United States General Accounting Office

GAO

Report to Congressional Requesters

May 1995

# HIGHWAY SAFETY

## Reliability and Validity of DOT Crash Tests

GAO/PEMD-95-5

GAO

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

**Program Evaluation and**
**Methodology Division**

B-250144

May 5, 1995

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

The Honorable Richard H. Bryan
United States Senate

Between 1980 and 1992, the annual death toll on America's highways dropped from more than 50,000 to less than 40,000. One factor contributing to this decline may have been the increasing concern among consumers about the safety of the vehicles they purchase. The crash test programs performed by the Department of Transportation (DOT) are the major source of the safety information available to today's automobile purchaser.

You asked us to supplement our earlier testimony and report to your Committee on the relationship between automotive design factors and safety by responding to a number of questions concerning automobile safety. As one component of this research, you requested that we review the automobile crashworthiness program of the National Highway Traffic Safety Administration (NHTSA) to determine whether NHTSA's crash test programs provide valid and reliable indicators of occupant safety in real-world crashes.

## Background

The Department of Transportation has four offices that conduct automobile crash tests: three within NHTSA and one in the Federal Highway Administration (FHWA). The activities of two programs run by NHTSA are the focus of this report. NHTSA's Office of Vehicle Safety Compliance performs a compliance testing program of 30-mile-per-hour full-frontal crashes of automobiles, light trucks, and vans into a fixed rigid barrier. This program was created under section 103 of the National Traffic and Motor Vehicle Safety Act of 1966, and it is designed to ensure that vehicles meet minimum safety requirements as specified in Federal Motor Vehicle Safety Standard No. 208 -Occupant Crash Protection (FMVSS 208).[1]

[1]Code of Federal Regulations, 49 C.F.R. Part 571: Federal Motor Vehicle Safety Standards (FMVSS), Standard No. 208 - Occupant Crash Protection. The Office of Vehicle Safety Compliance also uses dynamic crash tests to determine vehicular compliance with FMVSS 212, Windshield Mounting; FMVSS 219, Windshield Zone Intrusion; and FMVSS 301, Fuel System Integrity.

Also under the authority of NHTSA is the New Car Assessment Program (NCAP), conducted by the Office of Market Incentives. This program, mandated under title II of the Motor Vehicle Information and Cost Savings Act of 1972, was created to provide information to consumers on the relative crashworthiness, or safety, of automobiles. This charge differs from the compliance test in that vehicles tested in NCAP are not required to meet specified safety standards, while the purpose of compliance tests is to ensure that vehicles meet a level of safety required by law. The NCAP test also differs from the compliance test in two important aspects: NCAP crashes its vehicles at 35 miles per hour, which translates to over one-third more energy than compliance tests, and NCAP engages all manual and automatic restraints, while the compliance test employs only passive restraints. By using **all** restraint systems, NCAP assesses the maximum crashworthiness of a vehicle in high-speed frontal crashes.

In addition to the two programs described above, NHTSA's Office of Crashworthiness Research conducts a variety of tests to study a wide range of individual safety issues that arise from specific crash configurations. FHWA conducts crash tests to study the interaction between automobiles and roadside obstacles and devices such as guard rails, telephone poles, and bridge abutments.

## Objectives, Scope, and Methodology

To respond to your request, we examined data from tests conducted for compliance with FMVSS 208 (compliance tests) as well as those conducted under the New Car Assessment Program. We chose to focus on these programs because both conduct tests that are similarly configured, employ standardized procedures, and have been assessing vehicle crashworthiness over a period of years. The two other crash test programs run by DOT are largely research based and, although important, have different purposes from those of our study.

Our analysis consisted of three parts: (1) an examination of trends over time in crash test results of both programs, (2) an assessment of the reliability of NCAP results, and (3) a review of the relationship between NCAP results and real-world traffic injuries and fatalities. We first reviewed the background, sample selection, and testing procedures of both NCAP and the compliance program. (See appendix I). We then examined what it is that crash tests measure, as wellas how well measurement devices used in crash tests simulate human biomechanics and physiological response by reviewing biomechanic, human tolerance, and automotive safety literature and by interviewing experts in those fields. (See appendixes II and III.)

Next, we analyzed changes in crash test results by year for both the compliance program and NCAP. (See appendix IV.)

To address the reliability of crash test results, that is, the degree to which consistent results are obtained through repeated trials, we examined research conducted by NHTSA and compared NCAP results with those obtained in crash tests conducted by manufacturers. (See appendix V.) Finally, we conducted analyses using two national databases that allowed us to relate real-world fatality rates for drivers with the predicted injury risks derived from NCAP results. (See appendix VI). For this analysis, we used Poisson regressions to assess the relationship between fatality rates, derived from the Fatal Accident Reporting System and the R.L. Polk Vehicle Registration System, and the combined injury risk calculated from the NCAP measurements that assess the potential for skeletal injuries to the head and chest. Analyses were conducted for restrained drivers in one- and two-car frontal crashes.

We did not include information from the compliance program in the analyses we conducted on either the reliability or the predictive validity of crash test results. In our assessment of the reliability of crash test results, we did not uncover a quantity of data sufficient enough to compare the results of two or more trials of vehicle models. In the case of the predictive validity of crash test results, we did not use compliance test data for two reasons. First, the compliance program had conducted only 145 tests between 1987 and 1992. Second, the variation among compliance results was relatively narrow and scores tended to cluster far below the ceiling values for the compliance tests. These two items resulted in a dataset that was insufficient for conducting detailed statistical analyses.

We conducted our review in accordance with generally accepted government auditing standards.

## Results in Brief

Recent trends in test results indicate that, first, the probability of sustaining a serious injury, as it is measured by the NCAP and compliance tests, has decreased substantially since the inception of these test programs. In addition, the variation in scores between automobile models has shrunk, indicating that cars marketed in the United States have become more uniformly crashworthy. Some of this improvement, we concluded, could appropriately be attributed to DOT crash test programs initiated by the National Traffic and Motor Vehicle Safety Act of 1966 and the Motor Vehicle Information and Cost Savings Act of 1972.

Second, the reliability, that is, the consistency of results derived from NCAP tests is questionable. Existing evidence suggests that large differences between crash test scores most likely reflect true differences between vehicles. However, we cannot be sure that even moderately large score differences between two vehicles might not disappear or be reversed if they were tested again. This is because, in general, only one unit of a specific vehicle line is tested, and this is not enough to say with confidence that the results are indicative of how other units of that same vehicle line would perform. The new star rating system incorporated by NCAP in recent years, a system that places vehicles into one of five categories on the basis of potential injury risk, could exacerbate this problem in some cases.

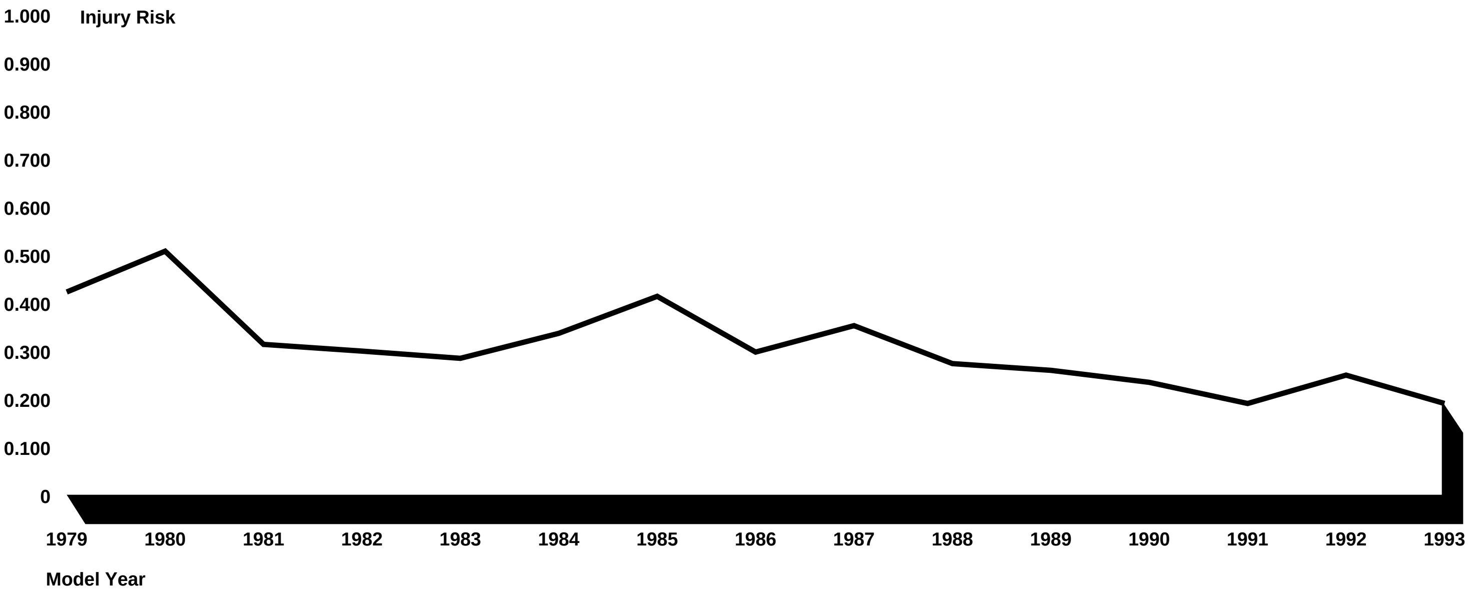
Third, we found that the ability of NCAP to predict a vehicle's occupant protection in real-world crashes is limited. By their nature, NCAP crash test results can be validly applied only to frontal collisions, which account for slightly more than half of all injury-producing accidents. We found a statistically significant relationship between fatality rates and NCAP predicted injury risk; however, this relationship derives from the high fatality rates associated with the poorest performers in NCAP (the 20 percent of vehicles with the highest potential injury risks derived from NCAP results).

# GAO's Analysis

## Automobiles Have Become More Crashworthy

In NCAP crashes, nearly all cars now meet the head and chest injury standards of the compliance tests, although they are 36-percent more violent than compliance crashes. The average probability of sustaining a serious injury in a 35 mile-per-hour crash as measured by NCAP has declined from over 0.5 in 1980 to less than 0.2 in 1993. (See figure 1.) Differences among the crashworthiness scores of vehicles tested in this program have experienced similar declines. The introduction of air bags has contributed significantly to this improvement. (For a complete discussion of the trends in crash test results, see appendix IV.)

**Figure 1: Mean NCAP Injury Risk by Model Year[a]**




[a]For all passenger cars. Does not include light trucks and vans.

A causal linkage between improved crash test scores and declining highway fatalities cannot be asserted with certainty because of both the many variables involved in a crash and the increased emphasis on traffic accident and injury prevention over the past decade. Nonetheless, it seems reasonable to conclude that manufacturers' successful efforts to improve their products' performance in NHTSA crash tests, particularly in NCAP, have contributed to improved occupant protection in real-world crashes, although we were unable to quantify that contribution. These improvements to performance have derived from a variety of efforts, with two examples being modernized manufacturing techniques and an increased emphasis on safety systems and designs.

In addition, in recent years, automotive designers have turned more to computer-based simulations to assist in the design of vehicles that meet crash test standards. Although we did not evaluate the state of the art in computer-based crash models, we learned from industry personnel that such modeling appears to accurately predict the results of actual crash tests. Indeed, one computer specialist informed us that the industry uses crash tests in part to validate their computer models.

Although simulated crashes are costly as they currently require access to supercomputers, they do allow the manufacturer to assess the crashworthiness of a vehicle in more trials, more quickly, and at impacts points other than the front (or side) of a car. These benefits over actual crash testing permit the identification of crash forces upon an occupant in a time frame that offers immediate redesign implications.

## Crash Test Results Have Questionable Reliability

To determine whether the result of any test is reliable, consistent results must be obtained through repeated trials of a specified procedure. In the case of crash tests, this means that consistent results of repeated tests of a specific vehicle model are required. This is particularly crucial when comparing the safety ratings of different vehicles. Both the NCAP and compliance programs generally conduct only one trial of a specific vehicle model; thus, insufficient data exist to accurately define the reliability of crash test results. That is, the ability to predict with confidence the likelihood of a tested model's receiving similar scores if tested again is low.

We found only two sources of information on which to assess the reliability of crash test results: a study conducted by NHTSA in 1984, which examined the variations in test results of 12 consecutively manufactured Chevrolet Citations, and our own analysis of the differences between results for vehicle models tested in NCAP and the results for those vehicles in corresponding tests conducted by automobile manufacturers. Our analysis of the data derived from NHTSA's 1984 study revealed wide variations in the head injury criterion (HIC) results, the measurement taken to assess potential skeletal head injuries. (See appendix V.) Although NHTSA ascribed the variation in results to a number of sources, including the test itself, it failed to discuss the implications of the combined effect of these sources on crash test results; namely, that even within a specific vehicle line, the result of one test may not be indicative of the model's performance from trial to trial, and large differences in the resultant HIC may occur.

We also examined the differences between the results of NCAP and manufacturers' tests provided to us by NCAP officials. The tests conducted by the manufacturers essentially duplicated the NCAP test procedures. We compared the results of the two tests using the star rating system recently developed by NCAP, hypothesizing that if the manufacturer test were considered a second trial for a model line, its results should be consistent with the NCAP, or first trial. The star rating system ranks cars from 1 to 5 stars, with 5 stars being the best rating, or safest car, and 1 star being the worst rating, or least safe car. These ratings are based on the risks of serious injury for vehicles, which are calculated from the head injury criterion and chest acceleration scores from NCAP tests. (For a discussion of crash test measurements and the star system, see appendix II.)

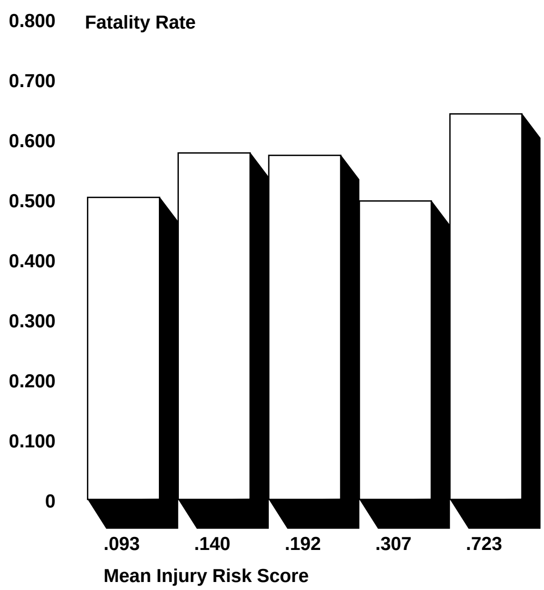
We found that in only about one-half of the paired comparisons would NCAP- and manufacturer-tested vehicles have received the same star rating. In 32 percent of the comparisons, the results of the second trial would have changed by 1 star, while in 8 percent of the cases, the ratings of the vehicles would have changed by 2 or more stars. When we compared the risks of serious injury (the base unit categorized into the five star ratings) derived from the manufacturer and NCAP data, we found that each star category was associated with a wide band in which the resultant risk scores of subsequent tests might fall. For example, the results of a second test of a vehicle rated as 4 stars by its first test could fall between 5 stars and 2 stars.

The analyses described above are based on the only two sources of information we could find. The quantity of data in each analysis was not enough for us to fully quantify the reliability of crash test results; however, we were able to determine that NCAP scores, whether reported in raw HIC and chest acceleration scores or as categories of injury probability, have associated levels of imprecision and that seemingly large differences in crash test results may not necessarily reflect true differences in a vehicle's safety potential. By not properly defining and publishing the degree of reliability, consumers may be misled into purchasing a vehicle purported to be more crashworthy than another when, in fact, it may be no more safe, or even less safe, than the comparison vehicle.

## NCAP Results and Real-World Fatality Rates

Since NCAP crash tests are designed to simulate full-frontal collisions, we restricted our analysis to those types of crashes and found the results of NCAP crash tests are generally reflected in real-world fatality rates. That is, on the whole, a statistically significant relationship exists between real-world highway fatality rates associated with vehicles tested in the NCAP program and their scores in crash tests. However, we concluded that this relationship derives mainly from the high fatality rates of vehicles with the worst NCAP scores. When we divided vehicles into NCAP score quintiles—that is, placed the vehicles into one of five 20-percentile categories based on their location in the distribution of NCAP results—we found that the quintile with the worst NCAP scores (those vehicles in the highest 20-percentile category) had significantly higher fatality rates than the remaining 80 percent of NCAP-tested vehicles. The remaining four quintile categories, however, had associated fatality rates that were not significantly different from one another. (See figure 2 and appendix VI.)

**Figure 2: Mean NCAP Injury Risk Scores and Fatality Rates[a]**




[a]By quintile, per 100,000 vehicles.

## Conclusions

Over time, the mean risk of injury in frontal crashes, as measured by NHTSA crash tests, has declined and indeed has mirrored a similar trend in the annual number of highway fatalities. While we cannot state with certainty that NHTSA crash tests are a causal factor in improved crashworthiness, we believe that efforts on the part of automobile manufacturers to produce vehicles that score well on these tests have contributed to the improvement of the overall safety of vehicles. At the very least, the results of NCAP and compliance tests provide indications that the vehicle fleet, on the whole, has become safer over the past 15 years.

These trends in the mean score of crash tests, however, do not necessarily suggest that individual vehicles have well-defined levels of safety, nor do they suggest that the relative rankings of two vehicles would be the same if subsequent trials were conducted. They also do not suggest that differing test results are reflected in data derived from real-world traffic collisions. Indeed, only the poorest performers in NCAP had associated fatality rates that were significantly different than other NCAP vehicles.

## Recommendations

On the basis of our findings, we make two recommendations to the Administrator of NHTSA. First, we recommend that information on NCAP reliability be updated and made available, in clear language, to the general public. Such an effort would require an update of the repeatability study the agency conducted in 1984 and could result not only in a better understanding of the reliability of crash tests for predicting injury risk, but also in discovering ways in which NHTSA can limit the error that derives from sources under its control.

We also recommend that NHTSA explore the feasibility of alternative means of testing the crashworthiness of new vehicles. Computer simulations may provide one such alternative. It may be possible to better assess the safety potential of a vehicle through computer-based modeling as this allows more trials, more quickly, and modeling is capable of simulating impacts at all points of a vehicle. In addition, this rapidly emerging technology has the added capability of providing immediate insights into redesigning vehicles in which the crashworthiness may not yet be optimal.

## Agency Comments

We received written comments on a draft of this report from the Department of Transportation. The Department concurred with our recommendation that it update its information on NCAP reliability. We are concerned, however, that the agency might believe it has already complied with this recommendation by developing the star rating format. As noted above and explained in detail in appendix V, this new format does not resolve our questions concerning NCAP reliability.

The Department interpreted our second recommendation as a recommendation to augment or replace “live” crash tests with computer simulations and did not concur with us. The Department cited concerns about the costs and predictive limitations of such simulations. We share these concerns, but we believe the Agency has misinterpreted the recommendation. We avoided recommending the adoption of any particular substitute for the current crash test procedures at this time. Rather, we urged the Agency to explore all possible means of reliably defining vehicle crashworthiness. Computer modeling is a potential alternative that deserves exploration and monitoring as the technology matures. Other alternatives could include extending testing programs to include side, rear, and frontal-offset impacts to gain a better understanding of the total safety of a vehicle or seeking greater sharing of crash test data developed by automotive manufacturers either through crash tests replicating NHTSA's or through their individual component testing programs.

The Department provided a number of other specific comments. They are reproduced in appendix VII, together with our response. We have also made modifications to the report as we deemed appropriate on the basis of these comments. After responding to our draft report, the Department also provided us with additional data relevant to NCAP reliability. The results of our analysis of these data can be found in appendix V.

We are sending copies of this report to the Secretary of Transportation, the Administrator of the National Highway Traffic Safety Administration, and to other interested parties. We will also make copies available to others upon request. If you have any questions or would like additional information, please call me at (202) 512-3092. Major contributors to this report are listed in appendix VIII.

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

# Contents

## Tables

## Figures



### Abbreviations

| | |
|---|---|
| AAAM | Association for the Advancement of Automotive Medicine |
| DOT | Department of Transportation |
| FARS | Fatal Accident Reporting System |
| FHWA | Federal Highway Administration |
| FMVSS | Federal Motor Vehicle Safety Standards |
| HIC | Head injury criterion |
| ISO | International Standards Organization |
| NASA | National Aeronautics and Space Administration |
| NASS | National Accident Sampling System |
| NCAP | New Car Assessment Program |
| NHTSA | National Highway Traffic Safety Administration |
| SAE | Society of Automotive Engineers |
| SUDAAN | Survey Data Analysis |

# Department of Transportation Crash Test Programs

## Introduction

The Department of Transportation has four offices that conduct automobile crash tests: three within NHTSA and one in the Federal Highway Administration. A compliance test program conducted by NHTSA's Office of Vehicle Safety Compliance consists of full-frontal crashes of automobiles, light trucks, and vans into a fixed rigid barrier to ensure that vehicles meet certain minimum safety requirements.[1] Also under the authority of NHTSA is the New Car Assessment Program, conducted by the Office of Market Incentives. NCAP tests are similar to compliance tests, but they are performed to provide consumer information on the relative crashworthiness of automobiles. The NHTSA Office of Crashworthiness Research conducts a variety of tests to study a wide range of individual safety issues that arise from specific crash configurations. Finally, FHWA conducts crash tests to study the interaction between automobiles and roadside obstacles and devices such as guard rails, telephone poles, and bridge abutments.

In this study, we focused on the crash tests run under the compliance program and the New Car Assessment Program because both tests are similarly configured, employ standardized procedures, and have been assessing vehicle crashworthiness over a period of years. The tests conducted by the Office of Crashworthiness Research and those of FHWA, although important, have different purposes from those of our study and could not provide a quantity of data sufficient for us to assess relationships between test results and real-world performance.

## Compliance Test Program

Section 103 of the National Traffic and Motor Vehicle Safety Act of 1966 charged DOT with establishing standards for vehicle safety. These safety standards are codified by NHTSA in the U.S. Code of Federal Regulations at 49 C.F.R. Part 571: Federal Motor Vehicle Safety Standards (FMVSS), Standard No. 208: Occupant Crash Protection. Beginning with the 1987 vehicle model year, FMVSS 208 required that passenger cars, light trucks, and vans sold in the United States be certified as meeting minimal safety levels as measured by anthropomorphic dummies in dynamic crash tests.[2] The purpose of this standard is

[1]Because these tests are required by NHTSA under standard 208 of the Federal Motor Vehicle Safety Standards, they are commonly called FMVSS 208 tests. NHTSA also uses dynamic crash tests to determine vehicular compliance with FMVSS 212, Windshield Mounting; FMVSS 219, Windshield Zone Intrusion; and FMVSS 301, Fuel System Integrity.

[2]Self-certification in a 30 mile-per-hour barrier collision and passive restraint requirements were phased in between 1987 and 1990 for passenger cars. Light trucks, vans, and sport utility vehicles were required to meet dynamic crash requirements beginning with the 1992 model year.

"to reduce the number of deaths of vehicle occupants, and the severity of injuries, by specifying vehicle crashworthiness requirements in terms of forces and accelerations measured on anthropomorphic dummies, and by specifying equipment requirements for active and passive systems."

These crash tests are conducted under the guidance of NHTSA's Office of Vehicle Safety Compliance.

## Sample Selection for Compliance Tests

The current compliance program relies, for the most part, on a certification process in which the manufacturer of a specific make and model vehicle states that the vehicle meets all safety requirements set forth in FMVSS 208. In addition, each year NHTSA selects a number of vehicles to test to ensure that the manufacturer's certification is justified. The criteria used to determine which specific makes and models to test are based on whether a vehicle is in its first or second model year, whether safety features have been added or redesigned, and how many units are on the road. In selecting models for testing, NHTSA also includes any evidence of poor crashworthiness derived either from consumer complaints filed about specific models or from other crash test programs (in particular, NCAP). Through these criteria, NHTSA compiles a preliminary list of about 50 candidate vehicles for testing and requests information on crash test performance from the manufacturer of each candidate model to determine the final list of vehicles to be tested. Though they are under no obligation to do so, manufacturers will normally provide one or two sets of results from their tests of the model NHTSA specifies. NHTSA uses these data not only as an input for determining the final list of test vehicles, but also as a baseline with which to compare its own results.

## Test Conditions

The compliance test consists of a full-frontal collision of a vehicle into a fixed rigid barrier at a velocity of 30 miles per hour. Anthropomorphic test dummies, fitted with instrumentation to measure forces and accelerations acting on the head, chest, and both femurs, are placed in the driver and front passenger seats. Only passive restraint systems—those that require no effort on the part of an occupant—are engaged. Examples of these are air bags and automatic seat belts. Seat belts that require active participation by the occupant are not used.[3] The underlying assumption is that if a vehicle meets the standards for those occupants who do not make

[3]This requirement applies to all passenger cars and those light trucks certified by the manufacturer as meeting the automatic occupant protection requirement and will be phased in for light trucks beginning with the 1995 model year.

use of all available safety restraint systems, it will also meet the requirements for those who do.

The test conditions further specify the forward placement of the seat, the angle of the seat back, the angle of the steering column (where the vehicle has tilt steering), and a number of other components. Some of these, such as adjustable backs for seats, are placed in the manufacturer's nominal design riding position—that is, the position the manufacturer says is the proper one for the average adult male (5 feet 9 inches, 167 pounds).

## New Car Assessment Program

A second crash test program we studied is the New Car Assessment Program conducted by NHTSA's Office of Market Incentives. This program was mandated under title II of the Motor Vehicle Information and Cost Savings Act of 1972 to provide consumers with an understanding of the relative crashworthiness of passenger motor vehicles. Since 1979, NCAP has conducted almost 500 crash tests of passenger cars, light trucks, and vans. From 1979 to 1986, NCAP was considered an indicant test for vehicle compliance with Federal Motor Vehicle Safety Standards 212, Windshield Mounting; 219, Windshield Zone Intrusion; and 301, Fuel System Integrity. That is, if a vehicle performed reasonably well on these tests, which required dynamic testing, then it would likely meet compliance test requirements because the NCAP test involves a more violent crash than the one required for the compliance test. If a vehicle performed poorly, the information would be transmitted to the Office of Vehicle Safety Compliance for testing its compliance with the safety standards. NCAP has not been an indicant program since the implementation of dynamic crash tests in the FMVSS 208 program in 1987; however, poor performance on the NCAP test typically leads to compliance testing of the same model.

## Comparison of the Two Tests

The NCAP crash test is generally similar to the compliance test. Both are full-frontal collisions into a fixed rigid barrier, and both use roughly the same criteria when determining which vehicles to test. However, three very important differences distinguish the two test programs. First, vehicles in the NCAP test are crashed at 35 miles per hour rather than 30 miles per hour, the velocity in the compliance test. This 5 mile-per-hour difference results in a 36-percent increase in the amount of energy in the system.[4]

[4]Kinetic energy is a function of mass and velocity ($E_k = 1/2 mv^2$). The additional energy in the NCAP test over the compliance test derives from the square of the velocity when the mass of the vehicle is held constant. Thus $(35 \text{ miles per hour})^2$ is 36-percent greater than $(30 \text{ miles per hour})^2$.

Second, all active as well as passive safety belts in the automobile are used in the NCAP test; that is, the test dummies are restrained by any manual seat belt furnished with the vehicle as well as any automatic belt or air bag. In the compliance test, as noted earlier, only passive restraints (automatic belts and air bags) are used.

The third and foremost difference between the two programs is the underlying purpose of the tests. NCAP is a market-based program that disseminates information to consumers on the relative safety of passenger vehicles. There are no minimum allowable safety performance criteria that vehicles must meet, although NCAP collects the same measurements as the compliance test. Despite the fact that NCAP is not a compliance program, industry personnel have expressed the opinion that the NCAP test has become the de facto regulation. That is, failure to meet compliance levels on this more stringent test involving a more forceful collision than the official compliance test could imply that a vehicle is unsafe. Currently, nearly all vehicles tested under NCAP meet the safety requirements specified in FMVSS 208.

# Crash Test Measurement and Reporting

## Crash Test Measurements

Both the compliance and NCAP tests use anthropomorphic test dummies to collect data related to injury potential by measuring accelerations and forces placed on an occupant's head, chest, and upper leg.[1] Specific levels for each measure, established under FMVSS 208, represent upper-bound limits for compliance with vehicle safety requirements. These ceilings were designed to correspond to the level at which there is a one-in-six chance of an occupant's sustaining an injury that poses a serious threat to life.

The head injury criterion, the measure used in crash tests to assess potential head injury, was adopted by NHTSA on the basis of research conducted to establish the likelihood of skull fractures under different velocity changes. HIC is measured as a composite of the axial accelerations of the head (in three dimensions). Specifically, HIC is the product of (1) the 2.5 power of the average of the resultant head acceleration over a time interval not more than 36 milliseconds and (2) that time interval. The equation for the function is

$$HIC=[1/(t_2-t_1)\int adt]^{2.5}(t_2-t_1).$$

A HIC score of 1,000, the highest allowable score for achieving vehicle compliance, is associated with a one-in-six chance of sustaining a serious skull injury.

For determining potential injury to the chest region, chest acceleration is measured in gravitational units (g's).[2] The potential for injury to the chest skeletal structure is measured by the actual resultant peak tridimensional acceleration of the upper thorax. Compliance with FMVSS 208 for this measure is set at 60 g's over 3 milliseconds, an acceleration level that has been associated with four or more fractured ribs.

Depending on the type of crash test dummy used, a second chest measurement, termed chest compression, is taken. (See appendix III.) This measures the amount of reduction in the distance between the sternum and the spinal column and is determined to assess the likelihood of injury to internal organs. Currently, only one of the two types of crash test dummies (Hybrid III) is capable of measuring this, and the choice of which dummy type to use in a test (either compliance or NCAP) is made by the

[1]Acceleration is the rate of change in velocity with respect to time, while force is the rate of change in velocity with respect to time for a given mass.

[2]One g is equal to 32 feet/second$^2$, or 9.8 meters/second$^2$.

manufacturer of the test vehicle. If the Hybrid III is used in a compliance test and the vehicle exceeds the 3-inch maximum reduction distance allowed (the limit associated with major lacerations to the spleen or kidneys), the vehicle is considered not to be in compliance with FMVSS 208.

The final measure taken in both the compliance and NCAP tests is the compressive force transmitted axially through the upper legs (femurs). The femur tolerance level of 2,250 pounds of force is based primarily on experimental impacts to the lower limbs and is associated with a one-in-six chance of sustaining a fracture to that bone.

When the results of a compliance test exceed the limit for any of the measures, an investigation is conducted to determine reasons for the failure and is typically accompanied by a recall or remedy campaign. If a determination of noncompliance is made, the model being tested may not be sold in the United States. This differs from NCAP as its tests are not conducted to assess vehicle compliance with federal regulations, and therefore, no punitive actions may be taken by NHTSA should a vehicle exceed any of the limits. Table II.1 lists the four measurements made in both test programs and presents the maximum allowable scores under compliance testing for each measure.

**Table II.1: Maximum Allowable Scores on FMVSS 208 Crash Tests for Vehicle Compliance**

| Measure | Maximum allowable score |
|---|---|
| Head injury criterion | 1,000 |
| Chest acceleration | 60 g's |
| Chest compression[a] | 3 inches |
| Femur load | 2,250 pounds |

[a]This measure applies to the Hybrid III dummy only. (See appendix III.)

## Reporting NCAP Test Results

In 1978 (for the 1979 model year), NHTSA began testing about 30 vehicles per year through its New Car Assessment Program. While no manufacturer is required to exceed 30 mile-per-hour standards, the program, using a 35 mile-per-hour crash test, is designed to inform customers of the relative crashworthiness of an automobile. Traditionally, NCAP reported the actual HIC, chest acceleration, and femur load scores with a disclaimer that only vehicles within 500 pounds of each other could legitimately be compared. Also, NCAP would cite the compliance ceiling levels (1,000 HIC, 60-g chest acceleration, and 2,250-pound femur load) as representing a one-in-six chance of sustaining a severe injury. Despite NHTSA's claim of overall success in providing information about how well or how poorly passenger vehicles protect their occupants in crashes, some critics argued that NCAP's method of reporting test results left consumers confused.

In response to fiscal year 1992 Senate Appropriations Committee requirements, NHTSA performed a user study and began implementing new methods of informing consumers of the comparative levels of the safety of passenger vehicles as measured by NCAP. This new method, a star chart rating system, is designed to provide consumers with a quick, simplified, single point of comparison to evaluate vehicles in the NCAP test.[3]

Based upon analyses of a variety of accident injury studies, NHTSA developed a scale, known as the "Level of Protection Scale," that relates the probability of sustaining an injury to the level of protection a vehicle provides its occupants from receiving such an injury.[4] This scale forms the basis of NHTSA's star chart method for releasing NCAP test results to the public. The star chart, which NHTSA began using in December 1993, reports a range of 1 to 5 stars, with 5 stars indicating the best crash protection for vehicles within the same weight class.

The number of stars a vehicle may be rated is derived from the injury probabilities associated with the HIC and chest g scores obtained in the crash tests. These probabilities are calculated using the following formulas:

$$P_{head} = [1 + \exp(5.02 - 0.00351 \text{ x HIC})]^{-1}$$

$$P_{chest} = [1 + \exp(5.55 - 0.0693 \text{ x Chest Acceleration})]^{-1}$$

$$P_{combined} = P_{head} + P_{chest} - (P_{head} \text{ x } P_{chest})$$

A vehicle is then assigned a star rating based on its combined injury risk, with the specific number of stars determined by the range in which the

[3]The star chart rating system applies only to vehicles tested in NCAP. In compliance tests, the actual HIC, chest acceleration, and femur load results are reported, as the primary purpose of the test is vehicular compliance to safety regulations and **not** a comparative assessment of the likelihood of occupants' sustaining serious injuries in different vehicle models.

[4]In SAE Paper No. 851246, "The Position of the United States Delegation to the International Standards Organization (ISO) Working Group 6 on the Use of HIC in the Automotive Environment," P. Prasad and H. Mertz presented an injury risk function curve (which this scale is based upon) that relates the probability of a severe head injury to HIC.

combined injury risk lies. The ranges for each star rating are shown in table II.2.

**Table II.2: Star Ratings and Their Combined Injury Probability Ranges**

| Rating | Range of probability |
|---|---|
| 5 stars | 0 - .10 |
| 4 stars | + .10 - .20 |
| 3 stars | + .20 - .35 |
| 2 stars | + .35 - .45 |
| 1 star | + .45 - 1.00 |

# Anthropomorphic Test Devices

## Introduction

Currently, two types of anthropomorphic test dummies are used in both the compliance and NCAP tests: the Hybrid II and Hybrid III 50th-percentile male dummies. Requirements for both types of dummies used in compliance testing are specified in 49 C.F.R. Part 572: Anthropomorphic Test Dummies.[1] The design and performance criteria specified for each dummy type

"are intended to describe a measuring tool with sufficient precision to give repetitive and correlative results under similar test conditions and to reflect adequately the protective performance of a vehicle or item of motor vehicle equipment with respect to human occupants."[2]

In this appendix, we discuss the characteristics and instrumentation of the Hybrid II and Hybrid III 50th-percentile anthropomorphic test dummies, under the provisions of the NHTSA standards pertaining to occupant crash protection. We also provide a comparison of the two dummy types' performance in NCAP and the compliance test programs. Finally, we summarize the 1993 decision to standardize the test dummy, requiring the mandatory use of the Hybrid III in all NHTSA crash test programs beginning in 1997.

In both the compliance and NCAP test, the manufacturer of the vehicle being tested has the option to choose which type of dummy will be used. While both dummies are designed to represent the physical characteristics of the average adult male, important differences between them exist. Despite the differences, the requirements for vehicular conformance to FMVSS 208 are not different for the two instruments, with the exception of the chest compression criterion, which applies only when the Hybrid III dummy is used.

## Comparison of Hybrid II and Hybrid III

### Differences in Construction

Part 572 of the Federal Motor Vehicle Safety Standards specifies the types of anthropomorphic dummies to be used in the FMVSS 208 compliance test. Currently, two specific types of anthropomorphic test dummies may be used in a compliance crash test: the Hybrid II and the Hybrid III. As

[1]NCAP is not required to follow the standards in part 572; nonetheless, it does.

[2]49 CFR Ch. V (10-1-91 Edition), Part 572-Anthropomorphic Test Dummies, p. 581.

specified in subpart B of 49 C.F.R. part 572, since 1973 the Hybrid II 50th-percentile male test dummy is 5 feet 9 inches tall and weighs approximately 164 pounds, and until 1986, this dummy was used when determining compliance to FMVSS 208. In 1986, 49 C.F.R. parts 571 and 572 were amended to adopt the Hybrid III 50th-percentile dummy as an alternative to the Hybrid II for FMVSS testing. This gave manufacturers the option of using either the Hybrid II or Hybrid III test dummy as the means of determining a vehicle's conformance to NHTSA's performance requirements. Like its predecessor, the Hybrid III is 5 feet 9 inches tall but weighs slightly more (167 pounds). Also, like the Hybrid II, each Hybrid III used in a compliance test must meet the specifications and performance criteria of part 572 before and after each vehicle test in order to be an acceptable compliance tool.

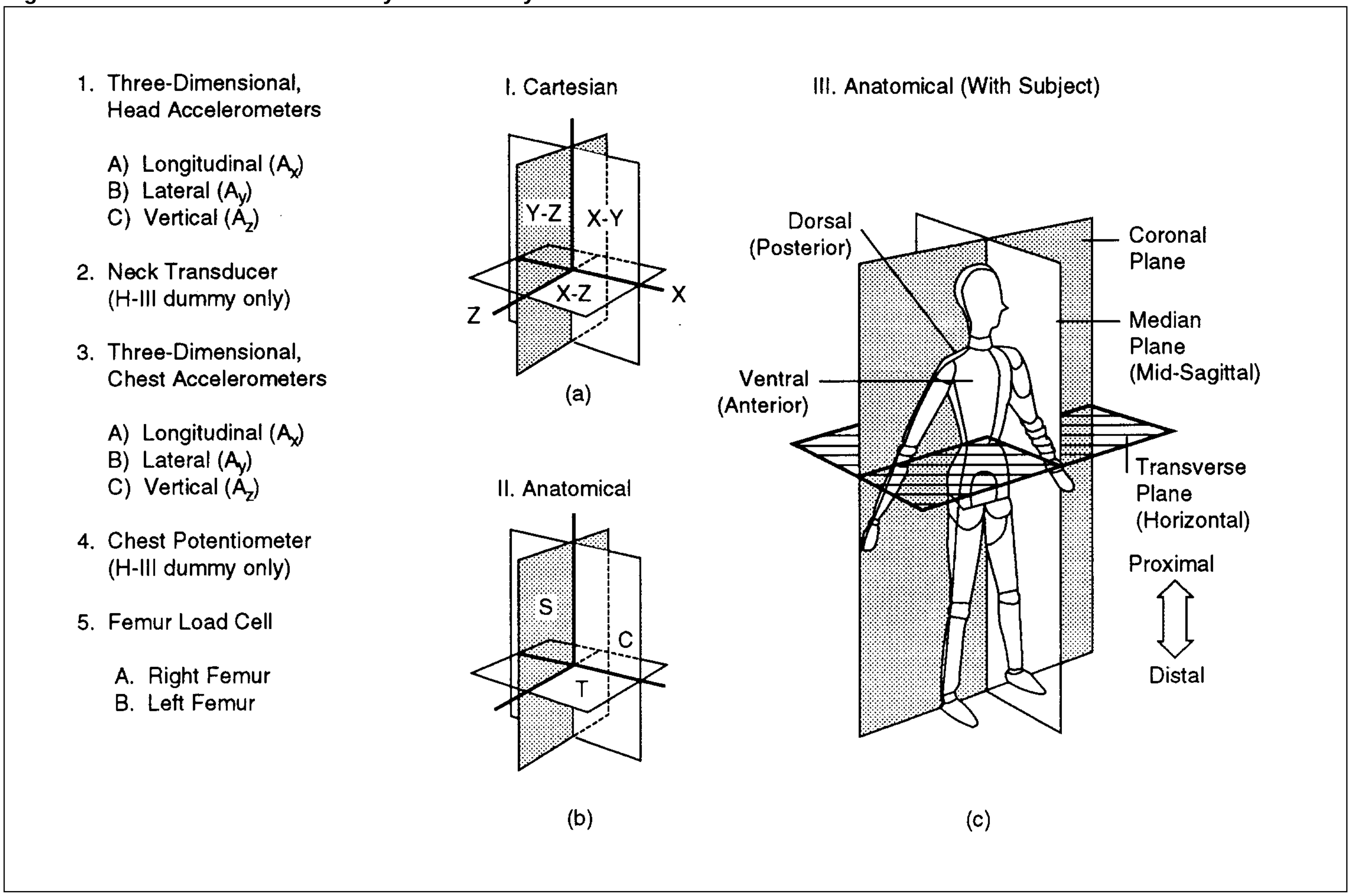
The Hybrid II and Hybrid III use the same instrumentation in the head, chest, and femurs. (See figure III.1.) However, according to General Motors, developer of the Hybrid III, its 50th-percentile male dummy was designed to improve on Hybrid II technology and biofidelity.[3] Most experts regard the Hybrid III test dummy as more biofidelic than the Hybrid II, having a more human-like seated posture, as well as head, neck, chest, and lumbar spine designs. The Hybrid III's responses to crash conditions more closely approximate the motions associated with human anatomy in crash situations and, therefore, more accurately evaluate injury risks. For example, the improved flexibility of the Hybrid III's neck over the Hybrid II allows researchers a greater ability to assess the injury potential of whipping motions.

In addition to the greater biofidelity, the experts we interviewed stated that the Hybrid III is more sophisticated technologically than the Hybrid II because it has more instrumentation for measuring potential injuries. Specifically, the Hybrid III is capable of measuring nearly four times as many forces and accelerations throughout the body as the Hybrid II. For example, not only does the Hybrid III measure injury potential to the skeletal structures, but it can also determine injury potential to the soft tissues in the upper thorax through the chest compression measure. Further, the Hybrid III has accelerometers and load cells placed in the neck and lower legs that can measure the potential for injuries caused to those anatomical areas. To date, no criteria have been established for meeting compliance for the additional measures other than chest compression. However, the measures do provide DOT with additional

[3]The biofidelity of a test dummy is mainly assessed by comparisons of its response to those of cadavers under the same test conditions.

information on the potential physiological responses associated with vehicular crashes.

**Figure III.1: Instrumentation in the Hybrid II and Hybrid III and the Cartesian and Anatomical Coordinates[a]**




[a]Reprinted with permission from SAE publication R-103, copyright 1990, Society of Automotive Engineering, Inc.

## Differences in Test Results

Currently, the determination of which dummy to use in a test is made by the manufacturer of the vehicle being tested. Through 1993, manufacturers chose to use the Hybrid III in 36 of the 133 tests of passenger cars in the compliance program and in 30 of the 86 tests of passenger cars in NCAP since the dummy became available for use in the two programs (1988 and

1990, respectively). One expert hypothesized that the reason so few compliance tests involve Hybrid III dummies is that they tend to receive higher HIC scores, especially in noncontact situations and that there is no guarantee that a car designed around the Hybrid II will pass a test using a Hybrid III.

The differences between the dummies used in NHTSA's tests, described above, led us to compare the driver-side HIC and chest acceleration scores from passenger-car compliance and NCAP tests in an attempt to quantify the potential effects on test reliability such differences could create. In this analysis, we controlled for the presence of a driver-side air bag in the car. In general, we found the Hybrid III dummy scores were lower than the Hybrid II scores, but that the presence of air bags strongly affects the relative performance of the two dummy types. (See tables III.1 and III.2.)

**Table III.1: Mean HIC and Chest Acceleration Scores From NCAP Tests for Hybrid II and Hybrid III Dummies[a]**

| Measure | Hybrid II | | Hybrid III | |
|---|---|---|---|---|
| | Score | Number of tests | Score | Number of tests |
| HIC | 793 | 56 | 637 | 30 |
| Air bag | 695 | 27 | 511 | 24 |
| No air bag | 903 | 29 | 1,141 | 6 |
| Chest acceleration (g's) | 50.7 | 56 | 47.1 | 30 |
| Air bag | 50.5 | 27 | 46.5 | 24 |
| No air bag | 50.9 | 29 | 49.3 | 6 |

[a]For passenger cars from 1990 to 1993, with and without air bags.

Specifically, we found that

- Vehicles tested with Hybrid III dummies had **lower** HIC and chest acceleration scores than those tested with Hybrid II dummies in both compliance and NCAP tests. In NCAP tests, Hybrid III dummies averaged 156 HIC and 3.6 g's less than Hybrid II dummies. (See table III.1) Similarly, in compliance tests, the mean head injury criterion score for cars tested with Hybrid III dummies was 97 HIC lower than the score for tests that used the Hybrid II, while the mean chest acceleration score was about 2 g's less for test cars that used the Hybrid III. (See table III.2)[4]
- In both the NCAP and compliance tests, Hybrid III had significantly **lower** HIC scores than Hybrid II dummies in vehicles equipped with air bags. In vehicles without air bags, Hybrid IIIs had significantly **higher** HIC scores

[4]All differences except for chest scores in compliance tests were statistically significant ($p < .05$).

than Hybrid IIs. The difference could occur because of the greater flexibility of the Hybrid III's neck.[5]

- In both the NCAP and compliance tests, Hybrid III dummies had significantly **lower** chest acceleration results than Hybrid II dummies in cars with air bags. There was little difference between the chest scores of Hybrid III and Hybrid II dummies in cars without air bags.

**Table III.2: Mean HIC and Chest Acceleration Scores From Compliance Program Tests for Hybrid II and Hybrid III Dummies[a]**

| Measure | Hybrid II | | Hybrid III | |
|---|---|---|---|---|
| | Score | Number of tests | Score | Number of tests |
| HIC | 487 | 97 | 390 | 36 |
| Air bag | 482 | 47 | 268 | 25 |
| No air bag | 492 | 50 | 667 | 11 |
| Chest acceleration (g's) | 45.4 | 97 | 43.3 | 36 |
| Air bag | 48.4 | 47 | 43.8 | 25 |
| No air bag | 42.6 | 50 | 42.3 | 11 |

[a]For passenger cars from 1988 to 1993, with and without air bags.

Manufacturers have been reluctant to use Hybrid III dummies for tests of cars not equipped with air bags because these dummies tend to produce higher HIC results, especially in cases where the dummy's head did not contact the interior components of the car. (Only 18 percent of compliance tests of cars without air bags from 1988 to 1993 and 18 percent of NCAP tests of cars without air bags from 1990 to 1993 used the Hybrid III.) Industry representatives stated that because HIC was developed to determine potential skull injuries—a condition that will not occur if the head does not contact the vehicle's interior—it should be applied only to cases in which the head actually makes contact. Although they agree that brain injuries can occur when the head does not contact the interior, they contend that the instrumentation in the dummy's head does not measure the potential for these types of injuries. Therefore, they conclude that in cases of "noncontact" HICs, the results are meaningless and misleading. Thus, rather than risk a spuriously higher HIC score in either the compliance or NCAP tests, manufacturers have tended to use the Hybrid II for vehicles that do not have air bags.

[5]Because the neck of the Hybrid III is more flexible, it offers less resistance to motion. Assuming two identical crash tests with the same amount of energy in each system, the Hybrid III's head will move through its forward arc of motion (flexion) at a greater velocity than the Hybrid II until it reaches the end of that arc, when it rapidly decelerates.

While these complex interactions of dummy type, safety equipment, and test conditions can be explained by biomechanical differences between the Hybrid II and Hybrid III dummies, they may also be explained by differences in the test vehicles themselves. As we noted earlier, the manufacturers specify which dummy type to use in NHTSA crash tests, and one may assume that, in the absence of other motivations, they would choose the dummy they anticipate will yield more favorable results.

## Exclusive Use of the Hybrid III

As noted above, each manufacturer undergoing a compliance test may specify either the Hybrid II or the Hybrid III test device. But in recent years, NHTSA has become more convinced that using the Hybrid III will help ensure that all new vehicles are designed with the benefit of the most human-like test dummy available. NHTSA regards the Hybrid III as more representative of human responses in frontal crashes, and it can monitor more types of potential injuries as well. Further, NHTSA has come to recognize that exclusive use of the Hybrid III for compliance testing under FMVSS 208 would result in greater comparability of test results among vehicles produced by different manufacturers.

For these reasons, the agency recently issued a Notice of Final Rule that requires the exclusive use of the Hybrid III for all compliance testing under standard no. 208.[6] The final rule takes effect September 1, 1997, to coincide with the date at which all passenger cars and 80 percent of light trucks must be equipped with air bags and all light trucks must have passive (automatic) restraint systems. NCAP will also switch to exclusive use of the Hybrid III test dummy beginning with the 1996 model year. These modifications to the two programs will create a greater degree of standardization of crash tests, thereby, in NHTSA's view, increasing the "comparability of test results among vehicles produced by different manufacturers, particularly those that now use different dummy types."[7]

[6]On Nov. 8, 1993, NHTSA published the final rule (Notice 83, 58 Fed. Reg. 59189), which requires the use of the Hybrid III test dummy for all compliance testing under standard no. 208, to be effective Sept. 1, 1997.

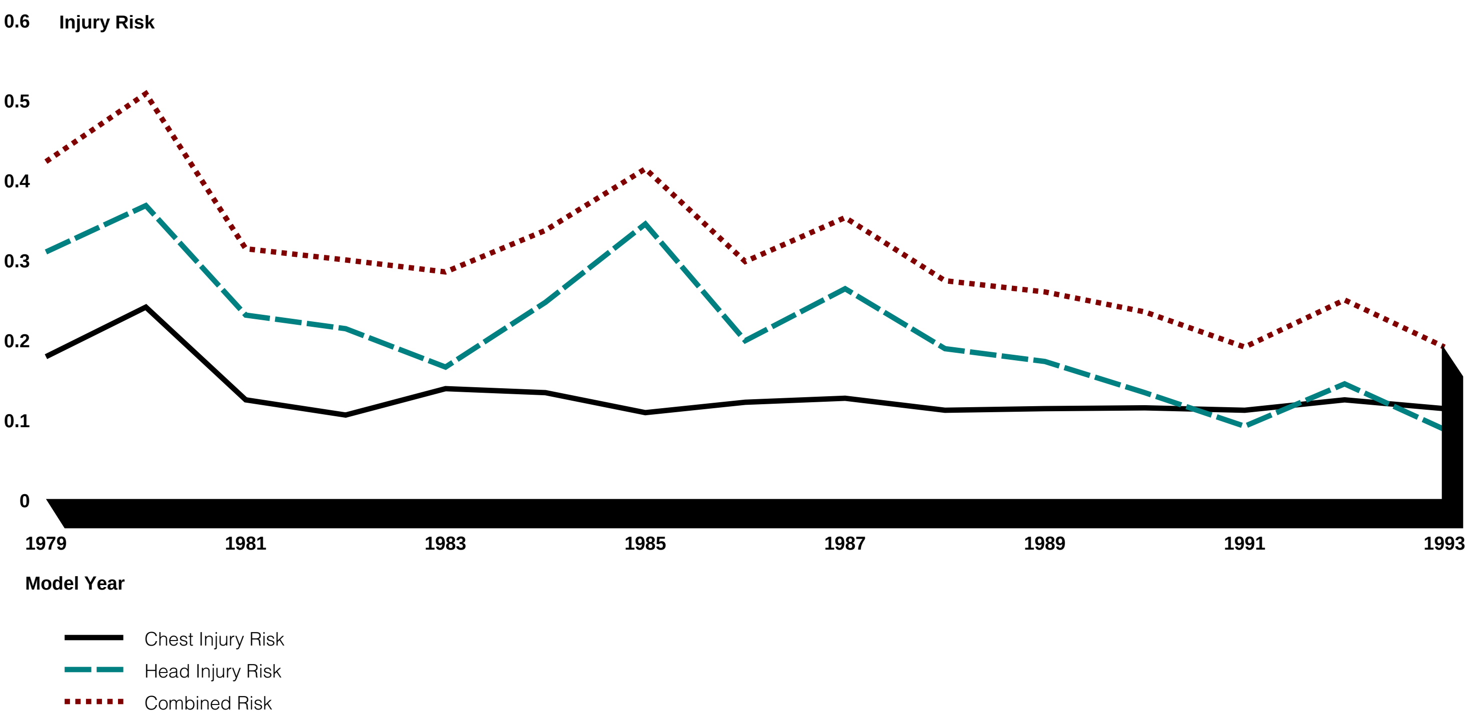
[7]DOT, NHTSA, 49 C.F.R. Part 571: Federal Motor Vehicle Safety Standards; Occupant Crash Protection; 58 Fed.Reg. 214, Nov. 8, 1993, p. 59189.

# Trends in Crash Test Results

## Changes in NCAP Injury Risk Over Time

We conducted analyses of the trends in NCAP test scores from 1979 through 1993 and found that scores have both improved and become more uniform during the period. We have expressed NCAP results in terms of the combined injury risk scores to which NCAP now translates its HIC and chest scores to produce its new “star system”.[1] Figure IV.1 shows the mean injury risks for the driver position, by year, for model years 1979 through 1993.

**Figure IV.1: Mean NCAP Combined, Head, and Chest Injury Risks by Model Year[a]**




[a]For all passenger cars. Does not include light trucks and vans.

The mean combined injury risk decreased significantly from a high of 0.507 in 1980 to a low of 0.190 in 1993. The figure also indicates that the significant reduction in the combined risk derives from a significant and consistent decrease in the mean head injury risk probability. While the mean chest injury risk declined significantly during the period, it has been

[1]For a complete discussion of the derivation of risk scores and star ratings see appendix II.

relatively stable since 1983. The variation between the individual test results has also decreased over the years. For example, NCAP head injury criterion scores for vehicles in 1979 ranged from 521 HIC to 4,513 HIC, whereas in 1993, the range was between 273 HIC and 1,459 HIC.

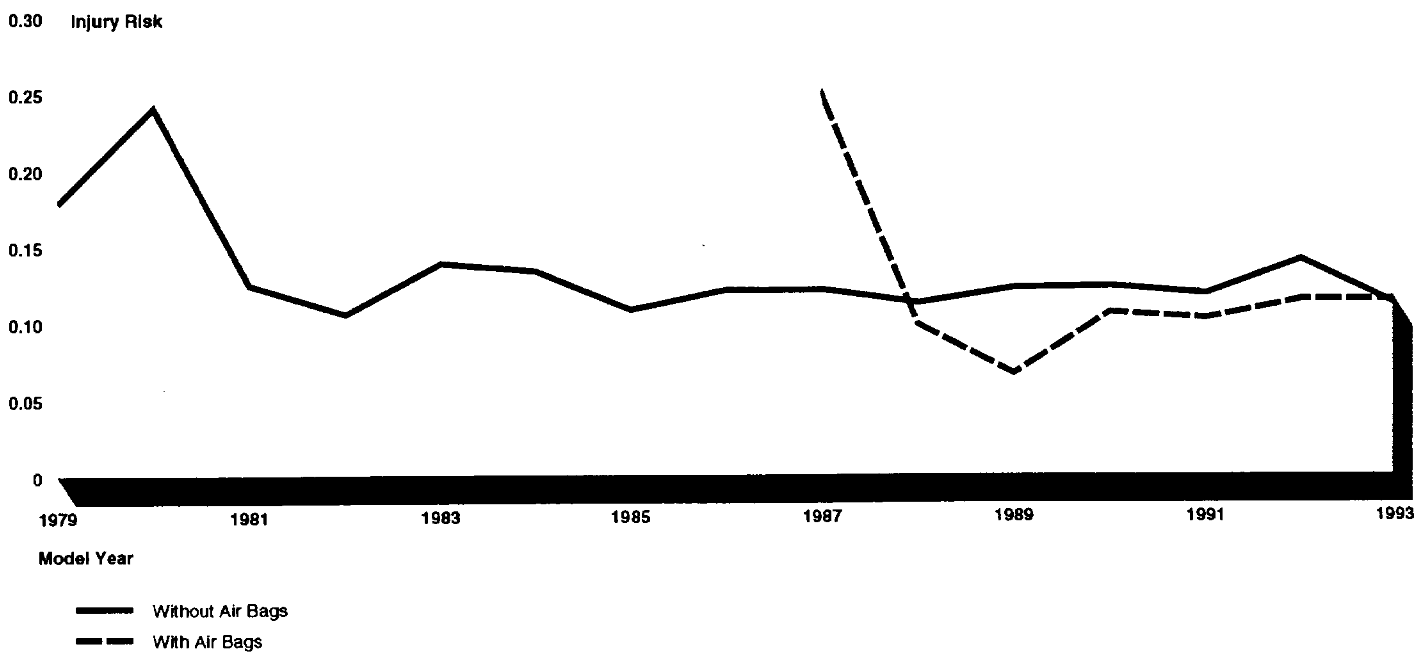
One reason for the decline in the mean combined injury risk and its accompanying variation over time is the increasingly widespread installation of air bags. Cars equipped with air bags had significantly lower head injury risk probabilities than cars without air bags. (See figure IV.2.) Since the first NCAP test of cars equipped with air bags in 1987, these vehicles have scored an average head injury risk of 0.063, while cars without air bags have averaged 0.216.

**Figure IV.2: Mean NCAP Head Injury Risk, With and Without Air Bags[a]**


[a]For passenger cars with and without driver-side air bags. Does not include light trucks and vans. Before 1987, no test vehicle was equipped with an air bag, and only one 1987 model-year test vehicle was equipped with an air bag.

There is little difference, however, between the mean chest injury risks for passenger cars equipped with air bags (0.108) and those that did not employ this type of restraint (0.120). (See figure IV.3.) Given the relatively flat chest injury risk shown in figure IV.1, it appears that this risk factor, regardless of the type of restraint, has contributed little to the declining trend for the combined injury risk.

**Figure IV.3: Mean NCAP Chest Injury Risk, With and Without Air Bags[a]**

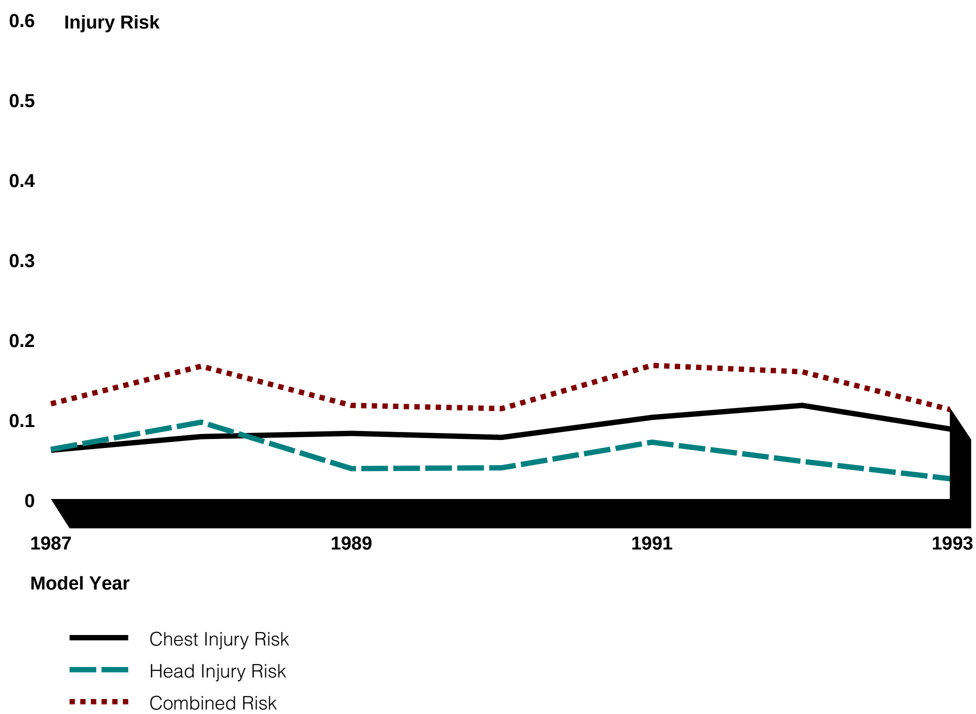



[a]For passenger cars with and without driver-side air bags. Does not include light trucks and vans. Before 1987, no test vehicle was equipped with an air bag, and only one 1987 model-year test vehicle was equipped with an air bag.

## Changes in FMVSS 208 Injury Risk Over Time

We also conducted similar analyses for passenger cars tested in the compliance program. Despite fluctuations from year to year, the combined injury risk did not change significantly from 1987 to 1993.[2] (See figure IV.4.)

**Figure IV.4: Mean FMVSS 208 Injury Risk by Model Year[a]**

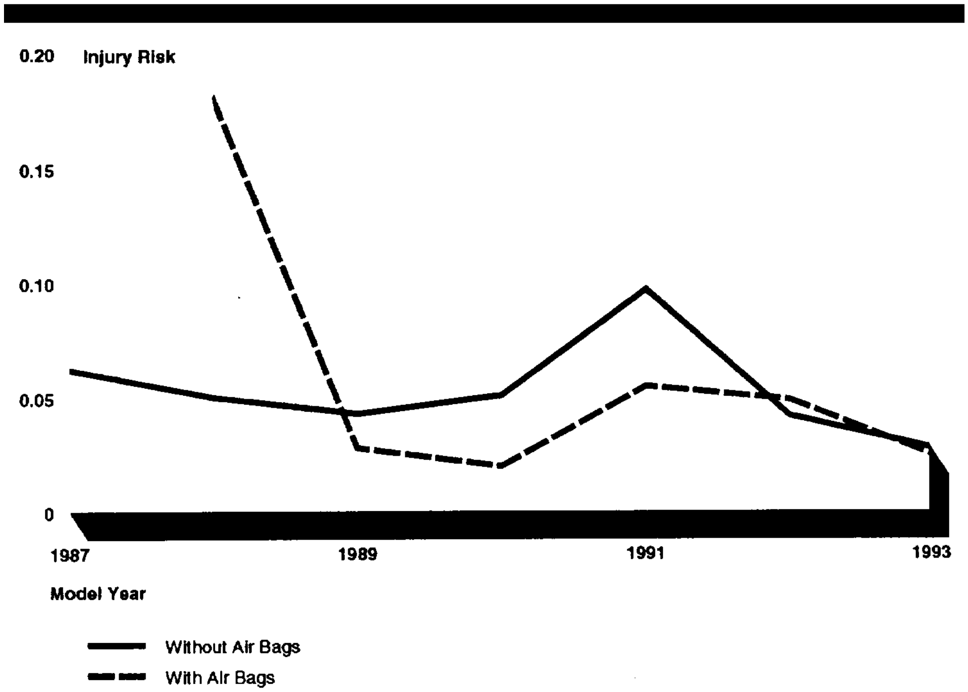



[a]For all passenger cars. Does not include light trucks and vans.

[2]Though the compliance program does not report scores in terms of risk, we translated the results to risk probabilities to give the reader a base scale from which to compare the differences in scores that might result from the differences in the velocities and restraint usage in the two crash test programs.

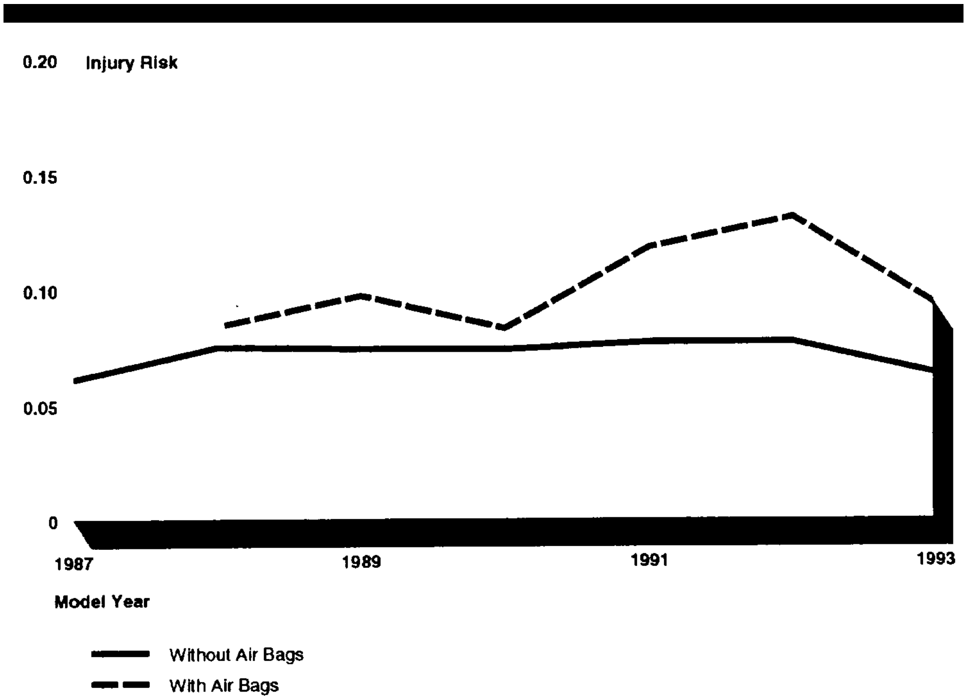
During the same period a steady, though not statistically significant, decline in the mean head injury risk was offset by a significant increase in the mean chest injury risk. These opposing trends are associated with the increased installation of air bags, which are associated with lower head injury risk probabilities and higher chest injury risk probabilities for compliance tests. (See figures IV.5 and IV.6.)

**Figure IV.5: Mean FMVSS 208 Head Injury Risk, With and Without Air Bags[a]**




[a]For passenger cars with and without driver-side air bags. Does not include light trucks and vans. Before 1988, no test vehicle was equipped with an air bag.

Figure IV.6: Mean FMVSS 208 Chest Injury Risk, With and Without Air Bags[a]




[a]For passenger cars with and without driver-side air bags. Does not include light trucks and vans. Before 1988, no test vehicle was equipped with an air bag.

The contrasting chest injury risk results between NCAP and compliance programs may have occurred because of the differences in the configuration of the two tests—largely from the determination of which restraint systems are used. In an NCAP crash that makes use of all available passive and manual restraint systems, the seat belt absorbs the dummy's kinetic energy over a gradual period (for a crash event) before the dummy contacts the air bag. However, cars with air bags are not required to have **automatic** seat belts. And since manual seat belts are not engaged in compliance tests, the dummy in the driver position is not likely to be restrained by a safety belt in compliance tests of cars with air bags. The dummy, therefore, is likely to move forward without a reduction in its kinetic energy, resulting in a more forceful collision with the air bag than if a seat belt was also in use. Over time, as more of the vehicles tested in the compliance program came equipped with air bags, the mean compliance chest score increased.

This is **not** to say that an air bag-equipped vehicle is less safe than one that does not have an air bag. Indeed, the chest g results of cars with air bags may not be directly comparable to those without the devices because the distribution of the force loading on the chest is different for air bags than for safety belts. Air bags distribute the load caused by chest contact across a larger surface area than safety belts. Nevertheless, the higher chest g result for an air bag-equipped vehicle is consistent with the view held by many traffic safety experts that safety belts alone (that is, without air bags) are much more effective than airbags alone (that is, without safety belts).[3]

The decreasing mean crash test results parallel a similar trend with annual fatalities, and part of that latter trend can rightfully be attributed to NHTSA crash tests. In discussions with industry representatives, we found that automobile manufacturers attempt to design vehicles to meet compliance levels for frontal collisions in the NCAP test, as well as ensure that all other safety criteria are met. In addition to "live" crash tests, manufacturers use computer models of frontal collisions, rear and side impacts, and roof crush to simulate NHTSA crash tests and to ensure that the cars meet NHTSA standards. The types of simulations range from models of specific components of the vehicle to nonlinear finite-element models, which incorporate all specifications of the automobile and can predict interactions between the car and its occupants during a collision. These simulations allow manufacturers to gain insight into the deformation of the vehicle, likely intrusions into the occupant compartment, and the force loads generated by various structural components. Though computer-simulated crashes are expensive because they generally require access to a supercomputer, they do allow manufacturers to gain knowledge of how the car will perform and how to correct problems before building prototypes. They are also much less costly and less time-consuming than building and crashing prototypes. In addition, one simulation expert stated that the results of finite-element simulations generally reflect the results obtained in actual crash tests and that the industry uses crash tests in part to validate its computer-based crash models.

[3]The relative protection afforded by safety belts and airbags is discussed in Highway Safety: Causes of Injury in Automobile Crashes (GAO/PEMD-95-4; May 1995).

# Reliability of NCAP Results

Our analyses of the trends in NCAP test results have shown that (1) injury risk probabilities have declined over time and (2) the variation between test results has lessened over time. (See appendix IV.) It would seem that cars have become more crashworthy—at least as measured by NCAP—and that this improved crashworthiness is more uniformly distributed across the passenger car fleet. Indeed, in 1979 the combined injury risk for NCAP-tested vehicles ranged from 0.106 to 1.0 (rounded), whereas in 1993, the ends of the distribution ranged from 0.096 to 0.581. Despite the decreased variation in test results, however, the difference between the highest and lowest risk probabilities is still substantial.

This variation is open to two quite different interpretations. It may indicate the sensitivity of crash tests to real differences between vehicle models, or it may reflect the imprecision of the test scores. In classical measurement theory, reliability is defined as the repeatability of test results. The reliability of crash tests would be estimated by comparing the results from repeated crash tests of the same model vehicle.

On only one occasion has NHTSA attempted to determine whether NCAP crash test results are, in fact, reproducible by crashing a single model on multiple occasions. In this study, 12 consecutively manufactured 1982 Chevrolet Citations were crash-tested by three test facilities, with each facility testing four vehicles, in an attempt "to quantify the degree of variation, as well as develop generalized statistical conclusions about test repeatability."[1] The mean HIC score for the 12 tests was 685, with the scores ranging from 495 HIC to 954 HIC. NHTSA identified several sources of variation in results derived from the test procedure, as well as from the testing facilities, the test instrumentation, the test dummy used, and the individual vehicles. NHTSA could not quantify the amount of variation attributable to each of these five areas because of the number of possible sources of error within each.

Although the amount of variation that can be attributed to any of the sources of error is incalculable, the confounding interactions of accumulated error lead to questions about the reliability of NCAP results. The variation between different units was artificially constrained by selecting 12 consecutively manufactured Chevrolet Citations, yet the head

[1]John M. Mackey and Charles L. Gautier, Results, Analysis and Conclusions of NHTSA's 35 MPH Frontal Crash Test Repeatability Program, SAE Paper No. 840201 (Warrendale, Pa.: SAE, 1984), p. 74.

injury results still had a range of 459 HIC.[2] This variation among the HIC scores implies that two scores with a range of less than 219 are not, in statistical terms, significantly different and that any score between 781 HIC and 1,219 HIC is not significantly different from 1,000.[3]

Table V.1 illustrates how this level of reliability could affect the interpretation of other crash test scores. The table displays the mean HIC score from the 12 Citation crash tests and the HIC scores of four other vehicles of similar weight.[4] It also indicates whether the NCAP HIC scores are significantly different from (1) the mean Citation score and (2) the putative ceiling of 1,000.[5] The HIC scores received by all the vehicles except the 1990 Lexus are significantly lower than 1,000. However, there is no statistically significant difference between the mean HIC score received by the Citation and three of the four other vehicles.

**Table V.1: Mean HIC Scores From NHTSA's Repeatability Test and NCAP Results for Vehicles of Similar Weight[a]**

| | | | Significantly different from | |
|---|---|---|---|---|
| Vehicle | Curb weight (pounds) | HIC | 1982 Citation | HIC = 1,000 |
| 1982 Chevrolet Citation | 3,260 | 685 | [b] | Yes |
| 1988 Oldsmobile Delta 88 | 3,460 | 710 | No | Yes |
| 1990 Lexus ES250 | 3,280 | 992 | Yes | No |
| 1991 Ford Taurus | 3,290 | 480 | No | Yes |
| 1991 Chrysler New Yorker | 3,310 | 511 | No | Yes |

[a]From NHTSA's 12 1982 Chevrolet Citation tests and NCAP results for vehicles of similar weight.

[b]Does not apply.

[2]This constraint was imposed by NHTSA to limit, as far as possible, sources of error outside the test process itself. The results of the 12 tests, however, revealed differences in the crush of the test vehicles, for example. (The "crush" is the distance by which the car is collapsed by the collision—from the front of the car moving rearward.) Thus, error derived from the manufacturing of the vehicles still influenced the overall variation among tests, but NHTSA could not quantify its influence.

[3]These ranges derive from a calculation of a 95-percent confidence interval using the mean and standard deviation of the results from NHTSA's repeatability study. A similar study of seven 1983 Volvo 760 GLEs conducted by the manufacturer revealed a somewhat narrower, though still substantial, range in HIC scores (from 697 to 1,004) and indicated that two HIC scores with a range of as much as 147 are not significantly different.

[4]NHTSA cautions that scores from vehicles more than 500 pounds different are not necessarily comparable.

[5]For these comparisons, we assume that the other models would demonstrate the same HIC score variations in multiple crash tests as the Citation. While this assumption may be subject to debate, we have no other data from which to form a different assumption. We use HIC scores here rather than combined injury risk since the chest g scores from the test were unavailable.

After reviewing a draft of this report, NHTSA provided us with a second set of data that could shed additional light on the reliability of NCAP scores. These data represent the results of crash tests of model-year 1991 through 1994 vehicles conducted by automobile manufacturers in tests that essentially duplicate the NCAP test conditions. The data were voluntarily submitted to NHTSA before planned NCAP tests.

We compared the manufacturer scores with those obtained from NCAP after translating them into the single injury probability score that serves as the basis for NHTSA's recently introduced star rating system. (See appendix II.) We found that a statistically significant first-order correlation exists ($r = .72$) between the two sets of injury risk probabilities.

We then compared the distributions of star ratings derived from NCAP and manufacturers' tests. Table V.2 compares the star ratings for the driver position, and table V.3, for the passenger position. If agreement between the two sets of tests had been perfect, all events in the tables would have fallen on the diagonals from upper left to lower right. In actuality, star ratings are the same for approximately one-half of the vehicle models tested (55 percent for the driver position and 45 percent for the passenger position). Differences of one or more stars exist between manufacturer and NCAP ratings for about one-half of the tests; 8 percent of the vehicle models have differences of two or more stars.

**Table V.2: Number of Vehicles Within Star Rating Categories, Driver Position**

| | Manufacturer star rating | | | | | |
|---|---|---|---|---|---|---|
| **NCAP star rating** | **1 star** | **2 stars** | **3 stars** | **4 stars** | **5 stars** | **Total** |
| 1 star | 6 | 0 | 3 | 3 | 0 | 12 |
| 2 stars | 0 | 1 | 3 | 1 | 0 | 5 |
| 3 stars | 1 | 1 | 16 | 14 | 0 | 32 |
| 4 stars | 0 | 1 | 15 | 37 | 6 | 59 |
| 5 stars | 0 | 0 | 0 | 6 | 5 | 11 |
| **Total** | **7** | **3** | **37** | **61** | **11** | **119** |

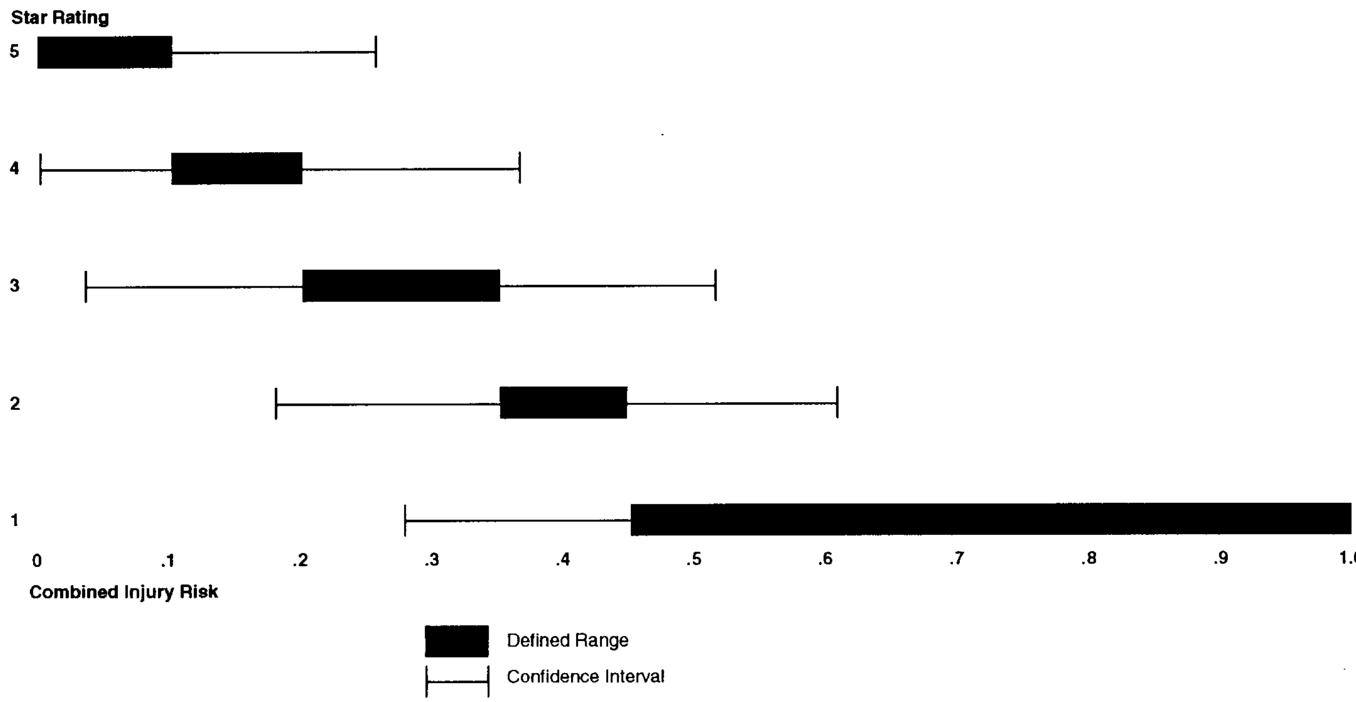
**Table V.3: Number of Vehicles Within Star Rating Categories, Passenger Position**

| | Manufacturer star rating | | | | | |
|---|---|---|---|---|---|---|
| **NCAP star rating** | **1 star** | **2 stars** | **3 stars** | **4 stars** | **5 stars** | **Total** |
| 1 star | 2 | 1 | 1 | 2 | 0 | 6 |
| 2 stars | 1 | 0 | 3 | 3 | 0 | 7 |
| 3 stars | 1 | 2 | 8 | 23 | 1 | 35 |
| 4 stars | 1 | 0 | 10 | 33 | 6 | 50 |
| 5 stars | 0 | 0 | 0 | 11 | 10 | 21 |
| **Total** | **5** | **3** | **22** | **72** | **17** | **119** |

As appendix II explains, each of NHTSA's star ratings represents a range of injury probability. For example, a rating of 4 stars indicates that in a crash situation similar to that tested by NHTSA the probability of serious injury to an individual is between 1 in 10 and 2 in 10. The solid bars in figure V.1 depict these probability ranges for each star rating. The lines attached to the bars represent "confidence intervals" that we estimated from the standard deviation of the absolute difference between the combined injury risks for drivers derived from manufacturer and NCAP tests. These confidence intervals represent the estimated range of injury probability within which a vehicle with a nominal rating could be expected to vary if tested again. For example, a 4-star rating could be associated with a vehicle with a "true" injury probability between zero and 0.363. This range overlaps the confidence intervals associated with the 5- and 2-star ratings.

Our estimates of the reliability of NCAP crash tests are based on the only two sources of relevant information we are aware of: the repeated crashes of the 1982 Chevrolet Citation, and a comparison of manufacturer and NCAP test scores from 1991 to 1994. Neither of these sources provides ideal information for precisely quantifying the measurement error associated with NCAP scores. We do not know how well the results of the Citation experiment can be applied to vehicles manufactured and crash-tested 10 years later. While the manufacturer-NCAP comparison applies to a large number of late-model cars, we cannot be sure how well the manufacturers succeeded in replicating NCAP crash test conditions in each case or to what extent the results from other manufacturer tests varied from the ones reported to NHTSA.

**Figure V.1: Risk of Serious Injury, With Estimated Confidence Interval, by NCAP Star Rating[a]**




[a]Confidence interval is based on 1.96 standard deviations of the absolute difference between the combined injury risk scores derived from manufacturer and NCAP tests of the same vehicle model.

Nevertheless, both analyses support the same conclusion that NCAP scores, whether reported in raw HIC and chest acceleration scores or as categories of injury probability, have associated levels of imprecision. As a result, substantial differences in scores between two test results (100-200 HIC, or a 1-star—or possibly 2-star—rating difference) may not represent true differences in crashworthiness.

# Methodology for Analyzing the Predictive Validity of NCAP

## Introduction

The overall purpose of both the compliance and NCAP crash tests is to determine the crashworthiness, or safety, of passenger vehicles. This implies, therefore, that a relationship exists between the results of crash tests and real-world injuries and fatalities. To examine this issue, we conducted two analyses comparing results derived from NCAP tests to those from national accident databases.[1] Specifically, our analysis compared NCAP results with traffic injury and fatality information from the National Accident Sampling System (NASS) and the Fatal Accident Reporting System (FARS). This appendix details the methodologies and results of both analyses.

## Relationships Between NCAP and NASS Data

Data from the National Accident Sampling System for 1988 through 1991 were combined to determine whether the results of New Car Assessment Program tests are good predictors of serious injuries and fatalities in real-world automobile crashes. The NASS is a sample of annual police-reported accidents involving passenger cars, light trucks, and vans that had to be towed because they were damaged. The NASS year corresponds to the calendar year rather than the automobile industry's model year, and emphasis is placed on the most recent 5 model-year vehicles. We chose this data system for two reasons: (1) It is a national database that contains information on all types of automobile collisions, and (2) it is the only national database that reports a vehicle's change in velocity (delta v)—the best available indicator of accident severity—resulting from the collision.

We reduced the NASS data sets to single-vehicle and two-vehicle accidents and then combined them into one data set. This resulted in a total of 14,253 vehicles in the final data set. We then matched the results of the NCAP crash tests to the vehicles in our NASS file. NCAP results from 1983 through 1992 were chosen for the analysis for two reasons: (1) We assumed that crash test results are applicable for a given time (as cars age, their crashworthiness may decrease owing to wear and tear) and (2) NASS data were available for the calendar years 1988 through 1991. Because NASS emphasizes for inclusion collisions involving vehicles from the 5 most recent model years, we chose 5 years as the period of time at which a test score no longer applies; therefore, NCAP scores from 1983 were the earliest

[1]For these analyses, we decided to use results only from NCAP for two reasons. First, NCAP has been crash-testing vehicles since 1979 and had conducted about 340 tests of passenger cars through 1992. The compliance program has been in existence only since 1987 and had conducted only 145 tests. Second, presumably because of the higher velocity used in the NCAP tests, the results of NCAP tend to vary to a greater extent than those of the compliance program, and a large portion of vehicles tested in NCAP have HIC and chest acceleration scores in excess of the ceiling values of 1,000 and 60 g's, respectively, while these scores tend to cluster far below the ceiling values for the compliance tests.

we included in the analysis. In addition, each NASS year has a half-year's data from the following model year, as the model year usually begins in late summer. Therefore, the 1991 NASS had some accidents involving 1992 model-year automobiles.

The NCAP results were matched to the NASS data set on the following criteria: the make of the vehicle (that is, the manufacturer), the model, the model year, and the body type (sedan, convertible, and so forth). In cases in which a specific make, model, model year, and body type were tested on more than one occasion, only the first test was used. Results for models with corporate twins (vehicles with platforms identical to one another but sold under different model names—for example, Ford Taurus and Mercury Sable) were projected to the twins. The resultant data set for our analyses contained 1,985 cases. When weighted by NASS sampling weights, these represented more than 9 million accidents.

We conducted logistic regression analyses to determine whether a relationship exists between serious injuries and fatalities in actual automobile collisions and results from crash tests conducted for NCAP. We analyzed crashes in which damage to the right front, left front, or full front of the vehicle occurred. Single-car and two-car crashes were examined separately.[2] We conducted two sets of analyses: one on the unweighted sample in our dataset and a second, which incorporated the NASS sampling weights.[3]

Our analyses of the relationship between real-world traffic injuries and fatalities and NCAP injury risk probabilities were limited to drivers of passenger cars (two-door and four-door sedans, coupes, hatchbacks, and convertibles).[4] We restricted our analyses to "restrained" drivers—that is, drivers who made proper use of either a manual or an automatic seat belt or whose air bag deployed during the crash.

The dependent variable used in the analyses was constructed from NASS injury codes to represent whether the driver of an NCAP passenger car involved in a crash either died or was hospitalized for at least 1 day

[2]"Two-car crashes" refers to collisions between two passenger cars, or a passenger car tested in NCAP and either a light truck or van as defined by the NASS.

[3]To analyze the weighted sample, we used the Survey Data Analysis (SUDAAN) statistical package, which takes into account the stratification and unequal selection probabilities inherent in the sampling design of NASS.

[4]Although we included light trucks and vans in determining the sample for two-car collisions, we analyzed the relationship between NCAP scores and traffic injuries and fatalities for passenger cars only.

specifically because of the crash. This was coded as a dichotomous variable, with those who died or were hospitalized receiving a 1 and all other nonmissing values receiving a zero. The independent variable of interest was the combined injury risk score associated with specific vehicle models as derived from the HIC and chest acceleration scores from NCAP tests. (See appendix II.) However, because characteristics of the driver and the vehicle and, most importantly, the severity of the crash (as measured by the total change in velocity, or delta v) are associated with the likelihood of injuries and fatalities, we included occupant characteristics (age, gender), vehicle characteristics (curb weight and, in two-car crashes, the weight of the other vehicle), and crash severity (delta v) in our logistic regression models.[5]

Tables VI.1 and VI.2 present the results of our analyses of the unweighted sample for one- and two-car crashes. The predictive power of delta v dominates both models, but the driver's age and the car's weight also appear as significant predictors of injury in two-car crashes. In these crashes, older drivers and drivers of lighter cars were more likely to suffer injury or death. In neither model was the NCAP injury risk significantly related to hospitalization or death in either one-or two-car crashes.

**Table VI.1: Logistic Regression Predicting Hospitalization or Death in Single-Car Frontal Crashes, Unweighted Sample[a]**

| Variable | Beta | Standard error | Wald statistic | Significance level |
|---|---|---|---|---|
| Injrisk | –6.8337 | 5.0594 | 1.8243 | .1768 |
| Age | 0.0004 | 0.0245 | 0.0002 | .9883 |
| Gender | 0.2689 | 0.4369 | 0.3790 | .5381 |
| Curbwgt | –0.1021 | 0.1404 | 0.5287 | .4672 |
| Dvtotal | 0.1580 | 0.0567 | 7.7579 | .0053 |
| Constant | 0.4380 | 3.9979 | 0.0120 | .9128 |

Legend

Injrisk = Driver injury risk
Age = Age of driver
Gender = Gender of driver
Curbwgt = Vehicle's curb weight
Dvtotal = Total change in velocity (mph)

[a]Represents 46 restrained drivers.

[5]See Highway Safety: Have Automobile Weight Reductions Increased Highway Fatalities? (GAO/PEMD-92-1; Oct. 1991); Highway Safety: Factors Affecting Involvement in Vehicle Crashes (GAO/PEMD-95-3; Oct. 1994); and Highway Safety: Causes of Injury in Automobile Crashes (GAO/ PEMD-95-4; May 1995).

**Table VI.2: Logistic Regression Predicting Hospitalization or Death in Two-Car Frontal Crashes, Unweighted Sample[a]**

| Variable | Beta | Standard error | Wald statistic | Significance level |
|---|---|---|---|---|
| Injrisk | 1.2413 | 1.0302 | 1.4519 | .2282 |
| Age | 0.0430 | 0.0142 | 9.1362 | .0025 |
| Gender | –0.0782 | 0.2388 | 0.1072 | .7434 |
| Curbwgt | –0.1689 | 0.0611 | 7.6327 | .0057 |
| Othvehwgt | –0.0014 | 0.0338 | 0.0017 | .9671 |
| Dvtotal | 0.2311 | 0.0516 | 20.0368 | .0000 |
| Constant | –2.7524 | 1.8460 | 2.2231 | .1360 |

Legend

Injrisk = Driver injury risk
Age = Age of driver
Gender = Gender of driver
Curbwgt = Vehicle's curb weight
Othvehwgt = Weight of other vehicle
Dvtotal = Total change in velocity (mph)

[a]Collisions with vehicles weighing less than 10,000 pounds. Represents 131 restrained drivers.

Tables VI.3 and VI.4 present the findings from the weighted sample and show very similar results to the unweighted sample. The strongest predictor remains the crash severity, and in two-car crashes, driver age is related to collision outcomes. The weighted sample presents two different conclusions from the unweighted sample, however. The curb weight of the vehicle falls short of statistical significance by traditional criteria, and more to the point, a significant relationship between NCAP risk scores and death or hospitalization appears.

**Table VI.3: Logistic Regression Predicting Hospitalization or Death in Single-Car Frontal Crashes, Weighted Sample[a]**

| Variable | Beta | Standard error | Wald statistic | Significance level |
|---|---|---|---|---|
| Injrisk | –14.2980 | 11.2770 | 1.6075 | .2230 |
| Age | –0.0300 | 0.0332 | 0.8135 | .3804 |
| Gender | 0.4028 | 1.5665 | 0.0661 | .8004 |
| Curbwgt | –0.2023 | 0.1546 | 1.7118 | .2092 |
| Dvtotal | 0.2672 | 0.0690 | 15.0409 | .0013 |
| Constant | 2.3599 | 6.3191 | 0.1395 | .7137 |

Legend

Injrisk = Driver injury risk
Age = Age of driver
Gender = Gender of driver
Curbwgt = Vehicle's curb weight
Dvtotal = Total change in velocity (mph)

[a]Represents 8,401 restrained drivers.

**Table VI.4: Logistic Regression Predicting Hospitalization or Death in Two-Car Frontal Crashes, Weighted Sample[a]**

| Variable | Beta | Standard error | Wald statistic | Significance level |
|---|---|---|---|---|
| Injrisk | 3.2379 | 1.3351 | 5.8817 | .0275 |
| Age | 0.0635 | 0.0200 | 10.0933 | .0059 |
| Gender | 0.9361 | 0.6075 | 2.3752 | .1429 |
| Curbwgt | –0.1298 | 0.0702 | 3.4238 | .0828 |
| Othvehwgt | –0.0433 | 0.0587 | 0.5435 | .4717 |
| Dvtotal | 0.2860 | 0.0891 | 10.3078 | .0055 |
| Constant | –7.1838 | 2.6253 | 7.4875 | .0146 |

Legend

Injrisk = Driver injury risk
Age = Age of driver
Gender = Gender of driver
Curbwgt = Vehicle's curb weight
Othvehwgt = Weight of other vehicle
Dvtotal = Total change in velocity (mph)

[a]Collisions with vehicles weighing less than 10,000 pounds. Represents 23,514 restrained drivers.

Some degree of doubt must be associated with these findings because of the nature of the sample on which they are based. NASS uses a highly complex stratified sampling design to achieve national representativeness for its relatively small sample of observations. The NASS database we used contained 21,377 observations, which when properly weighted, represent more than 9 million accidents. Unfortunately, we found only 366 instances of NCAP-tested cars that met our criteria of properly restrained drivers, and only about one-third of these could be used because of missing values on one or more variables. This drastic reduction in sample size, when combined with the highly uneven distribution of missing values across sampling strata, makes the sampling weight associated with any observation of doubtful validity.

## Relationships Between NCAP and Fatality Rates

To overcome the statistical limitations of our NASS database, we turned to the Fatal Accident Reporting System (FARS). By using FARS, we looked to substantially increase the number of usable cases in the analysis in that FARS contains information on all accidents in a given year that involve at least one fatality (about 45,000 cases per year), while NASS contains only a sample of all accidents (about 3,000 cases per year). In addition, we reasoned that while FARS lacks the information on crash severity provided by NASS' estimate of the total change in velocity, its severity of crashes was relatively homogeneous because the database is restricted to fatal—presumably severe—crashes.

For this analysis, only passenger cars from the 1982 to 1991 model years were included.[6] In addition to the actual test vehicles, our analysis included vehicles that had no substantial structural changes in model years following the tested model year. That is, if a 1984 model-year vehicle were tested in NCAP and no structural changes were made to the 1985 version of the vehicle and it was not retested, the 1985 model year was assigned the same combined injury risk score as the 1984 vehicle.

We then matched the vehicles to the FARS and the R.L. Polk Vehicle Registration System (Polk) databases based on the make (that is, the manufacturer), model, model year, and body type of the vehicle. The FARS database is a compilation of all automobile accidents in the United States in any given calendar year in which at least one fatality occurred. The Polk system is a database that contains information on the types, numbers and weights of vehicles registered in a given calendar year. Data for both systems are for the calendar years 1987 through 1991. As with the analysis of NASS data, we restricted this analysis to one- and two-car frontal collisions in which the driver of the NCAP-tested vehicle was restrained by either a seat belt or an air bag.

[6]Convertibles, light trucks, vans, and multipurpose vehicles were excluded from the analysis.

Having matched the NCAP vehicles to the FARS and Polk systems, we then calculated the fatality rates for the vehicles. This was done simply by dividing the number of fatalities by the number of registered vehicles. The fatality rates in our analysis are expressed in terms of fatalities per 100,000 registered vehicles.

We then correlated the driver combined injury risk scores and fatality rates associated with vehicle models in a number of ways. First, we calculated a simple correlation using just information on those elements. Next, we regressed fatality rates on additional characteristics associated with vehicles using a Poisson model, which allows one to compare rates of individual cases, especially when the sample size is moderately large and the probability of an event occurring is either very low or very high. This type of analysis fit our needs in that a large number of cases (884) were included in our analyses while the fatality rates of the vehicles included were low (overall, there were 1,036 deaths for approximately 19 million registered vehicles).

The variables added to the model held information on the model year and body style of the vehicles in the dataset. We controlled for the model year as a proxy for certain driver, vehicle, and roadway characteristics that could not be included in the model. We controlled for the body style for two reasons: (1) as a surrogate for the relationships found between specific body styles and certain driver characteristics and (2) as a rough surrogate for the weight of the vehicle.[7]

As a final analysis, we divided the NCAP injury risk distribution into quintiles and compared the fatality rates of the different groups. Each quintile represented one-fifth of the passenger cars tested in NCAP from 1982 to 1991.

We found that a first-order correlation between NCAP injury risk and fatality rates exists ($p = .007$). When information on the body style and model year of the vehicle was included in the analysis, the strength of the relationship increased ($p = .001$). However, the relationship appears to be the result of the high fatality rates associated with the poorest performers in NCAP. Indeed, vehicle models within the highest quintile of injury risk (those in the highest 20 percent of the distribution) had significantly higher fatality rates than all other quintile categories. Further, we found that the worst performers on the NCAP test had injury risk probabilities

[7]For a discussion of vehicle and driver characteristics associated with accident involvement, see Highway Safety: Factors Affecting Involvement in Vehicle Crashes (GAO/PEMD-95-3, Oct. 1994).

approximately eight times higher than the best-scoring cars, while their fatality rates were almost 28 percent higher. The remaining four quintile groups, on the other hand, were not significantly different from one another. (See tables VI.5 and VI.6.) Thus, it seems that the relationship between driver fatality rates and predicted injury risk stems from the significantly higher fatality rates associated with vehicles that have very high NCAP injury risk probabilities.

**Table VI.5: Mean Combined Injury Risk Scores and Fatality Rates for All Body Styles[a]**

| Quintile | Mean combined injury risk (NCAP) | Fatality rate (FARS/Polk) |
|---|---|---|
| 1 | .093 | .503 |
| 2 | .140 | .577 |
| 3 | .192 | .573 |
| 4 | .307 | .497 |
| 5 | .723 | .642 |

[a]Fatality rate calculated by dividing the number of fatalities in a given risk quintile by the total number of registered vehicles in that quintile. Fatality rates are expressed as fatalities per 100,000 registered vehicles. The mean combined injury risk is the average NCAP injury risk for the quintile. Body styles include coupes, sedans, two- and four-door hatchbacks, and station wagons.

**Table VI.6: Z-Scores of Differences Between Mean Fatality Rates of Combined Injury Risk Quintiles[a]**

| | Z-score | | | | |
|---|---|---|---|---|---|
| | Comparison quintile | | | | |
| Quintile | 1 | 2 | 3 | 4 | 5 |
| 1 | [b] | .363 | .224 | .813 | 2.745 |
| 2 | -.363 | [b] | -.150 | .520 | 2.684 |
| 3 | -.224 | .150 | [b] | .635 | 2.690 |
| 4 | -.813 | -.520 | -.635 | [b] | 2.119 |
| 5 | –2.745 | –2.684 | –2.690 | –2.119 | [b] |

[a]Mean fatality rates as reported in table VI.5. A z-score of 1.96 or greater shows significant differences between the means at a probability level of at least p = .05.

[b]Not applicable.

# Comments From the Department of Transportation

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

U.S. Department of Transportation

Assistant Secretary for Administration

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

July 13, 1994

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

Dear Ms. Chelimsky:

Enclosed are two copies of the Department of Transportation's comments concerning the U.S. General Accounting Office draft report titled, "Highway Safety: NHTSA Crash Tests Have Limited Reliability and Validity."

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

Sincerely,

Paul Weiss
for Jon H. Seymour

Enclosures

DEPARTMENT OF TRANSPORTATION (DOT) REPLY

TO

GENERAL ACCOUNTING OFFICE (GAO) DRAFT REPORT

ON

HIGHWAY SAFETY:

"NHTSA Crash Tests Have
Limited Reliability and Validity"

**SUMMARY OF GAO FINDINGS AND RECOMMENDATIONS**

The GAO draft report concluded that the worst performers in the National Highway Traffic Safety Administration's (NHTSA) New Car Assessment Program (NCAP) had statistically significant higher fatality rates than the best performers in the test. In addition, persuasive evidence exists that NHTSA's crash tests have had a salutary effect on the safety design of cars sold in this country. The improvement of crash test scores over the past decade most likely mirrors a real improvement in safety design which might not have occurred in the absence of NHTSA's crash testing program. However, the draft report expresses concern regarding the validity and reliability of the test results since only one vehicle of each make and model is tested and the results may not be scientifically reproducible.

The draft report recommends that NHTSA:

- o update information on NCAP reliability and make this information available in clear language to the general public, and
- o explore the feasibility of alternative means of testing the crashworthiness of new vehicles.

**DEPARTMENT OF TRANSPORTATION POSITION**

The Department is pleased that the GAO draft report recognized the positive contribution that NCAP has had on automotive safety. This program effectively accomplishes its intended purpose of disseminating information to consumers on the relative safety of passenger vehicles. Recent initiatives by the Department will further strengthen the effectiveness of the program. The transition to the easily interpreted star rating system effectively addresses the draft report's concerns regarding the implied precision of the previous numerical system. In addition the Department recently announced a rulemaking to expand the availability of consumer information to include other types of information, particularly the propensity for vehicle rollover.

See comment 1.

The Department is concerned, however, that the title of the draft report does not comport with its primary conclusions and is potentially misleading. We understand that the title's characterization of the NCAP program as being of limited reliability and validity is intended to refer to the narrow scientific sense of results reproducibility. However, common use lends these terms to much broader interpretation and could have damaging effects on the credibility of a program which the draft report recognizes as having salutary effects on passenger vehicle safety. The NCAP program as presently implemented uses valid and reliable data to provide useful information to consumers regarding the **relative** safety performance of vehicles.

**NCAP Testing Variability is Reasonable and Acceptable**

See comment 2.

The Department maintains that an inaccurate assumption in the draft report's analysis raises questions regarding its conclusions on data reliability and validity.
The draft report discussion of reliability was contingent on the results of NHTSA's repeatability and reproducibility testing of the 1982 Citation, and assumes that other makes and models would have similar variability. The draft report states the assumption on page 35 that other models would exhibit the same test result variation as the Citation; however, additional data provided by another manufacturer on a 1983 model showed that variations in test results will not be consistent from model-to-model and in fact were much less. As a result, this assumption is not borne out by fact, and the related portion of the draft report's analysis needs to be readdressed.

In 1984, NHTSA thoroughly analyzed the data from both test series and implemented changes that further reduced test variability. These changes, which are described in a paper published by the Society of Automotive Engineers (SAE), include more consistent placement of the test dummy and the initiation of an instrument auditing system.[1] The draft report could more fully describe NHTSA's reactions to the tests and the changes that have been implemented as a result. We maintain that any remaining variability does not influence the rating or relative ranking of the vehicle and does not affect the veracity of the information provided to the consumer.

**NCAP Provides Consumers with Reliable Relative Crashworthiness Information**

See comment 3.

NHTSA provides thorough oversight for NCAP testing to ensure that the program provides consumers with useful and reliable information regarding the relative crashworthiness of passenger vehicles. Testing for the NCAP program is conducted according to rigorous protocols with careful and well thought out quality assurance. NHTSA carefully reviews NCAP test data prior to accepting the final reports to ensure that the results are valid and reliable. This information is further verified by

[1]"Results, Analysis and Conclusion of NHTSA's 35 mile per hour Frontal Crash Test Repeatability Program," SAE Paper 840201.

2

comparison to the physical phenomenon observed on the post-crash vehicles and dummies and in the high-speed film. In addition, the results of most tests are further verified by cross checking our results against those of the manufacturers. Although conducting additional tests to enhance the scientific reproducibility of the data may narrow its standard deviation, the costs of such a program would be prohibitive. Given the present level of confirmed data accuracy and its application to a system of relative ranking, the additional costs are not justified in light of the limited benefits.

**NCAP Results Correlate with Crash Performance**

See comment 4.

Recent NHTSA technical analyses concluded that 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.[2] In a head-on collision between a car with "good" NCAP performance and a car of equal mass with "poor" NCAP performance, the driver in the acceptable car has on average a 15 to 25 percent lower fatality risk. The correlation between NCAP scores and actual fatality risk is statistically significant. One cannot expect perfect correlation between NCAP test results and actual fatality data because of the many variables which influence real-world crashes. Nonetheless, the significant correlation between NCAP and real-world crashes suggests that the NCAP scores reflect actual crashworthiness.

**NCAP Has Contributed to Improved Crashworthiness**

The GAO draft report and NHTSA's internal studies have concluded that NCAP scores and actual vehicle crashworthiness in head-on collisions have improved steadily since the inception of the NCAP program in 1979. Overall, passenger cars have become significantly safer in head-on collisions as NCAP scores improved. The passenger vehicle fleet has achieved a 20 to 25 percent reduction of fatality risk for belted drivers in actual head-on collisions between model years 1979 and 1991. Concurrently, the mean head injury score in NCAP testing has significantly decreased from a high of 1,308 in 1980 to 669 in 1993. In addition, NCAP results mirror the improvement of the passenger vehicle fleet as a whole as the variation between test results for different cars has decreased over the years. For example, NCAP head injury scores for 1979 ranged from 521 to 4,513 where in 1993 the range had narrowed to 273 to 1,459.

See comment 5.

While the Department has not conducted the requisite analysis necessary to verify any causal relationship between NCAP testing and improved actual crashworthiness, we note with interest the draft report's conclusion that manufacturers' successful efforts to

[2]"Correlation of NCAP Performance with Fatality Risk in Actual Head-On Collisions," January 1994, DOT HS 808 061. In addition, the Department provided information to the Congress regarding NCAP in a report titled, "Response to the NCAP FY 1992 Congressional Requirements," December 1993.

3

improve their products' performance in NHTSA crash tests have contributed to improved occupant protection in real world crashes. Further, the GAO draft report concludes that improvements in crash test scores over the past decade most likely mirror a real improvement in safety design which might not have occurred in the absence of NHTSA's testing program.

**NCAP Results Presentation Has Been Refined**

During the past year NHTSA has substantially improved the NCAP results presentation to facilitate its use by the general public and address concerns regarding the implied precision of describing specific quantitative results. NHTSA previously published the specific numerical results of head, chest, and femur injury testing. Numerical test results for each category would be displayed along with a graphic depiction intended to convey relative rank in a given vehicle category. After analyzing the previous system, we concluded that consumers found it confusing. Experts expressed concerns similar to those identified in the draft report regarding the implied precision of the numerical ratings. Beginning this year, NHTSA implemented a revised rating system providing a combined relative rank in the form of one to five stars, with five stars being the best rating. The combined system more directly addresses consumer concerns regarding the potential for injury regardless of type and eliminates the numerical standings and their implied precision.

See comment 6.

**RESPONSE TO GAO DRAFT REPORT RECOMMENDATIONS**

**Recommendation**: Update information on NCAP reliability and make this information available in clear language to the general public.

See comment 7.

**Response:** Concur. NHTSA has completed an extensive analysis regarding NCAP reliability. These reports, cited in footnote 2 of this response, concluded that NCAP provides reliable information to the general public regarding relative crashworthiness. Based on our work with consumer focus groups, these reports also concluded that the public would be better served by a transition to a simplified data presentation format, exemplified by the five star rating system.

**Recommendation:** Explore the feasibility of alternative means of testing the crashworthiness of new vehicles.

See comment 8.

**Response:** Nonconcur. The text accompanying the recommendation suggests the potential utility of computer simulations to augment or replace crash testing. It further suggests that these computer models are less costly than building and crashing prototypes and appear to accurately predict the results of actual crashes.

We maintain that, based on existing computer hardware and software, augmenting or replacing actual crash testing with computer simulations would be extremely costly,

4

require extensive manufacturer confidential data, and produce results with limited, if any, predictive ability. Although several automotive manufacturers are using crash simulation computer modeling, we understand that its use is limited to comparative evaluations for improving performance, not for determining absolute values. These calculations require expensive supercomputer equipment to perform the complex calculations. The models require hundreds of data inputs for vehicle subsystems such as chassis and body structure, energy management subsystems such as air bags and the steering column, seating system and safety belt influence on dummy kinematics, interior surface impact characteristics, and other factors integral to the vehicle's design. Such data are all company confidential and manufacturers would be extremely reluctant to share it with NHTSA due to competitive concerns. Finally, the predictive ability of these models is in question because despite manufacturers' best efforts at predicting crash test performance by performing complex interrelated calculations at millions of nodes, the head and chest injury scores measured in actual vehicle testing can be considerably different than predicted. In several investigations, both computer and crash test data from manufacturers have been examined and found in some instances to have widely different results. Therefore, a vehicle's crash test performance can only be determined from an actual test performed in strict accordance with the procedures specified in the Federal Motor Vehicle Safety Standards or NHTSA protocols.

**SPECIFIC COMMENTS**

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

See comment 1.

o **Cover:** As previously described, the title of the draft report does not comport with its primary conclusions regarding the program's beneficial effect on vehicle safety and statistically significant link to real-world performance.

See comment 9.

o **Page LET-2:** The draft report states that it found little evidence to support the predictive validity of NCAP in a model-by-model comparison. NHTSA has long recognized that there is insufficient real-world data to support NCAP's predictive validity relative to individual make and models. However, NCAP's predictive validity is well supported by trend analysis of crashworthiness performance, as stated in NHTSA Technical Report DOT HS 808 061, January 1994 and NHTSA's Report to Congress, "Response to the NCAP FY 1992 Congressional Requirements." The GAO draft report could acknowledge these analyses and explain the real-world data limitations and their impact on its analysis.

Now p. 4.
See comment 10.

o **Page LET-2:** The draft report makes apparently contradictory statements indicating that, "Sizable differences between the crash test scores of different test vehicles may have little meaning because of the unreliability of the test results." Several sentences later, the draft report states that "the improvement in crash test scores

5

most likely mirrors a real improvement in safety design...." Recent NHTSA studies have concluded based on extensive analysis that the NCAP tests produce valid and reliable data for assessing relative crashworthiness. Both the GAO draft report and NHTSA studies have concluded that there are statistically significant reductions in fatality risk for the occupants of vehicles that perform well in NCAP as compared to vehicles that perform poorly. Given these findings, the identified statements need to be consistently structured.

- **Page LET-3, note 1**: FMVSS 208 is only one of numerous standards which require crash tests under the Federal Motor Vehicle Safety Standards (FMVSS).

Now p. 16.

- **Page LET-5:** The statements on the enactment of FMVSS 208 are inaccurate. The passive restraint requirements and the self certification to a 30 mph barrier test became mandatory for model year 1987 passenger cars. This requirement was phased-in, with 10 percent of the passenger cars in model year 1987, 25 percent in model year 1988, 40 percent in model year 1989, and 100 percent in model year 1990. Light trucks, vans, and sport utility vehicles were required to meet the dynamic test requirements for restrained occupants but not necessarily automatic occupant protection, beginning with model year 1992. (See 49 CFR §571.208)

Now p. 17.

- **Page LET-6 and LET-7:** The draft report states that manual belts are not used in compliance tests. This condition only applies to passenger cars and to light trucks certified by the manufacturer as meeting the automatic occupant protection requirement of the standard. Most light trucks are tested with manual belts. Automatic occupant protection for light truck vehicles are required to be phased-in beginning with 20 percent on and after September 1, 1994, and 100 percent on and after September 1, 1997.

  The Hybrid II and Hybrid III crash test dummies are designed to represent the average sized adult male but weigh 164 pounds and 167 pounds respectively, not the 172 pounds identified in the draft report.

Now p. 18.

- **Page LET-8:** The report states that vehicles exhibiting poor performance in NCAP testing are typically tested for compliance in the following year. NCAP vehicles are generally tested early enough in a model year that if the results so warrant, a compliance test can be conducted in that same year.

  During the period between 1979 and 1986, NCAP served as indicant testing for FMVSSs Nos. 212, Windshield Mounting; 219, Windshield Zone Intrusion; and 301, Fuel System Integrity. It did not serve as indicant testing for FMVSS 208 because dynamic testing for this standard was not required during that time.

See comment 11.

- **Page LET-9, paragraph 3:** Manufacturers do not "participate" in NCAP except through an optional program that allows them to request and fund an NCAP test on

6

a specific make and model. NHTSA selects and tests approximately 35 different vehicles each year and releases the data to the public. Most manufacturers have opted to produce vehicles that perform well in NCAP because of perceived or actual market forces; however, it is inaccurate to state that manufacturers must design their vehicles to pass this more stringent test.

Now p. 21.

o **Page LET-10:** The draft report's discussion on compliance testing does not accurately portray the self-certification process. The draft report states, "When the results of a compliance test exceed the limit for any of the measures, the vehicle model being tested may not be sold in the United States." A determination of noncompliance must be made by the manufacturer or NHTSA. Such a determination typically follows an investigation into why the vehicle failed and is accompanied by a recall and remedy campaign for vehicles already sold and those in the manufacturer's vehicle distribution network.

Now p. 20.

o **Page LET-11 and LET-12:** The dummy chest requirement states that the resultant acceleration measured by X, Y, and Z axes accelerometers shall not exceed 60 Gs for a cumulative duration in excess of 3 milliseconds. The statement in the draft report needs clarification to indicate this 3 millisecond duration.

Deleted.

o **Page LET-12, paragraph 1:** The draft report states that the head injury criterion (HIC) was developed by NHTSA. The HIC was developed by the auto industry and adopted by NHTSA. Also, the draft report states, "HIC is measured as a composite...resulting from the head contacting components...." HIC is calculated from the resulting components of acceleration whether or not head contact occurs.

Term changed.

o **Page LET-13, paragraph 1:** The draft report uses the term "chest depression." This should be changed to "chest compression."

Now appendix III.

o **Page LET-14, paragraph 2:** The draft report states that, "The two types of dummies are similar in only a few aspects." GAO does not provide details on what is considered similar aspects. In a 1986 rulemaking, NHTSA allowed manufacturers the option to use either the Hybrid II or the Hybrid III dummy in determining certification to FMVSS 208. This option was allowed since NHTSA concluded from studies at that time that the two dummies were essentially equivalent in their response in the FMVSS 208 crash environment. Directly comparable tests have been conducted on only one passenger car since 1986. In the two comparable NCAP tests, the HIC was 1,018 and chest Gs were 46 for the Hybrid III and 1,063 and 47 for the Hybrid II. Although the results are from only one test, this tends to support NHTSA's original findings. The NCAP report to Congress previously cited provides an extensive analysis and comparison of Hybrid II and Hybrid III.

7

The draft report again uses the "average" male dimensions as "172 pounds," and states that both the Hybrid II and Hybrid III are designed to represent this dimension. As noted previously, the Hybrid II dummy is 164 pounds and the Hybrid III dummy is 167 pounds.

Now p. 27.

o **Page LET-16:** The draft report states that, "One expert explained that the reason so few tests involve Hybrid III dummies is because it is more difficult to pass the compliance test using it." This hypothesis is not supported by data. The NHTSA analysis in the draft report to Congress shows that since the Hybrid III has been used by manufacturers from model year 1990, 71 percent of the passenger cars that used Hybrid II and 70 percent of those that used Hybrid III passed FMVSS 208 requirements in the higher speed NCAP tests. Similarly, for light trucks with belt restrained dummies, 33 percent of those that used Hybrid II and 30 percent of those that used Hybrid III passed FMVSS 208 in NCAP tests.

Now pp. 26-29.

o **Page LET-22, paragraph 2:** The draft report states that, "...it remains true that scores from the Hybrid II may not be directly comparable to Hybrid III scores." As stated in the previous comment, this hypothesis is not supported by data. Further, direct statements on dummy comparisons are not valid because of the influence of numerous vehicle parameters. The only way to determine if the dummies are comparable is to run an experiment in which the dummies are the only variable. NHTSA conducted such tests before accepting the Hybrid III dummy as a FMVSS 208 test device and concluded that the dummies are "equivalent" in measuring frontal crash protection, although the numerical values of the test results may vary, particularly in cars without air bags.

Deleted.

Note 11 is incorrect. Nearly all NCAP testing in 1994, regardless of manufacturer, used the Hybrid III dummy. Of the 78 total seating positions tested, only 5 used the Hybrid II dummy.

Now p. 29.

o **Page LET-22, paragraph 3 and LET-23:** The statement that "NHTSA has recently responded to this threat to the reliability of its crash tests," is incorrect. NHTSA's basis for mandating the Hybrid III is not related to the reliability of its crash tests, but to the consensus that the Hybrid III is an improved anthropometric test device as described in the November 1993, final rule mandating the use of the Hybrid III.

See comment 12.

o **Page LET-28:** The draft report indicates that the mean head injury score decreased from a high of 594 points in 1987 to a low of 374 points in 1993, but concluded without explanation that this is not significant. We disagree. The statement is addressing the mean of the data points and a drop of 220 points, or about 40 percent is significant.

Now pp. 35-36.

o **Page LET-32, paragraph 1:** In the fourth sentence, the draft report offers a theory to explain higher Gs experienced by unbelted dummies restrained with an air bag.

8

We disagree with the explanation given. The difference is probably attributable to more loading of the chest by the air bag over a larger surface area and occurring later in the crash event than a belt restraint.

Now pp. 35-36.

o **Page LET-32:** It is stated that unrestrained dummies have a higher chest G score in crash tests of vehicles equipped with an air bag because "...the bag will forcefully hit him in the chest as it is deploying." In every compliance crash test film reviewed by NHTSA, the air bag is fully deployed prior to it being contacted by the dummy.

Now pp. 37-38.
See comment 2.

o **Page Let-33 through Let-35:** The draft report discusses variation in HIC based on Citation repeatability tests conducted by the agency. It is not appropriate to extrapolate this variation to today's NCAP results. First, as the draft report properly points out, the quality of the manufacturing process has increased significantly, thus reducing the variability between "identical" models. Secondly, the results from the Citation test series cannot be extrapolated to the variance for all test results, because the results are unique to that particular vehicle. At approximately the same time the Citations were tested, Volvo conducted an independent test series showing the test variability in their cars was much less than the Citation experienced. In 1984, NHTSA thoroughly analyzed the data from both of these test series and implemented changes that are intended to further reduce variability related to the testing process. These changes, which are described in a paper published by the Society of Automotive Engineers include more consistent placement of the dummy and the initiation of an instrument and dummy auditing system. NHTSA continues to monitor test results to identify and implement appropriate modifications of the testing procedures with the goal of reducing test variability to the lowest possible levels.

Now p. 38.
See comment 2.

o **Page LET-35:** The draft report assumption as stated in note 17 is not correct and will result in a significantly flawed analysis. Based on recent model years NHTSA found that any variation will be minimal.

Now pp. 39-41.
See comment 6.

o **Page LET-37, paragraph 2:** The new star rating system will eliminate concern about implied accuracy. For example, a vehicle in which the dummy responds with a chest G of 40 can have a HIC ranging from 200 to 550 and still receive a five star rating.

Now p. 42.

o **Page LET-39, paragraph 1:** The draft report states, "The overall purpose of both the compliance and NCAP crash tests is to determine the **relative** crashworthiness, or safety, of passenger vehicles." Only NCAP tests are conducted to determine the relative crashworthiness. Compliance tests are conducted to verify that the vehicle meets the minimum requirements of the safety standard.

Now pp. 42-43.
See comment 13.

o **Page LET-40:** In note 20, the statement "as cars age, their crashworthiness decreases due to wear and tear," is hypothetical and has not been proven.

9

Now pp. 44-47.
See comment 9.

o **Page LET-43, paragraph 1:** We agree with the first sentence in that it cites "...the relatively small size of our usable data base...," in referring to the National Accident Sampling System (NASS) frontal data. The NASS data are especially sparse for crashes where the change in velocity is 35 mph or greater. These are the crash severities at which the NCAP data are most applicable. NHTSA conducted analyses of the NASS files to determine the potential for use in comparing NCAP test results to real-world NASS cases. As stated in the report to Congress, "...the amount of crash information on individual makes and models remains inadequate for studying correlations to NCAP results. The major importance of NASS is the nationally representative detailed information on types and causes of injury, crash speeds, and crash configurations." As a result, the lack of correlation as described in the draft report between NCAP and the NASS data is due primarily to the application of inappropriate methodology and not, as the draft report maintains the limited reliability of the NCAP scores.

See comment 4.

o **Page LET-43, paragraph 2:** In the report to Congress, NHTSA concluded based on extensive analysis that FARS provides adequate data to determine whether the premise of improved safety with lower dummy responses on which both FMVSS 208 and NCAP are based is valid in the spectrum of real-world frontal crash events. However, NHTSA also concluded that the number of applicable FARS cases is not adequate to determine absolute make and model correlations to NCAP test results. NHTSA concluded that the fatality risk reduction for restrained drivers of "good" cars may be as much as a 30 percent lower than the fatality risk of restrained drivers in "poor" cars.

Now p. 5.
See comment 14.

o **Page LET-48:** Please clarify the basis for the draft report's statement that "we found no evidence of a causal linkage between improved crash test scores and declining highway fatality rates." We are unaware of any analysis that has been conducted which could provide the data to establish a linkage between cause and effect. As the draft report states, this type of linkage is almost impossible to determine.

o **Page LET-49, paragraph 1:** We agree with the statement that, "...manufacturers' successful efforts to improve their products' performance in NHTSA crash tests have contributed to improved occupant protection in real world crashes...." This contribution is "quantifiable" relative to fatality reductions when comparing "good" and "poor" performing vehicles.

See comment 6.

o **Page LET-49, paragraph 2:** We agree with the statement that variations such as those used in the draft report's example of 550 versus 700 probably have little effect on fatality reduction. We maintain that such potential variations are insignificant and consumers have received valid and reliable information regarding the relative crashworthiness of vehicles. There is no question that the public is well served by making this information available. In addition, the implementation of the

10

star system for displaying crash test results will eliminate concerns related to minor potential variations.

Now p. 37.

- **Page LET-49, paragraph 3:** The statement that "...cars, as we reported, have become homogeneous in their crashworthiness.", is an incorrect statement as worded. From our research and NCAP testing, we know only that cars have become more homogeneous in their crashworthiness.

Now p. 49.

- **Page LET-51, paragraph 1**: On page LET-45 the draft report states that fatality rates were 22 percent lower comparing the low group to the high group. The converse of this comparison would be 28 percent higher rather than "...22 percent higher...," as used in the second sentence on page LET-51.

Now p. 29.

- **Page APP1-18:** The section titled "Manufacturer's Mandatory Use of the Hybrid III" implies that a manufacturer must use the Hybrid III in its compliance tests. The self-certification process allows the manufacturer to choose the means by which it will certify that its vehicles meet all applicable FMVSS. NHTSA, however, must use the Hybrid III in its FMVSS 208 compliance testing beginning with model year 1998 vehicles.

Now pp. 21-23.

- **Page APP1-21:** Note 20 is incorrect. The "Level of Protection Scale" as shown in the focus group studies was subsequently changed to the star system that uses a combined HIC and chest G function.

Now pp. 42-47.
See comment 15.

- **Page APP2-2, paragraph 2:** The analysis described utilizes logistic regression to determine whether a relationship exists between the likelihood of serious injury/fatality in actual automobile collisions and results from NCAP crash tests. While logistic regression is well suited for this type of analysis, it is important to note that NASS is a complex survey, with stratification on such variables as vehicle age and injury severity. Specifically, the sampling protocol oversamples from the more serious crashes which result in hospitalization or fatality. Therefore, specialized statistical methods need to be employed to achieve statistically valid conclusions. The standard statistical packages such as SAS, SPSS, BMDP, while permitting the use of proper case weights in their analytical procedures, generally do not contain routines that recognize the sample design characteristics of complex survey data such as NASS. The analysis of complex survey data has received much attention recently in the statistical literature. NHTSA employs the SUDAAN (SUrvey DAta ANalysis) programs to conduct its analyses of NASS data.

  The variances and/or standard errors obtained from the use of standard statistical packages underestimate the true variance, since they do not account for the variance due to sampling error. Also, the draft report does not indicate whether the weighted NASS data were used in their analysis; the implication would be that the unweighted cases were used. Without properly weighing the data, each vehicle is

11

treated on an equal basis with all others, with the possibility that injury rates could be biased by a single "unlucky" observation which, when properly weighted, contributes far less to overall information for the aggregated make/model data. In order to fully understand the analysis in the draft report, we recommend that it present greater detail of the analytical methods employed.

Now pp. 44-47.
See comment 4.

- **Page APP2-4:** In any attempt to correlate NCAP test results to NASS data: (a) only frontal collisions with crash severities reasonably close to the NCAP crash severity can be considered, and (b) only crashes with restrained front seat occupants can be used. NHTSA has shown that little frontal crashworthiness difference exists between vehicles, for restrained occupants, in crashes with velocity changes at or below the FMVSS 208 test speed. The main reason for elevating the speed to 35 mph in NCAP was to be able to more readily distinguish any crashworthiness differences. When NASS data are limited to (a) and (b), NHTSA found that the number of remaining NASS cases was insufficient to conduct any type of statistical analysis.

Now pp. 45-46.
See comment 16.

- **Page AAP2-13, Table APP2-8:** In modeling injury risk in two-vehicle collisions, the results of which are presented in this table, we have found that the ratio of the two vehicle weights, rather than the explicit representation of each vehicle's weight, provided far greater explanatory power. This would also be applicable to the remaining models of two-vehicle collisions.

Now pp. 47-49.
See comment 17.

- **Page AAP2-16 through AAP2-21:** Registered vehicles are an inadequate measure of vehicle use, and adjusting fatalities for registered vehicles alone may not provide useful information on vehicle crashworthiness.

See comment 18.

- **Page APP2-16:** Although the draft report's analysis used only HIC scores, we maintain that as a minimum, both HIC and chest Gs should be used in the FARS analysis to define the "best" and "worst" vehicles.

Now pp. 47-49.
See comment 17.

- **Page APP2-17, paragraph 2:** In comparing the fatality rates per registered vehicle to determine differences in crashworthiness risk, it is important to recognize that these rates are not pure measures of crashworthiness, but also include crash involvement risk. For example, the number of fatalities per registered vehicle may differ between two vehicles because: (1) vehicle A may be more crashworthy than vehicle B, everything else being equal, and/or (2) vehicle B may be involved in more potentially fatal crashes, but is just as crashworthy as vehicle A. Specifically, if larger, heavier cars are generally driven by older, more cautious drivers than are lighter, sportier cars, inferences regarding their relative crashworthiness based on observed differences in their fatality rates per registered vehicle may overestimate the true crashworthiness effect. On the other hand, comparisons between midsize, high performance cars and lighter, lower horsepower cars may underestimate the true crashworthiness effect, due to the potentially more aggressive driving behavior

12

of their drivers. It would be very difficult to separate these two effects, especially in the relatively simple statistical models employed in the draft report.

See comment 19.

- **Page APP2-21, third sentence:** This sentence refers to Table APP2.12 for 2 door hatchbacks, pointing out that the poorer performing NCAP cars had lower fatality rates. The data in the table shows that this phenomenon is due to the fact that in 7 of the 10 years there were no fatalities in the worst performing NCAP vehicle, due to a lack of sufficient registered vehicles of that body style meeting the NCAP performance criteria to result in a fatality.

13

The following are GAO's comments on the letter from the Department of Transportation dated July 13, 1994.

## GAO Comments

1. We agree with NHTSA that the terms "reliability" and "validity," as used in the report, refer to their statistical meanings, test repeatability, and predictive validity, respectively. We share the concern that common usage of the terms "could have damaging effects" on NCAP's credibility and mislead a casual user to conclude that the tests have not had positive effects on the crashworthiness of the U.S. passenger car fleet. For this reason, we have modified the subtitle of the report.

NHTSA is correct in asserting that we use the term reliability in its traditional meaning of reproducibility. We disagree with NHTSA's implication that this meaning is not relevant to "the **relative** safety performance of vehicles." Indeed, as our report indicates, the band of uncertainty that surrounds crash test scores (as it does any test results) can affect the relative ranking of vehicles.

2. We recognize that model lines would have NCAP result variations unique to themselves, and the report clearly states this caveat (see p. 38, footnote 5). After completing the Citation experiment, NHTSA made changes in its test procedures to improve their reliability. Unfortunately, no equivalent test of how effective these changes were in reducing the variability between test scores was subsequently performed.

After NHTSA had provided its official comments on our draft report, the agency also provided us with crash test results from automobile manufacturers for model-year 1991 through 1994 vehicles. These data were provided to NHTSA in preparation for tests conducted under NCAP and were results of tests that essentially duplicated the NCAP testing procedure. The agency had previously declined to provide this information because they considered it proprietary. We analyzed these data and have included our findings in the body of the report (see pp. 39-41). Though this information cannot define the boundaries of NCAP reliability, the difference between manufacturer and NCAP results reinforces our conclusion that the reliability of NCAP results is limited.

3. We do not disagree with NHTSA that rigorous protocols for crash testing are followed and that NHTSA verifies the results of the crash test with high-speed film. However, this process merely verifies that the accelerometers placed in the test dummy accurately recorded the data

from the specific trial. It does not address the issue of reliability, which in classic statistical theory holds that test data are reliable if consistent results are obtained through repeated trials of an experiment using specific procedures. In the case of NCAP crash testing, a specified procedure exists, but the model of the test vehicle changes. Unless multiple trials of the same model line are conducted, we cannot determine the reliability of test results.

We are sensitive to the costs the Agency could incur in addressing our recommendation that it update and publish, in clear language, its knowledge of the reliability of NCAP results. For this reason, we suggest that the Agency explore alternative means for accomplishing this goal, in particular by making use of the knowledge base developed by manufacturers. Regardless of the method the Agency uses to address the recommendation, its purpose will not be, as NHTSA suggests, to "enhance the scientific reliability of its data" or "narrow its standard deviation," but to assure American consumers that they are provided with accurate information about the relative crashworthiness of vehicles.

4. We agree that NHTSA has found a statistically significant difference between the fatality risk of belted drivers involved in two-car frontal collisions in cars with "good" NCAP performance and those that were "poor" performers. In our analysis of the fatality rates, though using a different methodology, we found similar results.

The analyses we performed shared a common weakness with NHTSA's; namely, they were both limited to a relatively small proportion of real-world crashes. NHTSA's estimate of reduced fatality risk for better scoring NCAP cars is derived from analyses using only two-car crashes in which both drivers were belted and at least one occupant was killed. These conditions limited their analyses to between 81 and 170 crashes. (NHTSA's database was drawn from the 1979 through 1991 FARS years, which represent between 40,000 and 50,000 highway fatalities annually.)

Our analysis was also limited to NCAP cars involved in fatal accidents with restrained drivers, although we also included single-vehicle crashes and FARS data from 1987 through 1991. These limitations reduced our sample to 884 cars. Both NHTSA and we agree that a statistically significant correlation between NCAP scores and real-world crashes can be found but, to use NHTSA's words, the correlation is "far from perfect."[1] Our analyses

[1]See NHTSA, Correlation of NCAP Performance With Fatality Risk in Actual Head-on Collisions, 1994, p. xviii.

suggested that this correlation derived from the fatality rates of the worst scoring cars, and not from crashworthiness differences among relatively good NCAP performers.

5. We generally agree with NHTSA's comment that the improvement in NCAP scores over time has contributed to an improvement in highway safety. However, many other influences unrelated to crash testing, such as safety belt usage laws, and the toughening of drunk driving laws, have also contributed to this trend.

6. Our report does not address the purported ease of interpretation associated with the new star rating system. However, we did incorporate NHTSA's new reporting system into our analysis of the new data provided us by NHTSA. Our findings provide detail to support the conclusion of our draft report: that a reporting system can be no more reliable than the scoring system on which it is based.

We disagree that the new star rating system "eliminates . . . [the] implied precision" of HIC, chest, and femur scores. It is true that some cars with nonsignificant differences in scores would end up in the same category under the new system, and thus correctly be presented to the public as roughly equal in crashworthiness. However, it is also true that other cars with nonsignificant score differences could be placed in different categories, a scoring artifact that incorrectly implies substantial differences in the relative levels of crash protection provided by the vehicles. For example, while a vehicle with a chest g of 40 and a HIC of 550 will receive 5 stars, one with a chest g of 40 and a HIC of 555 will receive 4 stars. We do not believe that this difference in HIC scores implies an actual crashworthiness difference.

7. NHTSA appears to suggest that it has already complied with this recommendation through the adoption of its new reporting system. We disagree. The new system, while seemingly clear, does not communicate to the public the band of uncertainty associated with star ratings. Our analyses of the manufacturers' NCAP test results suggests that this band is sizable and illustrates its potential effects. However, additional information needs to be collected and analyzed before the precision of crash test results can be adequately defined.

8. The recommendation is to explore alternative methods for determining the crashworthiness of vehicles. We cited computer simulations as one promising avenue to explore. We recognize the limitations of the current

capabilities of computer simulations (the high cost of supercomputers, the complexity of programming, and so on), and we agree with NHTSA that this technology could not replace actual crash tests in the near future. However, NHTSA's comment suggests that it has examined the potential benefits of this rapidly emerging technology and has dismissed them. We believe that it should continue to monitor and periodically reassess them.

It appears to us that as the technology develops and becomes less costly, the potential benefits of such a system extending to a much larger set of crashes than NCAP now considers at some point in the future may outweigh its costs.

Other possible approaches include ones that NHTSA is already considering, such as extending the range of tests to include both side-impact crashes and frontal-offset crashes. While such tests would expand the applicability of NCAP tests to a larger portion of real-world events, they would also substantially increase the costs of the program. This consideration reinforces our belief that the costs and benefits of alternative approaches such as computer modeling need to be revisited regularly over the next decade.

9. We are aware of the limitations of the NASS database, and we agree with NHTSA that the number of cases that mimic the NCAP configuration are few. Indeed, in the 4 years of NASS data we analyzed for the project, only 46 of over 14,000 cases applied to our model most closely resembled the NCAP configuration.

We disagree that the methodology was inappropriate. The models used in the analyses were designed not only to simulate the NCAP conditions, but also to discover the sensitivity of NCAP for predicting other frontal collisions, and thereby maximize the number of frontal crash configurations for which the test was applicable and meaningful. It seems reasonable to expect that, given enough cases, NCAP should predict real-world traffic injuries and fatalities in collisions that essentially duplicate the test conditions; however, given the small number of actual events that apply to this configuration, the meaning of any unweighted statistically significant relationship is questionable.

10. There is no contradiction between these statements. NCAP's relative rankings of different models may be inaccurate in some cases. Nevertheless, it is unlikely that the parallel improvement in NCAP scores and highway safety statistics over the past 15 years is totally coincidental.

11. We agree with NHTSA's comment about manufacturer participation in NCAP, and the language has been changed. With respect to the quasi-regulatory nature of NCAP, manufacturers repeatedly stated that they must design automobiles to meet this test as if it were the standard. In oral commentary on our draft report, one NHTSA official pointed out that almost all passenger cars meet compliance standards in the NCAP test, and that "in effect, it's a de facto standard."

12. The 220-point reduction in the mean HIC falls short of statistical significance (p = .171). This, we believe, is a function of the low number of cases and high variations in the early years of compliance testing. Although the decline is not statistically significant, we would reiterate that, on average, vehicles tested for compliance with FMVSS 208 tend to have HIC and chest acceleration scores that are far below the maximum allowable levels.

13. Although "it has not been proven," it seems reasonable to assume that a 1979 vehicle that was involved in an accident in 1991 no longer had the same level of structural integrity as it did when it was new, owing to the rusting of the frame, for example, or the weakening of welds. It also seems reasonable to assume that a 1979 model-year vehicle would not perform at its full original safety potential in a collision that occurred in 1991.

14. As our report states, and NHTSA cites, no statistical analysis can, by itself, establish cause-and-effect linkage, and we do not demand this result of our analyses.

15. We accepted NHTSA's suggestion and used SUDAAN to perform additional analyses of the NASS data. The results were inconclusive. In one case (two-vehicle collisions), we found a statistically significant relationship between NCAP scores and serious injury. (See table VI.4.) However, this result could easily be spurious since the application of NASS sampling weights (which vary substantially and can be quite large) to the small subset of cases that both fit our criteria and have no missing data can greatly distort the analysis. If, as NHTSA suggests, a subset of NASS data is "insufficient to conduct any type of statistical analyses," applying sampling weights to a nonrandom selection of variously weighted cases is potentially misleading.

16. We are aware that in some analyses, NHTSA has used the combination of subject vehicle weight and its ratio to the other vehicle weight as predictors of injury instead of simply using the weight of the two vehicles.

We did not feel it necessary to reanalyze the data using weight and weight ratio since, as NHTSA has pointed out, they “are mathematically equivalent to the information provided by the two individual vehicle weights.”[2]

17. We used the traditional adjustment, the ratio of fatalities to the number of registered vehicles, to correct for the variations in exposure to accident involvement among the NCAP-tested vehicles. We agree with NHTSA that other factors, such as driver age and driving history, are also important predictors of accident involvement and are not captured by this adjustment. [3] Our goal here, however, was to answer the simple question: Are proportionately more drivers killed in poor scoring NCAP cars than in better scoring cars? Our answer is “yes.”

18. Based on NHTSA’s comment, we converted HIC and chest g scores to the combined injury probability, which forms the basis for NHTSA’s new rating system, and used it as a variable in the analyses conducted and presented in this report.

19. The section is no longer in the report.

[2]“NHTSA, A Collection of Recent Analyses of Vehicle Weight and Safety,” 1991, p. 6.

[3]See Highway Safety: Factors Affecting Involvement in Vehicle Crashes (GAO-PEMD-95-3; Oct. 1994).

# Major Contributors to This Report

## Program Evaluation and Methodology Division

Robert E. White, Assistant Director
David G. Bernet, Project Manager
Edward J. Logsdon, Project Staff
Martin T. Gahart, Project Staff
Dale Harrison, Project Staff
Beverly A. Ross, Project Staff
Venkareddy Chennareddy, Referencer

# Bibliography


Alem, Nabih M., Guy S. Nusholtz, John W. Melvin. "Head and Neck Response to Axial Impacts." Proceedings of the Twenty-Eighth Stapp Car Crash Conference (SAE Paper No. 841667). Warrendale, Pa.: Society of Automotive Engineers, 1984. Pp. 275-82.

Association for the Advancement of Automotive Medicine and the International Research Council on the Biomechanics of Impact. The Biomechanics of Impact and Motor Vehicle Crash Performance: A Global Concern—Conference Materials. Des Plaines, Ill.: AAAM, 1993.

Backaitis, Stanley H., ed. Biomechanics of Impact Injury and Injury Tolerances of the Head-Neck Complex (SAE Book No. PT-93-43). Warrendale, Pa.: SAE, 1993.

Blalock, Hubert M., Jr. Social Statistics. New York: McGraw-Hill, 1979.

Dixon, Camille M. "Automotive Crashworthiness Rating: Legislation and Testing." Proceedings of the Thirtieth Stapp Car Crash Conference (SAE Paper No. 862046). Warrendale, Pa.: SAE, 1986.

Eiband, A.M. Human Tolerance to Rapidly Applied Accelerations: A Summary of the Literature (NASA Memorandum No. 5-19-59E). Washington, D.C.: National Aeronautics and Space Administration, 1959.

Eppinger, Rolf H., and Susan C. Partyka. "Estimating Fatality Reductions With Safety Improvements." Proceedings of the Eighth International Technical Conference on Experimental Safety Vehicles. Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration, 1980. Pp. 432-38.

Foster, J. King, James O. Kortge, and Michael J. Wolanin. "Hybrid II: A Biomechanically-Based Crash Test Dummy." Proceedings of the Twenty-First Stapp Car Crash Conference (SAE Paper No. 770938). Warrendale, Pa.: SAE, 1977. Pp. 975-1014.

Fraser, T.M. Human Response to Sustained Acceleration" (NASA-SP-103). Washington, D.C.: NASA Scientific and Technical Information Division, 1966.

Gadd, C.W. "Criteria for Injury Potential." Impact Acceleration Stress Symposium, 27-29 November 1961, Brooks AFB, Texas (National Research Council Pub. No. 977). Washington, D.C.: National Academy of Sciences, 1962. Pp. 141-44.

Gadd, Charles W. "Use of a Weighted-Impulse Criterion for Estimating Injury Hazard." Proceedings of the Tenth Stapp Car Crash Conference (SAE Paper No. 660793). Warrendale, Pa.: SAE, 1966. Pp. 164-74.

General Motors. MVSS: A Guide To Federal Motor Vehicle Safety Standards & Regulations. Detroit: GM, Environmental Activities, 1989.

Gennarelli, Thomas A. "The State of the Art of Head Injury Biomechanics." Proceedings of the Twenty-Ninth Conference of the American Association for Automotive Medicine. Des Plaines, Ill.: AAAM, 1985. Pp. 447-63.

Gillis, Jack. The Car Book: The Definitive Buyer's Guide to Car Safety, Fuel Economy, Maintenance and More, 1993 ed. New York: Harper Perennial, 1993.

Gurdjian, E.S., et al. "Significance of Relative Movements of Scalp, Skull, and Intracranial Contents During Impact Injury of the Head." Journal of Neurosurgery, 29:1 (1967), 70-72.

Gurdjian, E.S., V.L. Roberts, and L.M. Thomas. "Tolerance Curves of Acceleration and Intracranial Pressure and Protective Index in Experimental Head Injury." Journal of Trauma, 6:5 (1965), 600-04.

Gurdjian, E.S., J.E. Webster, and H.R. Lissner. "Observations on the Mechanism of Brain Concussion, Contusion, and Laceration." Surgery, Gynecology, and Obstetrics, 101:12 (1965), 680-90.

Hirsch, Arthur E., and Rolf H. Eppinger. "Impairment Scaling From the Abbreviated Injury Scale." Proceedings of the Twenty-Eighth Conference of the American Association for Automotive Medicine. Morton Grove, Ill.: AAAM, 1984. Pp. 209-24.

Hodgson, V.R., L.M. Thomas, and P. Prasad. "Testing the Validity and Limitations of the Severity Index." Proceedings of the Fourteenth Stapp Car Crash Conference (SAE Paper No. 700901). Warrendale, Pa.: SAE, 1970.

Hodgson, Voigt, and L.M. Thomas. "Comparison of Head Acceleration Injury Indices in Cadaver Skull Fracture." Proceedings of the Fifteenth

Stapp Car Crash Conference (SAE Paper No. 710854). Warrendale, Pa.: SAE, 1971. Pp. 190-206.

Holbourn, A.H.S., and M.A. Edin. "Mechanics of Head Injuries." Lancet, 245 (1943), 438-41.

Joksch, Hans C. "Velocity Change and Fatality Risk in a Crash—A Rule of Thumb." Accident Analysis and Prevention, 25:1 (1993), 103-04.

Jones, Ian S., and R.A. Whitfield. "Predicting Injury Risk With New Car Assessment Program Crashworthiness Ratings." Accident Analysis and Prevention, 20:6 (1988), 411-19.

Jones, Ian S., R.A. Whitfield, and Diane M. Carroll. "New Car Assessment Program Results and the Risk of Injury in Actual Accidents." Proceedings of the Tenth International Technical Conference on Experimental Safety Vehicles. Washington, D.C.: DOT, NHTSA, 1985. Pp. 371-80.

Koch, M., et al. "Car Model Safety Rating—Further Development Using the Paired Comparison Method." Proceedings of the Eighth International Technical Conference on Experimental Safety Vehicles. Washington, D.C.: DOT, NHTSA, 1980. Pp. 432-38.

Kornhauser, Murray. Structural Effects of Impact. Baltimore: Spartan Books, 1964.

Langwieder, K. "Passenger Injuries in Collisions and Their Relation to General Speed Scale." Proceedings of the Seventeenth Stapp Car Crash Conference. New York: SAE, 1973. Pp. 1-34.

Langwieder, Klaus, Maximilian Danner, and Walter Schmelzing. "Comparison of Passenger Injuries in Frontal Car Collisions With Dummy Loadings in Equivalent Simulations." Proceedings of the Twenty-Third Stapp Car Crash Conference (SAE Paper No. 791009). Warrendale, Pa.: SAE, 1979. Pp. 201-31.

Mackey, John M., and Charles L. Gauthier. Results, Analysis and Conclusions of NHTSA's 35 MPH Frontal Crash Test Repeatability Program (SAE Paper No. 840207). Warrendale, Pa.: SAE, 1984.

McCormick, Earnest James. Human Factors in Engineering and Design. New York: McGraw-Hill, 1982.

Marquardt, James F. "Collision Severity—Measured By V." Proceedings of the Twenty-First Conference of the American Association for Automotive Medicine. Morton Grove, Ill.: AAAM, 1977. Pp.379-90.

Newman, James A. "Head Injury Criteria in Automotive Crash Testing." Proceedings of the Twenty-Fourth Stapp Car Crash Conference (SAE Paper No. 801317). Warrendale, Pa.: SAE, 1980. Pp. 703-47.

Partyka, Susan C. "A Comparison of AIS and ISS Predictions of Fatality on NCSS." Proceedings of the Twenty-Fourth Conference of the American Association for Automotive Medicine. Morton, Grove, Ill.: AAAM, 1980. Pp. 156-69.

Pike, Jeffrey A. Automotive Safety: Anatomy, Injury, Testing and Regulation. Warrendale, Pa.: SAE, 1990.

Prasad, Priya, and Harold J. Mertz. 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.: SAE, 1985.

Salvendy, Gavriel. Handbook of Human Factors. New York: John Wiley & Sons, 1987.

Society of Automotive Engineers. Crash Avoidance SP-544: International Congress & Exposition, 1983 (SAE Pub. No. SP-83-544). Warrendale, Pa.: 1983.

Society of Automotive Engineers. Human Tolerance to Impact Conditions as Related to Motor Vehicle Design (SAE Pub. No. SAE-J885-APR80). Warrendale, Pa.: 1980.

Swearingen, John J. Tolerances of the Human Face to Crash Impact (FAA Pub. No. AM-65-20). Oklahoma City: Federal Aviation Agency Office of Aviation Medicine, 1963.

Thomas, C., et al. "Crashworthiness Rating System and Accident Data Convergences and Divergences." Advances in Belt Restraint Systems: Design, Performance and Usage (SAE Paper No. 840200). Warrendale, Pa.: SAE, 1984.

U.S. Department of Transportation. Appendix AO (A-Optional): Part 572E Dummy Performance Calibration Test Procedure. Washington, D.C.: NHTSA, 1993.

U.S. DOT. A Collection of Recent Analyses of Vehicle Weight and Safety (DOT-HS-807-677). Washington, D.C.: NHTSA, 1991.

U.S. DOT. Correlation of NCAP Performance With Fatality Risk in Actual Head-On Collisions (DOT-HS-808-061). Washington, D.C.: NHTSA, 1994.

U.S. DOT. Federal Motor Vehicle Safety Standards and Regulations (DOT-HS-805-674). Washington, D.C.: NHTSA, 1989.

U.S. DOT. Head and Neck Injury Criteria: A Consensus Workshop (DOT-HS-806-434). Washington, D.C.: NHTSA, 1983.

U.S. DOT. Laboratory Indicant Test Procedure: New Car Assessment Program. Washington, D.C.: NHTSA, 1990.

U.S. DOT. Laboratory Test Procedure for: FMVSS 208, Occupant Crash Protection; FMVSS 212, Windshield Mounting; FMVSS 219, Windshield Zone Intrusion; and FMVSS 301, Fuel System Integrity. Washington, D.C.: NHTSA, 1993.

U.S. DOT. New Car Assessment Program: Plan for Responding to the Fiscal Year 1992 Congressional Directives. Washington, D.C.: NHTSA, 1992.

U.S. DOT. New Car Assessment Program: Response to the NCAP Fiscal Year 1992 Congressional Requirements—Report to Congress. Washington, D.C.: NHTSA, 1993.

U.S. DOT. The New Car Assessment Program—Status and Effect. Washington, D.C.: NHTSA, 1982.

U.S. DOT. 1991 Traffic Fatalities Preliminary Report. Washington, D.C.: NHTSA, 1992.

U.S. DOT, National Transportation Safety Board. Safety Effectiveness Evaluation of the National Highway Traffic Safety Administration's Rule-Making Process. Vol. 11: Case History of Federal Motor Vehicle Safety Standard 208: Occupant Crash Protection (NTSB-SEE-79-5). Washington, D.C.: U.S. Government Printing Office, 1979.

Unterharnscheidt, F.J. "Translational Versus Rotational Acceleration—Animal Experiments With Measured Input." Proceedings of the Fifteenth Stapp Car Crash Conference (SAE Paper No. 710880). Warrendale, Pa.: SAE, 1971. Pp. 767-70.

Versace, John. "A Review of the Severity Index." Proceedings of the Fifteenth Stapp Car Crash Conference (SAE Paper No. 710881). Warrendale, Pa.: SAE, 1971. Pp. 771-96.

## Ordering Information

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

Orders by mail:

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

or visit:

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

Orders may also be placed by calling (202) 512-6000
or by using fax number (301) 258-4066, or TDD (301) 413-0006.

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

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

Bulk Mail
Postage & Fees Paid
GAO
Permit No. G100

Official Business
Penalty for Private Use $300

Address Correction Requested