</u> chest g's $\leq$ 60, and the "poor" performers as the ones with either HIC > 1300 <u>or</u> chest g's > 60. The fatality risk is a statistically significant 19 percent lower for the drivers of the cars with the better NCAP scores (t for RELEXP is 2.31, $p < .05$).

The second page of Table 5-7 shows, not surprisingly that the average HIC and chest g's of the good performers are substantially lower than for the poor performers. The femur loads, however, tend to be somewhat higher for the good HIC-chest g performers. This page of Table 5-7 shows quite similar trends

## TABLE 5-7

## COLLISIONS OF CARS WITH "GOOD" NCAP HIC AND CHEST G SCORES INTO CARS WITH "POOR" HIC OR CHEST G'S

| | | Comparison of Real-World Performance | | | | |
|---|---|---|---|---|---|---|
| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **N of Crashes** | **% Fat Red for Good Car** | **Mean Relexp** | **Sum Relexp** | **T-Test for Relexp** |
| HIC $\le$ 800 AND Chest g's $\le$ 48 | HIC > 800 OR Chest g's > 48 | 165 | 3.3 | -.020 | - 3.23 | .39 |
| HIC $\le$ 800 AND Chest g's $\le$ 48 | HIC > 1100 OR Chest g's > 64 | 63 | 8.2 | -.047 | - 2.97 | .58 |
| HIC $\le$ 900 AND Chest g's $\le$ 55 | HIC > 1250 OR Chest g's > 60 | 92 | 23.0 | -.155 | -14.23 | 2.64 |
| HIC $\le$ 900 AND Chest g's $\le$ 55 | HIC > 1250 OR Chest g's > 62 | 86 | 26.7 | -.181 | -15.57 | 3.07 |
| HIC $\le$ 900 AND Chest g's $\le$ 55 | HIC > 1250 OR Chest g's > 65 | 81 | 26.7 | -.184 | -14.91 | 3.02 |
| HIC $\le$ 900 AND Chest g's $\le$ 56 | HIC > 1300 OR Chest g's > 60 | 93 | 24.7 | -.167 | -15.58 | 2.83 |
| HIC $\le$ 1000 AND Chest g's $\le$ 60 | HIC > 1000 OR Chest g's > 60 | 170 | 13.5 | -.084 | -14.36 | 1.82 |
| HIC $\le$ 1000 AND Chest g's $\le$ 60 | HIC > 1200 OR Chest g's > 70 | 104 | 19.1 | -.121 | -12.58 | 2.16 |
| HIC $\le$ 1100 AND Chest g's $\le$ 64 | HIC > 1100 OR Chest g's > 64 | 155 | 11.2 | -.070 | -10.79 | 1.43 |
| HIC $\le$ 1100 AND Chest g's $\le$ 60 | HIC > 1300 OR Chest g's > 60 | 125 | 18.9 | -.122 | -15.32 | 2.31 |
| HIC $\le$ 1200 AND Chest g's $\le$ 70 | HIC > 1200 OR Chest g's > 70 | 120 | 12.3 | -.076 | - 9.12 | 1.39 |

TABLE 5-7 (Continued)

COLLISIONS OF CARS WITH "GOOD" NCAP HIC AND CHEST G SCORES INTO CARS WITH "POOR" HIC OR CHEST G'S

| | | Mean NCAP Scores and Model Year ("Good" vs. "Poor" Car) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Definition of a "Good" Car | Definition of a "Poor" Car | HIC | | Chest G's | | L Femur | | R Femur | | Model Year | |
| | | Good | Poor | Good | Poor | Good | Poor | Good | Poor | Good | Poor |
| HIC ≤ 800 AND Chest g's ≤ 48 | HIC > 800 OR Chest g's > 48 | 613 | 1056 | 39.8 | 52.8 | 1106 | 890 | 1026 | 897 | 86.6 | 84.5 |
| HIC ≤ 800 AND Chest g's ≤ 48 | HIC > 1100 OR Chest g's > 64 | 600 | 1387 | 39.9 | 55.3 | 1075 | 840 | 1010 | 801 | 87.0 | 85.1 |
| HIC ≤ 900 AND Chest g's ≤ 55 | HIC > 1250 OR Chest g's > 60 | 711 | 1432 | 43.3 | 59.0 | 980 | 918 | 946 | 793 | 86.2 | 84.2 |
| HIC ≤ 900 AND Chest g's ≤ 55 | HIC > 1250 OR Chest g's > 62 | 712 | 1465 | 43.3 | 58.8 | 968 | 855 | 936 | 785 | 86.1 | 84.3 |
| HIC ≤ 900 AND Chest g's ≤ 55 | HIC > 1250 OR Chest g's > 65 | 713 | 1503 | 43.3 | 58.5 | 972 | 863 | 944 | 760 | 86.2 | 84.4 |
| HIC ≤ 900 AND Chest g's ≤ 56 | HIC > 1300 OR Chest g's > 60 | 709 | 1426 | 43.4 | 59.0 | 990 | 923 | 946 | 801 | 86.2 | 84.3 |
| HIC ≤ 1000 AND Chest g's ≤ 60 | HIC > 1000 OR Chest g's > 60 | 748 | 1339 | 45.6 | 55.9 | 1001 | 847 | 1003 | 801 | 85.5 | 84.6 |
| HIC ≤ 1000 AND Chest g's ≤ 60 | HIC > 1200 OR Chest g's > 70 | 742 | 1608 | 45.3 | 55.0 | 984 | 871 | 995 | 835 | 85.4 | 84.4 |
| HIC ≤ 1100 AND Chest g's ≤ 64 | HIC > 1100 OR Chest g's > 64 | 760 | 1395 | 46.2 | 56.0 | 994 | 809 | 1018 | 745 | 85.5 | 84.5 |
| HIC ≤ 1100 AND Chest g's ≤ 60 | HIC > 1300 OR Chest g's > 60 | 769 | 1414 | 46.0 | 59.3 | 1009 | 846 | 1016 | 779 | 85.3 | 84.2 |
| HIC ≤ 1200 AND Chest g's ≤ 70 | HIC > 1200 OR Chest g's > 70 | 757 | 1609 | 47.4 | 55.4 | 967 | 869 | 982 | 809 | 85.2 | 84.2 |

to the second halves of Tables 5-3 (partition by chest g's) and 5-5 (partition by HIC). To the extent that HIC and chest g's are fairly correlated, they are somewhat redundant measures in a statistical sense. They act as a check on one another, and using both of them together enhances the reliability of their information. Cars with low HIC tend to have low chest g's, and vice-versa, but the easiest way to find cars with low HIC and chest g's is to look at both of the variables. Neither variable, however, conveys the information that is contained in the femur load variable.

Since femur load is rather orthogonal (statistically uncorrelated) with HIC and chest g's, it might be expected that the combination of femur load with one of the other two variables is exceptionally useful for partitioning the fleet. Table 5-8 confirms that chest g's-and-femur load, or HIC-and-femur load can be used to differentiate the safer and the less-safe cars. If "good" performance is defined as chest g's $\leq$ 56 and femur load $\leq$ 1400 on each leg and $\leq$ 2400, total, while chest g's > 60 or femur load > 1700 on either leg or > 2700, total delineates "poor" performance, the fatality risk in 134 collisions between good and poor performers is a statistically significant 22 percent lower for the drivers of the cars with good NCAP scores (t for RELEXP is 2.93, $p < .01$). The lower half of Table 5-8 shows that the "good" performers, in this analysis, have lower average scores on all three body regions, not just the chest and femurs.

Most interestingly, the second analysis in Table 5-8 shows that HIC and femur load, <u>without chest g's</u>, can be used to partition the safer from the less safe cars. When the criterion for "good" performance is HIC $\leq$ 900 and femur load $\leq$ 1400 on each leg and $\leq$ 2400, total, and the criterion for "poor" performance is HIC > 1300 or femur load > 1700 on either leg or > 2700, total,

TABLE 5-8

COLLISIONS OF CARS WITH "GOOD" NCAP SCORES FOR TWO BODY REGIONS
INTO CARS WITH "POOR" NCAP SCORES FOR AT LEAST ONE OF THOSE BODY REGIONS

**Comparison of Real-World Performance**

| Definition of a "Good" Car | Definition of a "Poor" Car | N of Crashes | % Fat Red for Good Car | Mean Relexp | Sum Relexp | T-Test for Relexp |
|---|---|---|---|---|---|---|
| Chest g's $\le$ 56 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400 | Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 134 | 22.1 | -.147 | -19.66 | 2.93 |
| HIC $\le$ 900 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400 | HIC > 1300 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 121 | 19.4 | -.128 | -15.44 | 2.30 |

**Mean NCAP Scores and Model Year ("Good" vs. "Poor" Car)**

| Definition of a "Good" Car | Definition of a "Poor" Car | HIC Good | HIC Poor | Chest G's Good | Chest G's Poor | L Femur Good | L Femur Poor | R Femur Good | R Femur Poor | Model Year Good | Model Year Poor |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Chest g $\le$ 56 AND<br>LFem $\le$ 1400 AND<br>RFem $\le$ 1400 AND<br>L+R Fem $\le$ 2400 | Chest g > 60 OR<br>LFem > 1700 OR<br>RFem > 1700 OR<br>L+R Fem > 2700 | 890 | 983 | 43.9 | 55.6 | 754 | 1311 | 742 | 1336 | 85.5 | 84.8 |
| HIC $\le$ 900 AND<br>LFem $\le$ 1400 AND<br>RFem $\le$ 1400 AND<br>L+R Fem $\le$ 2400 | HIC > 1300 OR<br>LFem > 1700 OR<br>RFem > 1700 OR<br>L+R Fem > 2700 | 698 | 1194 | 46.8 | 51.5 | 738 | 1311 | 766 | 1294 | 85.7 | 85.1 |

the fatality risk in 121 head-on collisions between good and poor performers is a statistically significant 19 percent lower for the drivers of the cars with low HIC and femur load (t for RELEXP is 2.30, $p < .05$). The lower section of Table 5-8 shows that the cars with good HIC and femur load have substantially lower chest g's (46.8 vs. 51.5) than the cars with poor HIC or femur load. That HIC with femur load works about as well as HIC with chest g's or femur load with chest g's illustrates the extent to which the three NCAP scores contain both overlapping and complementary information.

## 5.6 Partitions of the fleet based on all three NCAP parameters

A reliable differentiation of safer and less-safe cars may be obtained by using NCAP scores for all three body regions, with separate boundary values ("pass-fail" criteria) for each body region. This method is perhaps not quite as efficient as a composite variable such as NCAPINJ, but is just as reliable, or more so, than the analyses based on one or two body regions. Table 5-9 illustrates five analyses using various boundary values for HIC, chest g's, and femur load. An accident sample close to the target of 120 crashes is obtained by defining "good" performance as HIC $\leq$ 900 <u>and</u> chest g's $\leq$ 56 <u>and</u> femur load $\leq$ 1400 on each leg and $\leq$ 2400, total. HIC > 1300 <u>or</u> chest g's > 60 <u>or</u> femur load > 1700 on either leg or > 2700, total defines "poor" performance. The fatality risk in 118 actual head-on collisions between a good and a poor performer is a statistically significant 21 percent lower for the drivers of the cars with good NCAP scores (t for RELEXP is 2.68, $p < .01$).

Table 5-9 shows that the boundary values can be varied by a moderate amount and the fatality reduction for the "good" performers will still be statistically significant, often at the .01 level. Such reductions are found in

TABLE 5-9

COLLISIONS OF CARS WITH "GOOD" NCAP SCORES FOR ALL THREE BODY REGIONS
INTO CARS WITH "POOR" NCAP SCORES FOR AT LEAST ONE BODY REGION

| | | | Comparison of Real-World Performance | | | |
|---|---|---|---|---|---|---|
| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **N of Crashes** | **% Fat Red for Good Car** | **Mean Relexp** | **Sum Relexp** | **T-Test for Relexp** |
| **HIC $\le$ 900 AND<br>Chest g's $\le$ 56 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400** | **HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700** | **118** | **21.2** | **-.139** | **-16.44** | **2.68** |
| HIC $\le$ 900 AND<br>Chest g's $\le$ 60 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1600 OR<br>R Femur > 1600 OR<br>L+R Femur > 2600 | 128 | 20.7 | -.137 | -17.54 | 2.65 |
| HIC $\le$ 1100 AND<br>Chest g's $\le$ 56 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1600 OR<br>R Femur > 1600 OR<br>L+R Femur > 2600 | 140 | 19.6 | -.128 | -17.91 | 2.72 |
| HIC $\le$ 1100 AND<br>Chest g's $\le$ 60 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1600 OR<br>R Femur > 1600 OR<br>L+R Femur > 2600 | 153 | 18.3 | -.118 | -18.05 | 2.46 |
| HIC $\le$ 1100 AND<br>Chest g's $\le$ 60 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1800 OR<br>R Femur > 1800 OR<br>L+R Femur > 2800 | 148 | 19.0 | -.123 | -18.24 | 2.50 |

TABLE 5-9 (Continued)

COLLISIONS OF CARS WITH "GOOD" NCAP SCORES FOR ALL THREE BODY REGIONS
INTO CARS WITH "POOR" NCAP SCORES FOR AT LEAST ONE BODY REGION

| | | Mean NCAP Scores and Model Year ("Good" vs. "Poor" Car) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Definition of a "Good" Car | Definition of a "Poor" Car | HIC | | Chest G's | | L Femur | | R Femur | | Model Year | |
| | | Good | Poor | Good | Poor | Good | Poor | Good | Poor | Good | Poor |
| HIC ≤ 900 AND<br>Chest g ≤ 56 AND<br>L Fem ≤ 1400 AND<br>R Fem ≤ 1400 AND<br>L+R Fem ≤ 2400 | HIC > 1300 OR<br>Chest g > 60 OR<br>L Fem > 1700 OR<br>R Fem > 1700 OR<br>L+R Fem > 2700 | 704 | 1154 | 43.1 | 53.6 | 778 | 1230 | 775 | 1200 | 85.9 | 84.8 |
| HIC ≤ 900 AND<br>Chest g ≤ 60 AND<br>L Fem ≤ 1400 AND<br>R Fem ≤ 1400 AND<br>L+R Fem ≤ 2400 | HIC > 1300 OR<br>Chest g > 60 OR<br>L Fem > 1600 OR<br>R Fem > 1600 OR<br>L+R Fem > 2600 | 703 | 1156 | 43.9 | 53.3 | 767 | 1229 | 776 | 1208 | 85.9 | 84.8 |
| HIC ≤ 1100 AND<br>Chest g ≤ 56 AND<br>L Fem ≤ 1400 AND<br>R Fem ≤ 1400 AND<br>L+R Fem ≤ 2400 | HIC > 1300 OR<br>Chest g > 60 OR<br>L Fem > 1600 OR<br>R Fem > 1600 OR<br>L+R Fem > 2600 | 739 | 1161 | 43.8 | 53.7 | 771 | 1201 | 775 | 1192 | 85.6 | 85.0 |
| HIC ≤ 1100 AND<br>Chest g ≤ 60 AND<br>L Fem ≤ 1400 AND<br>R Fem ≤ 1400 AND<br>L+R Fem ≤ 2400 | HIC > 1300 OR<br>Chest g > 60 OR<br>L Fem > 1600 OR<br>R Fem > 1600 OR<br>L+R Fem > 2600 | 744 | 1161 | 45.0 | 53.8 | 755 | 1185 | 763 | 1198 | 85.5 | 84.9 |
| HIC ≤ 1100 AND<br>Chest g ≤ 60 AND<br>L Fem ≤ 1400 AND<br>R Fem ≤ 1400 AND<br>L+R Fem ≤ 2400 | HIC > 1300 OR<br>Chest g > 60 OR<br>L Fem > 1800 OR<br>R Fem > 1800 OR<br>L+R Fem > 2800 | 745 | 1173 | 45.1 | 54.0 | 756 | 1180 | 770 | 1194 | 85.4 | 84.8 |

analyses with HIC boundary values reasonably close to or slightly above the FMVSS 208 value of 1000, with a modest gap; chest g boundary values close to the FMVSS 208 value of 60 g's; and femur load boundaries in the 1400-1800 pound range. The analyses show fatality reductions of 18-21 percent in accident samples ranging up to 153 crashes, and sum RELEXP values up to 18.24, approaching the sum RELEXP values of 20-21 found in the analyses based on NCAPINJ (see Table 5-1). The second page of Table 5-9 shows that the "good" performers had, on the average, substantially lower HIC, chest g's and femur loads than the poor performers.

## 5.7 Sensitivity test: collisions of two cars with similar mass

None of the analyses, so far, placed any limits on the relative masses of the two cars in the head-on collisions. The data included some crashes in which the two cars had a severe weight mismatch - e.g., 1800 and 3800 pounds. A question could be raised if the cases with severe mismatch are "driving" the results. In those crashes, where the driver of the lighter car is almost certain to be a fatality, the difference between actual and "expected" performance may not be as meaningful as in crashes where both drivers have a good chance of survival. Would the results be different if the sample were limited to collisions of cars with similar weights?

Table 5-10 limits two of the earlier analyses to the subset of head-on collisions in which the weights of the two cars differ by no more than 1000 pounds. In the subsample of 86 collisions of a car with NCAPINJ $\leq$ 0.6 into a car with NCAPINJ > 0.6, both cars having curb weights within 1000 pounds of one another, the fatality reduction for the good NCAP performers is 23.9 percent, which is about the same as the 26.4 percent fatality reduction in the unrestricted sample of 117 collisions (see Table 5-1). In the subsample of 94

TABLE 5-10

COLLISIONS OF CARS WITH "GOOD" NCAP SCORES INTO CARS WITH "POOR" SCORES
WHERE THE "GOOD" AND "POOR" CARS HAVE SIMILAR MASS

**Comparison of Real-World Performance**

| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **N of Crashes** | **% Fat Red for Good Car** | **Mean Relexp** | **Sum Relexp** | **T-Test for Relexp** |
|---|---|---|---|---|---|---|
| NCAPINJ ≤ .6 | NCAPINJ > .6<br>and weight difference ≤ 1000 pounds | 86 | 23.9 | -.164 | -14.14 | 2.33 |
| HIC ≤ 900 AND<br>Chest g's ≤ 56 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700<br>and weight difference ≤ 1000 pounds | 94 | 20.7 | -.135 | -12.72 | 2.17 |



**Car Weight, Driver Age, Sex and Model Year ("Good" vs. "Poor" Car)**

| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **Avg Weight** | | **Avg Age** | | **% Female** | | **Model Year** | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** |
| NCAPINJ ≤ .6 | NCAPINJ > .6<br>and weight difference ≤ 1000 pounds | 2872 | 2929 | 42.8 | 41.8 | 49 | 48 | 86.0 | 83.2 |
| HIC ≤ 900 AND<br>Chest g's ≤ 56 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700<br>and weight difference ≤ 1000 pounds | 2840 | 2842 | 42.3 | 43.5 | 50 | 40 | 86.0 | 84.9 |

collisions of a car with low HIC, chest g's and femur load into a car with high HIC, chest g's or femur load, the fatality reduction for the good performers is 20.7 percent, which is almost identical to the 20.8 percent reduction in the unrestricted sample of 118 cases (see Table 5-9). The lower half of Table 5-10 shows that, in both analyses, the "good" and "poor" NCAP performers have nearly the same average curb weights and driver ages.

## 5.8 Sensitivity test: weighted vs. unweighted composite score

The composite measure of NCAP performance,

NCAPINJ = .21 HEADINJ + 2.7 CHESTINJ + 1.5 (LFEMURINJ + RFEMURINJ)

was calibrated in Section 4.1 as the weighted combination of logistic injury probability functions for the head, chest and femurs that has maximum correlation with fatality risk in the set of head-on collisions where both cars match an NCAP test at the '3A or 4A level.' However, it was also shown in Section 4.1 that this particular weighted sum did not "magically" enhance correlation; other weighted sums had almost equally high correlations, and even an unweighted sum of the injury probabilities had significant correlation with fatality risk.

Similarly, the approach of this chapter showed that NCAPINJ was efficient for separating the safer cars from the less safe cars, maximizing the difference in relative fatality risk when a "good" NCAP performer hits a "poor" performer head-on. Here too, however, NCAPINJ is merely "first among equals" for the purpose of identifying safer and less safe cars. Table 5-11 shows that even an unweighted sum of the logistic injury probabilities,

INJ = HEADINJ + CHESTINJ + LFEMURINJ + RFEMURINJ

accomplishes the same purpose with just slightly less efficiency. The top half of Table 5-11 presents six analyses of head-on collisions between cars with

TABLE 5-11

COLLISIONS OF CARS WITH A LOW NCAP COMPOSITE SCORE INTO CARS WITH A HIGH NCAP COMPOSITE SCORE:
WEIGHTED VS. UNWEIGHTED COMPOSITE SCORE

**WEIGHTED:** NCAPINJ = .21 HEADINJ + 2.7 CHESTINJ + 1.5 (LFEMURINJ + RFEMURINJ)

| Definition of "Low" NCAPINJ | Definition of "High" NCAPINJ | N of Crashes | % Fat Red for Low NCAPINJ | Mean Relexp | Sum Relexp | T-Test for Relexp |
|---|---|---|---|---|---|---|
| NCAPINJ ≤ .4 | NCAPINJ > .4 | 186 | 17.0 | -.108 | -20.17 | 2.56 |
| NCAPINJ ≤ .5 | NCAPINJ > .5 | 147 | 17.3 | -.114 | -16.83 | 2.44 |
| NCAPINJ ≤ .6 | NCAPINJ > .6 | 117 | 26.4 | -.181 | -21.13 | 3.22 |
| NCAPINJ ≤ .7 | NCAPINJ > .7 | 108 | 27.2 | -.189 | -20.46 | 3.34 |
| NCAPINJ ≤ .8 | NCAPINJ > .8 | 80 | 28.7 | -.201 | -16.10 | 3.13 |
| NCAPINJ ≤ .9 | NCAPINJ > .9 | 55 | 31.9 | -.216 | -11.88 | 2.73 |

**UNWEIGHTED:** INJ = HEADINJ + CHESTINJ + LFEMURINJ + RFEMURINJ

| Definition of "Low" INJ | Definition of "High" INJ | N of Crashes | % Fat Red for Low INJ | Mean Relexp | Sum Relexp | T-Test for Relexp |
|---|---|---|---|---|---|---|
| INJ ≤ .3 | INJ > .3 | 191 | 14.6 | -.092 | -17.60 | 2.16 |
| INJ ≤ .4 | INJ > .4 | 149 | 17.7 | -.115 | -17.13 | 2.36 |
| INJ ≤ .5 | INJ > .5 | 121 | 21.6 | -.145 | -17.54 | 2.77 |
| INJ ≤ .6 | INJ > .6 | 97 | 18.1 | -.117 | -11.33 | 2.00 |
| INJ ≤ .7 | INJ > .7 | 78 | 21.2 | -.139 | -10.86 | 2.08 |
| INJ ≤ .8 | INJ > .8 | 57 | 16.7 | -.115 | - 6.55 | 1.56 |

"good" and "poor" NCAPINJ scores, recapitulating material from Table 5-1. The lower half of Table 5-11 presents results of six similar analyses based on the unweighted composite score INJ. In the NCAPINJ analyses, the fatality risk is 17-32 percent lower for the driver of the "good" car than for the driver of the "poor" car; the reductions in the first two analyses are significant at the .05 level, and in the last four analyses they are significant at the .01 level. In the INJ analyses, fatality reductions range from 15 to 22 percent; the first five analyses show statistically significant reductions, with the third analysis significant at the .01 level. In general, INJ works about as well as NCAPINJ when the boundary between "good" and "poor" performance is set at a fairly low level - i.e., the first two or three analyses. Only when the boundary is set at a high level does NCAPINJ become visibly more efficient than INJ.

In the third analysis with the unweighted score INJ, the fatality reduction is 21.6 percent with a sample of 121 crashes. Thus, at the target sample size, INJ is slightly more efficient than any single NCAP parameter, and works about as well as combinations of two or three NCAP parameters.

## 5.9 Sensitivity test: analyses on a different calibration data set

The preceding analyses were conducted with the principal data set of head-on collisions in which both cars matched an NCAP test vehicle at levels "3A" or "4A": the model year on FARS is within the range of model years considered valid for the NCAP test, and the make-models on FARS and NCAP are identical or true corporate cousins. In Section 3.4, an alternative data set was defined in which both vehicles not only match an NCAP test vehicle at levels 3A or 4A, but also have the same number of doors as the NCAP test vehicle (310 collisions, 620 cars). In the multiple regression approach of Section 3.4, the alternative data set produced higher regression coefficients for HIC, and lower coefficients for femur load than the principal data set. In Section 4.7 it was shown that NCAPINJ did not have optimal correlation with fatality risk on the alternative data set (although it was close to the optimum), and that the unweighted score INJ was just slightly less correlated with fatality risk than NCAPINJ.

When the methods of this chapter are applied to the alternative data set, the results closely parallel the earlier findings. Table 5-12 considers head-on collisions between a "good" and a "poor" NCAP performer, based on the values of a single NCAP parameter. The left three columns of numbers are the sample size, fatality reduction for the good car and t-test result for RELEXP in the principal accident data set (level 3A/4A matches). The right three columns are the corresponding analysis results on the alternative data set (level 3A/4A and N of doors matching). Since the principal data set contains the alternative set, the N's on the right are always smaller; when N is smaller, the same percentage of fatality reduction will produce a weaker t-test result.

The first section of Table 5-12 presents six analyses of crashes of cars with low NCAP chest g's into cars with high chest g's. The fatality reductions are virtually identical in the two data sets, ranging from 11 to 30 percent in the principal data set and 10 to 28 percent in the alternative set. In four of the six analyses, the reductions are within 2 percent on the two data sets; in the first analysis, the alternative data set produces a slightly higher reduction, while in the fifth analysis, the principal data set produces a greater effect. The findings are consistent with Section 3.4, where chest g's had nearly the same regression coefficient in the two data sets.

TABLE 5-12

COLLISIONS OF CARS WITH "GOOD" NCAP SCORES INTO CARS WITH "POOR" NCAP SCORES
COMPARISON OF TWO ACCIDENT DATA SETS

| | | FARS Matches NCAP at Level 3A or 4A | | | FARS Matches NCAP at Level 3A/4A and N of Doors | | |
|---|---|---|---|---|---|---|---|
| Definition of a "Good" Car | Definition of a "Poor" Car | N | % Fat Red | T-Test | N | % Fat Red | T-Test |
| Chest g's ≤ 44 | Chest g's > 44 | 172 | 16.2 | 2.26 | 139 | 18.9 | 2.36 |
| Chest g's ≤ 48 | Chest g's > 48 | 182 | 10.9 | 1.56 | 145 | 9.9 | 1.21 |
| Chest g's ≤ 52 | Chest g's > 52 | 145 | 15.0 | 2.02 | 113 | 14.0 | 1.59 |
| Chest g's ≤ 56 | Chest g's > 56 | 125 | 18.7 | 2.32 | 99 | 18.2 | 1.88 |
| Chest g's ≤ 60 | Chest g's > 60 | 92 | 24.2 | 2.74 | 71 | 20.4 | 1.91 |
| Chest g's ≤ 64 | Chest g's > 64 | 49 | 29.7 | 2.60 | 38 | 28.3 | 1.94 |
| HIC ≤ 800 | HIC > 900 | 133 | 4.2 | .44 | 108 | 12.1 | 1.23 |
| HIC ≤ 900 | HIC > 1000 | 127 | 10.7 | 1.23 | 96 | 20.5 | 2.25 |
| HIC ≤ 900 | HIC > 1200 | 92 | 19.0 | 2.13 | 68 | 26.0 | 2.73 |
| HIC ≤ 1000 | HIC > 1200 | 113 | 14.2 | 1.68 | 82 | 16.9 | 1.73 |
| HIC ≤ 1100 | HIC > 1200 | 118 | 13.1 | 1.54 | 86 | 17.1 | 1.80 |
| HIC ≤ 1200 | HIC > 1300 | 87 | 14.0 | 1.35 | 64 | 15.4 | 1.30 |
| L Fem ≤ 1500 AND<br>R Fem ≤ 1500 AND<br>L+R Fem ≤ 2500 | L Fem > 1500 OR<br>R Fem > 1500 OR<br>L+R Fem > 2500 | 142 | 18.0 | 2.18 | 121 | 12.9 | 1.40 |
| L Fem ≤ 1600 AND<br>R Fem ≤ 1600 AND<br>L+R Fem ≤ 2600 | L Fem > 1600 OR<br>R Fem > 1600 OR<br>L+R Fem > 2600 | 132 | 20.1 | 2.36 | 112 | 15.8 | 1.64 |
| L Fem ≤ 1700 AND<br>R Fem ≤ 1700 AND<br>L+R Fem ≤ 2700 | L Fem > 1700 OR<br>R Fem > 1700 OR<br>L+R Fem > 2700 | 128 | 20.2 | 2.30 | 108 | 15.7 | 1.58 |

The middle section of Table 5-12 contains six analyses based on HIC, each with a small gap between "good" and "poor" performers. Here, the fatality reduction is greater for the alternative data set, especially in the first three analyses, where the boundary values of HIC are relatively low. In the third analysis, the fatality reduction is statistically significant at the .01 level in the alternative data set. That level of significance was never achieved in the principal data set with HIC, despite larger sample sizes. In the last three HIC analyses, where the cars are partitioned at higher levels of HIC, the results for the two data sets more or less converge, and the reductions drop out of the statistically significant range, even in the alternative data set.

Conversely, the lower section of Table 5-12 shows that femur load does not work as well on the alternative data set as on the principal data set. Fatality reductions are 4-6 percent lower on the smaller data set, and they are not statistically significant. Whereas femur load seems to be more efficient than HIC on the main data set, HIC works slightly better on the alternative set, consistent with the regression coefficients in Section 3.4.

Table 5-13 compares the effects of the weighted composite score NCAPINJ and the unweighted sum of logistic injury probabilities, INJ for the principal and alternative data sets. The first half of the table presents six analyses for NCAPINJ. In the principal data set, the reductions are always statistically significant, and in the last four analyses the reductions range from 26 to 32 percent and are significant at the .01 level. NCAPINJ does not work so efficiently for the alternative data set, although it still produces fatality reductions up to 24 percent. Three of the six analyses produce statistically significant reductions at the .05 level.

TABLE 5-13

COLLISIONS OF CARS WITH A LOW NCAP COMPOSITE SCORE INTO CARS WITH A HIGH NCAP COMPOSITE SCORE:
COMPARISON OF TWO ACCIDENT DATA SETS: WEIGHTED VS. UNWEIGHTED COMPOSITE SCORE

**WEIGHTED: NCAPINJ = .21 HEADINJ + 2.7 CHESTINJ + 1.5 (LFEMURINJ + RFEMURINJ)**

| | | FARS Matches NCAP at Level 3A or 4A | | | FARS Matches NCAP at Level 3A/4A and N of Doors | | |
|---|---|---|---|---|---|---|---|
| Definition of "Low" NCAPINJ | Definition of "High" NCAPINJ | N | % Fat Red | T-Test | N | % Fat Red | T-Test |
| NCAPINJ ≤ .4 | NCAPINJ > .4 | 186 | 17.0 | 2.56 | 146 | 15.4 | 2.06 |
| NCAPINJ ≤ .5 | NCAPINJ > .5 | 147 | 17.3 | 2.44 | 113 | 14.4 | 1.75 |
| NCAPINJ ≤ .6 | NCAPINJ > .6 | 117 | 26.4 | 3.22 | 88 | 18.8 | 1.88 |
| NCAPINJ ≤ .7 | NCAPINJ > .7 | 108 | 27.2 | 3.34 | 82 | 21.2 | 2.06 |
| NCAPINJ ≤ .8 | NCAPINJ > .8 | 80 | 28.7 | 3.13 | 61 | 23.9 | 2.18 |
| NCAPINJ ≤ .9 | NCAPINJ > .9 | 55 | 31.9 | 2.73 | 41 | 21.8 | 1.53 |

**UNWEIGHTED: INJ = HEADINJ + CHESTINJ + LFEMURINJ + RFEMURINJ**

| | | FARS Matches NCAP at Level 3A or 4A | | | FARS Matches NCAP at Level 3A/4A and N of Doors | | |
|---|---|---|---|---|---|---|---|
| Definition of "Low" INJ | Definition of "High" INJ | N | % Fat Red | T-Test | N | % Fat Red | T-Test |
| INJ ≤ .3 | INJ > .3 | 191 | 14.6 | 2.16 | 149 | 18.5 | 2.46 |
| INJ ≤ .4 | INJ > .4 | 149 | 17.7 | 2.36 | 115 | 20.0 | 2.35 |
| INJ ≤ .5 | INJ > .5 | 121 | 21.6 | 2.77 | 97 | 18.5 | 2.02 |
| INJ ≤ .6 | INJ > .6 | 97 | 18.1 | 2.00 | 76 | 15.7 | 1.54 |
| INJ ≤ .7 | INJ > .7 | 78 | 21.2 | 2.08 | 61 | 17.0 | 1.50 |

The lower half of Table 5-13 presents five analyses with the unweighted composite score. It works about equally well on the principal and alternative data sets: with low boundary values, the fatality reduction is slightly greater on the alternative data set, but in the last three analyses, the results are more favorable on the principal data set. It is especially interesting to compare the upper and lower half of the table. With the principal data set, NCAPINJ does a visibly better job than INJ when the cars are partitioned at a relatively high score: NCAPINJ pushes the fatality reduction up to the 30 percent range, but INJ does not. On the alternative data set, NCAPINJ does not do as well as INJ on the first two analyses, and only slightly better on the subsequent ones.

These sensitivity tests with an alternative accident data set illustrate two points rather clearly: (1) The FARS data show significant relationships between each of the three NCAP parameters and fatality risk, but they are not really sufficient to rank-order the strength of the three relationships; small changes in the accident data set can change the rank order. (2) The FARS data show that a composite score based on all three parameters, such as NCAPINJ, has excellent correlation with fatality risk, but they are not sufficient to establish "ideal" relative weights for the three parameters; small changes in the data set will change the optimum relative weights.

## 5.10 Summary

Table 5-14 extracts from Tables 5-1 through 5-9 the analyses that maximized fatality reduction and sum RELEXP, in the preceding tables, with sample sizes close to the target of 120 crashes. They are the "best in their class" analyses, based on various ways of partitioning "good" and "poor" NCAP

TABLE 5-14

SUMMARY: COLLISIONS OF CARS WITH "GOOD" NCAP SCORES INTO CARS WITH "POOR" NCAP SCORES
(N of crashes approximately 120 in each comparison)

| | | | Comparison of Real-World Performance | | | |
|---|---|---|---|---|---|---|
| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **N of Crashes** | **% Fat Red for Good Car** | **Mean Relexp** | **Sum Relexp** | **T-Test for Relexp** |
| NCAPINJ ≤ .6 | NCAPINJ > .6 | 117 | 26.4 | -.181 | -21.13 | 3.22 |
| Chest g's ≤ 56 | Chest g's > 56 | 125 | 18.7 | -.123 | -15.42 | 2.32 |
| HIC ≤ 1000 | HIC > 1200 | 113 | 14.2 | -.090 | -10.22 | 1.68 |
| L Femur ≤ 1600 AND<br>R Femur ≤ 1600 AND<br>L+R Femur ≤ 2600 | L Femur > 1600 OR<br>R Femur > 1600 OR<br>L+R Femur > 2600 | 132 | 20.1 | -.131 | -17.30 | 2.36 |
| HIC ≤ 1100 AND<br>Chest g's ≤ 60 | HIC > 1300 OR<br>Chest g's > 60 | 125 | 18.9 | -.122 | -15.32 | 2.31 |
| Chest g's ≤ 56 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 134 | 22.1 | -.147 | -19.66 | 2.93 |
| HIC ≤ 900 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | HIC > 1300 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 121 | 19.4 | -.128 | -15.44 | 2.30 |
| HIC ≤ 900 AND<br>Chest g's ≤ 56 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 118 | 21.2 | -.139 | -16.44 | 2.68 |

TABLE 5-14 (Continued)

SUMMARY: COLLISIONS OF CARS WITH "GOOD" NCAP SCORES INTO CARS WITH "POOR" NCAP SCORES
(N of crashes approximately 120 in each comparison)

| | | Mean NCAP Scores and Model Year ("Good" vs. "Poor" Car) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **HIC** | | **Chest G's** | | **L Femur** | | **R Femur** | | **Model Year** | |
| | | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** |
| NCAPINJ ≤ .6 | NCAPINJ > .6 | 898 | 1106 | 45.0 | 62.6 | 878 | 1161 | 846 | 1117 | 85.8 | 83.1 |
| Chest g ≤ 56 | Chest g > 56 | 914 | 1088 | 44.6 | 63.4 | 949 | 964 | 945 | 930 | 85.8 | 83.5 |
| HIC ≤ 1000 | HIC > 1200 | 739 | 1616 | 47.3 | 54.8 | 964 | 861 | 995 | 807 | 85.3 | 84.3 |
| LFem ≤ 1600 AND<br>RFem ≤ 1600 AND<br>L+R Fem ≤ 2600 | LFem > 1600 OR<br>RFem > 1600 OR<br>L+R Fem > 2600 | 913 | 898 | 48.4 | 51.0 | 759 | 1531 | 734 | 1659 | 85.5 | 85.6 |
| HIC ≤ 1100 AND<br>Chest g ≤ 60 | HIC > 1300 OR<br>Chest g > 60 | 769 | 1414 | 46.0 | 59.3 | 1009 | 846 | 1016 | 779 | 85.3 | 84.2 |
| Chest g ≤ 56 AND<br>LFem ≤ 1400 AND<br>RFem ≤ 1400 AND<br>L+R Fem ≤ 2400 | Chest g > 60 OR<br>LFem > 1700 OR<br>RFem > 1700 OR<br>L+R Fem > 2700 | 890 | 983 | 43.9 | 55.6 | 754 | 1311 | 742 | 1336 | 85.5 | 84.8 |
| HIC ≤ 900 AND<br>LFem ≤ 1400 AND<br>RFem ≤ 1400 AND<br>L+R Fem ≤ 2400 | HIC > 1300 OR<br>LFem > 1700 OR<br>RFem > 1700 OR<br>L+R Fem > 2700 | 698 | 1194 | 46.8 | 51.5 | 738 | 1311 | 766 | 1294 | 85.7 | 85.1 |
| HIC ≤ 900 AND<br>Chest g ≤ 56 AND<br>LFem ≤ 1400 AND<br>RFem ≤ 1400 AND<br>L+R Fem ≤ 2400 | HIC > 1300 OR<br>Chest g > 60 OR<br>LFem > 1700 OR<br>RFem > 1700 OR<br>L+R Fem > 2700 | 704 | 1154 | 43.1 | 53.6 | 778 | 1230 | 775 | 1200 | 85.9 | 84.8 |

performers. Table 5-14 shows that NCAPINJ does a slightly better job than any of the original NCAP scores, singly or in combinations, in separating the safer cars from the less safe cars, at a target sample size of 120 crashes (with the caveats, as noted in Sections 5.8 and 5.9: an unweighted combination of the injury scores did nearly as good a job as NCAPINJ, especially on an alternative data set). The fatality reduction of 26 percent is higher than any of the others, which range from 14 to 22 percent. Sum RELEXP and the t-test values are also higher. A composite score such as NCAPINJ is more efficient than the other methods because it allows excellent performance on two body regions to compensate for moderately poor performance on the third. Intuitively, a car with HIC = 999, chest g's = 59 and femur loads = 1500 each did not perform as well on NCAP as a car with HIC = 1001, chest g's = 40 and femur loads = 500 each. NCAPINJ (or other composite scores) will put the first car in the "good" group and the second car in the "poor" group, consistent with intuition, while the other methods, if they had a boundary value of 1000 for HIC, would do the reverse.

The majority of cars, however, do not have unusual NCAP scores like the two examples above. They tend to be really good NCAP performers, or quite poor. All of the methods developed in this chapter will assign them to the correct group. The most important finding conveyed by Table 5-14 is that <u>any</u> reasonable partitioning of the fleet, based on HIC, chest g's and/or femur load will work. In every case, there are significantly fewer fatalities in the "good" cars than in the "poor" cars, when they collide head-on.

# CHAPTER 6

# FATALITY RISK INDICES FOR "GOOD" AND "POOR" NCAP PERFORMERS

All of the analyses so far examined head-on collisions in which both cars matched up with an NCAP test. When one of the cars in the collision was a good NCAP performer and the other had poor NCAP scores, there was a significant safety advantage for the car that performed well in NCAP. Ideally, though, a car with good NCAP performance should be safer-than-average for belted drivers over the full range of head-on collisions - regardless of whether the NCAP performance of the other car in the crash was poor, good, or unknown. This chapter presents a more generalized analysis, based on a larger sample of head-on collisions. The "case" vehicle in these collisions has to match up with an NCAP test, but the "other" vehicle in the crash can be any 1976-91 passenger car with a belted driver, not necessarily matching with any NCAP test. Fatality risk indices are calculated separately for the case vehicles that are good NCAP performers and for the case vehicles that are poor performers; the difference in the indices represents the safety benefit of good NCAP performance over the full range of head-on collisions.

## 6.1 Procedure for computing fatality risk indices

Chapter 2 defined a file of 926 head-on collisions, comprising 1,852 distinct vehicles. Both vehicles in a collision had to be 1976-91 passenger cars, with belted drivers; at least one of the drivers was a fatality. Some of the vehicles match up only weakly or not at all with an NCAP case; Section 2.5 presents criteria for assessing the quality of the match and Section 3.4 demonstrates that FARS and NCAP cases should match at least at the "3A or 4A" level: the FARS and NCAP vehicles should be of the same make-model or true corporate cousins, and the FARS model year should be within the "valid" range of the NCAP test. NCAPINJ, a composite measure of NCAP performance, was defined in Section 4.1. There are 392 head-on collisions, comprising 784 distinct vehicles, in which both cars match up with an NCAP test at the 3A or 4A level and NCAPINJ can be calculated for that test (i.e., missing NCAP data for no more than one body region). Those 392 collisions were the basis for the analyses in Chapters 4 and 5. However, there are an additional 405 head-on collisions in which only one vehicle matches an NCAP test at the 3A or 4A level with known NCAPINJ, and the other vehicle is a 1976-91 car with a belted driver.

A more generalized analysis, which allows a much larger sample size of 1189 crashes, applies to head-on collisions in which the "case" vehicle of interest is a 1979-91 car that matches up with an NCAP test, whose driver wore belts, but the "other" vehicle in the crash can be any 1976-91 passenger car with a belted driver, not necessarily matching closely with an NCAP test. Thus, there are a total of 1189 individual vehicles (784 + 405) that have level 3A or 4A NCAP matches with known NCAPINJ. They were involved in 797 distinct head-on collisions (392 + 405).

The accident analysis file, comprising 1189 head-on collision records, is created as follows. If both cars in a collision were level 3A/4A matches with known NCAPINJ, that collision contributes two records to the analysis file: one record with car 1 as the "case" vehicle and car 2 as the "other" vehicle, and the other record, vice-versa. If only one car in a collision was a 3A/4A match with known NCAPINJ, the analysis file contains one record, with that car as the "case" vehicle. Each record on the analysis file contains NCAP scores for the "case" vehicle and the curb weight, driver age and sex for both vehicles.

The first step in the analysis is a regression on these 1189 head-on collision cases whose dependent variable is the outcome for the driver of the case vehicle (fatality = 1, survival = 0) and whose independent variables are W, A and S - relative vehicle weight, driver age and sex:

Case Vehicle Is 3A/4A Match; Other Vehicle Is Any 1976-91 Car

| | Reg. Coeff. | Chi Square | Stat. Sig.? | Partial Corr. |
|---|---|---|---|---|
| INTERCEPT | .525 | 39.44 | RR | |
| W (car weight) | -5.214 | 219.92 | RR | -.379 |
| A (driver age) | .0516 | 195.72 | RR | .358 |
| S (driver sex) | .38 | 9.80 | RR | .072 |

The intercept and coefficients for W, A and S are similar to those obtained in Section 4.2, but the chi-squares are larger because there are more accident cases. The model is used to predict the expected fatality risk for each driver in the collision, in the absence of NCAP information. The expected fatality risk $E_{case}$ for the driver of the case vehicle is

$$\frac{\exp[.525 - 5.214(\log W_{case} - \log W_{other}) + .0516(A_{case}-A_{other}) + .38(F_{case}-F_{other})]}{1 + \exp[.525 - 5.214(\log W_{case} - \log W_{other}) + .0516(A_{case}-A_{other}) + .38(F_{case}-F_{other})]}$$

where $W_{case}$ is the curb weight of the case vehicle, $A_{case}$ is the age of the driver of the case vehicle and $F_{case}$ is 1 if the driver of the case vehicle is female, 0 if the driver is male. The expected fatality risk $E_{other}$ for the driver of the other vehicle is

$$\frac{\exp[.525 + 5.214(\log W_{case} - \log W_{other}) - .0516(A_{case}-A_{other}) - .38(F_{case}-F_{other})]}{1 + \exp[.525 + 5.214(\log W_{case} - \log W_{other}) - .0516(A_{case}-A_{other}) - .38(F_{case}-F_{other})]}$$

A fatality risk index can be computed for any subset of case vehicles (e.g., the case vehicles with poor NCAP scores), as follows. Each collision has an actual outcome $A_{case}$ for the driver of the case vehicle (fatality = 1, survival = 0) and $A_{other}$ for the driver of the other car. The actual and expected fatalities are summed over all the crashes included in the subset: sum($A_{case}$) and sum($A_{other}$) are the actual numbers of driver fatalities in the case and the other vehicles; sum($E_{case}$) and sum($E_{other}$) are the numbers of driver fatalities that would be expected in the case vehicles and the other vehicles, given the relative weight, age and sex in each crash. The fatality risk index for that subset of case vehicles is

$$\textbf{Index} = 100\ [\textbf{sum}(A_{case})/\textbf{sum}(E_{case})]\ /\ [\textbf{sum}(A_{other})/\textbf{sum}(E_{other})]$$

The risk index for any subset of case vehicles measures the fatality risk for this subset of case vehicles relative to the "average car on the road." The critical assumption here is that the "other" vehicles in these crashes are an essentially random sample of 1976-91 passenger cars with belted drivers: a representative cross-section of the "average car on the road." The assumption will be tested later in the chapter. If this particular group of case vehicles is as crashworthy as the average car on the road, sum($A_{case}$) will approximately equal sum($E_{case}$) and sum($A_{other}$) will approximately equal sum($E_{other}$): the risk index will be close to 100. The lower the risk index, the more crashworthy the subset of case vehicles (for belted drivers in actual head-on collisions).

The fatality reduction for one group of cars as compared to another, taking into account the full range of head-on collisions that can occur on the highway, is measured by the relative difference in the risk indices. For example, a risk index is computed for a subset of case vehicles with "good" NCAP scores and also for a subset with "poor" NCAP scores. The fatality reduction for good NCAP scores relative to poor NCAP scores is

$$\text{Fatality Reduction} = 1 - (\text{Index}_{good}/\text{Index}_{poor})$$

$$\mathbf{RELEXP = (A_{case} - E_{case}) - (A_{other} - E_{other})}$$

is defined for each individual collision, and it measures the actual safety performance of the vehicles "relative to expectations," as in Section 4.2. In the 1189 crashes, and in most subgroups of the crashes, RELEXP has a population standard deviation of 0.64. That makes it easy to test if the average value of RELEXP is significantly less than zero for a specific group of crashes (i.e., the case vehicles were significantly safer than the other cars), or if the difference in average RELEXP for two groups of crashes is statistically significant. A significance test for the difference in the risk indices for case vehicles with good NCAP scores and for case vehicles with poor NCAP scores is based on

$$Z = [\text{avg } RELEXP_{poor} - \text{avg } RELEXP_{good}] \ / \ [.64*(1/N_{poor} + 1/N_{good})^{.5}]$$

As in Chapter 5, the case vehicles are partitioned into "good" and "poor" NCAP performance groups (with possibly an in-between "borderline" group) by the composite score NCAPINJ, or by the actual NCAP test results for a single body region, two body regions, or all 3 body regions.

## 6.2 Risk indices for good and poor NCAP performers

Table 6-1 presents the results of nine analyses comparing the risk indices of good NCAP performers and poor NCAP performers. In the first of those analyses, the performance criterion is the composite score NCAPINJ. As discussed in Sections 4.1 and 5.8, NCAPINJ is the weighted sum of logistic injury probabilities for the head, chest and femurs that has maximum correlation with fatality risk in the set of head-on collisions where both cars match an NCAP test at the '3A or 4A level.' However, other weighted (or unweighted) sums had almost equally high correlations on this data set, and, in some cases, higher correlations on other data sets.

## TABLE 6-1

## FATALITY RISK INDICES FOR CARS WITH "GOOD" NCAP SCORES VS. CARS WITH "POOR" NCAP SCORES

| Crashes with "Good" Case Vehicles | | | Crashes with "Poor" Case Vehicles | | | Risk Comparison | |
|---|---|---|---|---|---|---|---|
| Definition | N of Crashes | Fatality Risk Index | Definition | N of Crashes | Fatality Risk Index | % Fat. Red. for "Good" Car | Z Test for Equal Relexp |
| NCAPINJ ≤ .6 | 951 | 93.90 | NCAPINJ > .6 | 238 | 119.35 | 21.3 | 3.01 |
| Chest g's ≤ 56 | 912 | 94.73 | Chest g's > 56 | 259 | 111.89 | 15.3 | 2.13 |
| HIC ≤ 900 | 737 | 95.75 | HIC > 1300 | 146 | 110.37 | 13.2 | 1.32 |
| L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | 858 | 93.43 | L Femur > 1400 OR<br>R Femur > 1400 OR<br>L+R Femur > 2400 | 325 | 113.63 | 17.8 | 2.62 |
| HIC ≤ 900 AND<br>Chest g's ≤ 55 | 654 | 94.49 | HIC > 1250 OR<br>Chest g's > 65 | 234 | 111.37 | 15.2 | 1.95 |
| Chest g's ≤ 56 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | 676 | 91.85 | Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 363 | 113.07 | 18.8 | 2.83 |
| HIC ≤ 900 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | 526 | 93.76 | HIC > 1300 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 349 | 113.49 | 17.4 | 2.34 |
| HIC ≤ 900 AND<br>Chest g's ≤ 56 AND<br>L Femur ≤ 1400 AND<br>R Femur ≤ 1400 AND<br>L+R Femur ≤ 2400 | 463 | 92.62 | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 446 | 109.58 | 15.5 | 2.21 |
| HIC ≤ 1200 AND<br>Chest g's ≤ 56 AND<br>L Femur ≤ 2000 AND<br>R Femur ≤ 2000 AND<br>L+R Femur ≤ 3000 | 694 | 94.37 | HIC > 1400 OR<br>Chest g's > 62 OR<br>L Femur > 2250 OR<br>R Femur > 2250 OR<br>L+R Femur > 3200 | 301 | 112.10 | 15.8 | 2.24 |

Intuitively, the boundary between "poor" and "acceptable/good" performance should be set so that approximately 20-25 percent of cars will be in the poor performance group: about 200-300 cars, given that the data file contains 1189 case vehicles. A single boundary value of NCAPINJ = 0.6, the same as used in Section 5.1, puts 238 cars into the "poor" performance group, leaving 951 cars in the "acceptable" group.

In the 951 crashes where the case vehicle had NCAPINJ $\leq$ 0.6, and the other vehicle could be any 1976-91 car with a belted driver, there were 572 driver fatalities in the case vehicles, but 590.1 were expected. There were 511 fatalities in the other vehicles, but 495.0 were expected. (The "other" cars average 200 pounds heavier than the case vehicles, so they have fewer expected fatalities.) The fatality risk index for the case vehicles is

$$\text{Index}_{good} = 100\ (572/590.1)\ /\ (511/495.0) = 93.9$$

In the 238 crashes where the case vehicle had NCAPINJ > 0.6, there were 150 actual and 133.2 expected fatalities in the poor NCAP performers. There were 132 actual and 139.9 expected fatalities in the "other" vehicles. The index is

$$\text{Index}_{poor} = 100\ (150/133.2)\ /\ (132/139.9) = 119.4$$

Over these 1189 collisions, the <u>fatality reduction</u> for good NCAPINJ scores relative to poor NCAP scores is

$$\text{Fatality Reduction} = 1 - (93.9\ /119.4) = 21 \text{ percent}$$

The average value of RELEXP is -.0359 in the 951 crashes involving "good" case vehicles and +.1038 in the 238 crashes involving "poor" case vehicles.

$$Z = [.1038 - (-.0359)]\ /\ [.64*(1/238 + 1/951)^{.5}] = 3.01$$

so the fatality reduction is statistically significant at the .01 level.

The other analyses of Table 6-1 partition the case vehicles into "good" and "poor" NCAP performance groups by the same criteria as in Chapter 5: first by a single NCAP parameter, then by a pair of NCAP body regions and, finally, by independent pass-fail criteria on all three body regions. In every analysis, the group of case vehicles with good NCAP performance has a lower fatality risk index than the poor performers. Most of the fatality reductions are statistically significant at the .05 level. Cars with chest g's $\leq$ 56 have a risk index 15 percent lower than cars with chest g's > 56 (Z for equal RELEXP is 2.13, $p < .05$). The fatality reduction for HIC $\leq$ 900, relative to HIC > 1300 is 13 percent, which comes close to statistical significance (Z for equal RELEXP is 1.32). Cars with low femur loads are 18 percent safer than cars with high femur loads (Z for equal RELEXP is 2.62, $p < .01$).

When the cars are partitioned according to NCAP scores for any two of the three body regions, or for all three body regions, the "good" cars always have significantly lower risk indices than the "poor" cars. The fatality reductions range from 15 to 19 percent; the sample of case vehicles with "poor" performance ranges from 234 to 446.

The results in Table 6-1 can be compared to those in Table 5-14, which summarized the analyses of those limited subsets of crashes where a "good" performer hit a "poor" performer. The reductions in the fatality risk indices, in the broad-based analyses of Table 6-1, range from 1 to 6 percent lower than the effects in the specialized, high-contrast analyses of Table 5-14, as follows:

| NCAP Performance Criterion | Fatality Reduction in Table 5-14 (%) | Fatality Reduction in Table 6-1 (%) |
|---|---|---|
| NCAPINJ | 26.4 | 21.3 |
| Chest g's | 18.7 | 15.3 |
| HIC | 14.2 | 13.2 |
| Femur load | 20.1 | 17.8 |
| HIC, chest g's | 18.9 | 15.2 |
| Chest g's, femur load | 22.1 | 18.8 |
| HIC, femur load | 19.4 | 17.4 |
| HIC, chest g's, femur load | 21.2 | 15.5 |

It is unknown why the analysis methods of Chapters 5 and 6 do not produce identical results. The small differences of 1-6 percent could easily be due to chance alone. Even though the effectiveness is always higher in the Chapter 5 analysis (and, at first glance, that resembles flipping a coin and getting "heads" 8 times in a row), the various analyses are hardly independent. They all use the same data set; the groups of cars with "good" performance largely overlap, and so do the groups of cars with poor performance. The 1-6 percent difference is well within the range of sampling error on a single analysis.

If the discrepancies are not due to chance alone, one possibility is that the "other" vehicles in the Chapter 6 analyses are not, as had been assumed above, essentially random samples of 1976-91 passenger cars with belted drivers. The basic assumption of this chapter was that the "good" cars and the "poor" cars are both hitting "average" cars, so their fatality risk indices are directly comparable. If, in fact, the "good" cars tend to hit other "good" cars and the "poor" cars tend to hit other "poor" cars, the difference in the risk indices would understate the fatality reduction for good NCAP scores. Detailed statistics, however, support the basic assumption. For example, in the first analysis of Table 6-1, the case vehicles with NCAPINJ $\leq$ 0.6 have average NCAPINJ

= 0.31 and the case vehicles with NCAPINJ > 0.6 have average NCAPINJ = 0.90. But the "other" cars that hit the "good" cars have nearly the same average NCAPINJ as the "other" cars that hit the "poor" cars: 0.453 vs. 0.466. They also have nearly the same average weight (2971 vs. 3010 pounds), driver age (42.7 vs. 43.2 years) and driver gender distributions (45 percent female vs. 44 percent female). The "other" cars, in both cases, are basically identical.

Another possibility is that the specialized analyses of Chapter 5 somehow intensify the fatality reductions associated with good NCAP scores. In the full range of head-on collisions, "good" scores reduce risk by X and "poor" scores increase it by Y, but when a "good" car specifically hits a "poor" car, the difference in risk may be even greater than X + Y. If so, the results of this chapter provide a more conservative assessment of the overall reduction of fatality risk for cars with good NCAP scores.

# CHAPTER 7

## ACTUAL CRASHWORTHINESS AND NCAP PERFORMANCE DURING 1979-91

It is well known that the performance of passenger cars on the NCAP test has greatly improved since the program was initiated in 1979. Substantial reductions in HIC and chest g's have been documented in NHTSA's 1992-93 Reports to the Congress [17], [23], [24] and in NHTSA presentations at ESV conferences during 1983-92 [7], [13], [14], [15], [16]. The studies cited specific changes in vehicle structures and occupant protection systems that improved NCAP performance. The first six chapters of this report demonstrated statistically significant associations between NCAP performance and the fatality risk of belted drivers in head-on collisions. Given that NCAP performance improved during 1979-91, and that good NCAP performers have lower fatality risk in actual crashes, it is logical to expect that cars became safer in actual crashes during 1979-91. This last chapter estimates the payoff: the reduction in the actual fatality risk of belted drivers in head-on collisions since 1979. (Of course, this report is a statistical study and it does not pin down cause and effect. Although it shows that cars became safer as NCAP scores improved, it does not prove that the NCAP program was a stimulus for each of the vehicle changes that saved lives during 1979-91. For example, Federal Motor Vehicle Safety Standard 208 has been an important stimulus for safety improvements during the NCAP era.)

In general, the fatality rates of cars of different model years are not directly comparable. There are two patterns in fatality rates that create the delusion that "cars are getting safer all the time." The overall fatality rate per 100 million vehicle miles has been declining for a long time - e.g., from 5 in 1969 to 3 in 1981 to 2 in 1992. But that improvement may primarily reflect long-term changes in driving behavior, roadway environments and demographics, not crashworthiness. For example, pedestrian fatalities, which are unaffected by improvements in vehicle interior crashworthiness, have declined as rapidly as occupant fatalities. The bias from long-term population trends could be avoided by using accident data from a single calendar year and comparing the fatality rates of cars of two different model years. However, in any single calendar year, the cars of earlier model years are older than the late-model cars. Because of their drivers' demographics and behavior, older cars typically accumulate fewer miles, but have more severe crashes; low-severity crashes of old cars are often unreported. Thus, the fatality rate per 100 million miles or per 100 reported crashes is lower for new cars than for old cars, even if both are equally crashworthy.

A head-on collision between cars of two different model years, however, reveals their relative crashworthiness. Both cars are in essentially the same frontal collision, on the same road, in the same year, on the same accident report. The behavior of each driver, prior to the impact, has little effect on who dies during the impact. After adjustment for differences in car weight, driver age and sex (vulnerability to injury), the model year with more survivors is more crashworthy. The methods of the preceding chapters, used there to compare good and poor NCAP performers, will now be used to compare cars of different model years. NHTSA's 1988 <u>Evaluation of Occupant Protection in Frontal Interior Impact</u> has already used this method to compare the fatality risk of cars of different model years for <u>unrestrained</u> drivers [19], pp. 111-140. It found that cars of model years 1970 through 1984 were about equally crashworthy for unrestrained drivers in head-on collisions. The remainder of this chapter studies the trend in fatality risk for <u>belted</u> drivers.

7.1 Cars with late model years hit cars with early model years

A straightforward way to estimate the improvement in crashworthiness over time is to study only those collisions where a late-model car hits an early-model car. Table 7-1 presents seven analyses of the same accident file that was used in Chapter 5: collisions of two 1979-91 cars, both of which match up with an NCAP test at level 3A or 4A - i.e., the FARS and NCAP vehicles are of the same make-model or true corporate cousins, and the FARS model year is within the "valid" range of the NCAP test (see Sections 2.5 and 3.4). In the first analysis, the "early" cars are MY 79-81 and the "late-model" cars are MY 82-91. The boundary between "early" and "late" is pushed forward, one year at a time, in the subsequent analyses.

Table 7-1 demonstrates statistically significant improvements in crashworthiness for belted drivers in head-on collisions. For example, the second analysis in the table is based on 121 actual head-on collisions between a model year 1979-82 car and a model year 1983-91 car. This analysis allows a comparison of cars built during the first four years of NCAP to subsequent cars, where manufacturers have had time to build in safety improvements. In the 121 older cars, 80 drivers died, whereas only 69.2 fatalities were expected, based on car weight, driver age and sex. In the newer cars, there were 61 actual and 71.2 expected driver fatalities. That is a fatality reduction of

$$1 - [(61/80) / (71.2/69.2)] = 26 \text{ percent}$$

for the 1983-91 cars, and it is statistically significant (t for RELEXP is 3.09, $p < .001$). Fatality reductions greater than 20 percent were also found in the first analysis (79-81 vs. 82-91) and the third analysis (79-83 vs. 84-91). When the boundary between "early" and "late" is pushed beyond 1983, the fatality reduction is diluted, because vehicles with safety improvements are taken out of

TABLE 7-1

COLLISIONS OF CARS WITH "LATE" MODEL YEARS INTO CARS WITH "EARLY" MODEL YEARS:
EFFECT OF MOVING THE BOUNDARY BETWEEN "EARLY" AND "LATE" MODEL YEARS
(MY 1979-91 cars; both cars match an NCAP test at level 3A or 4A)

| | | Comparison of Real-World Performance | | | | |
|---|---|---|---|---|---|---|
| **Definition of "Late" MY Car** | **Definition of "Early" MY Car** | **N of Crashes** | **% Fat Red for Late MY Car** | **Mean Relexp** | **Sum Relexp** | **T-Test for Relexp** |
| MY 82-91 | MY 79-81 | 98 | 29.9 | -.205 | -20.06 | 3.66 |
| MY 83-91 | MY 79-82 | 121 | 25.9 | -.174 | -21.02 | 3.09 |
| MY 84-91 | MY 79-83 | 146 | 21.3 | -.138 | -20.11 | 2.69 |
| MY 85-91 | MY 79-84 | 177 | 14.7 | -.092 | -16.34 | 1.93 |
| MY 86-91 | MY 79-85 | 183 | 9.0 | -.055 | -10.15 | 1.15 |
| MY 87-91 | MY 79-86 | 183 | 18.8 | -.120 | -21.97 | 2.53 |
| MY 88-91 | MY 79-87 | 143 | 14.8 | -.090 | -12.92 | 1.69 |

| | | Mean NCAP Scores and Model Year ("Late" vs. "Early" Car) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Definition of** | **Definition of** | **HIC** | | **Chest G's** | | **L Femur** | | **R Femur** | | **Model Year** | |
| **"Late" MY Car** | **"Early" MY Car** | **Late** | **Early** | **Late** | **Early** | **Late** | **Early** | **Late** | **Early** | **Late** | **Early** |
| MY 82-91 | MY 79-81 | 858 | 1029 | 47.4 | 55.9 | 859 | 942 | 904 | 999 | 86.1 | 80.0 |
| MY 83-91 | MY 79-82 | 821 | 1021 | 46.4 | 54.9 | 905 | 976 | 915 | 1048 | 86.8 | 80.6 |
| MY 84-91 | MY 79-83 | 853 | 1000 | 46.2 | 53.6 | 902 | 1004 | 891 | 1040 | 87.0 | 81.1 |
| MY 85-91 | MY 79-84 | 872 | 959 | 46.5 | 50.8 | 956 | 944 | 945 | 983 | 87.4 | 82.0 |
| MY 86-91 | MY 79-85 | 883 | 943 | 46.7 | 50.4 | 975 | 926 | 955 | 939 | 87.8 | 82.6 |
| MY 87-91 | MY 79-86 | 864 | 942 | 47.0 | 49.5 | 1022 | 948 | 1022 | 911 | 88.3 | 83.2 |
| MY 88-91 | MY 79-87 | 808 | 963 | 47.1 | 49.1 | 1014 | 885 | 1045 | 873 | 89.0 | 83.7 |

the "late" group and placed in the "early" group.

Each of the cars analyzed in Table 7-1 matched up with an NCAP case. Based on these match-ups, the lower section of Table 7-1 compares the average NCAP performance of the early and late-model cars. (NCAP scores are averaged over the various cars on the accident file - i.e., each NCAP-tested model is, so to speak, weighted by the number of fatal crashes involving that model.) HIC has improved from an average of 1029 in 1979-81 cars to 808 in 1988-91 cars. HIC was reduced from 1030 to the mid-800's in 1982-83, stayed close to that level for the next 5 years, and dropped below 800 after 1988. Chest g's were reduced from 56 in 1979-81 to 47 in 1982-91. Most of the reduction was achieved in the first 4 or 5 years of NCAP; chest g's have been close to 47 since 1982. Average femur loads dropped from about 1000 to 900 in the mid 1980's, but crept back to 1000 in the late 1980's. However, in the second analysis of Table 7-1, the 121 1983-91 cars performed substantially better than the 121 1979-82 cars on every NCAP parameter. Average HIC declined from 1021 to 821, chest g's from 54.9 to 46.4, left femur load from 976 to 905 and right femur load from 1048 to 915. The composite NCAP score, NCAPINJ (defined in Section 4.1), declined by a statistically significant 0.206 ($t = 5.85$, $p < .0001$).

The preceding analyses were based on cars that matched up with NCAP tests: the same data base as in Chapter 5. However, if the objective is merely to compare the crashworthiness of early vs. late-model cars, without regard to their NCAP performance, it is not necessary to limit the data to cars with matching NCAP information. In Table 7-2, the analysis has been extended to include any head-on collision between two 1979-91 cars, with both drivers belted. That is a set of 723 collisions (1,446 distinct vehicles) - nearly double the

TABLE 7-2

COLLISIONS OF CARS WITH "LATE" MODEL YEARS INTO CARS WITH "EARLY" MODEL YEARS:
EFFECT OF MOVING THE BOUNDARY BETWEEN "EARLY" AND "LATE" MODEL YEARS
(all MY 1979-91 cars - not necessarily matching an NCAP test)

| | | Comparison of Real-World Performance | | | | |
|---|---|---|---|---|---|---|
| **Definition of "Late" MY Car** | **Definition of "Early" MY Car** | **N of Crashes** | **% Fat Red for Late MY Car** | **Mean Relexp** | **Sum Relexp** | **T-Test for Relexp** |
| MY 82-91 | MY 79-81 | 196 | 23.0 | -.147 | -28.80 | 3.33 |
| MY 83-91 | MY 79-82 | 241 | 22.2 | -.143 | -34.54 | 3.43 |
| MY 84-91 | MY 79-83 | 274 | 20.5 | -.130 | -35.64 | 3.30 |
| MY 85-91 | MY 79-84 | 321 | 15.6 | -.096 | -30.94 | 2.62 |
| MY 86-91 | MY 79-85 | 321 | 11.8 | -.071 | -22.91 | 1.88 |
| MY 87-91 | MY 79-86 | 311 | 13.5 | -.083 | -25.68 | 2.14 |
| MY 88-91 | MY 79-87 | 248 | 8.9 | -.053 | -13.12 | 1.24 |

sample available for Table 7-1. Of course, Table 7-2 does not have a "lower section" like Table 7-1, since NCAP information is unavailable for many of the vehicles.

The pattern of fatality reductions in Table 7-2 is similar to Table 7-1, with the oscillations smoothed by the larger sample size. In the second analysis, the fatality reduction for 1983-91 cars, relative to 1979-82 cars is 22 percent, and it is statistically significant (t for RELEXP is 3.43, $p < .001$). The fatality reductions in the first three analyses are over 20 percent, as in Table 7-1, and here they are all significant at the .001 level.

7.2 Fatality risk index and average NCAPINJ by model year

A more generalized analysis of crashworthiness trends over time is achieved by computing fatality risk indices for cars of different model year groups. The procedure for estimating risk indices was developed in Section 6.1, and it was applied in Section 6.2 to compare the index for cars with good NCAP scores vs. cars with poor scores. However, a risk index can be calculated for any group of cars, such as all cars of a specific model year, or a group of model years. As in Chapter 6, the data base comprises all head-on collisions in which the "case" vehicle of interest is a 1979-91 car that matches up with an NCAP test, whose driver wore belts, but the "other" vehicle in the crash can be any 1976-91 passenger car with a belted driver, not necessarily matching with an NCAP test (1189 accident records). The actual and expected fatalities are tallied in the "case" and "other" vehicles; the fatality risk index is

$$100\ [(\text{actual}_{\text{case}}/\text{actual}_{\text{other}})\ /\ (\text{expected}_{\text{case}}/\text{expected}_{\text{other}})]$$

One advantage of this approach, unlike the method in Section 7.1, is that 1979-91 cars can be partitioned into more than two model-year groups. Specifically, case vehicles are assigned to three model-year groups: 1979-82, 1983-86 and 1987-91. The initial years of NCAP were 1979-82. By 1983-86, manufacturers had leadtime to address major deficiencies in the initial NCAP test results. The 1987-91 cars were often equipped with air bags or other automatic protection.

In the 280 crashes with a 1979-82 case vehicle, there were 181 driver fatalities in the case vehicles, but 166.2 were expected. There were 141 fatalities in the other vehicles (any 1976-91 car with a belted driver), but 154.0 were expected. The fatality risk index for 1979-82 cars is

$$\text{Index}_{79\text{-}82} = 100\ (181/166.2)\ /\ (141/154.0) = 119.0$$

The fatality risk index for the 452 1983-86 cars is just 95.0, and the risk index for the 457 1987-91 cars drops to 90.9 (a risk index of 100 corresponds to the "average" 1976-91 car on the road with a belted driver). The fatality reduction from 1979-82 to 1983-86 is

$$1 - (95.0/119.0) = 20 \text{ percent}$$

and it is statistically significant (Z for equal RELEXP is 2.60, $p < .01$). The additional fatality reduction from 1983-86 to 1987-91 is 4 percent, which is not statistically significant. The net fatality reduction from 1979-82 to 1987-91 is a statistically significant 24 percent (Z for equal RELEXP is 3.18, $p < .01$).

It is especially interesting to compare the trend in the actual fatality risk index with the trend in NCAP performance. Each of the case vehicles in the preceding analysis matched up closely with an NCAP test and has a composite NCAP score, NCAPINJ. The composite scores are averaged for each of the three model-year groups on the <u>accident</u> file (i.e., each NCAP-tested model is, so to speak, weighted by the number of fatal crashes involving that model). The risk indices and average NCAP performance for each model-year group are as follows:

| | Model Years | | |
|---|---|---|---|
| | **1979-82** | **1983-86** | **1987-91** |
| Fatality risk index in actual head-on collisions | 119 | 95 | 91 |
| Average value of NCAPINJ | .59 | .40 | .37 |
| Percent of cars with NCAPINJ > 0.6 | 49 | 14 | 9 |
| Average HIC | 1052 | 915 | 827 |
| Average chest g's | 54.9 | 46.8 | 46.5 |
| Average left femur load | 928 | 883 | 1002 |
| Average right femur load | 1079 | 784 | 1018 |

The trends in the actual fatality risk and the average value of NCAPINJ are almost identical. The risk index decreased from 119 in model years 1979-82 to 95 in 1983-86, to 91 in 1987-91, a large reduction followed by a much smaller reduction. In parallel, NCAPINJ greatly improved from an average of .59 in model years 1979-82 to .40 in 1983-86, with an additional, modest improvement to .37 in 1987-91. The percentage of cars with poor NCAP performance (NCAPINJ > 0.6, a yardstick established in Chapters 5 and 6) also took a big drop, from 49 percent in 1979-82 to 14 percent in 1983-86, followed by a small drop to 9 percent in 1987-91. As discussed in Sections 4.1 and 5.8, NCAPINJ is the weighted sum of logistic injury probabilities for the head, chest and femurs that has maximum correlation with fatality risk in the set of head-on collisions where both cars match an NCAP test at the '3A or 4A level.' However, other weighted (or unweighted) sums had almost equally high correlations on this data set, and, in some cases, higher correlations on other data sets. While NCAPINJ nicely portrays the trend of improved NCAP performance, other weighted (or unweighted) sums will show quite similar trends. Average HIC was substantially reduced from 1979-82 to 1983-86, and again from 1983-86 to 1987-91. Chest g's were greatly reduced from 1979-82 to 1983-86, but stayed about the same after that. The average femur loads did not change much during 1979-91.

Figure 7-1 graphs the actual fatality risk index, by model year, from 1979 to 1991 (data grouped into two-model-year cohorts, to smooth the results). Figure 7-2 graphs the average value of NCAPINJ, and Figure 7-3, the percentage of cars with NCAPINJ > 0.6, by model year from 1979 to 1991. The three figures have nearly identical patterns: little, if any, improvement from 1979 to 1981; impressive reductions from 1982 to 1984; leveling off after 1984, with a possible trend of further improvements after 1988.

As in the preceding section, the computation of risk indices does not have to be limited to case vehicles which closely match an NCAP test, but can be extended to <u>all</u> 1979-91 cars that collided head-on with a 1976-91 car, with both drivers belted. That has the advantage of extending the sample size from 1189 to 1632 vehicles, although, without the NCAP matches, the trend in risk indices cannot be compared with the trend in NCAPINJ. The risk indices in this extended sample are about the same as in the preceding analysis:

| | **3A/4A Matches** | | **Extended Sample** | |
|---|---|---|---|---|
| | N | Risk Index | N | Risk Index |
| MY 1979-82 | 280 | 119.0 | 425 | 117.5 |
| MY 1983-86 | 452 | 95.0 | 610 | 96.5 |
| MY 1987-91 | 457 | 90.9 | 597 | 92.6 |

FIGURE 7-1: FATALITY RISK INDEX IN ACTUAL HEAD-ON COLLISIONS, BY MODEL YEAR

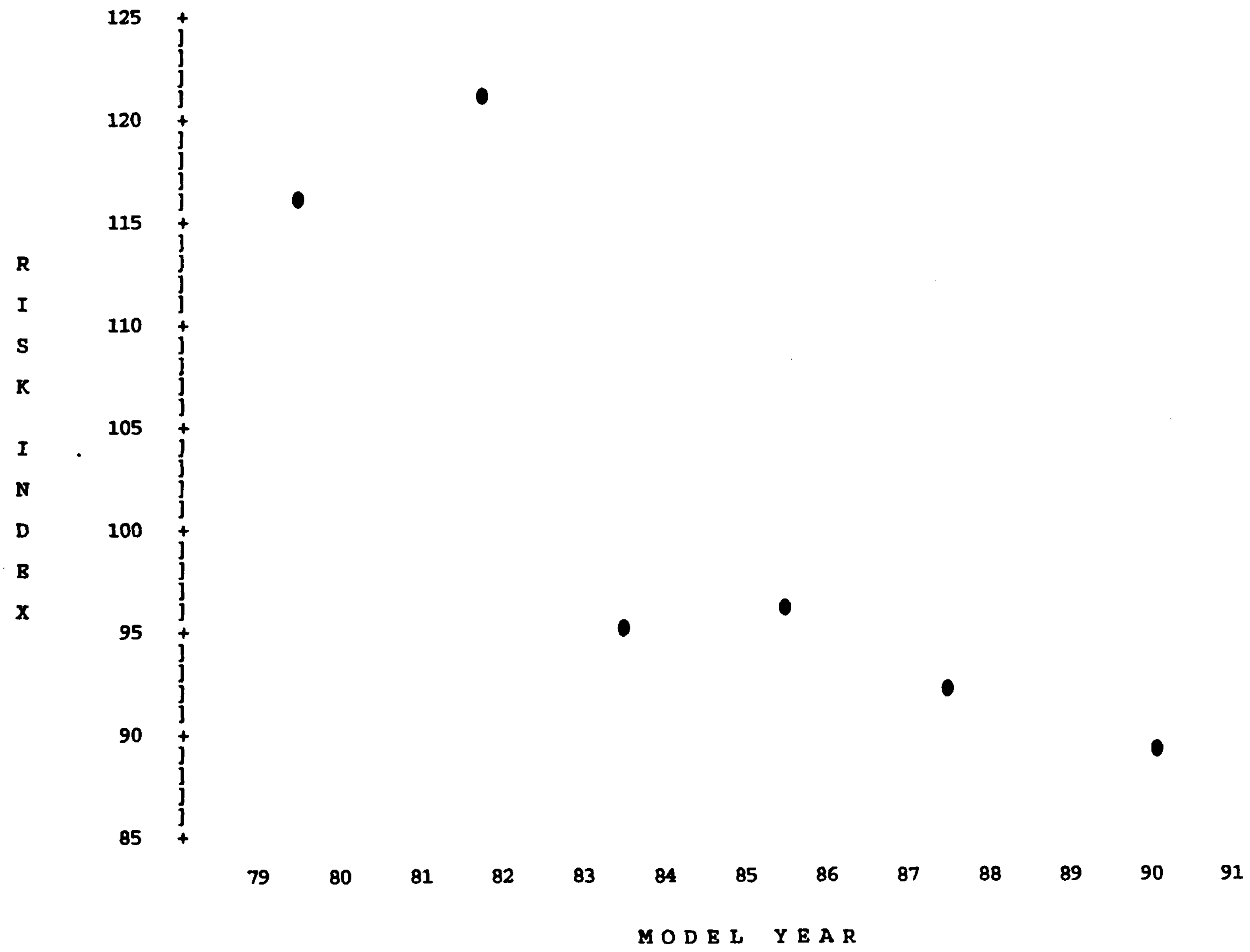


FIGURE 7-2: AVERAGE VALUE OF NCAPINJ, BY MODEL YEAR

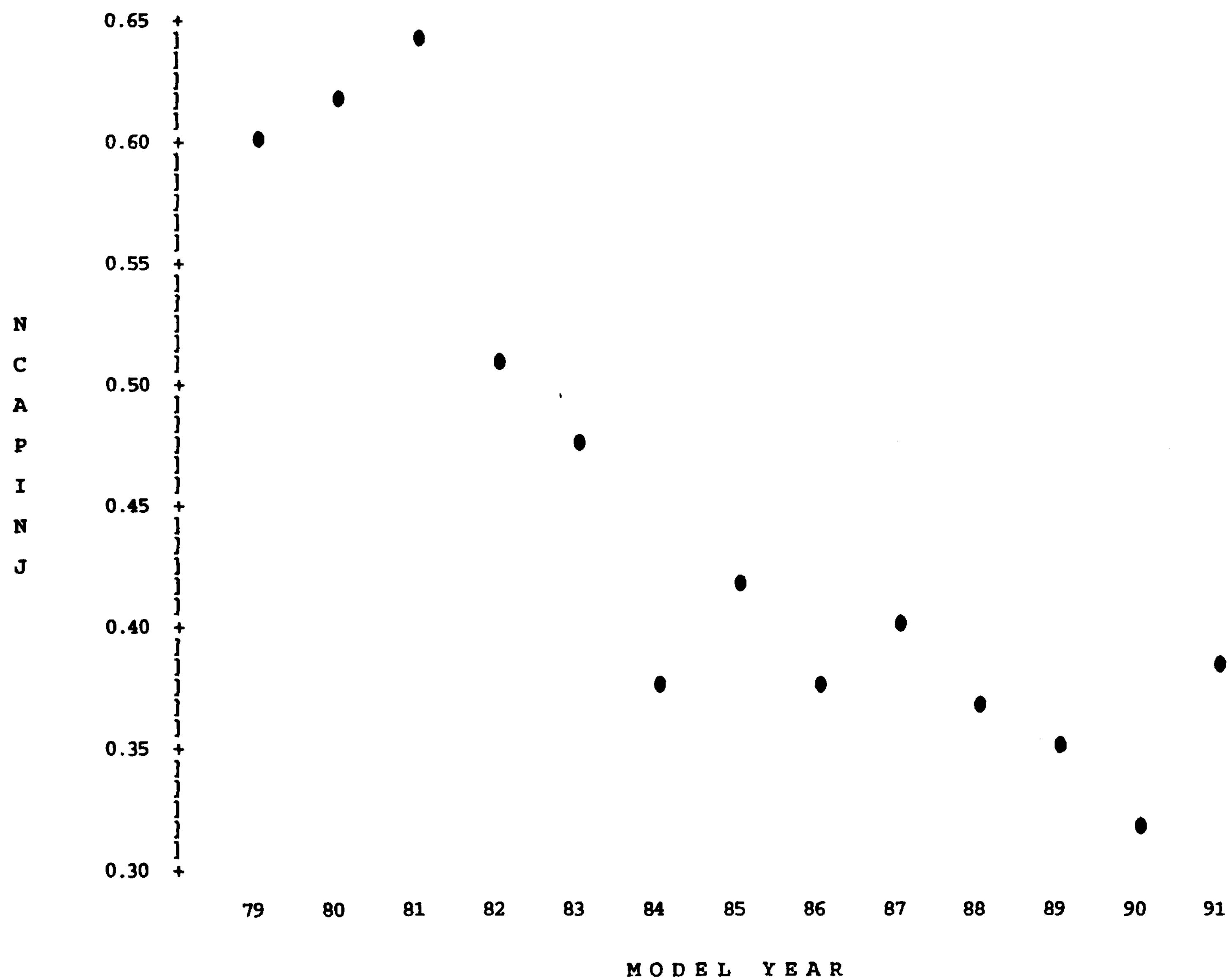
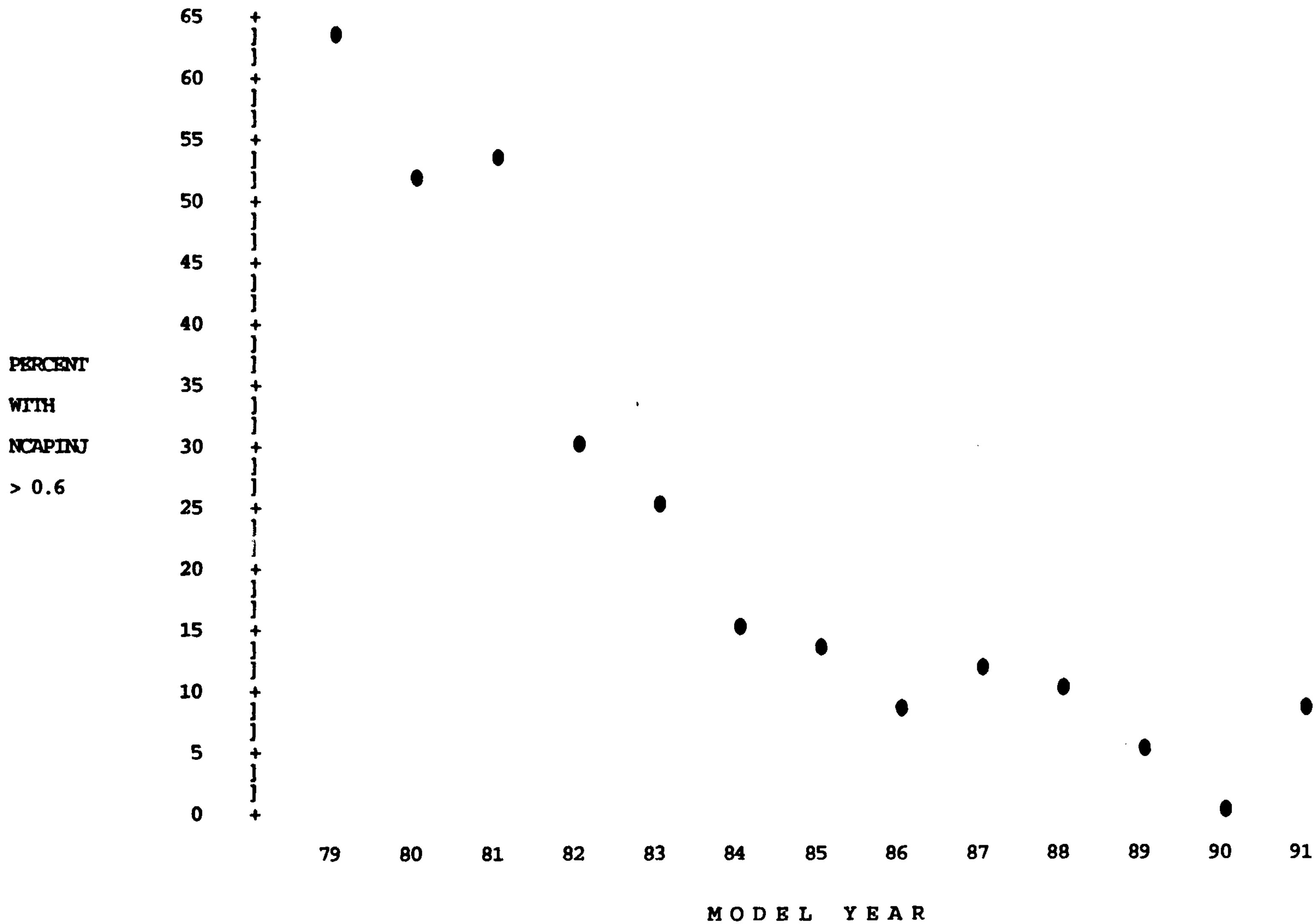


FIGURE 7-3: PERCENT OF CARS WITH NCAPINJ > 0.6, BY MODEL YEAR



The fatality reduction from 1979-82 to 1983-86 is a statistically significant 18 percent (Z for equal RELEXP is 2.69, $p < .01$). The additional fatality reduction from 1983-86 to 1987-91 is a nonsignificant 4 percent. The net fatality reduction from 1979-82 to 1987-91 is a statistically significant 21 percent (Z for equal RELEXP is 3.29, $p < .001$), slightly less than the 24 percent in the preceding analysis.

## 7.3 Comparison of the NCAPINJ and model-year effects

The principal finding in Chapters 4-6 is that cars with good NCAP performance are about 20 percent safer in head-on collisions than cars with poor NCAP performance. The principal finding here is that late-model cars are likewise about 20 percent safer than early-model cars. These two findings don't quite "add up." Although late-model cars have, on the average, substantially better NCAP performance than earlier models, the late models are not all "good" and the early models are certainly not all "poor." Thus, the 20 percent fatality reduction for late models cannot be fully explained by the 20 percent fatality reduction for good vs. poor NCAP performance. There has been some "residual" improvement, during 1979-91, which is not "explained" by a composite score such as NCAPINJ, or by other variables derived from NCAP scores. The three remaining analyses of this chapter compare the fatality reductions associated with NCAPINJ and the "residual" model-year effect.

An important reminder: the analyses that follow describe statistical associations, not cause-and-effect relationships. The portion of the 1979-91 fatality reduction "attributable to the reduction of NCAPINJ" is not necessarily "caused by NCAP." The "residual" reduction is not necessarily "caused by factors other than NCAP." Just because a vehicle change reduced NCAP scores does not prove that it was implemented purely in response to NCAP. Conversely, if a vehicle change improves actual safety without having much of an effect on NCAP scores, that does not prove that it was unrelated to NCAP: manufacturers usually don't know, in advance, exactly how a vehicle change will affect scores; this change might have been motivated, in part, by a hope that NCAP scores would improve. No claim is made here to include <u>or</u> exclude any portion of the actual fatality reduction as "lives saved by the NCAP program."

Chapter 5 demonstrated that when a car with "good" NCAP scores hits a car with "poor" scores, the "good" cars have a significant safety advantage. However, most of the comparisons in Chapter 5 showed that the "good" NCAP performers also had a later model year, on the average, than the "poor" performers. For example, in Table 5-1, the average model year of cars with NCAPINJ $\leq$ 0.6 was 85.8, and the average model year of cars with NCAPINJ > 0.6 was 83.1. The first analysis asks whether the fatality reduction for good NCAP scores exists independently of model year, or whether it is merely an artifact of the better NCAP performers being more recent cars.

The comparisons in Chapter 5 did not place any limits on the relative model years of the two cars in the head-on collisions. The data included some crashes in which one car might have been 10 or even 12 years older than the other. Table 7-3 limits two of the Chapter 5 comparisons to subsets of head-on collisions in which the model years of the two cars are close to one another. In the subsample of 61 collisions of a car with NCAPINJ $\leq$ 0.6 into a car with NCAPINJ > 0.6 in which $-5 \leq MY_{GOOD} - MY_{POOR} \leq 3$, the fatality reduction for the good NCAP performers is 32 percent, which is statistically significant (t for RELEXP is 3.03, $p < .01$) and, in fact, slightly higher than the 26 percent

TABLE 7-3

COLLISIONS OF CARS WITH "GOOD" NCAP SCORES INTO CARS WITH "POOR" SCORES WHERE THE "GOOD" AND "POOR" CARS HAVE SIMILAR MODEL YEARS

| | | Comparison of Real-World Performance | | | | |
|---|---|---|---|---|---|---|
| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **N of Crashes** | **% Fat Red for Good Car** | **Mean Relexp** | **Sum Relexp** | **T-Test for Relexp** |
| NCAPINJ $\le$ .6 and $-5 \le (MY_{GOOD} - MY_{POOR}) \le 3$ | NCAPINJ > .6 | 61 | 32.1 | -.238 | -14.53 | 3.03 |
| HIC $\le$ 900 AND<br>Chest g's $\le$ 56 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400<br>and $-5 \le (MY_{GOOD} - MY_{POOR}) \le 5$ | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 93 | 23.3 | -.137 | -12.76 | 2.27 |

| | | Car Weight, Driver Age, Sex and Model Year ("Good" vs. "Poor" Car) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Definition of a "Good" Car** | **Definition of a "Poor" Car** | **Avg Weight** | | **Avg Age** | | **% Female** | | **Model Year** | |
| | | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** | **Good** | **Poor** |
| NCAPINJ $\le$ .6 and $-5 \le (MY_{GOOD} - MY_{POOR}) \le 3$ | NCAPINJ > .6 | 2909 | 2864 | 43.5 | 41.8 | 46 | 46 | 84.6 | 84.6 |
| HIC $\le$ 900 AND<br>Chest g's $\le$ 56 AND<br>L Femur $\le$ 1400 AND<br>R Femur $\le$ 1400 AND<br>L+R Femur $\le$ 2400<br>and $-5 \le (MY_{GOOD} - MY_{POOR}) \le 5$ | HIC > 1300 OR<br>Chest g's > 60 OR<br>L Femur > 1700 OR<br>R Femur > 1700 OR<br>L+R Femur > 2700 | 2898 | 2740 | 43.2 | 41.9 | 48 | 35 | 85.5 | 85.6 |

fatality reduction in the unrestricted sample of 117 collisions (see Table 5-1). The range of allowable model years, $-5 \leq MY_{GOOD} - MY_{POOR} \leq 3$, serves to equalize the average model year of the good and poor NCAP performers at 84.6. In the subsample of 93 collisions of a car with low HIC, chest g's and femur load into a car with high HIC, chest g's or femur load in which $-5 \leq MY_{GOOD} - MY_{POOR} \leq 5$, the fatality reduction for the good performers is a statistically significant 23 percent (t for RELEXP is 2.07, $p < .05$), which is almost identical to the 21 percent reduction in the unrestricted sample of 118 cases (see Table 5-9). The last line of Table 7-3 shows that the "good" and "poor" NCAP performers have nearly identical average model years. Thus, the strong association between NCAP performance and actual fatality risk exists independently of the model year.

Conversely, the second analysis searches for a model year effect independent of NCAPINJ. In Section 6.2, fatality risk indices were computed for case vehicles with NCAPINJ $\leq$ 0.6 and NCAPINJ > 0.6; in Section 7.2, for late-model and early-model cars. But a risk index can be calculated for any group of case vehicles, including groups defined by their NCAP performance <u>and</u> model years:

| **NCAPINJ** | **Model Years** | **N of Cases** | **Risk Index** |
|---|---|---|---|
| $\leq$ 0.6 | 1979-82 | 144 | 111.2 |
| | 1983-91 | 807 | 91.1 |
| > 0.6 | 1979-82 | 136 | 128.0 |
| | 1983-91 | 102 | 109.2 |

The effect of NCAPINJ and the "residual" effect of model year are both strong and nearly independent in these risk indices. Controlling for NCAPINJ $\leq$ 0.6, the late-model cars are almost 20 percent safer than the early-model cars. Controlling for NCAPINJ > 0.6, the late-model cars again have close to 20 percent lower risk indices than the early model cars. In other words, there is a consistent "residual" model year effect, after controlling for NCAPINJ. But these risk indices also show a consistent effect close to 20 percent for NCAPINJ within model year groups: e.g., for 1979-82 cars, the good NCAP performers had a risk index of 111, and the poor performers, 128. The "N of Cases" column shows a dramatic shift from poor to good NCAP performance in the late-model cars. Thus, the net reduction in the risk index for late-model cars is associated with shift from poor to good NCAP performance plus a "residual" model-year effect.

The third analysis compares, in statistical terms, the relative "strength" of the NCAPINJ effect and the residual model-year effect in the file of 392 head-on collisions where both cars match up with an NCAP test at level 3A or 4A. In Section 4.3, it was shown that DELNCAP = $NCAPINJ_{case}$ - $NCAPINJ_{other}$ has a strongly significant correlation with RELEXP, actual safety performance relative to expectations; the Pearson correlation coefficient was .166 ($p$ = .001, N = 392). However, if another variable, DELMY = $MY_{case}$ - $MY_{other}$ is defined on that file, it also has a significant correlation of -.133 with RELEXP ($p$ = .008). In other words the <u>net</u> correlation of model year with actual fatality risk is significant. But if DELNCAP and DELMY are <u>simultaneously</u> entered as independent variables in a linear regression, with RELEXP as the dependent variable, the regression equation is

$$\text{RELEXP} = .036 + .23\ \text{DELNCAP} - .0148\ \text{DELMY}$$

The coefficient for DELNCAP is statistically significant at the .01 level ($t$ = 2.78), whereas the coefficient for DELMY is barely significant at the .05 level (t = 1.96). That suggests the association of NCAPINJ with fatality risk is strong, while the residual association of model year with fatality risk, after controlling for NCAPINJ, is not quite as strong.

## REFERENCES


[1] The Abbreviated Injury Scale (AIS) - 1990 Revision. Des Plaines, IL: American Association for Automotive Medicine, 1990.

[2] Automotive News Market Data Book Issue. Annual Publication. Detroit: Crain Automotive Group, Inc.

[3] Code of Federal Regulations, Title 49. Washington: Government Printing Office, 1990, Parts 571, 575 and 581.

[4] Comparison of New Car Assessment Program Crash Test Results with Real-World Crash Data: Summary and Analysis of Existing Studies. Submission to Docket No. 79-17-GR-060. Washington: National Highway Traffic Safety Administration, [1988].

[5] Department of Transportation and Related Agencies Appropriations Act for 1992. Conference Report No. 102-243. Washington: United States House of Representatives, October 7, 1991.

[6] Evaluation of the Effectiveness of Occupant Protection: Federal Motor Vehicle Safety Standard 208. Report No. DOT HS 807 843. Washington: National Highway Traffic Safety Administration, [1992].

[7] An Evaluation of the Effects of the New Car Assessment Program on Safety. Washington: National Highway Traffic Safety Administration, [1990].

[8] Evans, Leonard. Risk of Fatality from Physical Trauma versus Sex and Age. Report No. GMR-5359. Warren, MI: General Motors Research Laboratories, [1986].

[9] "FARS Head-On Collision Cases (1976-91 Cars) and Matching NCAP Test Results." Submission to NHTSA Docket No. 79-17-N40-002. Washington: National Highway Traffic Safety Administration, [1994].

[10] Fatal Accident Reporting System, 1981 Coding and Validation Manual. Washington: National Highway Traffic Safety Administration, [1980].

[11] Fatal Accident Reporting System, 1987 Coding and Validation Manual. Washington: National Highway Traffic Safety Administration, [1986].

[12] Fatal Accident Reporting System, 1991 Coding and Validation Manual. Washington: National Highway Traffic Safety Administration, [1990].

[13] Hackney, James. "The Effects of FMVSS No. 208 and NCAP on Safety, as Determined from Crash Test Results.." Thirteenth International Technical Conference on Experimental Safety Vehicles. Washington: National Highway Traffic Safety Administration, 1992.

[14] Hackney, James, and Carlin Ellyson. "A Review of the Effects of Belt Systems, Steering Assemblies, and Structural Design on the Safety of Vehicles in the New Car Assessment Program." Tenth International Technical Conference on Experimental Safety Vehicles. Washington: National Highway Traffic Safety Administration, 1985.

[15] Hackney, James, and Vincent Quarles. "The New Car Assessment Program - Status and Effect." Ninth International Technical Conference on Experimental Safety Vehicles. Washington: National Highway Traffic Safety Administration, 1983.

[16] Hackney, James R., William T. Hollowell and Daniel S. Cohen. "Analysis of Frontal Crash Safety Performance of Passenger Cars, Light Trucks and Vans and an Outline of Future Research Requirements." Twelfth International Technical Conference on Experimental Safety Vehicles. Washington: National Highway Traffic Safety Administration, 1989.

[17] Historical Performance of Different Auto Manufacturers in the New Car Assessment Program Tests. Report to the Congress. Washington: National Highway Traffic Safety Administration, [1993].

[18] Joyner, Stephenie P., ed. SUGI Supplemental Library User's Guide, 1983 Edition. Cary, NC: SAS Institute Inc., 1983.

[19] Kahane, Charles J. An Evaluation of Occupant Protection in Frontal Interior Impact for Unrestrained Front Seat Occupants of Cars and Light Trucks. Report No. DOT HS 807 203. Washington: National Highway Traffic Safety Administration, [1988].

[20] The Motor Vehicle Information and Cost Savings Act of 1973. Public Law 92-513, as amended, 15 United States Code 1901-2012.

[21] National Vehicle Population Profile. Annual Publication. Detroit: R. L. Polk.

[22] "NCAP Test Results for Drivers of Passenger Cars, 1979-91." Submission to NHTSA Docket No. 79-17-N40-002. Washington: National Highway Traffic Safety Administration, [1994].

[23] New Car Assessment Program: Plan for Responding to the NCAP FY 1992 Congressional Directives. Report to the Congress. Washington: National Highway Traffic Safety Administration, [1992].

[24] New Car Assessment Program: Response to the NCAP FY 1992 Congressional Requirements. Report to the Congress. Washington: National Highway Traffic Safety Administration, [1993].

[25] "New Car Assessment Program Results, Model Years 1987-1991." Washington: National Highway Traffic Safety Administration, Office of Market Incentives, September 1991.

[26] "NHTSA Releases Crash Test Results Applicable to 77 Model Year 1993 Vehicles." Press release 65-92. Washington: National Highway Traffic Safety Administration, December 23, 1992

[27] Prasad, Priyaranjan, and Mertz, Harold J. The Position of the United States Delegation to the ISO Working Group 6 on the Use of HIC in the Automotive Environment. SAE Paper No. 851246. Warrendale, PA: Society of Automotive Engineers, [1985].

[28] "Status Report Special Issue: Occupant Death Rates by Car Series." Insurance Institute for Highway Safety Status Report 24 (November 25, 1989).

[29] Viano, David C., and Arepally, Sudhakar. Assessing the Safety Performance of Occupant Restraint Systems. Report No. GMR-7093. Warren, MI: General Motors Research Laboratories, [1990].

[30] Zador, Paul L.; Ian S. Jones; and Marvin Ginsburg. "Fatal Front-to-Front Car Collisions and the Results of 35 mph Frontal Barrier Impacts." Proceedings of the Twenty-Eighth Conference of the American Association for Automotive Medicine. Morton Grove, IL: American Association for Automotive Medicine, 1984.