TRANSPORTATION RESEARCH RECORD 844

Automotive Technology, Information Needs of Highway Users, and Promotion of Safety Belt Usage

TRB
TRANSPORTATION RESEARCH BOARD
NATIONAL ACADEMY OF SCIENCES

The **Transportation Research Record** series consists of collections of papers in a given subject. Most of the papers in a Transportation Research Record were originally prepared for presentation at a TRB Annual Meeting. All papers (both Annual Meeting papers and those submitted solely for publication) have been reviewed and accepted for publication by TRB's peer review process according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The views expressed in these papers are those of the authors and do not necessarily reflect those of the sponsoring committee, the Transportation Research Board, the National Academy of Sciences, or the sponsors of TRB activities.

Transportation Research Records are issued irregularly; approximately 50 are released each year. Each is classified according to the modes and subject areas dealt with in the individual papers it contains. TRB publications are available on direct order from TRB, or they may be obtained on a regular basis through organizational or individual affiliation with TRB. Affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, DC 20418.

TRANSPORTATION RESEARCH RECORD 844

Automotive Technology, Information Needs of Highway Users, and Promotion of Safety Belt Usage

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1982

Transportation Research Record 844
Price $9.20
Edited for TRB by Scott C. Herman

mode
1 highway transportation

subject areas
51 transportation safety
52 human factors
53 vehicle characteristics

Library of Congress Cataloging in Publication Data
National Research Council (U.S.). Transportation Research Board.
Automotive technology, information needs of highway users, and promotion of safety belt usage.

(Transportation research record; 844)
Reports presented at the 61st annual meeting of the Transportation Research Board.
1. Automobiles–Safety measures–Addresses, essays, lectures.
I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 844 [TL242] 380.5s [363.1'256] 82-14106
ISBN 0-309-03322-5 ISSN 0361-1981

Contents

Authors of the Papers in This Record

Abramson, Paul, KLD Associates, Inc., 300 Broadway, Huntington Station, NY 11746
Barbaresso, James C., Oakland County Road Commission, 31001 Lahser Road, Birmingham, MI 48010
Brittain, Douglas B., Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139
Chase, Andrew, Booz, Allen and Hamilton, 4330 East West Highway, Bethesda, MD 20814
Cohen, Jay W., KLD Associates, 300 Broadway, Huntington Station, NY 11746
Collin, Lars Th, Department of Internal Combustion Engineering, Chalmers University of Technology, S-412 96, Göteborg, Sweden
Duda, James L., U.S. Department of Transportation, Code DPA-26, RSPA 400 7th Street, S.W., Washington, DC 20590
Flak, Mark A., Progressive Consultants Corporation, 17320 W. Eight Mile Road, Southfield, MI 48075
Green, Robert N., Immediate Past President, American Association of Automotive Medicine, and University of Western Ontario, London, Ontario, Canada
Hanscom, Fred R., Transportation Research Corporation, 27110 Ridge Road, Haymarket, VA 22069
Heathington, K.W., Transportation Center, University of Tennessee, Knoxville, TN 37916
McKnight, James A., National Public Services Research Institute, 123 North Pitt Street, Alexandria, VA 22314
McNees, Roger W., Texas Transportation Institute, Texas A&M University, College Station, TX 77843
McPherson, Kenard, National Public Services Research Institute, 123 North Pitt Street, Alexandria, VA 22314
Messer, Carroll J., Texas Transportation Institute, Texas A&M University, College Station, TX 77843
Parsonson, Peter S., School of Civil Engineering, Georgia Institute of Technology, Atlanta, GA 30332
Perry, Randy L., College of Engineering, Tennessee Technological University, Cooksville, TN 38507
Philpot, John W., Department of Statistics, University of Tennessee, Knoxville, TN 37916
Pulling, Nathaniel H., Liberty Mutual Insurance Company, 71 Frankland Road, Hopkinton, MA 01748
Rinalducci, Edward J., School of Civil Engineering, Georgia Institute of Technology, Atlanta, GA 30332
Schroer, Bernard J., Kenneth E. Johnson Environmental and Energy Center, University of Alabama in Huntsville, Huntsville, AL 35899
Sharpe, Gilbert S., Ontario Ministry of Health, and Faculty of Health Services, McMaster University, Hamilton, Ontario, Canada
Sheffi, Yosef, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139
Stein, Howard, KLD Associates, Inc., 300 Broadway, Huntington Station, NY 11746
Vaillancourt, Donald R., Liberty Mutual Insurance Company, 71 Frankland Road, Hopkinton, MA 01748
Werner, John V., State Farm Insurance Companies, Bloomington, IL 61701

Transportation R&D Technology: The Diagnostic Motor Vehicle Information Concept, 1969-1981

JAMES L. DUDA

An overview of the role of safety and economics of the diagnostic motor vehicle inspection (DMVI) concept is presented. The purpose of providing diagnostic inspection separate from repair was to provide information to automobile owners and operators about their vehicles. They would use this information as a basis for authorizing only those repairs that were really needed and thus avoid authorizing unnecessary or fraudulent repairs. End users would also benefit from reduced life-cycle costs, reduced pollution, and increased safety. However, the safety focus of the program has led to a DMVI facility configuration that is uneconomical. These factors are discussed and a systems analysis methodology is suggested that can be used to optimize the design of the DMVI facility. The methodology applies the technique of maximizing the marginal effectiveness of available resources. These resources consist of the appropriate equipment and labor complements that support specified levels of diagnostic inspection.

The number one consumer complaint of the past decade has been about the automobile repair industry. Testimony in Senate hearings by the Federal Trade Commission indicated that consumer complaints related to the automobile stood at 30 percent compared with 7 percent for all consumer products (Automotive Repair Industry Hearings before the Subcommittee on Antitrust and Monopoly of the Committee on the Judiciary, U.S. Senate, 90th and 91st Congresses, Parts 1-6, July 1972). They also reported that automobile-related complaints take longer to settle than complaints for other consumer products. The federal government reported that end users spend from $42 to $50 billion/year on the repair and maintenance of their automobiles. Of this, several studies estimated that from $13 to $17 billion is spent on repairs that were unnecessary, improperly done, or not done at all (1). The economic, social, and political implications of this situation are wide ranging and complex.

The Motor Vehicle Information and Cost Savings Act of 1972 (P.L. 92-513) represented the federal government's response to end-user concerns about the high cost of automotive maintenance and repair and their dissatisfaction with the repair industry. The Act was based on information obtained, in part, from an investigation started in 1968 by the Antitrust and Monopoly Subcommittee of the Senate Judiciary Committee. It was a four-year investigation that heard from dozens of witnesses who provided more than 4000 pages of testimony (see Legislative History of P.L. 92-513, p. 3962). In addition, the staff of the Senate Judiciary Committee and the Subcommittee on Antitrust and Monopoly collected more than 60 000 exhibits and interviewed several hundred individuals who were not witnesses in the formal hearings. Testimony and evidence was received from every segment of the automobile industry, from manufacturers to independent service establishments. The exhibits included thousands of letters from irate motorists. The Act was based on a great deal of information that documented the enormity and complexity of the repair and consumer problem. The Act was passed by Congress in 1972 in response to the great tide of end-user complaints in the late 1960s.

The centerpiece of this legislation was embodied in Title III, Diagnostic Demonstration Projects, which called for motor vehicle diagnostic inspection demonstration projects to be conducted by the National Highway Traffic Safety Administration (NHTSA). There were two critical aspects of the diagnostic concept as envisioned. First, the vehicle was to be diagnosed by an independent inspection facility with no vested interest in automobile repairs. Second, the vehicle owner was to have the option of taking the vehicle to a repair facility of his or her choice with full assurance that the repairs would be performed properly and at a reasonable cost.

Although in a complex scenario that involved manufacturers, the repair industry, parts suppliers, equipment manufacturers, and automobiles of different makes, models, and years of manufacture in an arena of steady technological and regulatory change, it was believed that it was practical to provide diagnostic information on the condition of the automobile to the end user. Such diagnostic information would help the end user avoid authorizing unnecessary repairs. In this way, a diagnostic facility would sell to the end user information on the condition of his or her vehicle. This information would then be used to authorize only those repairs that were truly needed and thus avoid questionable expenditures.

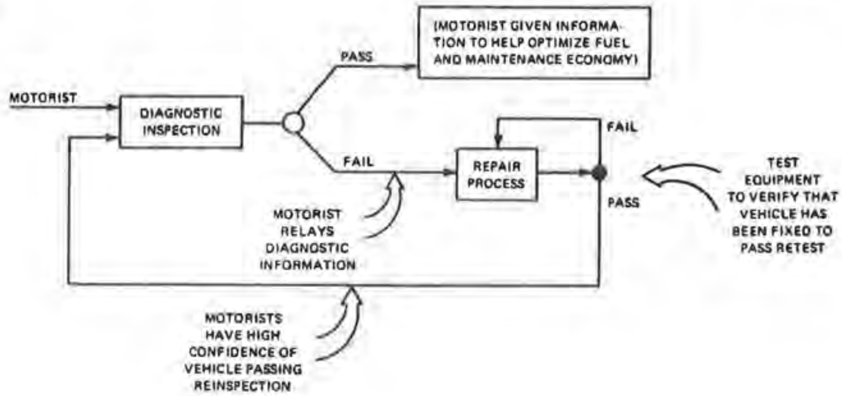
DMVI CONCEPT

As envisioned in the legislative history, the ideal high-volume diagnostic inspection station would supplement the existing automotive repair and service industry. The quality of the process would thereby be improved and the overall costs would perhaps be reduced. As shown in Figure 1, the availability of diagnostic motor vehicle inspection (DMVI) facilities would encourage separation of the diagnostic and repair functions that currently are not now separated in the service industry (2).

Ideally, a diagnostic inspection facility would use standardized and highly automated inspection equipment and data-handling techniques to pinpoint the vehicle components that caused failure of the safety, emissions, noise, or fuel-efficiency inspection standards. The motorist would then take his or her vehicle to a repair establishment and relay instructions to the mechanic concerning the necessary work. On completion of the work, the mechanic would have access in his or her own shop to suitable diagnostic equipment and/or procedures to check the results of the work, consistent with inspection standards. The end user would then be able to return to the inspection facility with confidence that his or her car would pass the motor vehicle reinspection. Through such a system, the motorist would be spared repeat trips between the inspection facility and a repair establishment trying to pinpoint the vehicle's malfunction and having it repaired adequately.

Diagnostic facilities in the late 1960s and early 1970s consisted of certain basic equipment that included a lift, engine analyzer, dynamometer, alignment tester, an assortment of hand-held tools, and an emissions analyzer. Inspections then, as now, consisted of visual checks, measurements, and automatic evaluations. The information was passed to the end user by means of an oral discussion and a piece of paper with varying amounts of relevant

Figure 1. DMVI concept.



diagnostic information. In some cases, the facility that performed the diagnosis also performed limited repairs and adjustments. The facility personnel also served as consultants to the repair industry and the end user. The facility personnel assisted the end user to perform his or her own repairs and reinspected the vehicle after it was repaired to ascertain that the repair was performed and to evaluate the quality of the repair. Figure 2 (3) presents a typical lane-type diagnostic inspection facility.

DMVI, 1969-1981

The configuration, equipment and labor complement, and diagnostic inspections recommended and in practice today are not fundamentally different from those that existed 12 years ago. Today's configuration of lanes and/or bays is essentially the same for brakes, alignment, body, lights, suspension, and exhaust system but with upgraded electronic equipment for engine analysis and emissions inspections, increasing numbers of smaller cars, diesels, and vehicles with front-wheel drive.

Another facet that has not changed in the past 12 years is the fact that facilities, as configured, are marginally economically viable. The economic success of these diagnostic facilities has been limited and temporary. Commercial diagnostic inspection fees are based on the estimated labor time required for the diagnosis plus, in some cases, an equipment amortization charge. In 1980 dollars, the commercial diagnostic inspection fee is on the order of $25, and it is not sufficient to operate the facility, pay all labor costs, and make a reasonable profit. In the few commercial facilities still operating, the diagnostic fee is considered a loss-leader that will draw repair work to the shop. Commercial facilities have reported that at least 50-75 percent of inspection customers usually request that repair work be performed by the diagnosing facility. Table 1 (4) lists diagnostic inspection fees reported by commercial establishments.

Not-for-profit facilities are subsidized by their membership; they operate at a loss or just break even. They report that patronage drops off after an initial spurt of interest. Occasional publicity campaigns are needed to revive interest. Without performing repair services, the diagnostic facilities are not economically viable.

To date, only a few inspection facilities have approached the ideal system described earlier. No state has implemented such a diagnostic inspection system. Also, many vehicles are not subject to any kind of mandatory periodic motor vehicle inspection (PMVI). We have reached a crisis in the DMVI concept as configured. Before the DMVI concept is put to rest, however, the following questions need to be answered.

SAFETY FOCUS

What is the objective of DMVI? The symptom of a problem is end-user dissatisfaction with the repair industry. Information on the condition of the vehicle was believed to be the answer. However, the information was to be based on the safety criticality and emissions of the vehicles; minimum cost and fuel economy were to be by-products of this safety focus. Safety is the primary role of NHTSA; consequently, there was a safety focus on the DMVI concept.

The fact that mechanical defects are frequently a contributing cause to about 10 percent of all traf-

Figure 2. Hypothetical diagnostic inspection lane.

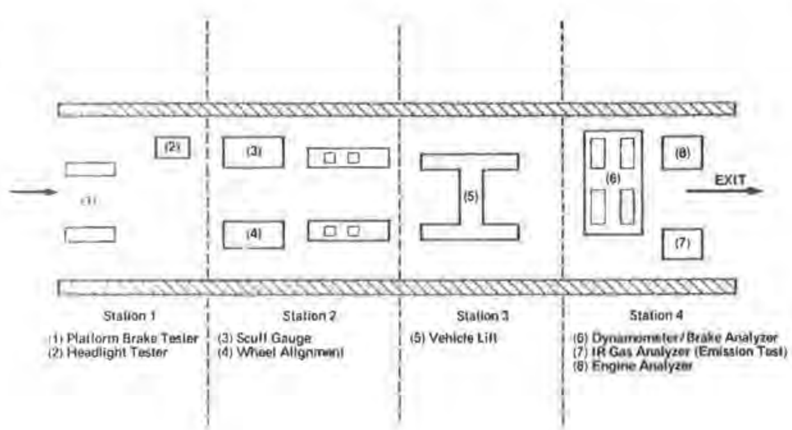


Table 1. Commercial diagnostic inspection fees.

Commercial Diagnostic Inspection Facility	Total Inspection Process Time (min)	Inspection Fee (1980 $)	Estimated Hourly Fee ($)
Auto Lab	45	24.95	33.27
Automotive Evaluation Center	60	21.00	21.00
Automotive Performance Specialists	52	30.00	34.62
Avocation, Ltd.	75	28.50	22.80
Call Carl	58	21.95	22.71
J.C. Penney	73	20.88	17.16
Montgomery Ward	60	18.95	18.95
Avg	58.5	22.95	23.54

fic accidents (rather than a sole cause) complicates the task of identifying the impacts of demonstration experiments on traffic safety (5). Furthermore, 90 percent of all accidents are caused by the driver--thousands yearly kill themselves and others through drunk driving, a majority shun the use of seat belts, and most states do not have even the most rudimentory form of PMVI. Consequently, programs aimed at getting people to wear seat belts, enforcement of the 55-mph speed limit, more strict laws that govern the driving habits of teenage drivers, and getting drunk drivers off the road may have a great deal more potential for improving safety than motor vehicle safety inspection.

This focus on safety may have placed constraints on the solution, perhaps to the detriment of the original intent of the legislation, i.e., to assist the end user to avoid excessive repair costs. On the one hand, a well-maintained automobile means a safe automobile; on the other hand, "too much safety" can lead to excessive maintenance.

Suppose we set aside the safety focus of the current DMVI program and focus instead on determining what form of DMVI can help the end user minimize the cost of operating and maintaining his or her automobile by using the DMVI concept. What impact would this change in emphasis have on the configuration of the diagnostic facility? How might this change the equipment and labor complement of the DMVI facility and the total cost to the end user of maintaining his or her automobile?

ECONOMIC FOCUS

Perhaps the question should be, What level of DMVI will save the end user money, provide a profit to entrepreneurs, and yield a vehicle that is safer, more fuel efficient, and with lower emissions?

Having identified the objective as saving the end user money with safety, fuel economy, and minimum emissions as the by-products of an economical and reliable automobile, the next step might be to approach the design of a DMVI facility from this point of view. The following questions should be answered. First take a look at where the end user is spending most of his or her after-market repair dollars. It may be that these components may offer the greatest potential for savings. The following lists the major automotive subsystems for end-user expenditures (6-9):

Item	Annual Expenditures (%)
Brakes	22.5
Tires and wheels	21.9
Engine	11.4
Suspension	11.1
Under the hood	9.0
Exhaust	5.8
Alignment	4.4
Steering	3.7
Electrical	3.2
Lighting	2.6
Body	2.1
Other	2.3

The first step is to estimate the fraction of consumer expenditures in the automobile after-market by item; that is, estimate the market share by automotive component. Of particular interest are high expenditure items. For example, if $50 billion were spent last year by end users for the above items, then 22.5 percent of $50 billion was spent on the brakes subsystem, or $11.25 billion.

The next step would be to estimate or calculate the amount for each item that is considered to be a questionable expenditure. For example, if $11.25 billion was spent on brakes and one-third of that is considered to be possibly misspent, then the questionable expenditures for brakes is one-third of $11.25 billion, or $3.75 billion/year.

However, not all of the $3.75 billion/year of questionable expenditures can be saved by a DMVI program. A certain amount of the questionable expenditures may be due to prudent repair practice, such as the replacement of wheel cylinders; some may be due to the desire or convenience of the owner or operator of the vehicle, and so on. Some analysis will be required to estimate what percentage of the questionable expenditures may be recaptured for each subsystem. For example, for brakes, suppose that 30 percent of the questionable expenditures can be avoided by an effective DMVI program; that would amount to $1.125 billion. If this $1.125 billion applies to a national fleet of 100 million passenger cars, then this would amount to $11.25/vehicle/year for the brakes subsystem.

A similar procedure would be applied to each subsystem to determine the percentage of questionable expenditures and the percentage that might be avoided by a DMVI. The high-ranking subsystems would be the focus of the DMVI to provide the end user with the greatest service with the objective of saving him or her money on the maintenance of their automobiles (see Table 2).

MAXIMIZING MARGINAL EFFECTIVENESS

A DMVI facility would then be configured by using the above information and applying the incremental effectiveness technique to maximize the marginal effectiveness. [Several applications of this concept can be found in Rudwick (10).]

Applying the technique entails the application of a very simple rule. The rule is that resources should be allocated to provide the maximum increase in effectiveness (or return, or benefits) per unit resource used. In this problem the marginal effectiveness might be measured by the increase in worth of the next level of DMVI. The unit of resources is the equipment and labor in the DMVI for a particular level of inspection. For example, for the brakes subsystem, level 1 might be a platform test, level 2 might be a platform test plus wheel pull, etc. The effectiveness of any level might be expressed by

$$E_i = nC_i + dD_i + (t-d)A_i \quad (1)$$

where

- i = level or depth of inspection,
- n = number of vehicles,

Table 2. Costs of repairs by subsystem.

	Cost ($ billions)		
Item	Annual Expenditures	Questionable Expenditures	Capture
Brakes	11.25	3.75	1.08
Tires and wheels	10.95	2.40	0.50
Engine	5.70	1.50	0.20
Suspension	5.55	3.25	0.94
Under the hood	4.50	0.80	0.07
Exhaust	2.90	0.40	0.20
Alignment	2.20	0.50	0.30
Steering	1.85	0.30	0.23
Electrical	1.60	0.20	0.14
Lighting	1.30	0.40	0.30
Body	1.05	0.15	0.08
Other	1.15	0.08	0.08
Total	50.00	13.80	4.12

d = number of discovered defects,
t = total number of defects per vehicle,
Ci = cost of DMVI per vehicle,
Di = average repair costs for discovered defects per vehicle, and
Ai = average repair costs for undiscovered defects per vehicle.

The incremental or marginal effectiveness, then, would be

$$\Delta Ei = Ei{+}i - Ei \qquad (2)$$

The baseline effectiveness, in this case, would be the situation with no inspection at all, which is level zero:

$$\begin{aligned} Eo &= 0\ Co + dDo + (t-d)Ao \\ &= dDo + (t-d)Ao \end{aligned} \qquad (3)$$

The application of the concept of maximizing the marginal effectiveness when assigning resources is described below.

Consider a structure that indicates the incremental effectiveness obtained by the assignment of the nth complement of equipment and labor to the subsystem or component that provides the next greatest return. Such a compilation will be the basis for applying the key decision rule to be followed in allocating resources; that is, always assign the next equipment or labor complement to that subsystem or component that will yield the highest marginal effectiveness of all of the assignment choices available. Thus, while there are many possible choices involved in the first allocation decision (i.e., assign the first equipment or labor complement to any of the many possible items), the highest marginal effectiveness is obtained by assigning the first complement to the first inspection item.

Hence, decision one will consist of allocating the first equipment or labor complement to the subsystem or component that will provide the greatest return to the end user and patron of the DMVI facility. This procedure can be continued as long as there are additional inspection equipment and labor complements to be allocated and their allocation does not exceed the economic feasibility constraints that provide economic incentives to users and returns a profit to the investors in the facilities.

The application of the concept of maximizing marginal effectiveness will do the following:

1. Establish the economic feasibility of a commercial DMVI,
2. Identify the key variables and interrelations,
3. Show whether an economic focus provides attractive safety benefits, and
4. Provide policy information required for consideration of DMVI by state and local governments.

I believe that data are available from NHTSA's Special Project and National Accident Sampling System, Hunter's Service Job Analysis, Chilton's, etc., to do a reasonable job of estimating the average repair costs for discovered and undiscovered defects. Economic data for the baseline case might be gleaned by comparing normalized area automobile after-market data for states with and without motor vehicle inspection. In a similar fashion, the total number of defects for a group can be estimated as well as the probability of detection at a specified level of inspection. Maybe the economic model should be restructured and parametric analyses performed to examine these factors and settle once and for all the issue of the sound and promising concept of a DMVI.

REFERENCES

1. Public and Private Coordination Needed If Auto Repair Problems Are to be Reduced. U.S. General Accounting Office, Rept. HRS-80-30, Jan. 11, 1980.
2. Evaluation of Diagnostic Analysis and Test Equipment for Small Automotive Repair Establishments--A Report to the Congress. National Highway Traffic Safety Administration, U.S. Department of Transportation, July 1978.
3. J.J. Innes and L.E. Eder. Motor Vehicle Diagnostic Inspection Demonstration Program--Summary Report. National Highway Traffic Safety Administration, U.S. Department of Transportation, Oct. 1977.
4. Booz-Allen and Hamilton. Characteristics of Motor Vehicle Diagnostic Inspection Facilities, Task 1 Report. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DTN H22-80-C-05018, Dec. 1980.
5. Institute for Research on Public Safety. A Study to Determine the Relationship Between Vehicle Defects and Failures and Vehicle Crashes. U.S. Department of Transportation, Rept. DOT-HS-800 902, Aug. 1973.
6. Computer Sciences Corporation. Program Evaluation Support for the Motor Vehicle Inspection Demonstration Program, In-Progress Evaluation Report No. 11. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT/NHTSA-HS-5-01036, April 1976.
7. E.M. Bonkoff. Survival in the Automotive Aftermarket. St. Albans Aftermarket Research Group, 1976.
8. The 1975 Service Job Analysis. Hunter Publishing Co., Chicago, 1976.
9. L. Eder. The National Vehicle Repair Cost Summary. National Highway Traffic Safety Administration, U.S. Department of Transportation, Sept. 1973.
10. B.H. Rudwick. Systems Analysis for Effective Planning: Principles and Cases. Wiley, New York, 1969, pp. 174-180 and 284-287.

Publication of this paper sponsored by Committee on Highway Vehicle Research.

Evaluation of Magnitude of Unnecessary Automobile Repairs

BERNARD J. SCHROER

The results of a four-year study to evaluate the magnitude of unnecessary automobile repairs are summarized. The sources of data were the Alabama Motor Vehicle Diagnostic Inspection Demonstration Project, the Missouri and California American Automobile Association diagnostic centers, and a seven-city undercover survey. The results indicate that users of diagnostic centers may experience a lower unnecessary repair rate and that the repair industry's knowledge of the after-repair inspection may have an effect on the quality of repairs.

The cost of owning and operating automobiles in the United States comprises the fourth largest item of personal expenditure--approximately $158 billion annually (1). The distribution of these expenditures is given in Figure 1 (2) (adjusted for consumer price index for transportation for 1977). Twenty-two percent, or $35 billion, is for maintenance and repairs. Light-truck expenditures and the value of economic loss due to inadequate repairs, accidents, wasted fuel, pollution, and reduced vehicle life amount to an additional $15 billion annually. When these expenditures are added to the cost of repairs and maintenance, this totals some $50 billion annually.

There has been much discussion and debate on the costs of maintaining and repairing an automobile and on how much of this $50 billion could have been saved. As a result, there has been much criticism of the repair industry. Beginning in 1968, the Senate Subcommittee on Antitrust and Monopoly began a four-year investigation of the automobile repair industry. These Senate hearings disclosed major areas where multibillion dollar economic losses occur to motorists. Foremost was the cost of unnecessary and unsatisfactory repairs. Testimony was given that the consumer loss may exceed $8-$10 billion annually. If this figure is adjusted for inflation and the increase in the vehicle fleet, the $8-$10 billion would exceed $20 billion (3). Testimony of other areas of consumer loss included the enormous damage suffered by vehicles in very low-speed crashes, used cars that had odometers turned back to enhance their value, and the economic losses that result from stolen vehicles.

Studies in eight states between 1973 and 1975, in which 200 vehicles with known faults were taken to repair shops, showed 40 percent of the shops charging for unnecessary repairs and 10 percent of them charging for work not performed. In addition, a survey of owner knowledge made by the National Highway Traffic Safety Administration (NHTSA) showed that close to half the vehicle owners lacked the rudimentary knowledge needed for correctly purchasing routine maintenance and repairs.

A survey published in the Harvard Business Review showed 35 percent of the respondents had recent complaints about faulty or unneeded automobile repairs and that 50 percent of owner complaints about repair quality are not satisfactorily resolved. Consumer complaint files from states and business organizations as well as other surveys provided similar data.

A study by NHTSA (3) came to the conclusion that approximately 40 percent ($20 billion) of the costs associated with automobile repairs was wasted. The NHTSA-estimated distribution of these consumer losses is given in Figure 2 (3).

Even though there has been this discussion and debate on the automobile repair problem, very few data have been collected that quantify the magnitude of the problem. The University of Alabama in Huntsville has conducted five studies funded by NHTSA (4-6), the Federal Trade Commission (FTC) (7), and the U.S. Department of Transportation (DOT) (8), which attempt to quantify the magnitude of unnecessary and unsatisfactory automobile repairs. This report compares one of the parameters common to each of these studies: the magnitude of the rate of unnecessary repairs.

DATA SOURCES

Automobile repair costs were collected and analyzed from four sources: the Alabama Diagnostic Inspection Demonstration Project, the Missouri American Automobile Association (AAA) diagnostic center, the California AAA diagnostic center, and an undercover survey conducted in seven cities throughout the country (5-8). The following sections briefly discuss each of these data sources.

Alabama Diagnostic Center

The results of the Senate hearings in the late 1960s were the justification for the passage of the Motor Vehicle Information and Cost Savings Act (P.L. 92-513) in 1972. Title III of the Act authorized the Secretary of Transportation to establish a number of motor vehicle diagnostic inspection and testing centers throughout the country. The objective of the program was to provide for the accumulation of data to determine if diagnostic inspections are cost effective in that public benefits would exceed program costs. Specific types of data collected by the inspection centers included vehicle outages, exhaust emission rates, repair costs, facility operation and staffing requirements, vehicle-in-use standards and feasible reject levels, equipment reliability and interchangeability, and the capability of the repair industry to correct diagnosed deficiencies.

The Alabama Motor Vehicle Diagnostic Inspection Demonstration Project (Auto Check) was one of five diagnostic centers established under the Act. The Auto Check facility is located on the campus of the University of Alabama in Huntsville. Under the initial program, from October 1974 through June 1976, only selected 1968-1973 vehicles were inspected. Since July 1976, the center has been inspecting all model years. To date, more than 19 000 vehicles have received more than 32 000 inspections.

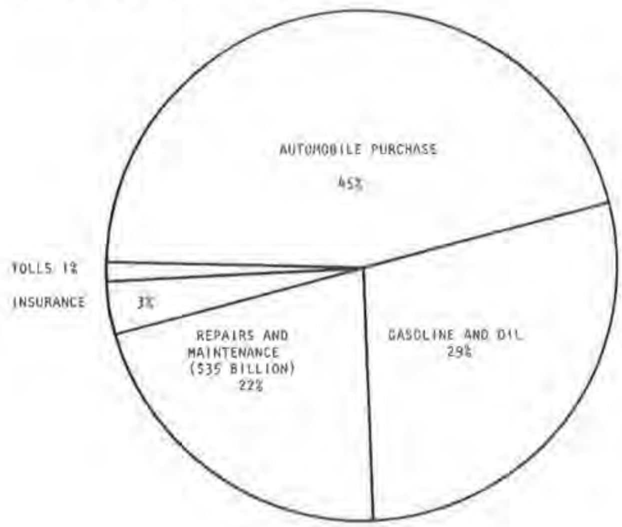
Federal funding for automobile inspections at the center ceased in October 1977. Since then, the center's inspection program has been supported by university funds and by a $10.00 inspection fee. Currently, only one of the three lanes is being maintained and is staffed to perform 10 inspections/day.

Each vehicle is given a thorough diagnostic inspection. After the inspection, the motorist is counseled concerning the condition of the vehicle. During the counseling the motorist is requested to have the necessary repairs performed and then to return to Auto Check for a repair inspection. In addition, the motorist is also requested to retain all repair receipts and to make these receipts available to Auto Check.

Beginning in 1977, the counselor gave all participants a prescription form that gave participants specific repair instructions to convey to the repair facility. Two prescription forms were used. One form was for engine-related outages while the second form was for brake, tire, steering, suspension, and wheel-alignment outages. The forms have a priority column where the counselor indicates the relative importance of each repair.

A sample of cars was selected from the Alabama center that failed the brakes, emission, suspension, steering, or alignment system; returned for an after-repair inspection; and provided the corresponding repair receipts.

Figure 1. How the automobile dollar is spent.



Missouri Diagnostic Center

The Missouri AAA diagnostic center is located at 3925 Lindell Boulevard in St. Louis. The center was originally established as a service to AAA members, although nonmembers may also use the center.

The center opened in the fall of 1967. The center occupies 10 000 ft^2 and is equipped with the latest equipment for static and dynamic analysis. Four types of inspections are performed: a complete diagnostic inspection at a cost of $30.00 for members and $40.00 for nonmembers, a system and component inspection, an after-repair inspection at a cost of $1.00, and the state motor vehicle safety inspection, required annually by Missouri. Since it opened, the facility has inspected more than 120 000 automobiles.

A sample of 444 cars was selected from the Missouri center that met the following conditions: inspected in 1978, failed the brakes or emissions inspection, returned for an after-repair inspection, and provided the corresponding brake or emissions repair receipts.

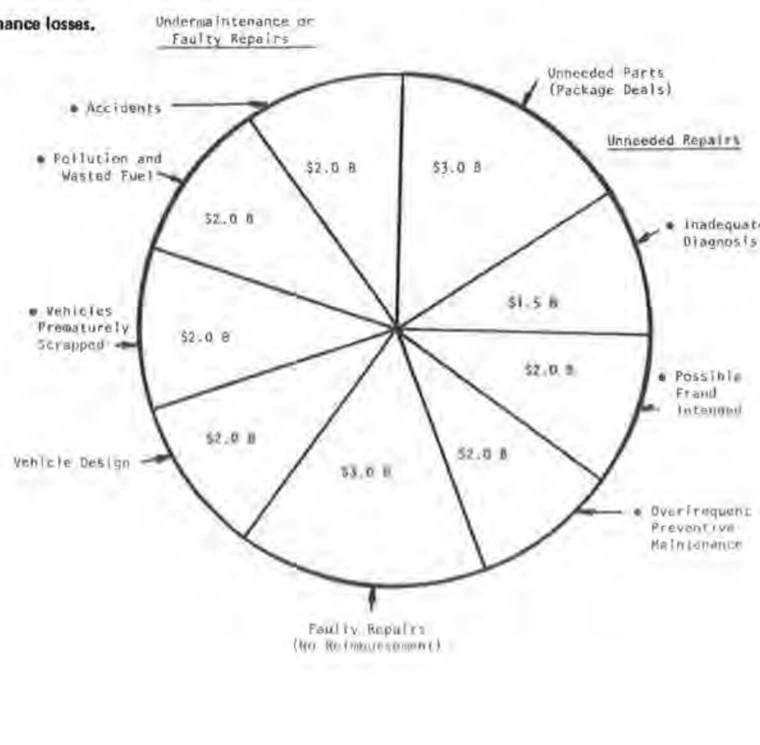
California Diagnostic Center

The California AAA diagnostic centers are located at 150 Hayes Street in San Francisco and at 2615 Keystone Avenue in San Jose. The San Francisco center opened in the fall of 1968 while the San Jose center opened in 1974.

The centers are equipped with the latest equipment for static and dynamic analysis. The centers perform complete diagnostic inspections at a cost of $30.00 for members and $40.00 for nonmembers. A system and component inspection costs from $6.00 to $22.00. Unlike the Missouri AAA diagnostic center, no post-repair inspections are normally performed. The San Francisco center has conducted more than 75 000 inspections while the San Jose center has conducted more than 33 000 inspections.

A sample of 513 cars was selected from these centers that met the following conditions: inspected in 1978, failed the brakes or emissions inspection, agreed to return for an after-repair inspection at no cost, and provided the corresponding brake or emissions repair receipts. During this

Figure 2. Estimated distribution of automotive repair and maintenance losses.



data-collection period, all motorists whose vehicles failed the brake or emissions inspection were asked to return for an after-repair inspection in order to collect repair receipts. There was no charge for the after-repair inspection.

Transportation Survey

The University of Alabama, under contract to DOT, conducted an undercover survey of repair facilities in the following seven cities: Atlanta; Philadelphia; Miami; Nashville; Houston; White Plains, New York; and Brooklyn, New York. The survey was conducted between January and March 1979.

In six of the cities, the survey was carried out in cooperation with the District Attorneys' offices. This effort was coordinated by the project director of the National District Attorney's Association Economic Crime Project. In the seventh city, Atlanta, the survey was carried out in cooperation with the Georgia Governor's Office of Consumer Affairs.

A reputable repair facility was identified in each of the cities and used as the secure facility. These secure facilities provided the inspection space and equipment and the master mechanics who assisted in the inspection of the vehicles. Sixty-two cars were inspected at the secure facilities and documented. Engine and suspension malfunctions were introduced into these cars, and the cars were then taken to randomly selected repair facilities in each city. No induced malfunctions were made in the brakes. After the repairs were made, the cars were inspected at the secure facilities and the repairs documented.

APPROACH

The detailed data-reduction procedures were initially developed during the analysis of the repair-cost data from the Alabama Diagnostic Inspection Demonstration Project. A description of these procedures is given elsewhere (4). These same procedures were used to reduce the data from the other sources. Therefore, the data could be readily compared among the four sources.

In summary, each repair action on a repair receipt was classified as being required, recommended, optional, or unnecessary. The criteria for determining the repair classification were as follows:

1. A repair was considered required if the item was found to be faulty (i.e., failed) during the inspection,
2. A repair was considered recommended if the repaired item is normally repaired as part of the repair of another faulty item repair even though nothing was found to be faulty with the subject item during the inspection,
3. A repair was considered optional if the repaired item may or may not be normally repaired as part of another faulty item repair even though nothing was found to be faulty with the subject item during the inspection, and
4. A repair was considered unnecessary if the repaired item passed the inspection and no other repair of another marginal or failed component would normally affect the decision to repair the subject item.

The determination of the repair classification was done by a team of individuals--an experienced diagnostic inspector and an experienced automotive parts specialist. Only the diagnostic inspector was used to classify the survey data.

UNNECESSARY REPAIRS

The repair actions and the corresponding costs for each of the four sites are given in Tables 1 and 2. In summary, 6075 repair actions that represent $129 217 were analyzed from Alabama, 680 repair actions that represent $18 475 were analyzed after Alabama introduced the prescription forms, 1454 repair actions that represent $67 444 were analyzed from California, 1014 repair actions that represent $31 610 were analyzed from Missouri, and 120 repair actions that represent $3163 were analyzed from the seven-city survey.

Unnecessary Repair Frequencies

Figure 3 contains a comparison of the unnecessary repair frequencies for each of the data sources. The seven-city survey had the highest unnecessary repair frequency of 27 percent, followed by 25 percent from the Alabama center during its first two years of operation. Statistically, there is no significant difference between the two frequencies ($\chi^2 = 0.18$).

After two years of inspections, the unnecessary repair frequency in Alabama was reduced from 25 to 18 percent. There may be several reasons for this reduction. One obvious reason is the classic learning-curve effect for project personnel, vehicle owners, and the repair industry. A second reason is the introduction of the prescription forms in 1977. A study of the data at the time (6) indicated that the overall unnecessary repair frequency was 26 percent between 1975 and 1977--24 percent during the first six months of 1976 and 15 percent during the first nine months of 1977. Therefore, it appears that the learning effect may be minimal and that the use of the prescription forms may have been a major reason for this reduction of the unnecessary repair frequencies. The California and Missouri centers have been using a similar procedure in relaying to the motorist what should be repaired. The AAA centers have been in operation since the late 1960s; therefore, this possible learning effect by the repair industry, which was observed in the Alabama data, may have already occurred and is reflected by the lower unnecessary repair frequencies for the California and Missouri centers.

The unnecessary repair frequency for Missouri (10 percent) was significantly lower than for the other sites. Even after the introduction of the prescription forms in Alabama, which reduced the unnecessary repair frequency to 15 percent, there was still a significant difference ($\chi^2 = 10.0$, $p < 0.005$), with Missouri significantly lower. One possible explanation for the low unnecessary repair frequency for Missouri (besides the learning effect) is the possible effect on the repair industry of the after-repair inspection being performed by the Missouri center. The repair industry is probably aware of this procedure. On the other hand, the California center is not performing after-repair inspections and the unnecessary repair frequency was 18 percent.

The unnecessary repair frequency for California (18 percent) was significantly lower than Alabama before the introduction of the prescription forms (25 percent) ($\chi^2 = 28.8$, $p < 0.001$). After the use of the prescription forms in Alabama, there was no significant difference in the unnecessary repair frequency (18 percent versus 15 percent) ($\chi^2 = 2.6$). These results suggest that unnecessary repair frequencies may be reduced by providing the motorist with understandable repair information for communicating with the repair industry.

In summary, the data in Figure 3 suggest that the magnitude of unnecessary automobile repairs may be as high as 25-27 percent. However, it appears that diagnostic centers may have an effect on reducing unnecessary automobile repairs. The data from the Alabama center, which has been in operation since 1975, indicated a high initial unnecessary repair frequency that was supported by the seven-city survey. After several years of operation, the Alabama unnecessary repair frequency was reduced significantly. Furthermore, it appears that the frequency may be reduced even further with time, as indicated by the Missouri center. This suggests that a definite learning effect does exist as the repair industry becomes aware of the diagnostic centers.

Coupled with this learning effect in decreasing unnecessary repairs is the after-repair inspection service that is provided by the diagnostic centers. The Missouri and Alabama centers have been providing this service while California has not. It appears that the repair industry's knowledge of the after-repair inspection may also have an effect on unnecessary repairs.

Unnecessary Repair Costs

Figure 4 contains a comparison of unnecessary repair costs. Overall, 53 cents of every dollar at the seven-city survey was spent on unnecessary repairs versus 29 cents for Alabama, 19 cents for Alabama after the prescription forms, 13 cents for California, and 11 cents for Missouri. There is close correlation between the unnecessary repair frequencies and the unnecessary costs for all sites with the exception of the seven-city survey. For the survey, 27 percent of the repair actions were unnecessary while 53 cents of each repair dollar was spent on unnecessary repairs. One explanation for this high unnecessary repair rate could be that the induced engine malfunction (shorting the number 4 spark plug) and the induced suspension malfunction (removing the stabilizer link) caused higher unnecessary repairs (8). No malfunction was induced for the brakes. A second possible explanation for this anomaly could be the relatively smaller sample size for the seven-city survey.

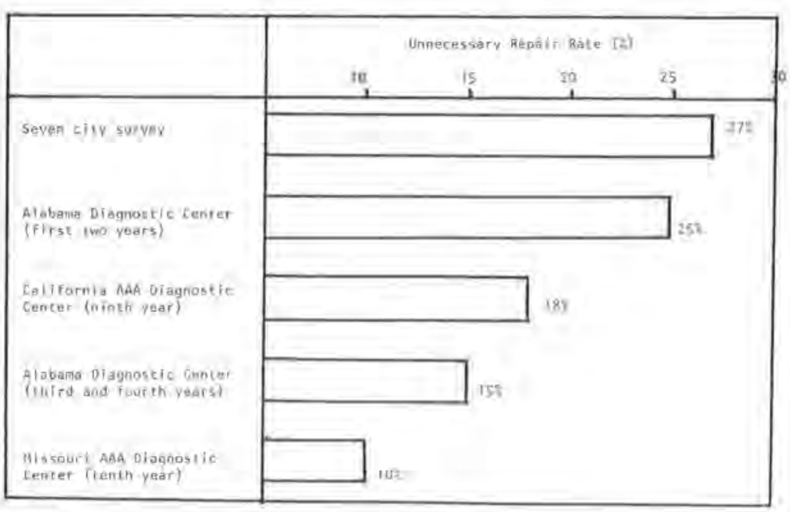
Table 1. Comparison of repairs.

	Alabama (overall)		Alabama (after prescription)		California		Missouri		Survey	
Repair	No.	Percent	No.	Percent	No.	Percent	No.	Percent	No.	Percent
Required	3932	65	575	85	1043	72	834	82	62	52
Recommended or optional	627	10			145	10	76	8	26	21
Unnecessary	1516	25	105	15	266	18	104	10	32	27
Total	6075		680		1454		1014		120	

Table 2. Comparison of repair costs.

	Alabama (overall)		Alabama (after prescription)		California		Missouri		Survey	
Repair	Cost ($)	Percent	Cost ($)	Percent	Cost ($)	Percent	Cost ($)	Percent	Cost ($)	Percent
Required	77 250	60	14 908	81	54 212	80	26 126	83	768	24
Recommended or optional	14 329	11			4 960	8	2 182	7	724	23
Unnecessary	37 638	29	3 567	19	8 272	12	3 302	10	1671	53
Total	129 217		18 475		67 444		31 610		3163	

Figure 3. Comparison of unnecessary repair rates.



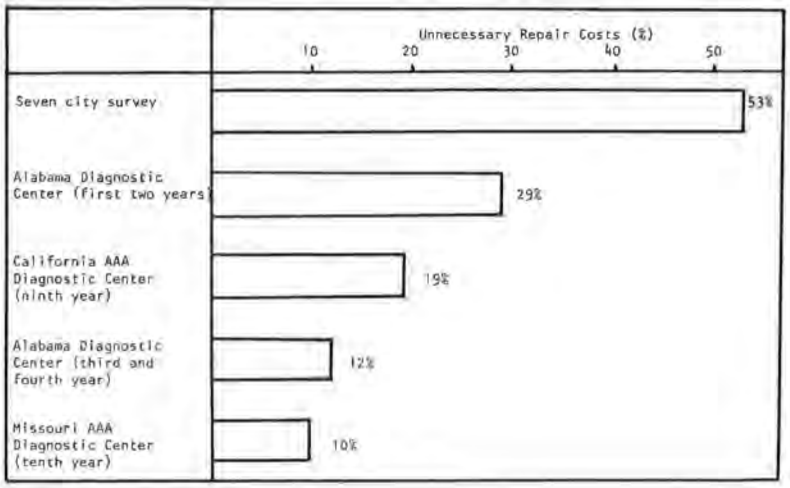


Figure 4. Comparison of unnecessary repair costs.

In summary, the data in Figure 4 suggest that the magnitude of unnecessary repair costs may be between 29-53 cents but probably closer to 29 cents. However, it is apparent that diagnostic centers reduce the rate of unnecessary repairs while also reducing the cost of unnecessary repairs. The magnitude of the reduction may be 10-12 cents/diagnostic center.

UNNECESSARY REPAIRS BY SYSTEM

Unnecessary Repair Frequencies

Table 3 contains a comparison of the unnecessary repair frequencies by system for each of the four sources. The unnecessary brake-repair frequencies were similar for California (28 percent), Alabama (28 percent), and the seven-city survey (26 percent). Likewise, the unnecessary brake-repair frequencies were similar for Missouri (19 percent) and Alabama after the prescription forms (15 percent).

There was a significant difference in the distribution of the unnecessary brake repairs ($\chi^2 = 22.04$, $p < 0.001$), with Alabama (after the prescription form) and Missouri having a significantly lower unnecessary brake-repair frequency. These lower unnecessary brake-repair frequencies may indicate the effect of conducting after-repair inspection at the Alabama and Missouri centers.

The unnecessary engine-repair frequencies were identical (30 percent) for Alabama and the seven-city survey. After the use of the prescription forms, the unnecessary engine-repair frequency for Alabama was reduced to 16 percent. Missouri and California had the lowest unnecessary engine-repair frequencies (5 and 8 percent, respectively).

There was a significant difference in the distribution of the unnecessary engine-repair frequencies ($\chi^2 = 174.19$, $p < 0.001$). The frequencies were significantly lower for California, Missouri, and Alabama (after the prescription form). Again, these lower unnecessary engine frequencies may indicate the long-term effect of conducting after-repair inspections at the Alabama and Missouri centers. Note that the unnecessary engine-repair frequency for Alabama fell from 30 to 16 percent during the third and fourth years, but not to the more long-term reduction as for the Missouri and California centers (5 and 8 percent, respectively).

The unnecessary engine-repair frequency was also low for the California centers, even though no after-repair inspections are conducted at the centers. One explanation may be the strong state legislation on emissions and, consequently, high public awareness and high awareness by the repair industry.

There was a wide variation in the unnecessary suspension-repair frequencies. Alabama (overall) had the highest (35 percent). California, Alabama (after prescription forms), and the seven-city survey were similar (21, 21, and 19 percent, respectively). Missouri had the lowest unnecessary suspension-repair frequency (11 percent).

Table 3. Percentage comparison of unnecessary repair frequencies.

Repair	California	Missouri	Alabama (overall)	Alabama (after prescription)	Survey
Brakes	28	19	28	18	26
Emissions	8	5	30	16	30
Alignment	6	4	8	4	0[a]
Suspension	21	11	35	21	19
Steering	20	0	22	-	67[b]
Total	18	10	25	15	27

[a]Sample of five repairs. [b]Sample of three repairs.

Unnecessary Repair Costs

Table 4 contains a comparison of the corresponding unnecessary costs by system for each of the four sources. The unnecessary repair costs by system closely followed the unnecessary repair frequencies. The anomaly is the unnecessary repair costs from the seven survey cities, where 43 cents of every dollar spent on brake repairs was unnecessary, 68 cents of every dollar spent on engine repairs was unnecessary, and 48 cents of every dollar spent on suspension repairs was unnecessary. These survey percentages for each system are the highest for all four sites. A possible explanation of these high costs could be that the induced malfunctions caused higher unnecessary repair costs.

CONCLUSIONS

The following conclusions are made based on the results of this study:

1. Users of diagnostic centers may experience a lower unnecessary repair frequency--on the order of 10-18 percent. But the higher 25 percent value was also experienced by users of the Alabama center. The final frequency is likely to be to some degree a function of comprehensiveness of the diagnostic inspection. The lower frequencies may be unique to

Table 4. Percentage comparison of unnecessary repair costs.

Repair	California	Missouri	Alabama (overall)	Alabama (after prescription)	Survey
Brakes	18	18	28	20	43
Emissions	7	6	35	20	62
Alignment	6	4	8	4	0[a]
Suspension	21	15	37	24	48
Steering	2	0	23	-	72[b]
Total	12	10	29	19	53

[a]Sample of five repairs. [b]Sample of three repairs.

the users of the centers, but may not have an effect on entire cities (9).

2. Diagnostic centers probably have an effect on reducing unnecessary automobile repairs. This effect probably increases with the length of operation of the facility, until some fairly stable level is achieved.

3. The industry's knowledge of the after-repair inspection may have an effect on the quality of repairs.

4. A learning effect probably exists while the repair industry becomes aware of diagnostic centers. The low unnecessary repair frequency for the Missouri center, which has been in operation for many years, by itself does not make that conclusion evident.

5. Unnecessary repair frequencies may be reduced by providing the motorist with understandable repair information for communicating with the repair industry.

ACKNOWLEDGMENT

This research was supported by contracts with the Consumer Affairs Division, DOT; the Office of State Vehicle Programs, NHTSA; and the Bureau of Economics, FTC.

REFERENCES

1. Survey of Current Business. U.S. Department of Commerce, Dec. 1977.
2. Personal Consumption Expenditures for Transportation: 1976. U.S. Department of Transportation, July 1976.
3. Auto Repair and Maintenance Program to Reduce Consumer Loss. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT HS-802 355, May 1978.
4. An Evaluation of Vehicle Repair Costs for Auto Check Participants. Univ. of Alabama in Huntsville, Huntsville, UAH Res. Rept. 197, Feb. 1977.
5. An Evaluation of Component Repair Costs for Auto Check Participants. Univ. of Alabama in Huntsville, Huntsville, UAH Res. Rept. 901, May 1977.
6. The Effects of Additional Consumer Information on Unnecessary Automobile Repairs. Univ. of Alabama in Huntsville, Huntsville, UAH Res. Rept. 208, Jan. 1978.
7. A Comparison of Automobile Repairs. Univ. of Alabama in Huntsville, Huntsville, UAH Res. Rept. 215, Aug. 1978.
8. Selective Survey of the Capability of Representative Automobile Repair Facilities to Diagnose and Repair Automobiles. Univ. of Alabama in Huntsville, Huntsville, UAH Res. Rept. 220, May 1979.
9. Evaluation of Diagnostic Analysis and Test Equipment for Small Automotive Repair Facilities. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT HS-803 536, July 1978.

Publication of this paper sponsored by Committee on Highway Vehicle Research.

Electric Vehicle Technology Update

ANDREW CHASE

Electric vehicles (EVs) may offer some advantages over gasoline and alternately fueled vehicles in terms of operating cost and as a hedge against future fuel shortages. However, existing EV technologies need to be advanced so that EVs will be as easy to operate and maintain as gasoline vehicles. An overview of some areas in which technology improvement is needed and is now being addressed by participants in the Electric Vehicle Demonstration Project of the U.S. Department of Energy is provided. These areas include state-of-charge monitoring, charging, battery capacity testing, electrolyte management, and battery connectors.

Given past experience with gasoline shortages and rising operating costs due to increased gasoline prices, vehicle owners, particularly fleet operators, have been looking into the potential of alternately fueled vehicles as a hedge against similar conditions in the future. Among the alternate-fuel options that have been tested are the following: diesel, propane, methane, methanol, and electricity. Of these, electricity may offer the most flexibility in that it is widely available, easily tapped, and not as susceptible to shortages as the others. Electricity may also perform well in terms of operating cost since it can be generated from a variety of fuels, and it should therefore not increase in price as rapidly as any one particular fuel.

Although ownership of an electric vehicle (EV) thus offers potential advantages, some obstacles need to be addressed. The transition from gasoline vehicles to EVs may not be as easy as the transition from gasoline to other fuels. Because electricity cannot be used in an internal-combustion engine, a significantly different propulsion system is required. Therefore, in a typical conversion of a gasoline vehicle to an EV, the propulsion system of the gasoline vehicle is removed and EV systems are added, which results in a purchase price about double that of the gasoline vehicle. As manufacturers gain production experience and demand allows large-scale production of EVs, this price gap can be expected to lessen.

In addition to its higher price, the operating and maintenance requirements of EVs differ more widely from gasoline vehicles than do those of other alternately fueled vehicles. For example, there are few differences between a gasoline vehicle and one converted to run on propane, vis-à-vis the differences between gasoline and electric vehicles, which makes a mechanic's transition to servicing EVs more difficult.

Such obstacles to EV ownership are being identified by 42 private companies, federal agencies, and state and local governments that comprise the Electric Vehicle Demonstration Project of the U.S. Department of Energy. Demonstration participants have operated about 500 EVs more than 900 000 miles during the past three years. Several of the participants have formed a task force to study problem areas and to introduce technological improvements for EVs, which will make them easier to own and operate in the gasoline-vehicle environment in which we live. This paper provides an update on some areas in which technological improvement is needed and is being addressed by the task force.

First, however, a brief description of EVs may be useful. EVs do not have a fuel tank--they use a pack of batteries to store electrical energy aboard the vehicle. Between 16-20 battery modules are arranged in series to provide a 96- to 120-V pack. EVs have no carburetor through which acceleration and velocity can be controlled; instead, a controller is relied on to regulate the draw of electricity from the battery pack in response to the driver's use of the accelerator pedal. As mentioned earlier, EVs do not have an internal-combustion engine but rather a highly efficient electric motor that converts the electrical energy to mechanical energy. The motor, used as a generator, can even convert movement of the vehicle back to electrical energy through regenerative braking. The electric motor does not require a cooling system or an ignition system. Finally, electric vehicles are not "fueled" with electricity by anything that resembles a gasoline pump. Instead, a charger, in circuit between an electrical outlet and the battery pack, "pumps" electricity into the batteries at a programmed rate.

In comparing an EV with a gasoline vehicle, we are comparing a product in its infancy with a mature one. The gasoline vehicle has benefited from almost a century of research, development, testing, user feedback, and refinement. EVs, however, must rely at the moment on a compilation of off-the-shelf technologies, originally designed for other applications. The result is that EVs are not yet as owner foolproof as gasoline vehicles.

An example of this can be seen by contrasting the fuel gauge of a gasoline vehicle with the equivalent in an EV. The gasoline fuel gauge is a simple device that provides relatively accurate information on the amount of gasoline remaining in the fuel tank. Since EVs have a shorter range (35-60 miles) and charging takes some time (about 8 h for a full charge), an accurate fuel or battery state-of-charge gauge is critical. Fully depleting the charge in an EV is not only inconvenient but can greatly reduce battery life. State-of-charge monitoring in currently produced EVs has proved to be inaccurate or misleading. The state-of-charge meter reading (based on measurement of the voltage across the battery pack) when the vehicle is stationary will differ from that when it is being driven and can also vary at different driving speeds. Further, even if the meter correctly indicates the state of the overall battery pack, it does not indicate the condition of individual battery modules. It is possible for some modules to be very weak while the overall state of charge of the pack appears adequate. In this case, the modules low on charge will impair the performance of the EV and are likely to suffer permanent damage if they become completely discharged.

Accurate monitoring of battery module charge is one of the key areas being addressed by the EV task force. The task force is considering the use of a microprocessor that uses individual battery module voltages and current measurements to determine state of charge. The next step would be to allow depleted battery modules to be removed from the circuit while driving, to protect them and the performance of the vehicle.

The differences that often exist in state of charge among battery modules in a pack has led the EV task force to take several other initiatives. First, a battery module can be ruined not only by overdischarging but also by overcharging. When an EV is plugged in to charge, the charger provides all battery modules with the same number of ampere hours, regardless of each module's state of charge. As a result, some battery modules are overcharged while others are insufficiently charged. This not only shortens battery life but also wastes electric power. A possible solution, which is being investigated by the EV task force, is "smart" chargers. In concept, these chargers, along with appropriate wiring of the battery modules, will be capable of metering the correct current to each module, thus equalizing the level of charge among them.

Another initiative to address the problem of battery modules with dissimilar characteristics in a pack is currently being implemented by several demonstration project participants. This effort involves the use of diagnostics to weed bad battery modules out of the battery pack. After such modules have been replaced, all the modules are brought up to full and equal charge. They should then all charge and discharge to the same level, thus greatly reducing the risk to battery life and also improving performance. The diagnostic tool being used to identify weak battery modules is a load bank. The testing method involves fully charging each battery module and discharging the pack into the load bank at a constant rate. The voltage of each module is monitored. By comparing the time it takes each module to reach a certain voltage to the battery manufacturer's specifications, the capacity of each module can be determined.

Another area of interest to the EV task force is battery electrolyte management. To constrast this area to gasoline vehicles, consider the single battery used to start a gasoline vehicle. Typically, this battery requires little attention--occasional checking of the electrolyte level in six cells is generally all that is done. EV batteries are different. They are deeply discharged, overcharged, and lose a lot of water by evaporation in the process. Regular inspection and replenishment of water is required, and inspecting and adding water to a 120-V battery pack that has 60 battery cell caps is a time-consuming task that needs frequent repetition. Filling batteries cannot be put off because their life depends on it. Cell caps with glass inserts are made by one manufacturer to simplify electrolyte inspection. The task force is looking to reduce the time required to add water; one possible solution may be to advance the state of the art of single-point watering systems. Such systems allow all battery cells to be filled by adding water to a single reservoir--somewhat like filling one corner of an ice cube tray and letting the water spill over to the other compartments.

In addition to electrolyte management, battery connectors constitute an important maintenance concern to EV owners. Although dirt and deposits on the battery of a gasoline vehicle may occasionally interfere with the starting circuit or loose cable clamps may result in an open circuit, these are relatively infrequent and minor problems. In an EV, however, they are serious. Dirty or loose connections on any of the 40 battery posts can significantly degrade performance, cause improper charging, and sometimes generate enough heat (because of the high resistance) to melt the battery posts. Currently, most connectors used in EVs are similar to those used for gasoline-vehicle starting batteries. They were not designed to carry high currents with a minimum voltage drop for long periods. Tightening and cleaning connections has been a regular and time-consuming maintenance item on EVs in the demonstration project. The EV task force efforts in this regard are directed toward testing new connectors and developing other alternatives. One recently manufactured type of connector is spring loaded to maintain a tight connection and has a plastic cap to keep the connection clean.

An EV that is as owner foolproof as, and competitive with, a gasoline vehicle will depend on an overall systems approach rather than modification of off-the-shelf components. Such an EV will probably be produced by the same companies that now produce gasoline vehicles. The efforts of the EV task force are important to the near-term improvement of EVs and very possibly may contribute to the ultimate system design. The incentives for task force members to contribute are great since all members foresee an important role for EVs in their own vehicle fleets.

Publication of this paper sponsored by Committee on Highway Vehicle Research.

Analysis of the Effectiveness of Bumper Standard FMVSS 215

PAUL ABRAMSON, HOWARD STEIN, JAY W. COHEN, AND JOHN V. WERNER

The primary objective of this paper is to evaluate the effectiveness of crash-protecting automobile bumpers as required by Federal Motor Vehicle Safety Standard 215—Exterior Protection, Passenger Cars. This study focused on three distinct versions of the standard: (a) the initial 1973 regulations that required compliance with a barrier test, (b) regulations after 1974 that required both a pendulum and an upgraded barrier test, and (c) regulations after 1978 that limited total vehicular damage as a result of the pendulum and barrier tests. Following the recommendations of several previous studies, insurance claims were used as the data base. A comprehensive data base provided by the State Farm Insurance Company was categorized by vehicle model years that represent four time periods—1972, 1973, 1974-1978, and 1979—and into vehicle-size classes, impact points, vehicle age, and repaired and replaced damaged bumper categories. By using reported claims and average cost of these claims as measures of effectiveness, it was shown that the model years with more protective bumper systems experienced significantly lower proportions of bumper-related claims relative to all property-damage claims. However, in general, these model years also had higher average repair costs. The reduced percentages of bumper-related claims were primarily attributable to decreased claims that involve bumpers being replaced rather than claims that involve bumper repair only.

In 1971, the National Highway Traffic Safety Administration (NHTSA) issued Federal Motor Vehicle Safety Standard (FMVSS) 215--Exterior Protection, Passenger Cars. The general purpose of this standard was to prevent low-speed accidents from impairing safe operation of the vehicle and to reduce the frequency of override and underride impacts in higher-speed collisions (i.e., collisions of two vehicles where the initial contacts slide over or below the bumper). It was also hoped that, as a consequence of FMVSS 215, the cost of repairs to vehicles involved in low-speed collisions would be reduced. Hence, an economic advantage to the consumer would be realized.

Bumper-performance tests used to determine that safety-related items (lights, fuel system, cooling system, etc.) are not rendered inoperable include pendulum and barrier-impact tests of the bumper system. The barrier tests consist of front and rear impacts against a flat rigid barrier at specified speeds. The pendulum-impact test consists of striking the bumper at specified heights and angles with a pendulum hammer. This test is designed to promote consistent bumper heights so as to reduce the likelihood of underride or override of bumpers in car-to-car collisions.

The performance testing for compliance with the safety-related requirements of FMVSS 215 has gone through various stages of development [see Table 1 (1)]. The initial standard model year (1973) was subject to barrier-impact testing only (5-mph front and 2.5-mph rear impacts). Beginning with the following model year (1974), the rear-barrier test was upgraded to 5 mph and the pendulum-impact test (longitudinal and corner impacts) was introduced. The pendulum-impact test was amended starting with the 1976 model year, which decreased the number of longitudinal pendulum impacts.

Title I of the Motor Vehicle Information and Cost Savings Act (P.L. 92-513, 1972) instructed NHTSA to develop property-damage bumper standards that would provide the maximum feasible reduction of costs to the public and to the consumer. The Part 581 standard issued under the authority of this Act required that, effective with the 1979 model year, front and rear bumpers must be capable of protecting vehicles from damage in barrier and pendulum longitudinal crash tests at 5 mph and pendulum corner impacts at 3 mph. In addition, damage criteria were upgraded to permit damage only to the bumper itself and the brackets, fasteners, etc., that attach the bumpers to the chassis framework and not to any other vehicle components or surfaces. For 1980 and future models, the standard also limited bumper face-bar damage.

In general, the objective of this study addresses the basic question, Has the imposition of a bumper standard resulted in reduced damage and overall cost to the motorist? The analysis attempts to answer this question by determining whether or not insurance claims and their estimated repair costs have been changed by the imposition of FMVSS 215. The major portion of the analysis is aimed at determin-

Table 1. Basic requirements of FMVSS 215 by model year.

Model Year	Test Requirement		
	Barrier	Pendulum	Impact
1973	5-mph front	None	
	2.5-mph rear	None	
1974[a]	5-mph front	5-mph front	6
	5-mph rear	5-mph rear	6
		3-mph corner at 20-in front	1
		3-mph corner at 20-in rear	1
1975	5-mph front	5-mph front	6[b]
	5-mph rear	5-mph rear	6[b]
		3-mph corner at 20-in front	1
		3-mph corner at 20-in rear	1
1976	5-mph front	5-mph front	2
	5-mph rear	5-mph rear	2
		3-mph corner front at 20 and 16 in[c]	2
		3-mph corner rear at 20 and 16 in[c]	2
1977, 1978	5-mph front	5-mph front	2
	5-mph rear	5-mph rear	2
		3-mph corner front at 20 and 16 in	2
		3-mph corner rear at 20 and 16 in	2

[a]Pendulum test requirements did not apply to cars with wheelbases of 115 in or less that have either convertible tops or no roof supports between the A-pillar and the roof-support structure or no rear designated seating positions.
[b]Reduced to 2 impacts for cars manufactured after May 12, 1975.
[c]Corner impacts at 16 in did not apply to cars with wheelbases greater than 120 in.

ing whether statistically significant variations in claim frequency percentages and average repair costs have occurred for vehicle model years of the post-standard periods. Comparisons are made after the data are stratified into appropriate categories. Additional analyses also include an estimate of the relative contribution of repaired versus replaced bumper claims to the total of all bumper claims.

METHODOLOGY

A basic measure of bumper effectiveness should be the changes in vehicle damage incurred in low-speed collisions. However, direct real-world observation cannot be made of low-speed, low-damage crashes. An added problem exists because there is no single data source that contains information on all crashes (i.e., reported and unreported crashes). It has been recommended that insurance claims be used as a measure for reported vehicle involvement and damage (2,3). These claims should include collision coverage involvement as well as property-damage coverage involvement claims.

Collision coverage provides insurance for damage to a policyholder's automobile. Under collision coverage, a deductible is applied, which means that the policyholder pays for some of the damage. (Typical deductible amounts are $100 and $150.) Liability coverage provides insurance for damage to other property, including other automobiles, caused by the policyholder. These coverages do not have deductibles.

Claims under these two coverages, both of which involve damage to the automobile, produce different patterns of damage due to the nature of the coverages. As an illustration, consider a typical two-car low-speed crash in which one car is struck in the rear by the other. The damage to the car struck in the rear will typically be paid for by the property-damage liability coverage of the striking car. The striking car's damage, if it exceeds the deductible, will be paid for by the collision coverage. Because of these sorts of differences, collision coverage claims more often involve front-end damage than property-damage liability claims, which more often involve rear-end damage.

The principal data source for this study is the computerized data base maintained by the State Farm Insurance Company, which includes files on both collision and property-damage (liability) coverage claims. The following sections describe the nature of the State Farm data base, the stratification applied to the data base, and the measures of effectiveness tested. Note that other stratifications (and analyses) are not presented here but are detailed elsewhere (4,5). These include coverage involvement (collision and liability), object struck (fixed object and other vehicles), and bumper design.

Data Base

The data base maintained by State Farm Insurance contains a sample of claim information obtained from their claim service centers. The data used in this study consist of tabulations of these data, subject to various factors. The factors and subsequent comparisons of the data were therefore restricted by the data-base content. The number of claim records for the individual model years was at least 10 000 claims. No totaled-vehicle claims were included in this study.

Stratification of Data Base (Independent Variables)

Stratification of State Farm's claim data base was dictated, in large part, by its own format and content. These stratifications constitute independent study variables in the study. Listed below are the study variables and their respective categories.

Model-Year Comparison

The model-year categories that generally describe changes and refinements of the bumper standard are grouped in the following manner:

1. Pre-1973 model year--These vehicles are not required to meet any performance standards;
2. 1973 model year--These vehicles were subject to the initial version of FMVSS 215;
3. 1974-1978 model years--For these years, most vehicles had to meet both barrier- and pendulum-test requirements (note that various versions of the performance requirements applied for certain model years); and
4. 1979 model year--Limited-damage criterion applicable for barrier and pendulum tests.

Age of Car

Both one-year-old and three-year-old vehicles were analyzed. These ages were selected since appropriate data were available for both the prestandard and poststandard vehicle models.

Impact Points

A bumper-related claim was defined as a claim in which damage occurred to the face bar and the direction of impact was one of four different categories: front, front corner, rear, or rear corner. Additional data were tabulated for other categories: side impacts, other impact points (includes front and rear impacts for which the face bar did not require repairs), and all claims regardless of impact point.

Vehicle-Size Class

Size class designations are those as defined by the Highway Loss Data Institute (HLDI). The four levels of stratification are as follows:

Vehicle Size	Wheelbase (WB) Length (in)
Subcompact	WB ≤ 101
Compact	101 < WB ≤ 111
Intermediate	111 < WB ≤ 120
Full sized	WB > 120

Bumper Damage

Two levels of bumper damage were tabulated: bumper repaired and bumper replaced. Bumper damage was defined as damage to the face bar of the bumper. If a vehicle had any part of the bumper assembly replaced or parts both repaired and replaced, then the claim was categorized as a replaced claim. Since vehicle speed is not directly obtainable from the data base, bumper damage was an attempt to define a surrogate measure of relative impact speed of crashes that resulted in claims. Claims that showed that the bumper had only been repaired indicate a relatively lower-impact speed than claims where the bumper was replaced. Although bumper damage itself is not a criterion of FMVSS 215, the effectiveness of the standard will be judged, in part, by any change in bumper-repair costs. Hence, the damage resistance of the bumper system itself becomes an important factor.

Measures of Effectiveness

Although the primary purpose of FMVSS 215 is to prevent low-speed collisions from impairing the safe operation of vehicle systems, standardized and strengthened bumpers should effect crash damage in general. Thus, vehicles that meet the bumper requirements should experience proportionally fewer incidents that require insurance claims. In addition, the total cost of repairs as a result of all crashes (including unreported) should be reduced.

Hence, the measures of effectiveness are (a) proportion of bumper-related reported claims and (b) costs of these reported claims. The analysis of reported claims will determine if the proportions of claims by vehicles in the different comparison periods have changed significantly for various stratifications. The costs of reported damage claims will determine if average repair costs for vehicles of the different comparison periods have changed due (in part) to the effect of FMVSS 215.

CLAIM ANALYSIS

By using claims as a measure of effectiveness, this analysis aims to determine whether or not the bumper standards have significantly altered the proportion of bumper-related claims. Hypothesis tests on the difference between claim proportions are applied to indicate the effectiveness of the standards.

Analysis of Bumper-Related Claim Trends

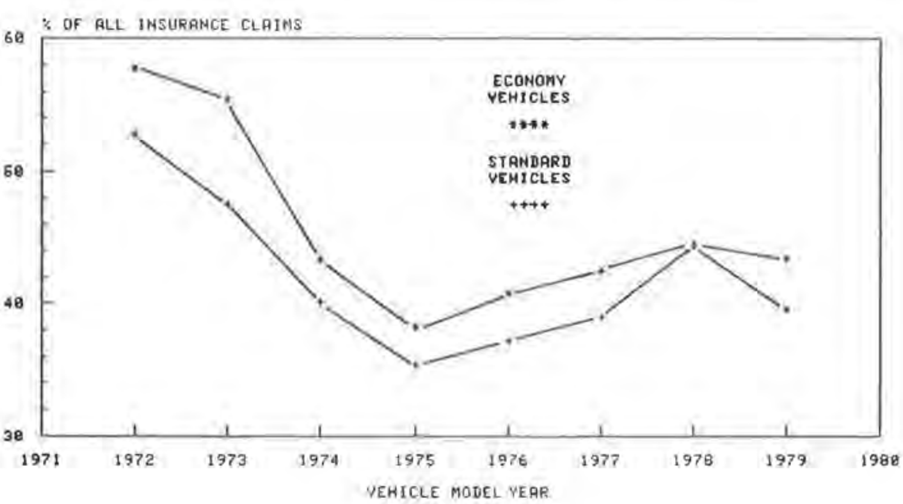
The percentage of total insurance claims that are bumper related for the individual model years during their first year of exposure is presented in Figure 1. For ease of presentation, the vehicle-size class stratification has been collapsed to form the economy (subcompact and compact) and standard (intermediate and full-sized) categories. Important observations from this trend graph include the following:

1. The prestandard (1972) model-year's proportion of bumper-related claims is greater than any poststandard (1973 or later) model year.
2. The initial standard (1973) model-year's proportion of bumper-related claims is greater than any subsequent model year. The initial version of the standard required only a barrier-impact test (5-mph front and 2.5-mph rear-impact speeds). For the following model years, the performance criteria were upgraded to include a pendulum test and a 5-mph impact speed for the rear bumper during the barrier test (see Table 1).
3. Economy and standard vehicles maintain the same trend characteristics between individual model years. The proportion of bumper-related claims for standard vehicles has been less than that of economy vehicles for each model year except 1979.

In general, the bumper-related claim trend of one-year-old vehicles can be described as decreasing with each model year from the prestandard (1972) model year until the poststandard (1975) model year. The following years (1976-1978) experienced increases with each model year and then declined for the 1979 model year.

These trends clearly indicate that the imposition of the bumper standard has influenced the claim experience of these model years. However, these trends are also influenced by other factors, such as the distribution of claims by coverage type (i.e., collision or liability), the influx of imported vehicles and an overall shift to a fleet of smaller vehicles, changes in vehicle design (particularly the use of elastomeric materials and one-piece multifunctional fabrication of bumpers and their related vehicle features), and other changes in the

Figure 1. Claim distribution—bumper-related claims.



driving environment in general.

The following analysis compares in detail the claim trends subjected to vehicle-size class and impact-point stratification. The 1974-1978 model years are grouped for this detailed analysis as they were all subject to some version of the barrierand pendulum-impact tests (see Table 1) with the same performance criteria.

Analysis of Claim Distributions

Table 2 presents the percentage of claims that are bumper related and the principal point of impact of these claims for the vehicles that comply with the different versions of FMVSS 215. The claim percentages are tabulated for the front-center, front-corner, rear-center, rear-corner, and side (not bumper-related) impacts. The total bumper category is the sum of the four front- and rear-impact points and constitutes the total proportion of claims that are bumper related. The hypothesis test for differences in proportions (percentages) relative to the prestandard (1972) model year has been applied, and statistically significant differences are indicated in the table.

(A possible bias is introduced because the sum of the percentages of the size class impact-point distribution must equal 100 percent. A bumper design change may result in an expected reduction in the percentage of front of rear damage claims and therefore increase the percentage of side involvement. No adjustment is applied at this time because it would tend to enhance any claim frequency reduction that may exist in the data.)

One-Year-Old Vehicles

Comparisons of claim distributions for the different model-year groupings for one-year-old vehicles are presented in Table 2. This table compares the 1973, 1974-1978, and 1979 model years with the prestandard (1972) model year. Comparing the claim percentages for the prestandard (1972) model year with the poststandard (1973) model year during their first year of exposure showed that the proportion of front-impact claims was significantly reduced for the poststandard vehicles for all market classes. The 1973 version of the standard introduced barrier-impact performance criteria for the front and rear bumper. However, the criterion for the front bumper was 5 mph while for the rear bumper it was at 2.5 mph. The 1973 model-year results for the other impact points (front corner, rear, and rear corner) were not consistent for all market classes, as explained below.

1. Subcompacts and compacts experienced an increase in both rear and rear-corner impact claims, where most of these increases are statistically significant. There were significant reductions for the front-corner impact point.
2. Intermediate vehicles indicated reductions for front-corner and rear impacts and an increase for rear-corner impacts, none of which were significant, although the total bumper-related claim percentage has been significantly reduced for the 1973 model year.
3. Full-sized vehicles exhibited a significant reduction for the rear-impact point claims and a significant increase for rear-corner claims for the 1973 model year. However, the total percentage of bumper-related claims has been significantly reduced for the 1973 model year.

Thus, for all bumper-related claims, the total claim percentages were reduced to a greater extent as vehicle size increased for the 1973 model year.

The upgraded versions if the standard that applied to the 1974-1978 model years introduced the pendulum-impact test and increased the performance requirement of the barrier test of the rear bumper to 5 mph. Table 2 also compares prestandard (1972) to poststandard (1974-1978) model years and showed significantly lower claim percentages in all front, front-corner, and rear-center claims for all size classes. In addition, all size classes showed significantly lower percentages for the total of all bumper-damage claims. In comparing the model years that represent two versions of the standard--the 1973 with the upgraded 1974-1978 version--vehicles of the later model years (1974-1978) indicated significant reductions in claim percentages primarily for the front-corner, rear-corner, and rear-center impact points. The front impact point experienced both increased and decreased percentages, although only the decrease for intermediate vehicles was significant. (Note, although not shown in Table 2, significance tests at the 5 percent level were conducted comparing the sets of poststandard model-year data.) Since the 1974-1978 version of the standard introduced the pendulum test and an improved rear bumper, these results are consistent

Table 2. Claim distribution for one-year-old vehicles by model year.

		Distribution of Claims by Impact Point (percentage of all claims)						
		Bumper Related (bumper repaired or replaced)						
Vehicle-Size Class	Model Year	Front	Front Corner	Rear	Rear Corner	Total	Side	All Claims[a]
Subcompact	1972	16	21	14	9	60	28	100
	1973	12[b]	19[b]	15	13[b]	59	28	100
	1974-1978	11[b]	17[b]	8[b]	8	45[b]	36[b]	100
	1979	12[b]	14[b]	8[b]	7	41[b]	35[b]	100
Compact	1972	13	21	13	8	56	32	100
	1973	9[b]	19[b]	15[b]	10[b]	55	34	100
	1974-1978	10[b]	16[b]	7[b]	9	42[b]	39[b]	100
	1979	11[b]	12[b]	7[b]	7	37[b]	37[b]	100
Intermediate	1972	13	20	11	9	53	36	100
	1973	10[b]	18	11	10	49[b]	38	100
	1974-1978	8[b]	14[b]	8[b]	9[b]	39[b]	42[b]	100
	1979	11	11[b]	10	11	42[b]	39	100
Full sized	1972	12	19	13	9	53	35	100
	1973	7[b]	17	10[b]	13[b]	46[b]	39[b]	100
	1974-1978	7[b]	14[b]	7[b]	10	38[b]	44[b]	100
	1979	8	16	9	12	45[b]	40	100

[a]All claims include all bumper-related claims, side-impact claims, and other impact-type claims.
[b]Null hypothesis of equal proportion relative to the prestandard (1972) model-year data rejected at 5 percent significance level.

with the intent of the standard. In addition, all vehicle classes experienced significant reductions for the total percentage of bumper-related claims from the 1973 to 1974-1978 model-year data.

Beginning with the 1979 model year, the performance criteria of the barrierand pendulum-impact tests were amended to limit damage of all vehicle surfaces, including the bumper itself. The previous versions of the standard only afforded protection to vehicle safety features. The claim data for the 1979 model year are included in Table 2 and indicated significantly lower claim percentages for the total of all bumper claims for all size classes relative to prestandard vehicles. In reviewing the bumper-related trend (Figure 1), the 1979 model year is significant in that it reverses a pattern of increases in the percentage of bumper-related claims between the 1975 and 1978 model years. A detailed comparison for the 1979 model year with the 1978 model year is presented in Table 3. This comparison indicates significant decreases in the proportion of claims that are bumper related for subcompact and compact vehicles between the two model years. These decreases are concentrated in the front-corner and rear-corner impact points. Comparisons of the intermediate and full-sized vehicle classes indicate no significant pattern of differences between the two model-year's claim experiences.

Three-Year-Old Vehicles

The detailed results for three-year-old vehicles can be found elsewhere (1). In general, the data exhibit similar trends as the one-year-old vehicle damage-claims comparisons: front-impact claim percentages were reduced in the 1973 version and rear-impact claim percentages were reduced in the upgraded 1974 version. However, fewer reductions were statistically significant. Yet, some statistically significant reductions remained evident relative to the prestandard model year (1969). These reductions included percentages of most front-impact claims of the 1973 model year, rear-impact claims of the 1974 model year (except subcompact), and the total claim proportion for all size classes for both the 1973 and 1974 model years. In comparing the claim percentages of the 1973 model year with that of 1974, few differences were statistically significant except for rear-impact claims for subcompact and compact vehicles and the total claim proportion of subcompact vehicles.

The results of the three-year-old vehicles may be influenced by the fact that the 1974 model year was the only available data to represent the upgraded versions of the standard. However, it appears that even after three years of exposure, the imposition of a bumper standard has influenced claim experience.

Repaired and Replaced Bumper Claims

In this analysis, the relative contribution of repaired and replaced bumper claims to the combined bumper-claim distributions were examined for the different model years. It might be hypothesized that these stratifications represent surrogate measures of the impact speeds. Claims that involve bumpers repaired are considered lower-speed impacts relative to claims that involved bumper replacement.

Table 4 indicates the percentage of claims that involve bumpers repaired versus bumpers replaced for all vehicle-size classes. It is clear that the overwhelming percentage of bumper claims for all size classes resulted from replaced rather than repaired bumper claims.

In comparing prestandard (1972) with poststandard (1973) model years for one-year-old vehicles, the percentage of repaired bumper claims generally increased while the percentage of replaced bumpers decreased significantly. The latter of these trends becomes more pronounced when comparing the latter poststandard (1974-1978 and 1979) vehicle claims with the prestandard (1972) data. Between the poststandard vehicles, replaced claim percentages significantly decreased.

For three-year-old vehicles (4), replaced bumper claims decrease for the improved bumpers on vehicles in almost all cases. Repaired bumper claims show an increase for some market classes between the prestandard (1969) and poststandard (1973) model years. In general, there is a shift from replaced bumper claims to repaired bumpers. This shift, however, cannot account for the total change in the two damage categories, since there is a larger reduction in replaced bumpers than there is an increase in repaired bumpers. Some of the shift may, therefore, be accounted for by an increase in accidents that sustain no damage or, at least, do not require that the bumper be repaired or replaced. The shift appears to be more pronounced in the poststandard model-year comparisons. Thus, the major effect of the reduced frequencies is due to a decrease in replaced bumper claims.

COST ANALYSIS

The cost analysis examined the average claim costs of the model years that represent the different versions of the standard. All costs were adjusted to the base economic year (1972) by using a discount rate of 10 percent/year (4). The statistical test applied was the t-test of significance between means. The experimental hypothesis is that there is no difference between the average repair costs being compared. However, an increase in average repair cost for reported crashes (as measured by this study) may in fact be expected. If a bumper stan-

Table 3. Claim comparison of one-year-old vehicles for 1978 and 1979 model years.

		Distribution of Claims by Impact Point (percentage of all claims)						
		Bumper Related (bumper repaired or replaced)						
Vehicle-Size Class	Model Year	Front	Front Corner	Rear	Rear Corner	Total	Side	All Claims[a]
Subcompact	1978	11	19	9	8	47	33	100
	1979	12	14[b]	8	7	41[b]	36	100
Compact	1978	10	15	8	9	42	37	100
	1979	11	12	7	7[b]	37[b]	37	100
Intermediate	1978	9	16	9	10	45	39	100
	1979	11	11[b]	10	11	43	39	100
Full sized	1978	9	16	9	10	43	45	100
	1979	8	16	9	11	45	40	100

[a] All claims include all bumper-related claims, side impact claims, and other impact-type claims.
[b] Null hypothesis of equal proportion rejected at 5 percent significance level.

Table 4. Comparison of repaired and replaced bumper claims for one-year-old vehicles.

Vehicle-Size Class	Model Year	Claim Proportion (percentage of all claims) Repaired	Replaced	Total Bumper Related	Avg Repair Cost (1972 $) Repaired	Replaced	Total Bumper Related
Subcompact	1972	3.5	56.8	60.3	172	331	322
	1973	6.0[a]	52.5[a]	58.5	230[b]	341	331
	1974-1978	5.7[a]	39.0[a]	44.6[a]	218	423[b]	398[b]
	1979	6.8[a]	34.4[a]	41.2[a]	188	410	379
Compact	1972	5.3	50.5	55.8	196	383	365
	1973	6.0	47.4[a]	53.4	230	345[b]	333[b]
	1974-1978	5.6	36.6[a]	42.2[a]	227	425[b]	400[b]
	1979	4.8	32.4[a]	37.2[a]	220	507[b]	470[b]
Intermediate	1972	5.6	47.3	52.8	150	371	348
	1973	6.2	42.6[a]	48.9[a]	219[b]	395	372
	1974-1978	4.0[a]	35.0[a]	39.0[a]	224[b]	412[b]	394[b]
	1979	4.5	38.0[a]	42.5[a]	183	395	373
Full sized	1972	4.8	48.1	52.9	163	376	357
	1973	5.3	40.9[a]	46.2[a]	193	400[b]	376
	1974-1978	4.2[a]	34.0[a]	38.1[a]	218	414[b]	393[b]
	1979	6.2[a]	39.1[a]	45.3[a]	230	424	402

[a]Null hypothesis of equal proportion relative to the prestandard (1972) model-year data rejected at 5 percent significance level.
[b]Null hypothesis of equal average costs to the prestandard (1972) model-year data rejected at 5 percent significance level.

dard is effective in reducing the percentage of insurance-reported incidents, those reported that remain may involve, on average, repairs to more severe crash damage.

Claim Cost Trends

Figure 2 presents the average total repair cost of all bumper-related claims for each individual model year during their first year of experience. Vehicles have been grouped as before into the economy or standard categories. Unlike the claim trends, the cost trends between the individual model years were not similar between economy and standard vehicles. The overall trend of higher costs for the later model years (after an inflationary adjustment) holds true for both vehicle categories. However, if the imposition of the bumper standard has influenced claim frequency by eliminating low-speed crash-damage claims from the distribution, then this result is to be expected. The previous analysis has indicated that a significant percentage of low-speed crashes were missing from the claim distribution. These missing claims probably were the result of insufficient damage for a claim or no vehicular damage being produced from a crash. Hence, the elimination of these low-speed, low-cost claims would tend to shift the distribution of claim costs and the average of the total repair costs higher.

One-Year-Old Vehicles

Table 5 shows the comparison of the average cost of repairs for one-year-old vehicles relative to the prestandard (1972) model year (all costs adjusted for 10 percent inflation rate in 1972 dollars). In comparing prestandard (1972) with poststandard (1973) model years, the costs for the poststandard (1973) vehicle claims were, in general, higher for all size classes (except compact). Claims for front impacts consistently showed significant increases in repair cost for the models after the initial implementation of the standard while changes in costs for other impact points varied from size class to size class.

Poststandard (1974-1978) model-year repair costs showed significantly higher average repair costs for most front-impact points relative to prestandard costs. Moreover, the average cost of total bumper-related claims was significantly higher for all market classes. Assessment of the incremental cost increases between the two versions of the standard was also statistically compared and indicated that subcompact and compact vehicles exhibited significant increases in total bumper-related claim costs. The poststandard (1979) model year average repair claim costs are significantly higher than the prestandard (1972) model year for the front-impact claims of subcompacts and compacts and the total bumper-related claims of compacts. In a comparison with the previous (1978) model year's experience (shown in Table 6), the only size class with consistently higher average claim costs was compact vehicles (all costs adjusted for 10 percent inflation rate in 1972 dollars). Intermediate and full-sized vehicles experienced lower average claim costs for most impact points relative to the 1978 model year. However, neither of these trends proved statistically significant.

Three-Year-Old Vehicles

Poststandard (1973) average repair costs compared with prestandard (1969) costs showed decreased repair costs (4). The decrease was significant only for the rear-corner impact and total of all bumper claims for full-sized vehicles. For poststandard (1974) vehicles, average costs were generally higher than the prestandard data, particularly for the subcompact or compact classes. However, there was a general trend (significant for compact total bumper claims) of increased costs for the total of all bumper claims from the 1973 to 1974 model year.

Thus, an initial decrease in average cost with the introduction of the standard (1973 model year) has been offset somewhat with the upgrading of the standard (1974 model year). The resultant effect of the two stages of the standard was a general increase (not statistically significant) in the average repair costs.

Repaired and Replaced Bumper Claims

Table 4 presented the average claim costs for repaired and replaced bumpers for one-year-old vehicles. The average costs are all in terms of a 1972 economic year, adjusted by using a 10 percent/year inflation rate.

The comparisons show that for one-year-old vehicles both replacement and repair costs were higher for the poststandard model years. Between poststandard vehicles, repaired bumper costs have remained about the same and replacement costs have increased.

For three-year-old vehicles (4), some repair costs have increased and all replacement costs have decreased from the prestandard to the initial poststandard (1973) model year. Between poststandard model years, repair costs have remained about the same but replacement costs have increased. Thus, the pattern of cost variation between poststandard model years is mainly attributed to increased costs of bumper-replacement claims.

FINDINGS AND CONCLUSIONS

The primary findings and conclusions obtained in this study are as follows:

1. Statistically significant reductions in the proportion of bumper-related front-impact claims were indicated for the initial poststandard vehicles (1973) that were subjected to barrier-test require-

Figure 2. Average repair cost—bumper-related claims.

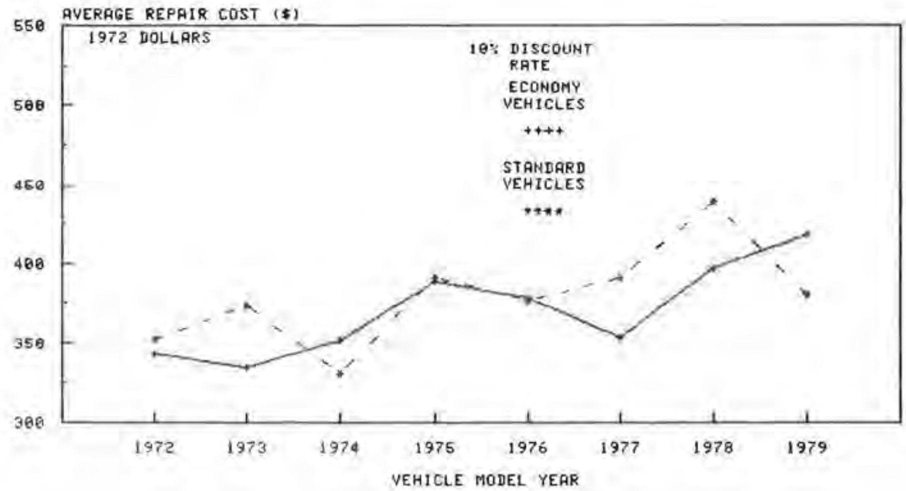


Table 5. Average repair cost comparison for one-year-old vehicles by model year.

		Bumper Related (bumper repaired or replaced)					
Vehicle-Size Class	Model Year	Front	Front Corner	Rear	Rear Corner	Avg	Side
Subcompact	1972	375	349	248	277	322	232
	1973	445[a]	344	300[a]	242	331	228
	1974-1978	525[a]	413[a]	293[a]	295	398[a]	227
	1979	519[a]	359	287	292	379	250
Compact	1972	405	425	272	292	365	240
	1973	485	341[a]	268	276	333[a]	254
	1974-1978	541[a]	414	293	306	400[a]	235
	1979	625[a]	457	325	409	470[a]	248
Intermediate	1972	409	376	293	286	348	247
	1973	469	402	285	316	372	245
	1974-1978	516[a]	411[a]	296	338[a]	394[a]	221[a]
	1979	468	382	330	314	373	235
Full-sized	1972	416	376	317	286	357	233
	1973	521[a]	386	346	308	376	255
	1974-1978	529[a]	425[a]	333	288	393[a]	226
	1979	514	448	307	331	402	294

[a]Null hypothesis of equal proportion relative to the prestandard (1972) model-year data rejected at 5 percent significance level.

Table 6. Average repair cost comparison of one-year-old vehicles for 1978 and 1979 model years.

		Average Repair Cost by Impact Point[a] ($)					
		Bumper Related (bumper repaired or replaced)					
Vehicle-Size Class	Model Year	Front	Front Corner	Rear	Rear Corner	Avg	Side
Subcompact	1978	525	409	285	275	389	228
	1979	519	359	287	292	379	250
Compact	1978	563	423	279	306	405	235
	1979	625	457	325	409	470	248
Intermediate	1978	592	414	372	345	428	241
	1979	468	382	330	314	373	235
Full Sized	1978	663	489	393	290	461	250
	1979	514	448	307	331	402	294

[a]Null hypothesis of equal proportion relative to the prestandard (1972) model-year data rejected at 5 percent significance level.

ments as compared with prestandard vehicles (1972 or earlier).

2. Consistent with the intent of the 1974 (and later) version of the standard, which introduced the pendulum test (front, rear, and corner impacts) and upgraded the rear-barrier test speed to that of the front, the 1974-1978 vehicle model years showed significant reductions in the proportion of bumper-damage claims, primarily for corner and rear impacts, when compared with the poststandard (1973) model year. In addition, the proportion of front and rear-center bumper-damage claims and the total proportion of bumper-damage claims for all vehicle-size classes of the poststandard (1974-1978) model years were significantly lower when compared with the prestandard (1972) model year.

3. The 1979 model year, which introduced the limited-damage criterion as a result of FMVSS 215 impact-test requirements, experienced statistically significant reductions for the total proportion of bumper-damage claims for the subcompact and compact vehicle classes when compared with the previous years' damage claims. When compared with the prestandard (1972) model year, all vehicle classes experienced significantly lower damage claim percentages.

4. Claims for poststandard model years showed higher average repair costs when compared with the prestandard (1972) model year. The increases were statistically significant for most front-impact claims and for the total of all bumper-damage claims of all vehicle-size classes of the 1974-1978 model year comparison. Comparison of claim costs between poststandard model years resulted in few significant differences.

5. Reduced claim trends of the model-year comparisons and increased average claim costs were generally attributable to claims that involve replacement of the bumper rather than those that involve only bumper repairs.

In reviewing these conclusions, several points must be noted. First, if the standard has, in fact, been effective in reducing low-speed, low-cost crash damage, the cost distribution would be influenced. As a result of protective bumpers, the low-cost crash-damage claims would disappear from the low end of the cost distribution of reported damage and the average claim dollar amount would tend to increase. Thus, the claim and cost trends noted are not necessarily independent. In addition, the mixture of claim types (collision with deductible of $100 or greater, collision with deductible less than $100, and property-damage liability) experienced by State Farm policyholders has changed in more recent years to include a larger percentage of collision claims with higher deductible amounts. This pattern would also produce higher average claim costs due to the more expensive nature of collision claims as well as the higher deductible amounts. A final important point is that the bumper standard only specified performance criteria, not design criteria. Manufacturers still determine the specific bumper design. Thus, this study truly only evaluates the performance of those designs chosen by the manufacturers.

ACKNOWLEDGMENT

The work reported in this paper was performed for NHTSA, U.S. Department of Transportation. We wish to extend our appreciation to Warren LaHeist, project director, for his guidance.

REFERENCES

1. Automobile Insurance Losses Collision Coverages: An Assessment of the Effects of the Federal Safety Bumper Standard on the Losses for 1972 through 1978 Models. Highway Loss Data Institute, Washington, DC, Res. Rept. HLDI A-12, June 1980.
2. R.L. Braun and others; Transportation Institute, Stanford Research Institute. Evaluation Methodology for Federal Motor Vehicle Safety Standards: Volume 2--Technical Findings. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT-HS-803-341, May 1977.
3. Center for Environment and Man, Inc. Final Design and Implementation Plan for Evaluating the Effectiveness of FMVSS 215: Exterior Protection. National Highway Traffic Safety Administration, U.S. Department of Transportation, May 1977.
4. P. Abramson, J. Cohen, and H. Stein; KLD Associates, Inc. Analysis of Insurance Claims to Determine Effect on Crash Damage. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT-HS-9-02207, March 1980.
5. P. Abramson, J. Cohen, and H. Stein; KLD Associates, Inc. Analysis of Insurance Claims to Determine Bumper Effect on Crash Damage--1979 Model Year. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT-HS-9-02207, Oct. 1980.

Publication of this paper sponsored by Committee on Highway Vehicle Research.

Present Status of Inertia-Collection System

LARS Th COLLIN

The inertia-collection (INCOLL) concept for engine emission checking is based on transients that use inertia forces as braking torque. Thus, the emission-control systems developed for cars during the late 1970s have necessitated a further development of the INCOLL system. It is shown that the method has a high potential of predicting federal test procedure results. Results from official tests of the INCOLL system in the United States are presented together with some indications of results from similar tests in Europe, as well as the prospects for the further development of the method.

The purpose of this paper is to describe the present status of the inertia-collection (INCOLL) method and to present results achieved by the U.S. Environmental Protection Agency (EPA) in Ann Arbor, Michigan, in 1980, as well as other activities during the past few years.

BACKGROUND

During the development of vehicle-propulsion systems in order to comply with the more stringent emission regulations in different countries, the systems have become more efficient but also more sensitive for malfunctioning components or systems. Therefore, a demand has developed to find a simple method, not only to check the individual car condition at delivery, but also to be able to follow its behavior from an emission point of view during its use. Through such a system, the high-emitting individual vehicles in a population could be discriminated.

The basic requirements for a short test can be expressed by the following criteria:

1. The short test should be able to establish the three main pollutants with satisfactory accuracy and repetitivity; this should be done essentially quicker than with the certification cycles; and
2. The procedure should be reasonably capable of being correlated with certification tests.

All proposed short tests can be divided into two categories: transient cycles and steady-state cycles. The transient cycles on a chassis dynamometer tend to get rather complicated and, as a result, relatively large handling and investment costs are required. The steady-state tests, both dynamometer and idle tests, usually fall far short of providing adequate correlatability with existing certification tests. The problem has been to develop a transient test without the complexity and cost involvement of a dynamometer.

The INCOLL method offers a transient test, which uses an entirely new concept, for emission inspection of light- and heavy-duty vehicles. The development of the INCOLL method has been described elsewhere (1-3), where early test results are also given.

REQUIREMENTS

During the late 1970s, more stringent legislation, especially in the United States, has forced the development of fuel systems that are able to control the air/fuel ratio exactly. Those systems are all of the closed-loop type, which means a feedback of the oxygen concentration in the exhausts to the fuel system. As a closed-loop system has a relatively long transient response, where the dominating time lag is due to gasoline transport through the engine, it is essential that the short-test transients do not substantially differ from the transients for which the system is constructed. This has meant that the INCOLL layout for closed-loop vehicles has had to be modified in order to adapt to the ability to handle transient processes.

Features

The INCOLL system is based on the principle that energy is stored in the moving parts of the engine and transmission during an acceleration of the engine while the transmission is in the neutral position. Therefore, it is possible to apply a transient load against inertia forces, where the load intensity is determined by the actual throttle setting. In Figure 1, a simplified equation of motion for a vehicle on a chassis dynamometer is given, where J equals actual moments of inertia, η equals gear efficiencies, and u equals gear ratio (moment of friction force that results from bearings, etc., is neglected). Since the angular acceleration $d\omega/dt$ is given from the actual driving cycle, the instantaneous engine torque could be calculated.

Figure 2 shows the simplified system equation for the INCOLL principle, where the parts involved in the inertia resistance are manufactured with narrow tolerances. The resulting loading can therefore be well controlled and is quite repeatable for vehicles belonging to the same family.

The accelerations and decelerations are accomplished through changes in the throttle setting of the engine. In this case, the throttle setting is guided by a movement with constant speed during acceleration and by a feedback system during deceleration.

The INCOLL test is built on four parameters:

1. Ramp speed and acceleration,
2. Deceleration speed,
3. Upper revolutions per minute (rpm) limit, and
4. Lower rpm limit.

The parameters are for a given vehicle family that was chosen so that the transients of the INCOLL test become equivalent to the transients of an official test.

Hardware

The INCOLL system consists of three major components that are illustrated in Figure 3. The central unit monitors the test, the servo amplifier amplifies the signals from the central unit, and the servo motor controls the throttle setting of the engine. Figure 4 shows the actual equipment used during the tests.

TESTS AT EPA

In order to investigate the suitability of using the INCOLL test procedure as an emission inspection and maintenance testing procedure compared with existing testing procedures, a test program was conducted at EPA in Ann Arbor, Michigan, from April 29 to June 13, 1980 (4).

Test Program

The following test sequence was used by EPA:

1. Federal test procedure (FTP) (1979 procedure),
2. INCOLL,
3. 50-cruise test,
4. Four-mode idle test,
5. Two-mode loaded test,
6. Los Angeles-4 cycle (LA-4),
7. INCOLL,
8. 50-cruise test,
9. Four-mode idle test, and
10. Two-mode loaded test.

The first round of short tests (2-5) is designated 1 in the evaluation and the second round (7-10) is designated 2. The FTP, INCOLL, and LA-4 emissions were diluted and analyzed with the same instruments while the others were analyzed by using a garage-type infrared hydrocarbon/carbon-monoxide (HC/CO) analyzer.

Test Vehicles

In order to control the ability of the short tests to detect high-emitting vehicles, a number of malfunctions were induced in the test vehicles. A description of the test vehicles and test configurations follows:

1. Dodge Dart, 1975, 3.69 dm^3 (California version)--Stock, 10 percent misfire, and rich idle;
2. Chevrolet Chevette, 1977, 1.39 dm^3--Stock, rich idle, and vacuum leak;

Figure 1. Simplified equation of motion for vehicle on chassis dynamometer.

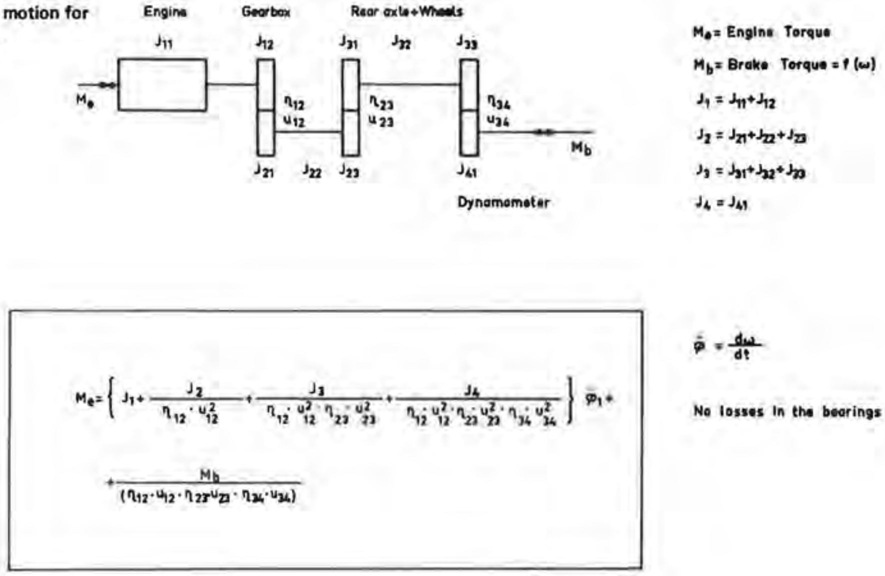


Figure 2. Simplified equation of motion for engine with its gearbox in neutral position.

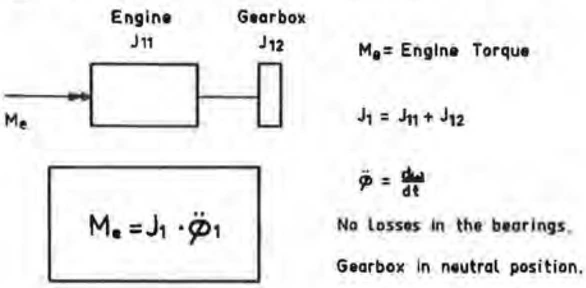
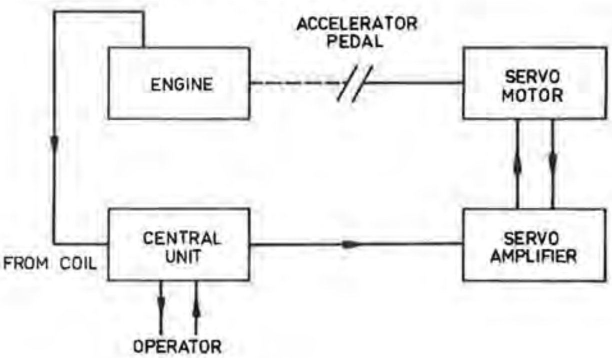


Figure 3. Relation between INCOLL components.



3. Ford Mustang, 1980, 3.28 dm^3--Stock, 10 percent misfire, and rich idle;

4. Chevrolet Nova, 1979, 4.10 dm^3--Stock;

5. Chevrolet Citation, 1980, 2.50 dm^3 (closed-loop, California version)--Stock (poisoned catalytic converter) and oxygen sensor disconnected (poisoned catalytic converter); and

6. Toyota Celica Supra, 1980, 2.56 dm^3 (closed-loop, 50-state version)--Stock and closed-loop disabled.

Evaluation of Test Results

Statistically, three different methods are associated with the concept of correlation: regression analysis, ranking analysis, and contingency-table analysis. These are briefly described below.

Figure 4. Present INCOLL equipment.

In a regression analysis (simple regression), to describe the predictive relation of two methods, a line is fitted to the test results (in the case of simple regression, a polynome of first degree). If there is a perfect linear relation between the variables, r (the correlation coefficient) will assume 1 or -1. If r is equal to zero, this does not mean that there is no relation between the two variables being correlated. Since r or r^2 measures only the goodness of fit of the regression line, this only indicates that the relation (if any) is not linear.

The Spearman rank correlation coefficient is a nonparametric test (i.e., valid independently of type of probability distribution of the observed stochastic variables). The two variables being correlated are each ranked in ascending size and subsequently compared to see whether the ranks of the two variables tend to be high or low together. The correlation coefficient assumes values between 1 and -1. It assumes the value 1 if the two rankings are identical and -1 if they are exactly in the opposite order.

The contingency-table analysis is based on the assumption that cut points can be found for the independent variable that will discriminate the groups of the dependent variable. In this analysis, results from the tests are grouped into four categories, as shown in Figure 5. In Figure 5, the letters take the following meanings:

a = Correctly passed--Vehicle does not exceed certification or short-test standard,
b = Error of omission--Vehicle exceeds certification standard for one or more emissions but does not exceed the short-test standard,
c = Error of commission--Vehicle does not exceed the certification standard but exceeds short-test standard for one or more emissions, and
d = Correctly failed--Vehicle exceeds both certification and short-test standard for one or more emissions.

Figure 5. 2x2 contingency table.

PREDICTOR VARIABLE \ PREDICTED VARIABLE		CERTIFICATION TEST	
		PASS	FAIL
SHORT TEST	PASS	a	b
	FAIL	c	d

From this figure, the short-test effectiveness (STE) can be defined, i.e., STE = d/(d + b), which is a measure of the fraction of vehicles that fail the FTP identified by the short test.

Results

For the linear regression, the correlation coefficients r, based on simple regression, are presented in Table 1.

The calculated values of the Spearman rank-order correlation are given in Table 2.

For the contingency-table analysis, the evaluation differs slightly from the regression and ranking analyses. The short tests following the FTP and the short tests following the LA-4 are added to each other in order to get a bigger sample size.

To determine the INCOLL cut points, all vehicles were treated together in a regression analysis. From the obtained regression line, cut points were determined as a function of the actual certification standard. The standards obtained are given in Table 3. The collection of cut points with scatter plots and regression lines are shown in Figures 6-11.

There are, of course, other ways of determining the cut points of a short test. One of the most common is to set errors of commission equal to 5 percent. Due to the small number of tests, with an equal certification standard in this test program, it is impossible to use this criterion here. By applying the regression technique there is no risk of

Table 1. Linear regression correlation.

Tests Compared	Sample Size (no.)	HC	CO	NO_x
FTP versus LA-4	14	0.99	0.99	0.99
INCOLL				
1	14	0.87	0.96	0.79
2	13	0.79	0.91	0.85
Four-mode, idle 1				
Idle A	16	0.83	0.66	
2500 rpm	16	0.73	0.81	
Idle B	16	0.50	0.64	
Drive	13	0.78	0.68	
Four-mode, idle 2				
Idle A	15	0.64	0.65	
2500 rpm	15	0.84	0.78	
Idle B	15	0.57	0.69	
Drive	12	0.57	0.69	
50-cruise				
1	16	0.55	0.72	
2	15	0.32	0.74	
Two-mode, loaded 1				
30 mph	15	0.68	0.85	
Idle	16	0.81	0.66	
Two-mode, loaded 2				
30 mph	15	0.60	0.73	
Idle	15	0.68	0.62	

Note: 1 designates the short test following the FTP and 2 designates the short test following the LA-4.

Table 2. Spearman rank-order correlation.

Tests Compared	Sample Size (no.)	HC	CO	NO_x
FTP versus LA-4	14	0.97	0.94	0.98
INCOLL				
1	14	0.79	0.87	0.86
2	13	0.68	0.82	0.81
Four-mode, idle 1				
Idle A	16	0.62	0.61	
2500 rpm	16	0.59	0.50	
Idle B	16	0.57	0.61	
Drive	13	0.63	0.50	
Four-mode, idle 2				
Idle A	15	0.80	0.68	
2500 rpm	15	0.77	0.64	
Idle B	15	0.83	0.68	
Drive	12	0.89	0.53	
50-cruise				
1	16	0.66	0.08	
2	15	0.74	-0.19	
Two-mode, loaded 1				
30 mph	15	0.66	0.62	
Idle	16	0.57	0.41	
Two-mode, loaded 2				
30 mph	15	0.77	-0.20	
Idle	15	0.80	0.62	

Table 3. Selection of cut points.

Emission	Certification Standard (g/mile)	INCOLL Standard (ppm)	Emission	Certification Standard (g/mile)	INCOLL Standard (ppm)
INCOLL 1[a]			INCOLL 1 + 2[b]		
HC	1.5	72.3	HC	1.5	68.6
	0.9	34.6		0.9	32.6
	0.41	3.9		0.41	3.2
CO	15.0	485.0	CO	15.0	464.0
	9.0	265.0		9.0	252.0
	7.0	191.0		7.0	181.0
NO_x	2.0	17.9	NO_x	2.0	20.9
	1.0	6.3		1.0	6.6

Note: ppm = parts per million.
[a]14 tests. [b]27 tests.

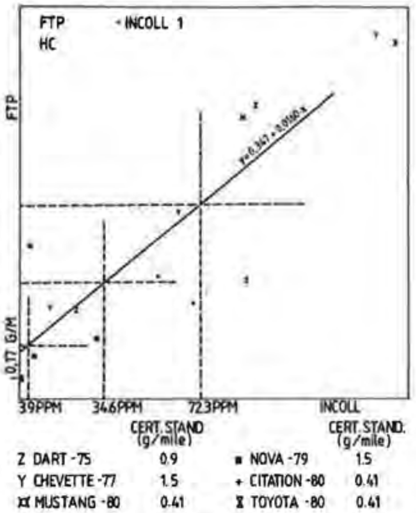


Figure 6. FTP versus INCOLL 1, HC.

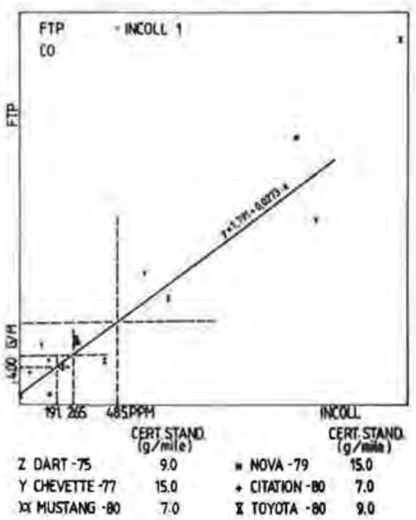


Figure 7. FTP versus INCOLL 1, CO.

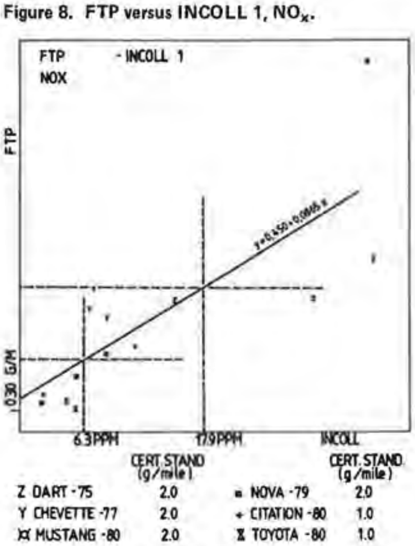


Figure 8. FTP versus INCOLL 1, NO_x.

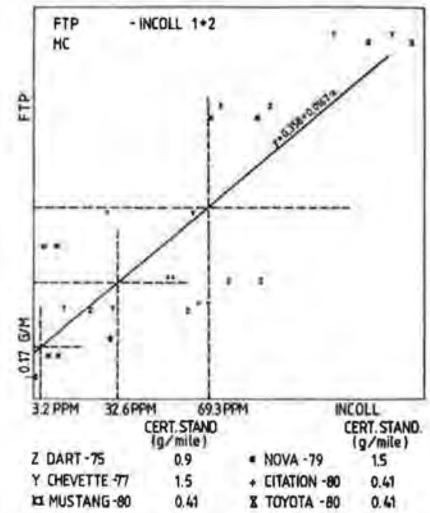


Figure 9. FTP versus INCOLL 1 + 2, HC.

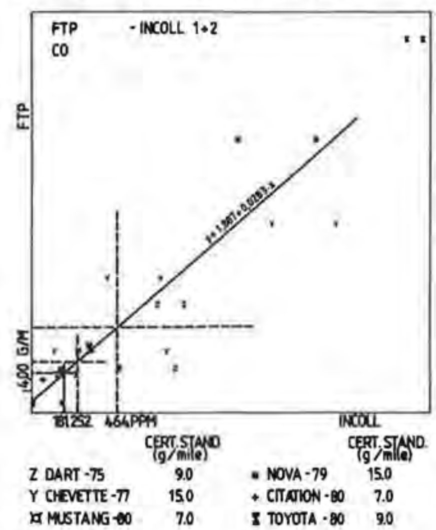


Figure 10. FTP versus INCOLL 1 + 2, CO.

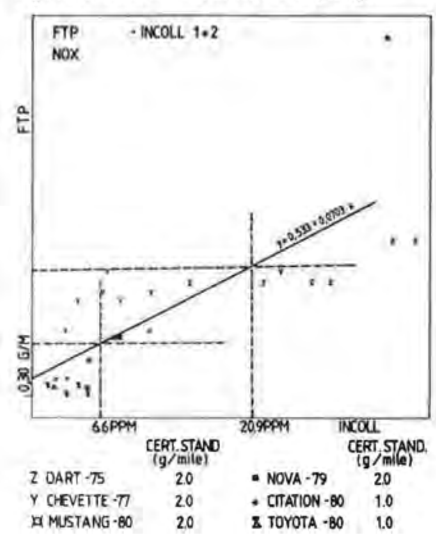


Figure 11. FTP versus INCOLL 1 + 2, NO_x.

making a subjective determination of cut points in order to get the best possible results.

For the other short tests, the following cut points are used:

1. LA-4--The same standard as the certification standard,
2. Four-mode idle--1 percent CO and 200 parts per million (ppm) HC, and
3. Two-mode loaded and 50-cruise--1.2 percent CO and 220 ppm HC.

The results of the analysis are given in Table 4.

RECENT TESTS AT CHALMERS UNIVERSITY OF TECHNOLOGY

Figures 12 and 13 are taken from a recent test program at Chalmers University of Technology, Sweden (5). Figure 12 shows the scatter plot of the nitrogen oxide (NO_x) result for the hot-transient phase of FTP versus INCOLL for a vehicle with induced malfunctions (exhaust-gas-recirculation tampering, Pulsair tampering, advanced/retarded ignition timing, and rich/lean idle). As can be seen, a very high correspondence can also be reached for vehicles equipped to comply with Swedish legislation (simple regression correlation = 0.93, Spearman correlation = 0.94). Figure 13 gives the same for CO at the same occasion (simple regression correlation = 0.93, Spearman correlation = 0.82).

During the summer and fall of 1981, INCOLL tests were included in a test program at Porsche AG, West Germany, under an Umweltsbundesamt contract. The final report is not yet ready, but the test sequence covered three cars of the same family being repeatedly manipulated as to idling CO, ignition timing, etc. The preliminary results indicate linear regression correlation values of the magnitude of 0.85-0.90 for the three investigated components--HC, CO, and NO_x--when compared with environmental control equipment cycle results.

SUMMARY

The tests at EPA have shown that the INCOLL test tends to correlate best with the FTP test of all the short tests investigated in the test program. The tested vehicles were all modern U.S. market cars equipped with catalytic converters (both oxidizing and three-way catalyst). As a high correspondence also can be reached with European vehicles, it should be possible to extend the INCOLL method to assembly-line testing and research activities, where the short time necessary for the test and its high

Table 4. Contingency table analysis.

Tests Compared	Sample Size (no.)	Correct Pass--a (%)	Correct Fail--d (%)	E_c--c (%)	E_o--b (%)	Short-Test Effectiveness
FTP versus LA-4	14	14	86	0	0	1.0
INCOLL						
1	14	14	79	7	0	1.0
1 + 2	27	11	74	11	4	0.95
Four-mode, idle 1 + 2						
Idle A	31	19	26	0	55	0.32
2500 rpm	31	19	13	0	68	0.16
Idle B	31	19	26	0	55	0.32
Drive	25	16	16	0	68	0.16
Two-mode, loaded 1 + 2						
30 mph	30	30	7	0	73	0.08
Idle	31	19	26	0	55	0.32
50-cruise, 1 + 2	31	19	10	0	71	0.12

Figure 12. Hot-transient phase of FTP versus INCOLL, NO_x.

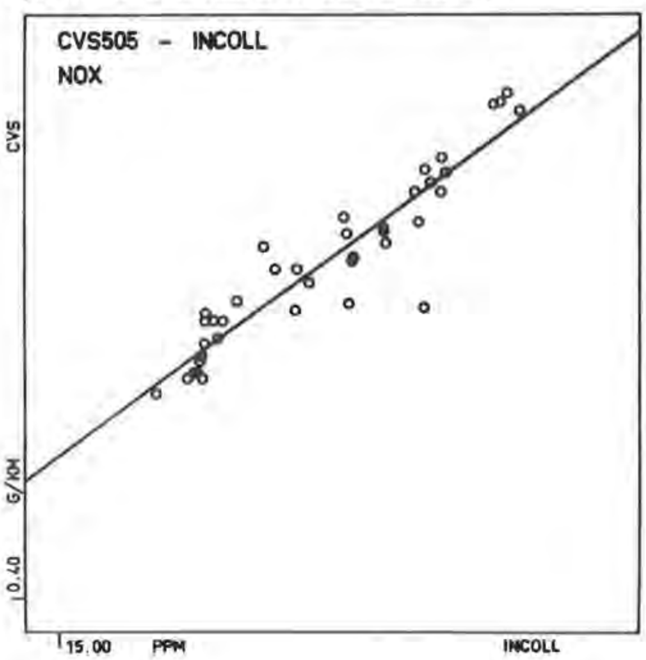
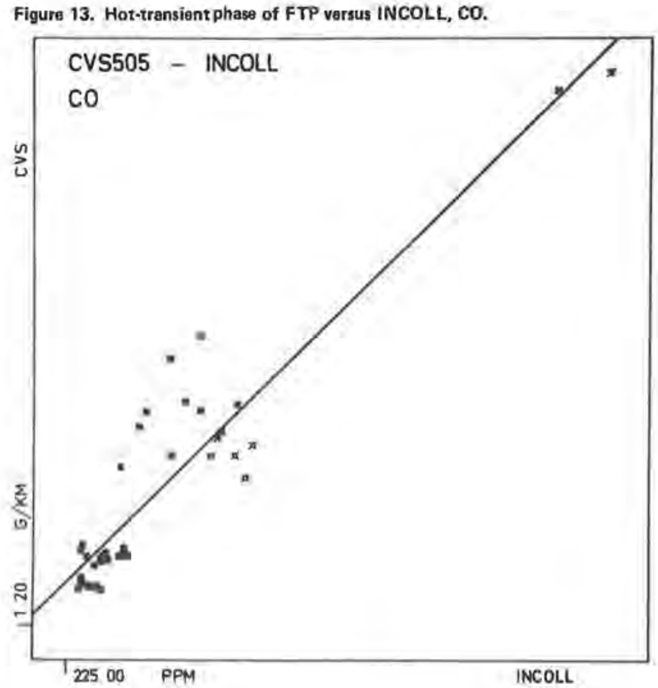


Figure 13. Hot-transient phase of FTP versus INCOLL, CO.



repetitivity would reduce costs. A further development toward fuel-efficiency tests would also be possible, as well as INCOLL cycles for diagnostic purposes.

The past few years of development of the INCOLL system have shown the following:

1. It is a transient test that does not need any chassis dynamometer;
2. It could be executed in 4-5 min, including adoption, automatic calibration, some automatic warming up, automatic testing, and dismantling;
3. It could be laid out to give test results comparable to certification tests that deal with emissions and consequent fuel consumption;
4. It could indicate and define errors or malfunctions within the engine system;
5. It could be laid out to sample from any load-rpm area of interest; and
6. It could (in a short time) supply large amounts of statistical data.

Thus, the main future area of use might not be emission investigations as such but quality and diagnostic investigations based on emission results.

REFERENCES

1. L. Th Collin. INCOLL--A New Technology in Emission Testing. Society of Automotive Engineers, Warrendale, PA, SAE Paper 780618, 1979.
2. L. Th Collin. INCOLL-Ein neues Messverfahren zur Emissions-kontrolle an Fahrzeugen. ATZ 81, 1979, p. 4.
3. L. Th Collin. A New Technology in Emission Testing. Kolloquium Periodische Überwachung der Schadstoffemission von Kraftfahrzeugen. TÜV-Akademie Rheinland, Düsseldorf, West Germany, May 17, 1979.
4. B. Smuda. Evaluation of Applicability of the INCOLL Procedure as an I/M Strategy. U.S. Environmental Protection Agency, Rept. EPA-AA-IMS/81-5, 1981.
5. L. Th Collin and I. Denbratt. Further Development of the Inertia Collection System. ISATA, 1981.

Publication of this paper sponsored by Committee on Highway Vehicle Research.

Transverse Air-Cushion Restraints for Commuter Vans and Small Buses

NATHANIEL H. PULLING AND DONALD R. VAILLANCOURT

The recent proliferation of commuter vans and small buses has directed attention to their safety deficiencies and the need for better restraint of their occupants. A promising alternative to lap belts, which are generally ignored by van and bus commuters, would be the patented transverse air cushions described in this paper. Details are given of full-scale tests on a deceleration sled, where these air cushions restrained anthropometric dummies in simulated frontal barrier crashes up to 48 km/h.

During the past few years, the problems of urban air pollution coupled with the soaring cost and prospective shortage of motor vehicle fuels have forced changes in the patterns of commuting in private motor vehicles. More commuters are driving small cars in order to save fuel and an increasing number of commuters are vanpooling for even greater savings (1-2).

Safety experience in vanpooling so far has been good (3,4). Most vanpooling arrangements use experienced and prudent drivers, optimum routing, and well-maintained equipment. Driving is done in daylight (for the most part) and during hours when there is minimal likelihood of encountering conflict with other drivers under the influence of alcohol or other impairments. These and other factors foster an excellent safety situation for commuting in vans and small buses.

Nevertheless, there is cause for concern because commuter vans carry up to 15 persons in each vehicle, and small buses even more, so that many more commuters are at risk in any accident than would be the case with automobiles. Furthermore, vans lack some of the safety features of passenger cars (5), and accident studies indicate that they are less crashworthy (6,7). An effective countermeasure would be the use of lap and shoulder belts (8,9), but there are no indications that their use by U.S. commuters in vans and buses is any greater than in passenger cars, where currently only one in nine car occupants uses restraints (10).

Air bags are an obvious solution to the restraint problem for people who will not wear seat belts. Conventional air-bag systems could be used for the front-seat occupants of a van, but for the other passengers they would have serious disadvantages. First, a conventional air bag deflates in a controlled manner after being triggered by the vehicle impact (11), so that afterwards there is little protection afforded during any subsequent collisions or rollover in a multiple-event accident. Second, there would have to be a large penalty in vehicle weight due to the massive seat structures required to withstand the enormous forces created by several human bodies in forward, high-gravity deceleration.

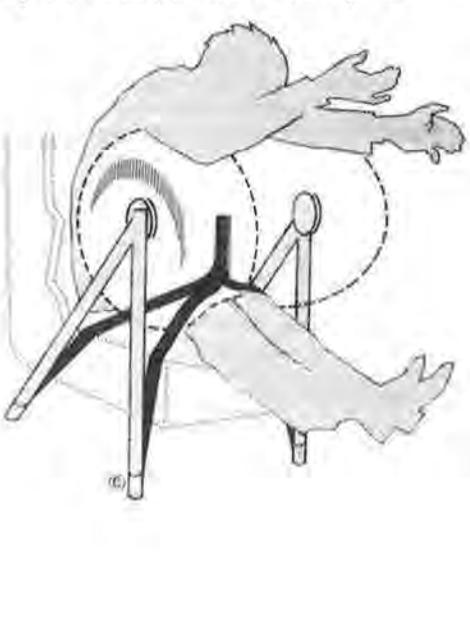
TRANSVERSE AIR-CUSHION RESTRAINTS

The transverse air-cushion restraints developed by Liberty Mutual Insurance Company circumvent these problems. On deployment during a collision they would form a rigid cylinder of air cells in front of each row of passengers and extend across the width of the van or bus. The air cushions are constrained (as described later) in such a way that no body impact forces are transmitted to seats in front. These air cushions contain air pressure that would be maintained for a number of seconds until the vehicle comes to rest after any multiple collisions or rollover. The full-coverage crash protection of this air-cushion system would be available to van and small-bus occupants without subjecting them to the nuisance and discomfort of strap restraints.

Transverse air-cushion restraints act more like very wide strap restraints than like conventional air bags. They are resilient and extend the deceleration time of the body while at the same time spreading the collision force over an extended area of the torso. The dynamic tests described in the next section show how the body tends to jackknife in a frontal collision and drape itself over the air cushion, as shown in Figure 1, while the thighs are forced tightly into the seat. The air cushions form a cylinder about 0.33 m in diameter so that a person's body does not move forward very much during a frontal collision, as it would with strap restraints that stretch or with conventional air bags that collapse. This means that the head and knees are fully protected from striking the back of a seat in front. Furthermore, while the air cushions are somewhat compliant, they maintain shape well enough during a collision to restrain an adult's body (whether large or small) centered in the seat, so that dangerous lateral movement would be prevented during an angle or side collision. Because the air cushions remain fully inflated for 5 s, the torso would be fully restrained during a rollover more effectively than with the air-bag restraints currently in use.

Prior to deployment, the air-cushion material would be folded compactly and stowed in armrest canisters 15-20 cm in diameter and 8-10 cm thick. These would be fastened to a steel framework, as shown in Figure 1. This framework can be formed of thin metal tubing since, when deployed in a collision, the air cushions would be braced together in a column across the width of the vehicle. A mainstay and a backstay would be folded and stowed with the air bags and securely anchored to strong cross members under the floor pan. Canisters, stays, and framework would be concealed behind the upholstery of the vehicle seat and rip out only when deployed during a dangerous collision. Although the air cushions are pressed together by gas pressure, there is a tendency in a frontal collision for the forward-moving torso of a seat occupant to force itself between the air cushions. This is prevented by restraining the bags with the diagonal backstays shown in Figure 1, as well as with the mainstays, which also serve to press the thighs down in the seat cushion and center them. In the preinflation tests described in the next section, this arrangement of occupant restraints dynamically adjusted itself to accommodate a range of body sizes from 5th percentile female to 95th percentile male anthropometric dummies. The air bags are flexible enough to form themselves over a passenger's legs if, prior to the collision, they were crossed or angled sidewise in the seat. Further details concerning the actions of these restraints are given in the U.S. patents (patent numbers 3,981,519 and 3,981,520).

Figure 1. Transverse air-cushion-restraint system.



The potentially active elements of the air-cushion-restraint system include a canister that contains the folded air bag and stays and a manifold for distributing the gas during inflation. These would be concealed in humps at the front of the armrests of the passenger seats, as shown in Figure 2. The driver would have a steering wheel air bag of a conventional kind (11). Either gas generators or bottles of stored air could be used for the passenger air bags. In order to provide protection from any direction, deployment of the air bags would be triggered by sensors that indicate vehicle deformation. These would be located in the front and rear bumper supports and in several door frames. They would signal deployment for collision velocities around 20 km/h. For redundancy, there also should be omni-directional inertial sensors in the front and rear of the vehicle. A signal from any one of these sensors would cause all the air cushions in the vehicle to inflate within 0.04 s. This means that even in a moderately fast collision (such as 50 km/h), the air cushions would blow up in front of a passenger before the torso moved forward significantly.

The inflation pressure in this air-cushion-restraint system is approximately 0.07 kg/cm². Ordinary air bags are porous or have venting valves so that their internal pressures drop rapidly within less than 1 s. The transverse air cushions, however, are constructed of a plastic-coated fabric in order to restrain high pressure for an extended period. It is envisioned that timers would automatically activate release valves after about 5 s. This is a time duration that examination of accident reports indicates would encompass complex, multiple-event accidents.

The strength of the transverse air-cushion system for keeping vehicle occupants in place is derived partly from the mainstays and backstays and partly from the air cushions being forced against each other so as to create a rigid cylinder between the door frames at the sides of the vehicle. An artist's rendering of this situation is shown in Figure 3. For lack of suitable load cells, lateral pressures were not measured during the sled tests of this system. They may be high enough to warrant stiffening the side-frame members in a van.

In addition to superior occupant protection for

Figure 2. Commuter van equipped with air-cushion restraints.

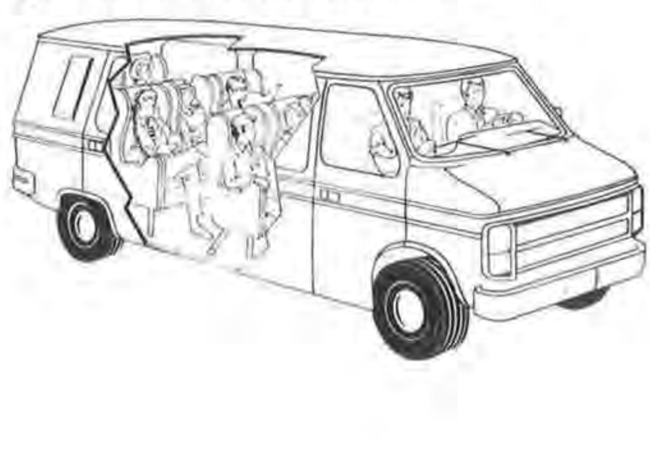


Figure 3. Commuter van with activated air cushions.

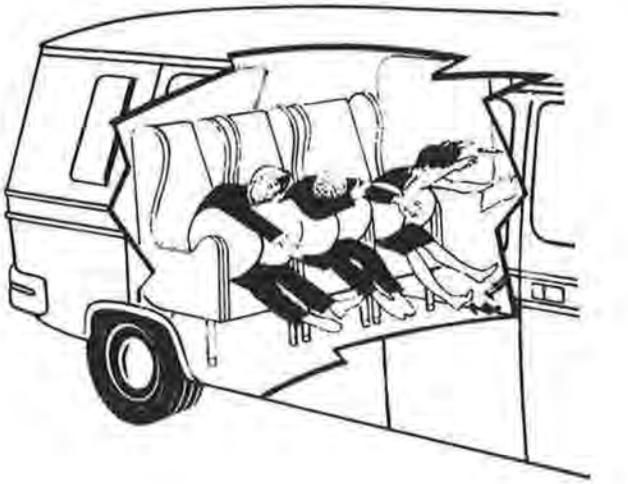
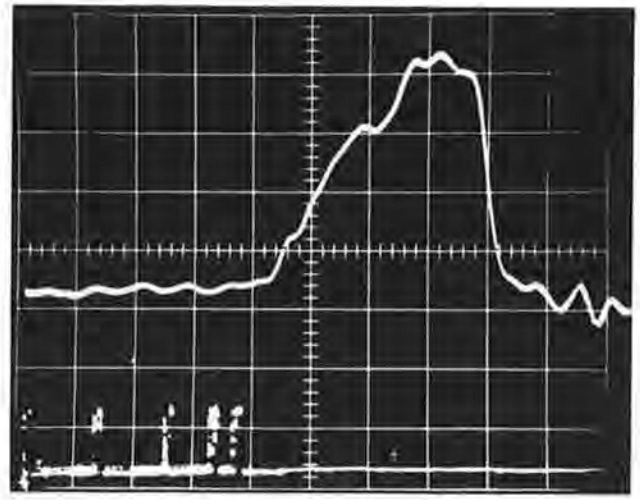


Figure 4. Sled deceleration pulse.



nearly all types of serious collisions, the transverse air cushions have the important advantage of not compromising passenger comfort in any way. The safety devices would be unobtrusive and would not limit the design of comfortable and commodious seating. One such possible arrangement is illustrated in Figure 2.

DECELERATION SLED TESTS

Valid and fully believable tests of the transverse air-cushion restraints would require the entire system to be deployed normally during a full-scale vehicle collision. Unfortunately, we did not have the expertise and resources available to develop all the necessary hardware. However, we did conduct sled tests with a safety seat mock-up that involved preinflated air cushions and anthropometric dummies.

Tests were performed with the Liberty Mutual automotive crash simulator (12). A 2-m sled was accelerated by compressed air over a distance of 2.7 m at up to speeds of 50 km/h (the upper limit of the sled system for realistic testing). After coasting free for 0.6 m, the sled was decelerated by a combination of an aluminum honeycomb and a hydraulic decelerator. Velocity was measured electrically and strain-gauge accelerometer data were transmitted through trailing cables to an instrumentation tape recorder. A typical deceleration pulse [filtered according to Society of Automotive Engineers, Inc. (SAE) standard J211b] measured at the sled frame is shown in Figure 4. It simulates a vehicle-to-barrier collision.

The air cushions employed for these tests were 25 cm thick and 38 cm in diameter (before inflation) and were sewn from plastic-coated cloth. About 5 cm of automotive seat belt material was sewn to the bags to form the mainstays and backstays. In order to maintain inflation pressure accurately while the sled propulsion system was being pressurized, inside bladders made from inner-tube material were employed. Tests indicated that an air-cushion pressure of 0.07 kg/cm^2 was adequate for torso retention at 50 km/h, and it was employed thereafter.

The air cushions in the test set-up were attached to 20-cm circular reaction plates bolted to the arms of the seat mock-up. The latter was constructed of a structural steel channel and fitted with plastic-foam cushions covered with vinyl. The air cushions were centered on an axis about 35 cm forward of the seat H-point (SAE standard J1100) and 20 cm higher. Sierra 5th percentile female, 50th percentile male, and 95th percentile anthropomorphic dummies were employed. The test seat is shown in Figure 5.

The dynamic action of the air-cushion restraints and dummy motions are best studied by high-speed motion-picture photography. For this purpose, a rotary-prism Fastax camera was employed at 400 frames/s. A significant observation from the test films is that rebound velocity of a dummy is less than one-third of the entrance speed into the restraint system. At 48 km/h, the maximum rebound observed was 13 km/h, which would present no safety hazard.

Manually synchronized still photographs were also made, and typical examples are shown in Figures 6 and 7. They illustrate how the upper thighs, abdomen, and chest of the human surrogates are wrapped around the air-cushion restraint; thus, the collision force is distributed harmlessly over a wide area of the torso. Pelvic and chest decelerations in the dummies were well within survivability limits and there were no lacerative injuries or any other skin damage.

DISCUSSION

The crash simulations just described are worst-case tests because frontal collisions are the most frequent and the most dangerous (13) of all automotive accidents. Therefore, it is strong evidence of their potential value that the transverse air cushions successfully restrained representative human surrogates in realistically simulated collisions at speeds up to 48 km/h and in a manner that most likely would not have harmed their human counterparts. Insofar as sled tests can be considered realistic, these were crucial tests of this restraint concept. Since tests for an out-of-position occupant, side collision, and rollover would by no means be as formidable as the frontal collision, our inability to perform them is not considered critical at this stage.

Most of the essential design elements in the transverse air-cushion-restraint system concept, notably collision sensors and air-bag inflators, have been fully developed elsewhere. The only important item as yet undeveloped is the air-cushion material itself, which must be (a) strong, (b) able to remain flexible while stowed tightly for years, (c) capable of deployment within 40 ms, (d) able to

Figure 5. Air-cushion test seat mounted on deceleration sled.

Figure 6. Sled test with 5th percentile dummy.

Figure 7. Sled test with 50th percentile dummy.

maintain adequate pressure for several seconds, and (e) manufacturable in the form required. Informed vendor opinion is that such a material is feasible, although casual inquiry did not develop a source. Availability of such a material is obviously a prerequisite to further development of this restraint system.

CONCLUSION

Transverse air-cushion restraints would offer crash injury protection to van and small-bus occupants without the necessity of having them buckle up, which experience indicates few riders will do.

The proposed system differs from conventional air bags in the following respects:

Difference	Benefit
Side-acting	Lack of forward body thrust obviates need for strong seats; tends to center occupant
Prolonged inflation	Protects in side impacts and rollovers
Backstays	Centers occupant in seat
Mainstays	Maintains torso position and low profile

Although the transverse air-cushion-restraint system has not been completely developed, its dynamic capabilities have been demonstrated in realistically simulated crash tests, and construction of fully operable hardware appears to be feasible. Accordingly, it is proposed as a practical occupant-restraint concept for passenger vans and small buses where large numbers of riders are at risk without adequate crash injury protection.

REFERENCES

1. Economic Feasibility of Independent Vanpool Operations. Graduate School of Management, Univ. of California, Los Angeles, Sept. 1976, pp. 1, 5. NTIS: PB-266 577.
2. D.A. Maxwell and J.P. McIntyre. Economics of Vanpooling. TRB, Transportation Research Record 724, 1979, pp. 52-57.
3. An Analysis of Commuter Van Experience. Urban Institute, Washington, DC, Feb. 1976, p. 24. NTIS: PB-252 304.
4. R.D. Owens. Getting to Work at 3M Center. Traffic Safety, Chicago, March 1977, p. 22.
5. Unwarranted Delays by the Department of Transportation to Improve Light Truck Safety. Comptroller General Report to the U.S. Congress, U.S. General Accounting Office, Rept. CED-78-119, July 6, 1978, pp. ii, 6.
6. H.W. Sherman and D.F. Huelke. Clinical Case Reviews of Light Truck and Van Crashes. SAE, Warrendale, PA, March 1979, pp. 1, 24.
7. Special Study: Safety of Multipurpose Vans. National Transportation Safety Board, U.S. Department of Transportation, Rept. NTSB-HSS-79-1, March 22, 1979, pp. 1, 37.
8. P.F. Waller and others. Safety Belts: The Uncollected Dividends. U.S. Department of Transportation, Rept. DOT HS-802 598, May 1977, pp. iii, 8-2.
9. The Case for Seat Belts. National Highway Traffic Safety Administration, U.S. Department of Transportation, Jan. 1973, pp. 2, 5.
10. Highway Death Rate Up: Safety Belt Use Ignored. National Highway Traffic Safety Administration, U.S. Department of Transportation, newsletter, Jan. 9, 1981, p. 1.
11. D.J. Romeo. Driver and Passenger Air Bag Unit

Assemblies. SAE, Warrendale, PA, Rept. 800293, Feb. 1980, p. 3.
12. N.H. Pulling and others. Liberty Mutual Automotive Crash Simulator. SAE, Warrendale, PA, Rept. 700405, April 1970.
13. Fatal Accident Reporting System. National Center for Statistics and Analysis, U.S. Department of Transportation, 1979, pp. 1, 13.

Publication of this paper sponsored by Committee on Highway Vehicle Research.

Positive-Guidance Demonstration Project at a Railroad-Highway Grade Crossing

PETER S. PARSONSON AND EDWARD J. RINALDUCCI

A rural railroad-highway grade crossing used exclusively by local drivers was selected for a positive-guidance demonstration project. Initially, and throughout the project, the crossing was controlled by STOP signs because of sight-distance restrictions. The project first upgraded the motorist-information system at the site by installing improved advance-warning signs and markings that conform to the Manual on Uniform Traffic Control Devices. (The crossing is on a minor county road, not on the state system.) Then, application of the positive-guidance procedure resulted in the addition of rumble strips on both approaches and LOOK FOR TRAIN and HIDDEN XING signs on the more restricted approach. The as-is condition and the two levels of improvement were evaluated by observing the extent to which drivers slowed down at a safe rate, stopped at a safe distance from the track, and looked both ways before crossing. It was found that the positive-guidance procedure was workable, and the project yielded field-tested schemes for evaluation that can be recommended to other agencies that desire to use the positive-guidance system at grade crossings. However, overall, the project did not produce an improvement in driver behavior. In fact, the rumble strips induced some swerving into the oncoming lane and may have been responsible for the observed increase in vehicles crossing at reckless speeds. These findings, while essentially negative, are important because they document the difficulty of influencing drivers who are thoroughly familiar with a road. Rumble strips should be reserved for nonresidential areas where unfamiliar drivers are numerous.

Positive guidance is a set of rational steps developed during the 1970s by the Federal Highway Administration (FHWA) (1,2) to provide drivers with sufficient information where they need it and in a form that they can best use to avoid hazards. It combines highway engineering and human factors technologies to produce an information system matched to facility characteristics and driver attributes. Positive guidance often provides high-payoff, short-range solutions to safety and operational problems at relatively low cost. The procedure consists of six major functions, which are as follows:

1. Data collection at problem locations,
2. Specification of problems,
3. Definition of driver-performance factors,
4. Definition of information requirements,
5. Determination of positive-guidance information, and
6. Evaluation.

SITE CONDITIONS

This paper describes a positive-guidance demonstration project funded by FHWA through a contract with the Georgia Department of Transportation (DOT). A railroad-highway grade crossing in rural Georgia was the site of the project. The crossing is used exclusively by local drivers and was controlled by STOP signs throughout the project. The project first upgraded the motorist-information system at the site by installing improved advance-warning signs and markings that conform to the Manual on Uniform Traffic Control Devices (MUTCD) (3). (The crossing is on a minor county road, not on the state system.) Then, the positive-guidance procedure was applied to determine the type and location of additional improvements and modifications in the overall information system. Rumble strips and certain nonstandard signs were identified and installed. The as-is condition and the two levels of improvement were evaluated for improvement in driver performance. This paper briefly summarizes the final report to the Georgia DOT (4).

Figure 1 is a condition diagram for level 1, the as-is condition. Stanley Road, located northwest of Atlanta in Kennesaw, is a rural 18-ft two-lane road that crosses the L&N Railroad with poor sight distance from both roadway approaches (Figure 2). Before any improvement, the crossing had warning and protective devices that consisted of two STOP signs, one stop-ahead sign, and a wood crossbuck that faced in both directions. Figures 3 and 4 give further indication of the sight-distance restrictions due to fences, trees, and hillocks in the four quadrants of the crossing. These figures also show the stations of the hidden observers who collected performance data during the project.

Stanley Road has an average daily traffic (ADT) of 1100 vehicles/day and is used only by local drivers. The speed limit is 35 mph. The crossing averaged only about 8 trains/day during daylight hours when data were collected. County records over a number of years showed no accidents at this site. Federal records listed a recent accident when a train struck an unoccupied car that had stalled on the track on a rainy night. The Peabody-Dimmick hazard-index formula, which considers traffic volume, train volume, and level of crossing protection, predicts over eight accidents over a five-year period. Train arrivals were entirely unpredictable; drivers had no expectancy as to when a train might arrive. Train speed varied widely from 5 to 30 mph, depending on block signals.

Pilot observations showed that a significant fraction of the motorists ignored the STOP sign and slowed down no more than necessary to negotiate the crossing, i.e., between 20-25 mph. With sight distance limited as it is, they relied entirely on the locomotive engineer's duty to sound the horn for the crossing. Those who stopped did so too close to the tracks. The site typified the classic problem of the inattentive local motorist who lacks respect for the danger of a crossing.

Figure 1. Condition diagram for level 1 (as-is) condition.

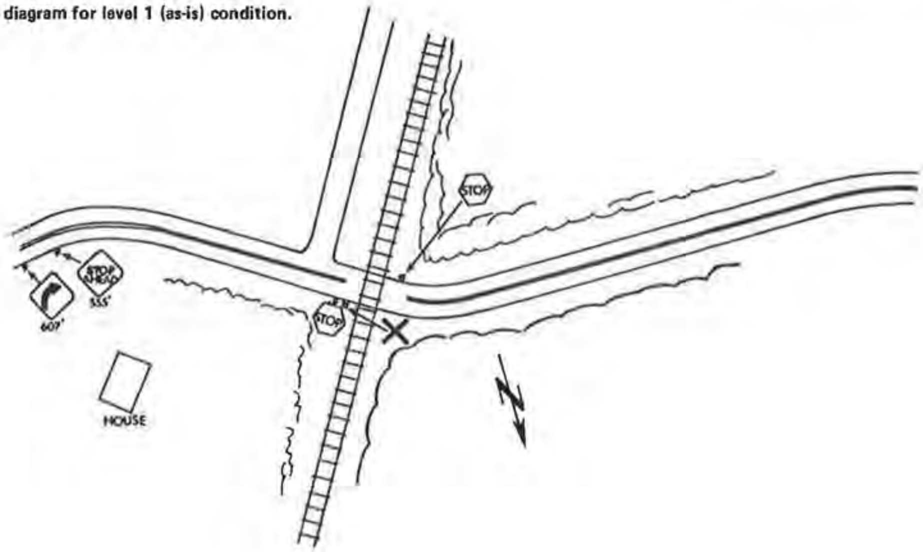


Figure 2. Sight distance graph.

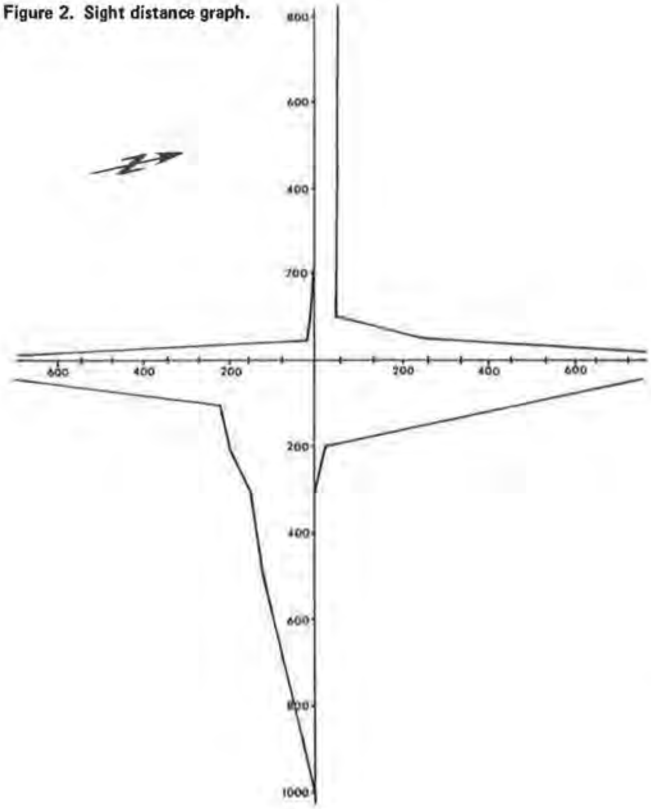


Figure 3. Sketch of west (eastbound) approach.

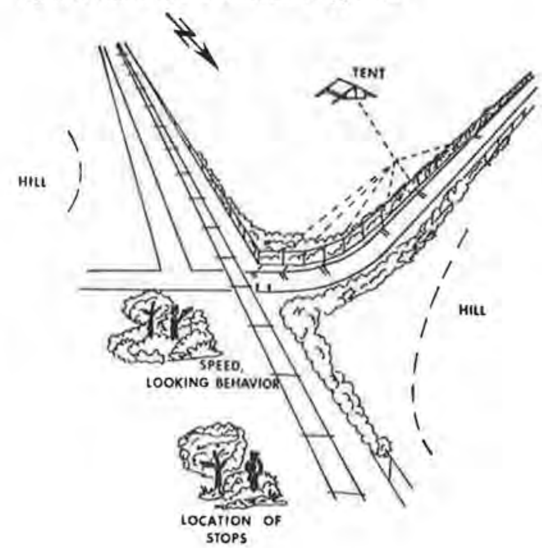
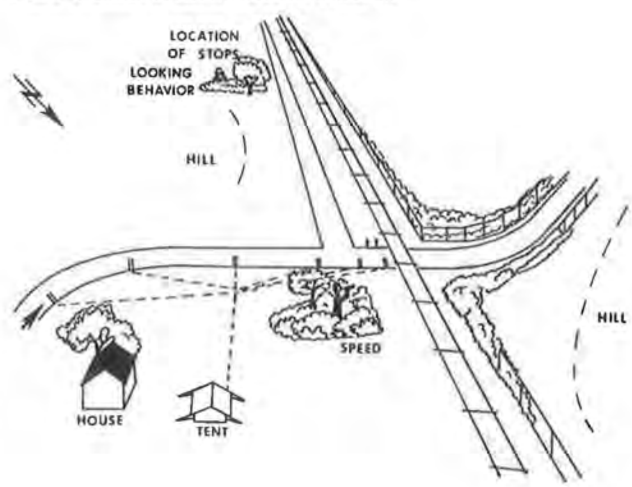


Figure 4. Sketch of east (westbound) approach.



EVALUATION DESIGN

Driver performance was evaluated for level 1 (the as-is condition) by obtaining the following data for each approach for two weekdays, a Saturday, and a Sunday, during daylight hours.

1. Driver looking behavior. Hidden observers with binoculars noted whether drivers turned their heads to the left and/or the right. It was not attempted to judge whether the head motion was for the purpose of seeing a train.

2. Location of stop, for those vehicles that do stop. Three zones (0-10 ft, 10-20 ft, and more than 20 ft) were identified on the assumption that at least 10 ft of clearance from the nearer rail to the front bumper of the vehicle are required for safety. The stop lines that would be added later in level 2 should be placed 15 ft from the track, according to MUTCD, so the zone of 10-20 ft represents normal and safe compliance.

3. Speed profiles. Speed profiles were obtained by using pairs of tapeswitches located 10, 50, 100, 200, 300, and 500 ft from the crossing. Figure 5 (5) illustrates the concept of speed profiles for grade crossings protected by a crossbuck or a STOP sign. The Georgia Institute of Technology used a RATEM II microprocessor, designed by Ken G. Courage of the University of Florida, to record the tapeswitch closures and print out statistical summaries on paper tape in the field. Figures 3 and 4 show the tapeswitches in place on each approach, with wires leading off to a small tent out of sight of the motorists, where the microprocessor and printer were located. Tapeswitches are all but invisible to motorists at speed and are quite silent and unobtrusive. Ignored by motorists during level 1 and 2 evaluations, the tapeswitches became a target for vandals after the positive-guidance solution was installed.

4. Crossing speeds. Crossing speeds of all lead vehicles were obtained by a hidden observer with a radar gun. The minimum speed as the vehicle approaches and crosses the tracks is recorded. The gun may be designed to blank out its display at speeds below 3 mph, in which case a zero is recorded.

The results of the level 1 evaluation are reported later, along with the level 2 and level 3 results.

The site was then upgraded to level 2 to conform with the MUTCD by adding stop lines and advance-warning signs and markings as shown in Figure 6. An acclimation period of 30 days was allowed to pass before driver-performance data were taken again. It was obvious to the project staff that driver performance under level 2 conditions was still inadequate. A positive-guidance solution (level 3) was clearly still needed to improve driver behavior.

POSITIVE-GUIDANCE SOLUTION

The eastbound approach offers an especially poor view of the track from a distance. It was decided to install a LOOK FOR TRAIN sign 150 ft from the crossing and a HIDDEN XING plate below the railroad (RR) advance-warning sign 435 ft from the track (Figure 7; note, shaded areas are proposed rumble strips). The word hidden was specifically considered

Figure 5. Concept of speed profiles for crossbuck and STOP sign grade crossings.

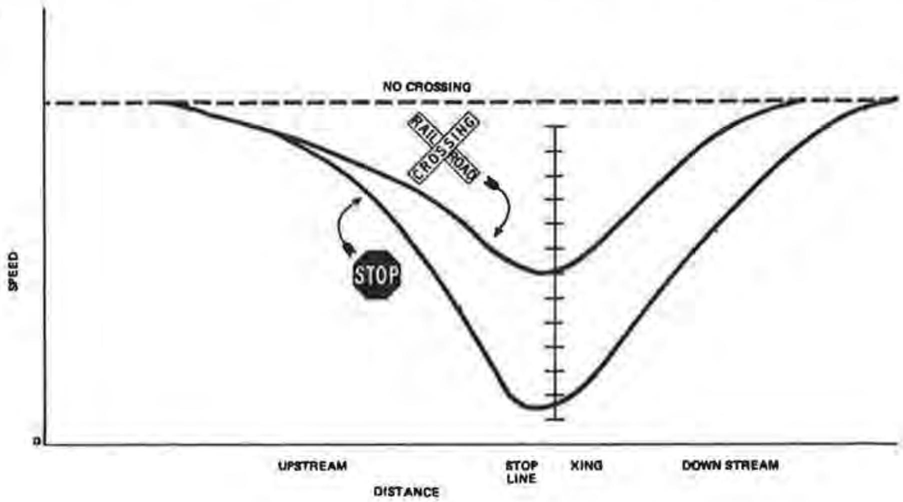
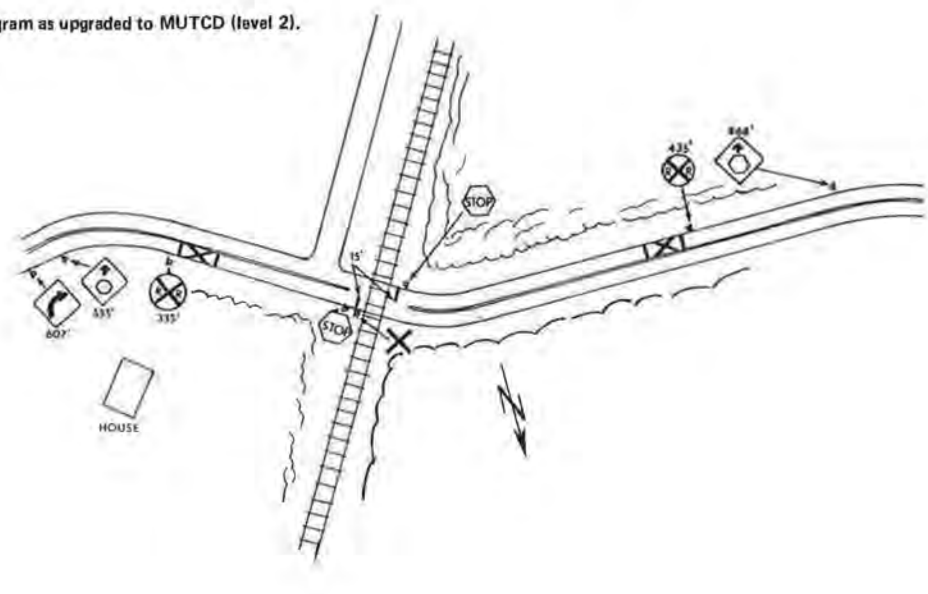


Figure 6. Condition diagram as upgraded to MUTCD (level 2).



preferable to blind. No new signs or markings were recommended for the westbound approach.

It was decided to install three rumble strips on each approach to the crossing in order to call attention to the warning signs. Rumble strips offer a cross-modality stimulation that is both tactile and auditory. They would appear to offer an attractive reinforcement to the visual stimulation of the signs (provided they are not used everywhere). They are formed of corrugations 0.75-in deep. They should not be used indiscriminately in residential areas, such as the project's westbound approach, as their noise can annoy and disturb neighbors. Furthermore, a local driver who uses the route frequently may avoid the noise and vibration by crossing the centerline into the lane used by oncoming traffic. Such a maneuver could pose a greater hazard than the trains.

The warning signs and rumble strips were installed at locations determined through the positive-guidance procedures. Figure 8 shows, for each approach, the three zones that correspond to the nature of the tasks the driver must perform when approaching the crossing. The approach zone corresponds to the decision sight distance (6) minus the desirable stopping sight distance (7). The nonrecovery zone begins at the point beyond which there is insufficient stopping sight distance. It was prepared for a design speed of 35 mph. An analysis of information needs and zone assignments led to the conclusion that, on the eastbound approach, a rumble strip was needed at the beginning of the nonrecovery zone (250 ft from the crossing) to call attention to the LOOK FOR TRAIN sign 100 ft ahead. Similarly, another rumble strip was installed 100 ft in advance of the RR advance-warning sign, with its HIDDEN XING plate. The third rumble strip was installed 100-ft upstream of the symbol stop-ahead sign. The same procedure was followed in locating the three rumble strips on the westbound approach.

Figure 7 shows a diagrammatic left-turn railroad-crossing sign recommended for the road that runs parallel to the track.

RESULTS

Table 1 reports the looking (head-turning) behavior for levels 1, 2, and 3, which were the as-is, MUTCD, and positive-guidance conditions, respectively. The table shows that looking behavior became progres-

Figure 7. Diagram of proposed positive-guidance solution.

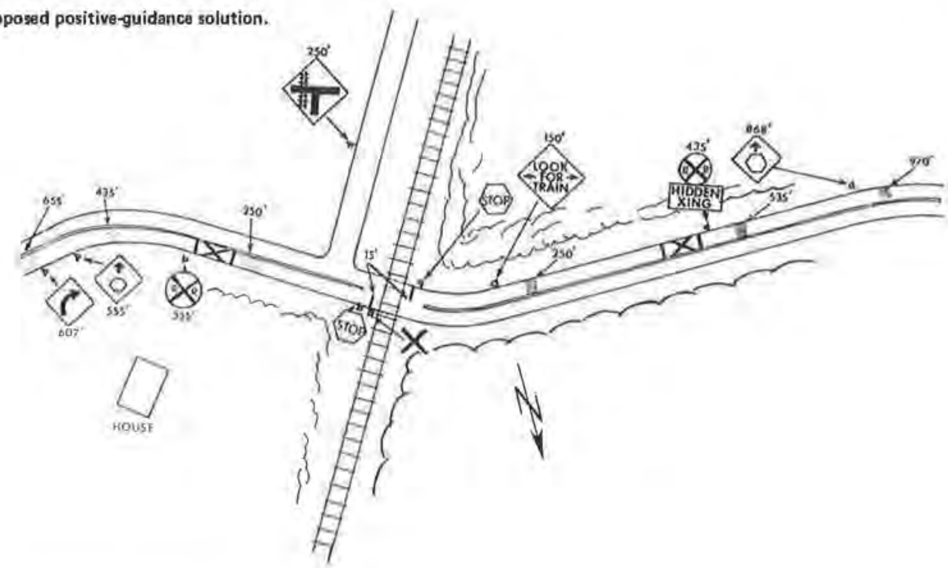
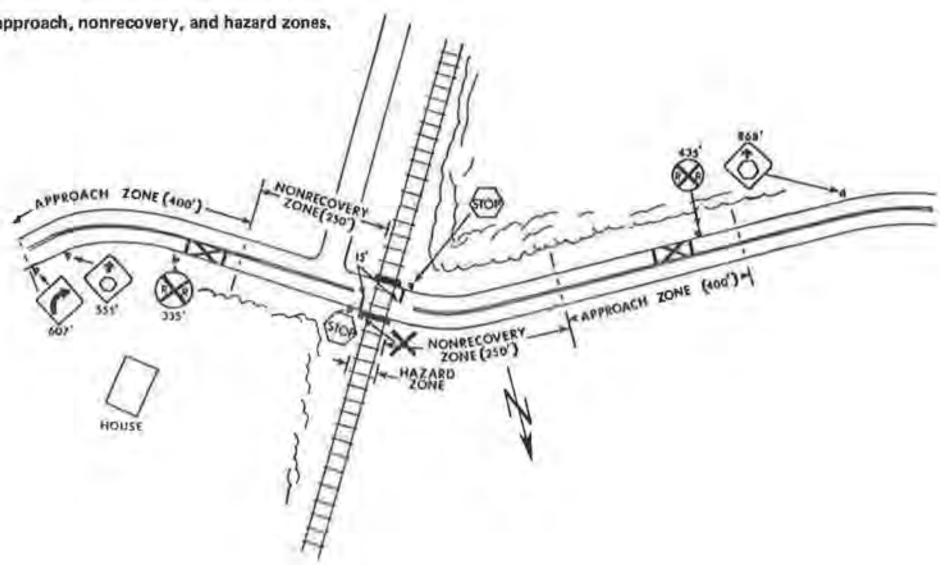


Figure 8. Plot of approach, nonrecovery, and hazard zones.



sively worse from level 1 to 2 to 3. A chi-square test showed that these results were highly significant statistically.

Table 2 shows the changes in stopping location and the percentage of vehicles not stopping at all for the three levels of control. For both eastbound and westbound vehicles, the percentage of vehicles not stopping decreased dramatically after the installation of the positive-guidance solution. The location of stop improved appreciably for the westbound flow but stayed about the same for the eastbound movement.

Speed profiles were determined by tapeswitch from 7:00 a.m. to 6:00 p.m. on two weekdays, a Saturday, and a Sunday, just the same as the other types of data. Each approach averaged about 300 vehicles during an 11-h period. No significant difference from day to day was found. The 85th percentile speeds for levels 1, 2, and 3 at each of the six distances from the track are summarized in Table 3. (The 85th percentile speed was preferred to the mean as an indicator of the performance of the relatively reckless drivers.) It is immediately apparent that there are no differences to speak of among the three levels. The speed at 200 ft from the crossing is considered to be especially important, as the safe driver will have reacted to the advance-warning signs before reaching that location (8). Table 3 shows that the speeds at 200 ft, on both approaches, were unaffected by the two levels of improvement and were approximately as high as the posted speed limit.

A concealed observer with a hand-held radar speed meter determined the lowest speed of each vehicle as it approached and crossed the track. The means of these speeds are not of much interest, as the average driver typically operates his or her vehicle in a safe manner. Tables 4 and 5 examine the speeds of the faster vehicles, which are those more likely to be involved in an accident. The tables show that the fast group of eastbound drivers (≥11 mph) held steady at only about 7 percent of the stream. However, the truly reckless drivers, crossing at more than 25 mph, increased eightfold to almost 1 percent of the total. The fast group of westbound drivers increased their proportion of the stream from 14 to 22 percent, and reckless drivers tripled to about 0.5 percent of the total.

This rural area is very conservative politically. Perhaps some of these reckless drivers were rebelling against noisy, annoying rumble strips by doing the opposite of what they know the authorities are trying to get them to do. About six drivers/day in each direction were observed to avoid the rumble strips by crossing the centerline into the opposing lane.

Table 1. Looking behavior for levels 1, 2, and 3.

Behavior	Percentage of Drivers					
	Level 1		Level 2		Level 3	
	Yes	No	Yes	No	Yes	No
Eastbound						
Looked left	85.9	14.1	82.7	17.3	82.1	17.9
Looked right	85.9	14.1	81.1	18.9	77.9	22.1
Looked one/both directions	89.9	10.1	86.2	13.8	85.9	14.1
Westbound						
Looked left	88.1	11.9	76.2	23.8	68.5	31.5
Looked right	86.7	13.3	75.2	24.8	63.8	36.2
Looked one/both directions	91.2	8.8	81.3	18.3	75.2	24.8

CONCLUSIONS AND RECOMMENDATIONS

The primary objective of this demonstration project was to test the positive-guidance procedures for applicability to the problem at a highway-railroad grade crossing with restricted sight distance. A secondary goal was to achieve an improvement in traffic operation at the site.

The project staff had no difficulty in applying the positive-guidance procedures. The procedures may seem unnecessarily comprehensive for this simple site; however, agencies concerned with mounting numbers of lawsuits should find that time spent documenting positive-guidance procedures will pay for itself many times over in reduced liability.

The evaluation procedures purposely were more ex-

Table 2. Chi-square analysis of stopping zone.

Zone (ft)	Percentage of Vehicles in Zone			Zone (ft)	Percentage of Vehicles in Zone		
	Level 1	Level 2	Level 3		Level 1	Level 2	Level 3
Eastbound[a]				Westbound[b]			
0-10	18.0	9.5	24.3	0-10	7.7	3.6	10.4
11-20	39.6	28.9	60.4	11-20	14.4	12.2	51.3
More than 20	3.7	5.6	9.8	More than 20	6.8	6.7	25.7
No stop	38.7	56.1	5.4	No stop	71.2	77.5	12.6

[a]Chi-square = 313.3, significant at 90 percent, and probability of error (type I error) less than 0.
[b]Chi-square = 368.6, significant at 90 percent, and probability of error (type I error) less than 0.

Table 3. Speeds from tapeswitch data.

Item	85th Percentile Speed[a] (mph)			Item	85th Percentile Speed[a] (mph)		
	Level 1	Level 2	Level 3		Level 1	Level 2	Level 3
Eastbound				Westbound			
Distance from track (ft)				Distance from track (ft)			
10	11	12	11.5	10	13.5	14	14.5
50	20	19	20.5	20	19.5	18	20
100	28	17.5	28	100	26.5	27	27.5
200	37	36	36	200	34	33	34
300	40	40	40	300	35	35	35
500	43.5[b]	43[b]	41[b]	500	38	37	37
Sample size	1999	1216	1222	Sample size	1333	1210	1147

[a]Avg of four days.
[b]Speeds above about 40 mph are not precise because of the speed range programmed for the microprocessor

Table 4. Percentage stratification of lowest approach speeds into speed groups.

Item	Speed Group (mph) 0-5	6-10	11-15	16-20	21-25	>25
Eastbound[a]						
Level 1	66.0	27.0	5.3	0.8	0.8	0.1
Level 2	64.3	29.2	4.8	1.4	0.4	0.1
Level 3	66.7	26.0	5.0	1.1	0.4	0.8
Westbound[b]						
Level 1	39.2	47.0	11.8	1.5	0.3	0.2
Level 2	48.5	37.4	11.3	2.4	0.5	0
Level 3	41.9	36.5	13.9	5.4	1.7	0.6

[a]Chi-square = 16.8 with 10 df; significance = 0.0784.
[b]Chi-square = 55.9 with 10 df; significance = 0.000.

Table 5. Stratification of lowest approach speeds into groups of fast and slow.

Item	Percentage of Vehicles: Slow (<11 mph)	Fast (>11 mph)	Sample Size[a]	Item	Percentage of Vehicles: Slow (<11 mph)	Fast (>11 mph)	Sample Size[a]
Eastbound[b]				Both approaches combined[d]			
Level 1	93.0	7.0	769	Level 1	90.0	10.0	1417
Level 2	93.4	6.6	1368	Level 2	90.0	9.6	2271
Level 3	92.7	7.3	857	Level 3	86.5	13.5	1525
Westbound[c]							
Level 1	86.2	13.8	648				
Level 2	85.9	14.1	903				
Level 3	78.4	21.6	668				

[a]These sample sizes pertain to Tables 1, 2, and 4, also. They are smaller than the sample sizes in Table 3 because only the lead vehicle in a platoon was recorded.
[b]Chi-square = 0.4264 with 2 df; significance = 0.808.
[c]Chi-square = 19.83 with 2 df; significance = 0.000.
[d]Chi-square = 15.92 with 2 df; significance = 0.0003.

tensive than any operating agency would use, as it was desired to determine which of the procedures is most cost effective. It is recommended that tapeswitches not be used for routine evaluation because of the complexity of the equipment and the susceptibility to vandalism. It is recommended that observers not be deployed to determine the percentage of motorists stopping and the zone of stop. These judgments are too subjective and raise serious questions of inter-observer repeatability.

It is recommended that evaluations of this type be performed by two observers located as close to the road as possible (that is, not at a side location down the tracks). One observer holds a radar speed meter and records the minimum speed of all oncoming vehicles and as many vehicles moving away from him or her as convenient. The second observer holds 8-power binoculars (or possibly a 20-power telescope for unusual distances) and observes head movements of oncoming vehicles only. This observer should be instructed to look for a "Fuzzbuster" on the dash.

Stanley Road is used entirely by local motorists. It may be impossible to change their behavior at the crossing short of installing gates, lights, and bells. Overall, driver performance did not improve as a result of the two levels of upgrading. In fact, the rumble strips induced some swerving into the oncoming lane and may have been responsible for the observed increase in vehicles crossing at reckless speeds. Recommendations in the literature (9) for rumble strips at grade crossings should be amended to apply only to nonresidential locations where unfamiliar drivers are numerous enough to form the target population.

ACKNOWLEDGMENT

We are grateful to Donald Mills of Georgia DOT and Harold Lunenfeld of FHWA for their advice and assistance on this project. Mills was the project monitor for the Georgia DOT

This paper reflects our views, and we are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Georgia DOT or FHWA. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. A Users' Guide to Positive Guidance. Office of Traffic Operations, FHWA, June 1977.
2. H. Lunenfeld. Evaluation of Traffic Operations, Safety, and Positive Guidance Projects. Human Factors Branch, Office of Traffic Operations, FHWA, Oct. 1980.
3. Manual of Uniform Traffic Control Devices. FHWA, 1978.
4. P.S. Parsonson and E.J. Rinalducci. Positive Guidance Demonstration Project, Final Report. School of Civil Engineering, Georgia Institute of Technology, Atlanta, April 1981, 107 pp.
5. Safety Features of Stop Signs at Rail-Highway Grade Crossings, Final Report. Offices of Research and Development, FHWA, March 1978, p. 72.
6. H.W. McGee and others; Biotechnology, Inc. Decision Sight Distance for Highway Design and Traffic Control Requirements, Final Report. Offices of Research and Development, FHWA, Rept. FHWA-RD-78-78, Feb. 1978, 65 pp.
7. A Policy on Design Standards for Stopping Sight Distance. American Association of State Highway Officials, Washington, DC, 1971, 15 pp.
8. J. Coleman and others. Railroad Grade Crossing Passive Sign Study. Transportation Engineering, Nov. 1977, pp. 15-18.
9. D.W. Schoppert and D.W. Hoyt. Factors Influencing Safety at Highway-Rail Grade Crossings. NCHRP, Rept. 50, 1969, pp. 100-101.

Publication of this paper sponsored by Committee on User Information Systems.

Effectiveness of Changeable Message Signing at Freeway Construction Site Lane Closures

FRED R. HANSCOM

The effectiveness of changeable message signing (CMS) devices in advance of freeway construction and maintenance zone lane closures was evaluated. Operational traffic behavior and driver interview data were gathered in four states. Before-and-after studies of baseline (no CMS) versus CMS application consistently demonstrated increased advanced preparatory lane-change activity, smoother lane-change profiles, significantly fewer late exits (within 100 ft of closure), and reduced speeds at the lane-closure point to be associated with CMS use. The most preferable CMS location was found to be 0.75 mile (1.2 km) in advance of the lane closures. Of three tested device types (one-line bulb matrix, two-line rotating drum, and three-line bulb matrix), the large, obtrusive three-line bulb matrix tended to produce more advance lane-change behavior; however, no difference in the hazardous late exit maneuvers was observed between types. All three were equally effective in eliciting speed reductions at the entrance to the lane closure. Driver interview data tended to favor the three-line device due to its greater information display capacity. Message combinations of speed, lane closure, and merge advisories were tested on the devices. Although lane-change behavior of the total traffic stream did not significantly differ between message conditions, interviewed motorists favored the speed and lane-closure message combination as being most helpful, providing most response time, and meeting information needs. The study recommends CMS applications as a supplement to standard device schemes but not as a substitution for the arrowboard. Suggested cost-efficient CMS applications involve (a) short-term closures characterized by decreased driver expectancy, (b) traffic volumes of 900 vehicles/h or greater, and (c) limited sight distance to the closure.

Highway construction activity or other types of incidents (e.g., accidents, unexpected road obstructions, and maintenance activities) frequently require the closure of one or more traffic lanes. Although the Manual on Uniform Traffic Control Devices (1) describes recommended treatments for typical lane closures, there is a need for improved methods of providing advance information to the motorist. The need for this research is emphasized by the current trend for highway rehabilitation projects, many of which require lane closures. Accident experience at lane-closure locations, especially on high-speed facilities, demonstrates the need for better guidance for the motorist and protection of the worker.

OBJECTIVES AND APPROACH

The objective of this research was to determine the effectiveness of changeable message signing (CMS) applications at lane closures on high-speed freeways. Right- and left-lane closure situations were studied and observations were made under day and night conditions. By using field studies at selected lane-closure sites, this research examined traffic performance effects of various changeable message displays. In addition, a sample (N = 489) of driver responses (detection, comprehension, and interpretation) was obtained. This applied methodology examined appropriate relations between driver information processing and vehicle behavior required for validating operational measures of CMS effectiveness.

FIELD-STUDY METHODS

Two separate procedures were applied to study CMS effects at planned lane closures. Manual coding of vehicle performance was applied to gather traffic operational responses to the CMS alternatives, and in-vehicle questionnaires were administered to test subjects to obtain sensitive measures of driver response.

Traffic-Operations Measurement

Manual observations of vehicle speed and lane distributions (proportions of traffic in the closed and through lanes) were obtained at the following data-collection points on the approach to freeway lane closures:

1. Advance--The advance point was selected in advance of the sight distance to the CMS, approximately 1 mile (1.6 km) before the lane closure. The purpose of collecting data at this point was to determine behavior of traffic not influenced by the CMS.
2. CMS Point--The CMS location was either 2000 ft (600 m) or 0.75 mile (1.2 km) (the two tested CMS placements) in advance of the taper. Data were gathered here to determine the advance effect of the CMS.
3. Intermediate--Midway between the CMS and taper, the intermediate point defined the lane-change profile effect of the CMS.
4. Taper--The most critical collection point was 100 ft (30 m) in advance of the first taper channelizing device. Data gathered here revealed the level of hazardous late exit behavior.

This uniformity of data-collection points between CMS test locations permitted limited combining of data for the purpose of comparing CMS effects. Time of day for data collection was also controlled in order to eliminate its possible confounding effect.

Both speed and lane-distribution data were sampled within 30-min data-collection intervals. This incremental observation procedure permitted the monitoring of interactive effects of speed and volume changes as conditions fluctuated throughout the data-collection day.

In-Vehicle Driver Response

A driver questionnaire was completed by subjects who participated in a controlled field study staged at construction sites observed in the traffic-operations study. These subjects were not aware that they were participating in a study specifically related to highway construction zone signing until they had completed a considerable portion of the questionnaire. Lane-change behavior and driving speeds were unobtrusively recorded and subsequently matched to questionnaire responses.

The applied questionnaire strategy involved first asking a series of general questions regarding observations of traffic-control devices that the drivers had passed. Although answers to these questions were provided prior to the subjects being directly asked about their CMS observations, the answers nevertheless reflected a direct impact of the CMS. This provision of the survey afforded an internal response-validation mechanism.

Completed questionnaires void of missing data items were obtained for a sample of 489 drivers. Age and sex distributions of the sample did not significantly differ between states. Age and sex distributions were controlled so as to approximate normal exposure rates. The sample included substantial proportions of drivers younger than 20 and older than 60 while maintaining a nearly even male-female distribution.

Tested CMS Conditions

A review of current use revealed general characteristics of available CMS devices applicable for construction zone traffic management. Three device types that represent a variety of available characteristics were applied in this study. Three message capacities (one, two, and three lines) were tested, which represent two display types (bulb matrix and rotating drum). Figure 1 summarizes tested conditions (note: 1 ft = 0.3 m, and 1 mile = 1.6 km).

Field-Test Scenario

The study procedure had to accommodate a variety of constraints. First, it was not possible to stage construction activity for the purpose of controlling necessary site conditions (e.g., highway geometry and traffic volume). Therefore, the study procedure could be applied only at existing construction sites. Second, it was not possible to test all CMS devices at one site. The research team was dependent on CMS manufacturers and state agencies for providing the devices and, therefore, constrained to specific locations and data-collection times. Finally, at one site it was not possible to test a baseline (no CMS) condition because of possible liability consequences to the state agency.

These locational and CMS device constraints required that the applied field-test scenario (see Table 1) use a variety of data bases. Existing differences between data bases (e.g., varying traffic-control-device standards between states) dictated complete reliance on within-site data analysis. Therefore, in the interest of statistical validity of the analysis, adequate sample sizes were gathered at each site.

Analysis of the data addressed the four CMS effect issues identified in Table 1. The effect of CMS device application was determined at sites that initially contained standard (no CMS) traffic-control-device schemes via a before-and-after study of each tested CMS device. Placement conditions (including use of more than one CMS) were tested as the result of the simultaneous availability of two devices at one site. Three placement alternatives that varied CMS location with respect to the lane closure were as follows.

1. 0.75-mile advance placement,
2. 0.75-mile and 2000-ft placements, and
3. 2000-ft placement.

A variety of message conditions were tested in one state that routinely applied CMS devices. The following message types were permitted to be specified:

1. Speed and closure advisory,
2. Speed and merge advisory,
3. Merge and closure advisory, and
4. Closure advisory.

FIELD-STUDY RESULTS

Results are separately discussed for the three applied field measure types: traffic distribution across lanes, speed measurements, and in-vehicle responses.

Lane-Distribution Results

The relative proportions of traffic in the through and closed lanes that approach construction zone lane closures were observed for a sample of more then 196 500 vehicles. Data gathered in three states (Georgia, Colorado, and California) were used to compare these lane distributions between baseline (no CMS) conditions and various CMS applications. A fourth data set, gathered in South Carolina, was used to determine the relative effects between certain CMS message alternatives (i.e., speed and closure, speed and merge, and closure and merge advisories) and various placement configurations (i.e., one CMS at 2000-ft or 0.75-mile advance placement, and two CMS devices, one at each advance location).

A number of findings evolved from this analysis. CMS application was consistently shown to improve lane-distribution profiles (e.g., increased advance preparatory lane changing) on the approach to construction sites, and certain findings evolved regarding specific CMS characteristics. Findings are now discussed for each of the CMS effects noted in the field-test scenario (previously summarized in Table 1).

Application

Consistent results between baseline (no CMS) and CMS conditions based on data collected in Georgia, Colorado, and California demonstrated improved lane-distribution profiles following the application of CMS at both right- and left-lane closures.

Figure 2 shows lane-distribution profiles of baseline CMS effects observed for one- and two-line CMS devices in Georgia and Colorado, respectively. Because distinctly different baseline profiles were noted, it would be inappropriate to combine these data across sites for illustrative purposes. Higher volumes noted for the Colorado sites likely explained the increased early exiting from the closed lane. As can be seen from the figure, application of CMS devices was associated with decreased closed-lane proportions of traffic at all three data-collection points within 0.75 mile of the closure.

Two CMS conditions were compared with baseline conditions at one site, the results of which are shown in Figure 3. Dramatic reductions in the proportions of vehicles that remain in closed lanes were observed for both conditions. Differences observed between specific CMS conditions of placement, message type, and format are discussed next.

Placement

Four CMS placement schemes were tested in the South Carolina data base. These were as follows:

1. Single CMS use--One device placed approximately 2000 ft in advance of the taper,
2. Advance CMS use--One device placed 0.75 mile in advance of the taper,
3. Two CMS devices--One device at each of the above noted locations, and
4. Advance CMS with supplemental arrowboard--One CMS placed 0.75 mile in advance of the taper and an additional arrowboard at the 2000-ft location.

Figure 4 depicts an apparent effect of CMS placement on lane-distribution profiles. Significantly smaller proportions of traffic were observed in the right (closed) lane for the three noted conditions that included as CMS at the 0.75-mile advance location. Data collected at the CMS location (2000 ft

Figure 1. Tested CMS conditions.

Site	CMS Format/ Placement	Message Type	Display
South Carolina	Three-line bulb matrix (2000 feet from taper)	Speed and Closure Advisory	MAX SPEED 45 MPH ⇄ RIGHT LANE CLOSED
		Speed and Merge Advisory	MAX SPEED 45 MPH ⇄ MERGE LEFT
		Merge and Closure Advisory	RIGHT LANE CLOSED ⇄ MERGE LEFT
	Supplemental One-line bulb matrix (3/4 mile advance)	Closure Advisory	RIGHT / LANE / CLOSED / AHEAD
		Speed and Merge Advisory	MERGE / LEFT / SLOW TO / 45 MPH
Georgia	One-line bulb matrix (3/4 mile advance)	Closure Advisory	LANE / CLOSED / AHEAD
Colorado	Two-line rotating drum (3/4 mile advance)	Closure Advisory	RIGHT LANE CLOSED AHEAD
California	One-line bulb matrix	Speed Advisory	SLOW TO / 18 MPH
	Two-line rotating drum	Speed Advisory	SLOW TO 45 MPH
		Closure Advisory	RIGHT LANES CLOSED AHEAD
	Three line bulb matrix (All 3/4 mile advance)	Speed and Closure Advisory	RIGHT LANES CLOSED ⇄ MAX SPEED 45 MPH

Table 1. CMS field-test scenario.

CMS Effect	Test Condition	Data Base
Application	Baseline versus one-line device	Georgia
	Baseline versus two-line device	Colorado
	Baseline versus three-line device	California
Placement location	2000-ft advance	South Carolina
	0.75 mile and 2000 ft	
	0.75-mile advance	
Message condition	Speed and merge advisories	South Carolina
	Speed and closure advisories	
	Merge and closure advisories	
	Closure advisory	
Format	Two-line versus three-line device	California

Note: 1 ft = 0.3 m; 1 mile = 1.6 km.

in advance of the taper) indicated a dramatic reduction in the proportion of closed-lane traffic for the advance CMS schemes. No statistical differences were noted between the advance CMS conditions. However, a tendency was seen in the data for the earliest preparatory lane changing to occur in the presence of a CMS at the 0.75-mile advance location and supplemental arrowboard at the 2000-ft location.

Message Condition

The four tested message conditions were speed and closure advisory, speed and merge advisory, merge and closure advisory, and closure advisory. All four conditions were tested at the South Carolina site and the results obtained were used as a basis for message application at subsequent sites.

The majority of this data base was gathered by using the two state standard speed and closure and speed and merge advisory messages. Only limited data were available for the remaining two conditions, as these represented deviations from state standards. Improved lane-distribution profiles were associated with the use of speed and merge advisory messages.

Display Type

Data gathered at the California site permitted a direct comparison between the two-line rotating-drum device and the three-line bulb-matrix sign because both devices could be separately tested at one lane closure. Figure 3, previously presented to show baseline versus CMS application, also plots the comparative effect of the CMS display types. A prominent difference in observed effects between the two- and three-line devices was a smaller proportion of vehicles present in the closed lane at the CMS point while the three-line device was displayed. This difference between device types did not continue, however, into the intermediate and taper observation points.

An explanation of the lower closed-lane volume at the CMS point during the presence of the three-line device likely resides in the fact that the device itself is large and highly visible at a substantially greater lead distance than is the two-line CMS. There was no possible sight-distance difference effect because both devices were deployed at the same location. The data strongly suggest, however, that increased obtrusiveness of the sign itself did not serve to reduce closed-lane occupancy closer to the taper.

Although no difference in late exit behavior was noted between the two devices for the total vehicle sample, a different effect was observed for the truck population. Fewer trucks were observed to perform late exits during the presence of the two-line device.

Speed Observation Findings

A sample of 41 463 vehicle speed observations revealed an effect of CMS application at the construction site lane closures. Data-collection points on the approach were the same as those previously discussed for lane-distribution results. A large data base (N = 30 790) was initially collected in South Carolina to examine relative effects associated with specific CMS conditions. A number of message and placement variations were compared. The South Caroline data-collection effort was limited due to state CMS use requirements and consequent liability concerns that precluded testing a baseline. Baseline conditions were subsequently compared with CMS application effects in Georgia, Colorado, and California.

Unlike the lane-distribution results, substantive conflicting findings were noted between sites. However, a number of distinct tendencies were found in the data to support the finding that certain speed effects did result from CMS use.

CMS Application

Comparison between baseline and CMS conditions revealed speed reductions to be associated with speed advisory messages under most circumstances. The only exception was one site that exhibited low preexisting speeds of approximately 47 mph (75 km/h). No reduction was noted at the taper in the presence of a speed advisory CMS that requested reduced speeds of 45 mph (72 km/h).

Extensive speed measurements during baseline versus CMS-application comparisons were made in Georgia by using a one-line bulb-matrix CMS. Although no speed advisory message was displayed on the device, generally lower speeds indicated a

Figure 2. Lane distribution profiles for right-lane baseline and CMS use conditions (Colorado and Georgia samples).

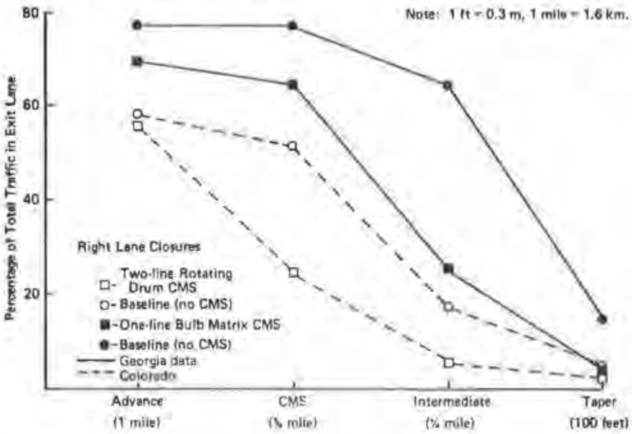
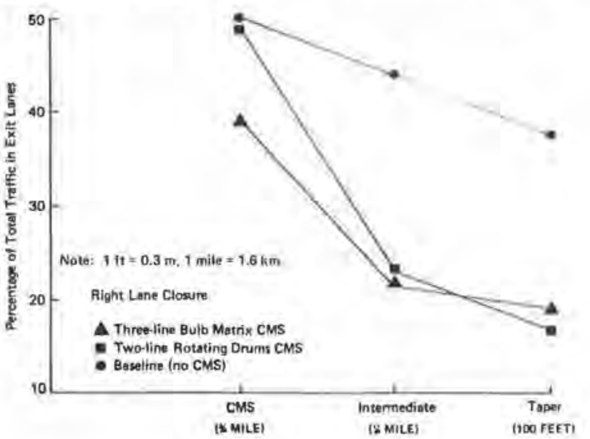


Figure 3. Lane distribution profiles for right-lane baseline and CMS use conditions (California sample).



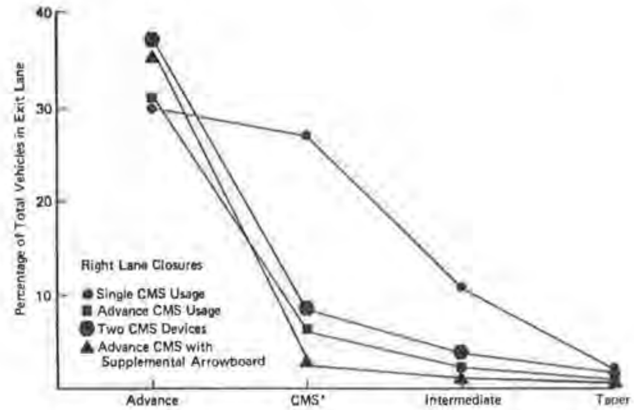
possible residual effect of motorists' increased awareness of the hazard. Because of the transient nature of construction activity, a more controlled experimental approach was applied at subsequent sites.

A modified procedure applied in Colorado and California entailed concurrent baseline and CMS-condition data collection within a period of a few hours. The advantage was that effects of geometry (previously noted on many occasions to obfuscate effects of the CMS) were eliminated by conducting both the before and after studies while construction crews were working at one point. Although sample sizes were obviously restricted by using this procedure, a sufficient number of observations were nevertheless obtained to support statistically reliable significance tests.

This procedure was used to compare the speed effects of all three CMS devices at one California site. The devices and displayed messages were as follows:

1. One-line bulb matrix; two-phase message (two words flashed at a time); "Slow to 45 mph";
2. Two-line rotating drum; single-phase message (all words continuously displayed); "Slow to 45 mph"; and
3. Three-line bulb matrix; single-phase message (all words, flashing display); "Reduce Speed, 45 mph".

Figure 4. Lane distribution for right-lane closures with advance CMS deployment.



*Standard CMS location; second when two were deployed.

The table below gives the results of the comparison (1 mile = 1.6 km):

Device	Speed (mph) Control	Experimental	Corrected Reduction
No CMS	62.5	63.7	N/A
One-line bulb matrix	62.1	56.3	7.0
Two-line rotating drum	63.0	56.6	7.6
Three-line bulb matrix	63.7	57.7	7.2

Note that in the absence of a CMS, speeds at the taper averaged 63.7 mph (101.9 km/h). Minor speed changes (day-to-day effects, etc.) were noted at the control location as average speeds varied from 62.1 to 63.7 mph (99.4-101.9 km/h). The table notes that reduced speeds were observed at the taper during the presence of each CMS device. These reduced speeds ranged from 56.3 to 57.7 mph (90.1-92.3 km/h), which indicates significant reductions of 6.0 to 7.5 mph (9.6-11.8 km/h) below speeds observed when no CMS was present.

The applied experimental design permitted the computation of a corrected speed reduction, which compensates for speed fluctuations observed at the control site. As an illustration of the speed-correction procedure, note that speeds at the control site dropped 0.4 mph (0.6 km/h) [62.5 to 62.1 mph (100.0 to 99.4 km/h)] between data-collection periods for the no-CMS and one-line CMS conditions. Thus, the observed 7.4 mph (11.8 km/h) [63.7 to 56.3 mph (102.0 to 90.1 km/h)] experimental site speed reduction, ostensibly elicited by the one-line CMS, was corrected by subtracting 0.4 mph to show the true effect.

The result of this experimental procedure indicated that each CMS device had a significant speed-reducing effect. Statistical tests indicated no difference between effects of the three devices.

Placement and Message Condition

Speed measurements gathered for a sizable vehicle sample (N = 30 790) proved to be more than adequate to statistically determine mean differences for a variety of CMS placement and message conditions. One notable tendency from the data was that placement seemed to affect speeds at the intermediate and taper data-collection locations. CMS deployments that use a device at the 0.75-mile advance location resulted in lower mean speeds. No similar trend was noted for message content.

Display Type

Results previously given in the above table support the finding that no appreciable differential speed effects were obtained between the tested one-, two-, and three-line formats. Although greater speed reduction was associated with the rotating-drum display than for either of the bulb-matrix signs, this difference was not statistically significant.

Questionnaire Findings

The human factors portion of the study involved application of a questionnaire to 489 subjects in order to gather measures of driver detection, recognition, and comprehension of the CMS devices. Characteristics of the subjects were controlled in order to ensure representativeness of the driving public.

Many statistically significant differences were found to distinguish between CMS conditions. In certain instances, questionnaire findings were seen to refute or clarify traffic-operation findings. In all cases of departure from traffic-operation results, findings based on questionnaire data were deemed highly credible because of the controlled nature of this experimental method. Questionnaire findings did not tend to refute the more convincing findings based on traffic-operation data (e.g., CMS improvement over baseline condition). Questionnaire findings are separately discussed, which reveal effects of CMS application, placement, message condition, and format.

Application

Certain questions were designed to determine whether or not drivers sensed general device improvement during the application of CMS devices. Two questions at the outset of the questionnaire requested drivers to provide a general rating of the overall adequacy of the traffic-control devices and to rate the signs as to how helpful they were. Each question was posed prior to any questionnaire reference to the CMS. These two questions were provided as follows:

1. In this driving test, you have just passed a highway area which is under construction. Please rate the overall adequacy of the construction warning devices (signs, barricades, etc.) according to the following scale.

Very poor	Poor	Borderline	Good	Very good
1	2	3	4	5

2. Please rate the signs as to how helpful you think they were.

Not at all helpful	Somewhat helpful	Extremely helpful
0	1	2

Comparisons between baseline and CMS-application conditions at all sites demonstrated an increase in the warning-device adequacy and sign helpfulness rating during the presence of any CMS device. In one isolated instance (e.g., two-line rotating-drum device based in Colorado) the increase was not statistically significant; however, high statistical significance was most frequently obtained.

Placement

Although mixed responses were obtained between CMS conditions that employ one and two devices, results more often favored the use of two devices. Higher detection rates and fewer complaints about providing inadequate information were associated with the use of two devices. In view of the fact that two-device arrays contained considerably greater amounts of information, lower average message recall scores were associated with their use. The trade-off between greater observation rate versus lower verbatim recall rate is interpreted to favor the use of two CMS devices.

Differences in questionnaire scores were noted between two message conditions (speed and closure and speed and merge), present with and without the use of supplementary advance devices. In each case a significant improvement in reported read and react time was noted in the presence of the advance device.

Message Condition

Questionnaire results heavily favored the use of speed and closure advisory messages. General device adequacy and sign helpfulness ratings, noted earlier to distinguish between baseline and CMS-application conditions, were also sensitive to CMS message differences. Higher ratings based on these two scores were associated with use of the speed and closure message than with others tested. However, in this case no increase was noted for the addition of the supplementary advance CMS.

Both with and without use of the supplemental CMS, the amount of information shown in the CMS array was approved most often during the presence of speed and closure advisories. Drivers reported that the speed and closure message was the easiest to read and that they most frequently modify their driving during its presence. This latter finding was validated by comparing vehicle behavior that had been matched for questionnaire responses. The validation procedure demonstrated that motorists interviewed during the presence of the speed and closure message made earlier preparatory lane changes and entered the taper area at lower speeds than those interviewed during the presence of other message conditions.

Another questionnaire finding, which refutes advantages shown in the traffic-operation data to be apparently associated with the speed and merge advisory, was that CMS devices were more often rated as being less helpful while this message was displayed. Moreover, low-CMS helpfulness ratings were indicated, and the amount of information shown was criticized as being inadequate for the Lane Closed Ahead message in the absence of specifying which lane was closed. The closure advisory message was most often correctly recalled.

Display Type

A number of differences were found between CMS display types. Lower overall device adequacy and sign helpfulness ratings were associated with the two-line rotating-drum sign. The one-line bulb-matrix sign drew less driver approval of the amount of information shown; moreover, drivers reported less available time to read and react to it both when used alone and in combination with the three-line bulb-matrix device. Moreover, when the three-line device was used alone, drivers more often reported seeing this device and rated it as being helpful.

The questionnaire item that provided the greatest distinction between CMS device types was the following question:

16. What changes would you want to see made to this sign?

Overall size:	Letter brightness:
______Larger	______Brighter
______Smaller	______Dimmer
______Neither	______Neither
Letter size:	Message length:
______Larger	______Longer
______Shorter	______Shorter
______Neither	______Neither

Table 2 summarizes the percentage of drivers who approved of (wanted no change in) overall CMS size, letter size, letter brightness, and message length for each of the tested devices under both day and night conditions. Certain significant differences were noted in these percentages. Lowest approvals of overall device size and message length (71 and 68 percent, respectively) were seen for the one-line bulb-matrix device. In addition, lesser approval of letter brightness was noted for the two-line rotating-drum sign both for day and night conditions. This latter comparison rated the difference between the bulb-matrix and rotating-drum CMS formats. Note also that significantly more drivers approved of the letter size associated with the three-line sign.

Validation of Questionnaire CMS Response

Two validations of subject questionnaire response were obtained on the basis of matched vehicle performance. South Carolina drivers interviewed during the presence of the speed and closure advisory message indicated that they responded to the CMS by slowing down and making earlier preparatory lane changes. Their self-reports were validated by matching observed lane-change behavior and comparing it with lane-change and speed behavior observed during other sign conditions. A positive validation was based on significant differences in average behaviors between the groups of drivers. This comparison indicated a significant tendency for drivers interviewed during the presence of the speed and closure advisory sign to exit the closed lane prior to reaching the intermediate data-collection point at an average speed of 49.6 mph (79.4 km/h), while interviewed drivers during other CMS conditions tended to change lanes beyond the intermediate point and their speeds averaged 53.4 mph (85.4 km/h).

Another statistical check on questionnaire validity, as well as CMS effectiveness, examined speed differences for drivers who saw the CMS versus those who did not see the CMS. Driving groups exposed to two different CMS conditions (one containing speed advisory information and not containing any speed message) were each taken from large homogeneous data bases (South Carolina and California). All of the South Carolina sample (N = 140) were exposed to a CMS speed advisory while the California group (N = 96) were exposed to merge or closure advisories. Of the total sample, 161 drivers saw the CMS and 75 did not. The matrix depicted in the table below indicates a significant speed reduction for drivers who saw the speed advisory CMS while no statistical difference was noted for the nonspeed advisory messages (note: 1 mile = 1.6 km; * = significant reduction):

	Average Speeds (mph)	
Item	Saw CMS	Did Not See CMS
Speed advisory	50.0*	52.0
No speed advisory	57.7	58.0

Table 2. Percentage of driver approvals of specific CMS design elements.

	Day			Night	
CMS Characteristic	One-Line Bulb Matrix	Two-Line Rotating Drum	Three-Line Bulb Matrix	Two-Line Rotating Drum	Three-Line Bulb Matrix
Overall size	71↓	81	88	100	80
Letter size	70	70	90↑	100	100
Letter brightness	77	54↓	76	50	80
Message length	68↓	83	82	75	80

Note: ↓ indicates existence and directionality of difference obtained between this and other CMS devices

SUMMARY

Before-and-after studies (CMS versus no-CMS conditions) conducted in three states consistently demonstrated beneficial traffic-operation effects that resulted from CMS application. Increased advance preparatory lane-change activity, smoother lane-change profiles, and significantly fewer late exits [exit from closed lane within 100 ft (30 m) of closure] were observed in each state. Reduced average traffic speeds approaching the taper were observed at locations characterized by preexisting speeds in excess of 48 mph (77 km/h). All tested CMS devices were nearly equal in their effectiveness. However, an observational study conducted in a fourth state demonstrated that advance placement 0.75 mile from the closure produced improved results by comparison with a 2000-ft advance placement.

Effectiveness differences between message conditions were not clearly discernible on the basis of lane-change behavior for the total traffic sample. However, driver interviews consistently favored the speed and closure (e.g., Right Lane Closed; Slow to 45 mph) message combination. Driver ratings of the adequacy of traffic-control devices were highest during the presence of this message. Drivers reported that this message was the most helpful of all tested, it was the easiest to read, it met their information needs, and they were most likely to change lanes early and reduce speed when the speed and closure message was displayed. Vehicle performance exhibited by interviewed drivers confirmed the validity of this latter claim.

A single traffic behavioral difference was observed between various CMS display types. More preparatory lane-change behavior was observed 0.75 mile in advance of the closure during the presence of a three-line bulb-matrix device. However, no lane-distribution differences were observed closer to the taper between this display type and the others tested, i.e., a two-line rotating-drum and a one-line bulb-matrix device.

Driver questionnaire data indicated a clear preference for CMS devices that displayed more information at a single glance. A three-line device was rated as being more helpful and more likely to provide necessary information than either a one- or two-line device. Sign letter brightness associated with the bulb-matrix format was favored by motorists over that of the rotating-drum format. Reported rates for drivers who saw the CMS did not differ between device types.

Vehicle performance data were coupled with driver interview responses to validate findings of the CMS evaluation. As previously noted, interviewed drivers reported more slowing and earlier lane changes in response to the speed and closure advisory, and they actually differed in those respects from interviewed drivers exposed to other message conditions. Also, separate analyses of drivers seeing versus those not seeing CMS devices that contained speed advisory messages verified the observed total traffic effect of reduced speed response to the appropriate CMS message.

While improved traffic behavior was convincingly demonstrated to occur with CMS use, it was repeatedly shown that beneficial effects can be overridden by such factors as roadway geometry. For example, CMS observation by drivers was shown to be affected by traffic volume and sight distance to the device. Effects of grade and interchange proximity were seen to obfuscate speed and lane-change responses otherwise elicited by CMS devices.

CONCLUSIONS

Numerous substantiated results that CMSs tend to improve traffic flow on the approach to construction zone lane closures support their limited application. The associated smoother lane-change profiles can potentially reduce side-swipe and rear-end accidents on the construction zone approach, and the reduced speeds may increase safety for construction zone workers.

Yet, that beneficial effects of CMS were often seen to be overridden by specific highway geometric conditions points out the need for their judicious application (as is the case with traffic-control devices in general). Furthermore, any conclusion regarding the effect of CMS devices must emphasize that these devices are to be considered supplemental in nature to standard traffic-control schemes currently in use rather than a substitution for any specific device.

Suggested cost-efficient CMS applications are as follows:

1. Short-term closures characterized by decreased driver expectancy,
2. Minimum traffic volumes of 900 vehicles/h, and
3. Limited sight distances to the closure.

Four specific guidelines for CMS application resulted from this research:

1. Device format should permit maximum amount of information display at a glance (i.e., use three-line presentation format with a maximum of two message phases),
2. CMS devices should be located 0.75 mile in advance of closure,
3. CMS devices are to be considered supplemental in nature to currently applied standard traffic-control device schemes, and
4. CMS devices are not to be considered as an alternative to the arrowboard; arrowboard placement and brightness have a considerably greater impact on operational safety than does CMS use.

ACKNOWLEDGMENT

This research was performed under NCHRP Project 3-21(2) under contract to BioTechnology, Inc. This work was sponsored by the American Association of State Highway and Transportation Officials, in cooperation with the Federal Highway Administration, and was conducted in the National Cooperative Highway Research Program, which is administered by the Transportation Research Board of the National Research Council.

REFERENCE

1. Manual on Uniform Traffic Control Devices. FHWA, 1978.

Publication of this paper sponsored by Committee on User Information Systems.

Reading Time and Accuracy of Response to Simulated Urban Freeway Guide Signs

ROGER W. McNEES AND CARROLL J. MESSER

The results and methodology used in a laboratory study to determine motorists' time required to read urban freeway guide signs and the accuracy with which they read the signs are presented. The study was performed by using licensed drivers as subjects. The subjects, ranging in age from 18 to 79 years, were taken along a hypothetical urban freeway route where 2-, 3-, 4-, and 5-panel signs were used. A sign bridge typically has between one and four sign panels that have a green background and a white border around each panel. Each panel contains one or two route designations, one or more destination cities, and additional action messages. Each panel contained either 2, 4, 6, 8, or 10 units of information. The results of this study indicate that the optimum accuracy level was about 6 units of information/panel. When the information level was less than 16 units, 100 percent of the subjects could read the signs acceptably; when the level was between 16-30 units, 51 percent could read the sign acceptably; and when the level was between 31-50 units, only 33 percent could read the sign acceptably. It is apparent that route-selection accuracy decreases as the number of route choices increases. On a large sign (3 or more panels), the information content should not exceed 16 units of information/sign bridge. The time required to read a sign also increases with the number of route choices and total information on the sign.

Extensive research in the area of sign reading began in the late 1930s and continues even today. These efforts have mainly been concerned with the physical dimensions of the lettering, types of sign, illumination and reflectorization, recognition and effectiveness, message content, and placement in relation to the driver's cone of vision and line of sight. These research efforts have led to the development of standards that apply to alphabets and numerals used on the signs. The Federal Highway Administration (FHWA) has a standard alphabet that dictates the letter series to be used in the design of exit-direction signs. The standard alphabet used on overhead exit-direction signs is the series E(M) alphabet. These letters are designed in such a way that they can be seen by a person with 20-20 vision at a distance equal to 60 ft/in of letter height.

This means that if the sign has 15-in letters, it can be read at 900 ft.

These factors have been related to both lateral and longitudinal sign-placement locations. The implication of research in this area for longitudinal sign placement is that an effective sign affords a longer sign-reading distance; thus maneuvering can be initiated earlier. Hence, the greater the distance at which the sign can be read, the shorter the distance it needs to be placed upstream of the decision point.

Forbes and Holmes (1) studied legibility with regard to letter height, width, and reflectorization in the laboratory situation. Tests of series B (narrow) and series D (wide) letters were made during daylight and at night. The daylight visibility distance for subjects with 20/40 visual acuity (average) was 33 ft/in of letter height for the B-series letters and 50 ft/in for the D-series letters. At night there was a 10-20 percent loss in legibility for both letter series. They also found that at night reflectorized letters against a dark background were as legible as floodlighted letters; however, there was a 30-40 percent loss of legibility distance when the signs were lighted by low beams.

Richards (2) reviewed research concerned with visibility of signs during night driving. Night visibility of signs depends on their size, reflectivity, and illumination. Reflectorized signs depend on the amount and the angle of illumination for the sign to be most effective with regard to legibility.

Straub and Allen (3) in a laboratory study investigated the effects of vehicle illumination on sign placement. The laboratory studies consisted of several types of reflective signs that were illuminated by a vehicle headlamp. The sign position was changed to correspond to different placement locations on a highway. The level of illumination reflected for each type of sign was determined by using a light meter. For overhead signs, the study showed that heights 5-8 ft above the pavement were more easily read than those placed higher. For roadside signs it was found that signs should be no more than 6 ft from the road edge for optimum headlight luminance. The authors found that both beams result in the highest sign luminance between 200 and 300 ft from the sign.

Allen (4) performed a field study to compare day and night legibility of different types of static highway signs. The subjects were placed in a vehicle that traveled on a straight flat section of rural highway in which a test sign had been erected. The sign contained several four-letter words that the subjects were told to read. These same 48 subjects participated in both the day and night observations. The letters were standard series E with a widened stroke. Letter heights were 8, 12, 15, and 18 in. These signs were illuminated by a rheostat-controlled bank of tungsten lamps placed on the side of the roadway at levels of 0.1, 1.0, 10.0, and 100.0 footlamberts. The findings were that average daytime legibility distance was about 88 ft/in of letter height. On average, optimally illuminated signs (10 footlamberts) were approximately 15 percent less than daylight legibility.

Roberts (5) also studied the effects of roadside placement of signs on legibility distance. He found that if the signs were placed less than 30 ft from the edge of the road, the legibility of the sign was greater than 50 ft/in of letter height. Signs farther than 30 ft from the edge had a legibility of less than 50 ft/in of letter height.

Ivey and others (6) performed a field test of the effects of rain on visibility and the operation of the motor vehicle. They used four different targets, selected to be representative of highway obstacles, to determine the visibility distances from inside an automobile during natural rainfall. The simulated rainfall represented 0.5, 1.0, 1.5, and 2.0 in of rain/h. This test indicated that object size and contrast with road surfaces were factors that have a significant effect on visibility during various levels of rainfall.

Messer and others (7) performed a controlled field study to determine the legibility distance for words on a matrix changeable message sign. Three four-letter words with 18-in letter heights were used in the study. All tests were performed during the day and the drivers were not task loaded. The results of the test indicated a mean legibility distance of 46.6 ft/in of letter height, or 830 ft, only slightly less than the 50 ft/in rule-of-thumb for static signs.

One of the critical elements in motorists' use of urban freeway guide signing is the time they require to read and react to navigational messages presented to them. Surprisingly, there is little literature specifically related to the subject. King (8) presented an analytical analysis of signing in 1970 in which the sign-reading literature was summarized. Two equations for determining the time required to read a sign were presented. King pointed out that Mitchell and Forbes recommended using Equation 1 when the sign contained more than three words. About 1 s was recommended as the time required to make a single glance. Forbes also stated that an ordinary person can read three or four familiar words during a single glance.

$$t = (N/3) + 1.0 \quad (1)$$

Work by the British Transport and Road Research Laboratory presents an alternative (Equation 2), which calculates reading time to determine letter size requirements. This equation is as follows:

$$t = 0.31N + 1.94 \quad (2)$$

In both equations, t is the reading time and N is the number of familiar words on the sign.

Abramson (9) expanded this definition of N to include numerals together with familiar shapes and symbols, such as route shields and lane-assignment arrows. No experimental evidence for this expanded definition was given. One may conclude that the time required to read a familiar word is assumed to be about 0.31 s/word.

As is evidenced by the previous equations, it is generally believed that unfamiliar motorists require more total time to read the information on a sign as more words are added to the message. It is assumed in the models that the increase in time is a linear constant with the number of words, although this assumption is questionable. By using King's expanded definition of N, the reading time required of a 4-panel overhead guide sign might require a total of 14.3 s to read, assuming each panel had 10 words on it. Personal driving experiences would suggest that 14.3 s is an unreasonably long required reading time.

Mourant and others (10) used an eye-marker camera to record a driver's visual search and scan patterns under three levels of route familiarity (mediated by instructions) and two driving conditions (open driving and steady-state car following). The drivers drove on an expressway route for which they had memorized a set of directions. They were instructed on trial 1 to read all road signs as a driver who is unfamiliar with the route must do; on trial 2 to read only those signs necessary; and on trial 3 try not to read any signs, as the driver who is very familiar with the route does. The results showed that for open driving the visual patterns shifted to the left and down and showed more compactness as a function of trials. In addition, the percentage of time spent viewing road signs and the saccadic travel distance to fixations on other traffic, road and lane markers, and bridges and road signs decreased as the driver became more familiar with the route. For car following, the increase in compactness of the visual patterns over trials was pronounced, but there was no change in the center of location of the visual pattern. Compared with open driving, the travel distances for car following were greater when looking ahead and at bridges, road signs, and other vehicles. However, drivers in the car-following condition spent less time reading road signs, which indicates that they used the lead car as an aid for route guidance. Possible visual aids for decreasing the driver's visual workload under today's driving conditions were discussed.

Bhise and Rockwell (11) conducted a study to develop a methodology to evaluate signs by determining the degree of match between the characteristics of the sign, the abilities of the driver, and other components of the highway such as traffic and road geometrics. An eye-marker camera was used to obtain data on driver eye movements while reading 400 Interstate highway signs. The authors employed both laboratory and field studies to obtain the required data. By use of the eye-marker camera, it was determined that the drivers eyes do not continually sample information but make successive discrete fixations. A driver acquires information only during a fixation that is between 100-600 ms in length. For the information to be available to the driver, it must be resolvable. A computer program [sign evaluation by analysis of driver eye movements (SEADEN)] was developed to determine the eye fixation that provides resolvable information about the sign to the driver by using eye-movement data, roadway geometry, velocity profile and path, sign characteristics, and visual acuity. SEADEN also computed various time-distance relations that concern the resolvability of the information on the sign.

The results indicated that the drivers first obtained this resolvable information from the awareness of the legibility of the maximum-sized letters. When the ratio between the distances at which the sign becomes legible and the distances of the driver's first fixation is 1.0, then the driver acquires information as soon as the sign becomes visible. This ratio decreases when the driver's visual load increases. When there is a sequence of signs, the driver spends more time in acquiring information on the first and last sign. As the driver becomes more familiar with the sign, he or she requires less time to obtain the information, and the driver time-shares between the sign and the road while obtaining the information. Bhise and Rockwell (11) also found that when drivers were searching for the mileage to a given destination, it required between 1.6 and 2.2 s when there was between 2-4 words on the sign and between 2.2 and 2.3 s when the information load increased from 4 to 10 words. When the driver was searching for a particular destination that was presented on the sign, he or she required between 1.7-2.2 s for 2-4 words and 2.2-2.4 s for 4-10 words, and when the driver was searching for a destination not presented on the sign, the driver required between 1.65-2.6 s for 2-4 words and 2.6-2.9 s for 4-10 words/sign.

These studies indicate that various researchers have done some studies in the area of information loading; however, a great deal more needs to be accomplished. It is not only sufficient to know how a driver acquires the information and the time and distance required. It still remains to be determined at what information level do drivers become so inundated with information that they cannot possibly acquire it and therefore start making errors in their control and navigation tasks of driving. This research focuses on that problem.

RESEARCH OBJECTIVES

The objectives of this study were (a) to determine the time required to find and read the correct test sign panel (that sign that gives the subjects information about their destination) embedded on a simulated urban freeway sign-bridge structure, and (b) to determine the accuracy of the selection process as related to sign design (information presented) and reading time. The research objectives were addressed in a laboratory environment by using licensed drivers as test subjects. The responses of these subjects to 35-mm slides of signs projected on a screen were recorded and evaluated.

STUDY VARIABLES

The specific magnitudes and variables studied during this phase of the research effort were as follows:

1. The number of panels per overhead-sign structure (2, 3, 4, and 5 panels),
2. The number of units of information on each panel (2, 4, 6, 8, and 10 units),
3. The display time available for subjects to read the signs (2.5, 4, and 6 s), and
4. The percentage of subjects giving the correct response.

A discussion of these variables follows.

Number of Panels

The number of panels selected for study includes almost all likely overhead sign designs. Most overhead guide signs in large cities have 3 or 4 sign panels/sign structure. A few signs have 5 panels on them. In the fringe area of cities and in smaller cities and towns, 2- and 3-panel signs are more common. A typical sign panel has a green background (route-guidance information) and a white margin around it and usually contains an exit number, route number, cardinal direction, two destinations, and an exit direction for a total of 6 units of information. Sign panels that have up to 10 units of information have been observed at major interchanges.

Units of Information

The following list illustrates what is defined in this study as a unit of information:

1. Place name (Denver),
2. Street name (Lamar Street),
3. Route number (I-95),
4. Cardinal direction (North),
5. Exit number (Exit 243A),
6. Command (Exit),
7. Distance (0.5 mile),
8. Lane-use arrows (↓),
9. Junction (Jct), and
10. Exit Only.

Some differences of opinion and need for discretion are to be expected in applying these measures. For example, all lane-use arrows to the same destination are counted as one unit of information. Some complex traffic facility names, particularly freeways like Central Expressway or Santa Barbara Free-

Figure 1. Three-panel, 4-units/panel test sign.

way, may be considered two units of information because of their size and possible confusion with a destination city.

The sign presented in Figure 1 illustrates a simulated sign used in this study that has 3 panels/sign structure with 4 units of information/panel. One should keep in mind that information rates in reality are only those messages that are needed and evaluated by the driver and may not be accurately reflected by the total content of all the words, numerals, and symbols on a sign.

Display Time

The projection or display times of the slides of the signs in the laboratory simulated the total time a motorist may have available to read freeway guide signs in a typical urban freeway traffic environment. The reading time is only a portion of the total time that the sign is visible. It is also less than the total static legibility distance (or time), which is less than the visibility distance (or time). The reason for this latter reduction is that motorists must time-share reading signs with other driving tasks, such as lane tracking and avoiding adjacent traffic. In addition, the last 150 ft or so immediately in advance of the sign is likely to be difficult to read due to the large vertical angle and relative motion of the sign with respect to the driver's visual scene.

The display times provided for reading the signs in the laboratory were selected to represent extreme, minimum, and desirable traffic (and design) conditions. High-quality guide signs are readable for most people in the absence of obstructions, beginning about 900 ft away, or 11 s of lead time. Deducting 2 s for clearance of the sign and 50 percent of the remaining time as required for conducting other driving tasks leaves 4.5 s available for sign reading. From a conservative design viewpoint, it would be desirable to provide freeway motorists with perhaps 6 s of unobstructed reading time in an unloaded driving-task condition for each overhead freeway guide sign. The motorists would take a portion of that time (perhaps 4 s) to select the appropriate sign panel by locating and reading the route number, cardinal direction, and destination. Some additional confirmation time might be allowed. A minimum acceptable design criterion might assume that the overhead guide signs were readable for at least 4 s, which reflects higher traffic congestion, more critical alignments, and higher probabilities for vehicle blockage of the signs. As the previous calculations showed, the laboratory display times of 2.5, 4, and 6 s seem to reasonably reflect extreme (unacceptable), minimum (acceptable), and desirable available reading times.

Accuracy

Accuracy of subject responses (selection of the proper sign panel that gives destination information) was measured in addition to reading times. The percentage of correct reponses based on the total laboratory subject population was determined for each test condition. It was expected that as the total information load increased and as the display time decreased for the same level of experience, accuracy levels would drop. An uninformed (or first-try) accuracy rate of 85 percent was arbitrarily selected as the minimum acceptable accuracy level.

RESEARCH METHODOLOGY

Two similar laboratory studies were conducted in an effort to accomplish the study objectives. In both studies, laboratory subjects were asked to follow a hypothetical route through an unfamiliar city based on (a) navigational directions provided by a schematic map of the area and (b) simulated guide signing presented by 35-mm slides at 22 locations along the route. A total of 87 subjects participated in the first laboratory study. The second study, conducted one year later, contained 70 subjects taken primarily from the same subject pool. The subjects ranged in age from 18 to 79 years, with a medium age of 34 years. Approximately 60 percent were male and 40 percent were female. We were not controlling for visual acuity and we wanted a representative sample of the driving public. We did not determine the visual acuity for all of the subjects. A discussion of the components of the research methodology follows.

Trip Scenarios

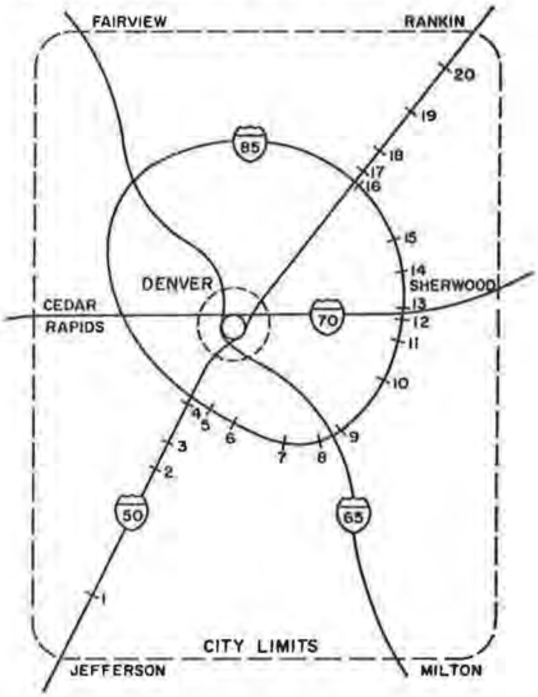
Subjects approached the city of "Denver" from the southwest on I-50. They were then directed to follow the south loop around the city, and then were directed to take I-25 to Omaha. The subjects were advised of their trip before testing began. The loop route was selected to maximize the number of interchanges that could be conveniently studied.

A set of 22 test signs that have preselected design attributes was developed for testing as the subjects "drove" along the route. The 22 signs were composed of 4 types of panels by 5 levels of information-unit rates (20 test signs) plus duplicates of the 4x10 and 5x10 panels. An artist developed the test signs following the style (to some extent) of overhead freeway guide signs found in the urban centers of Texas. Photographs of the simulated signs were taken and converted into 35-mm 2x2-in slides. One example of the 22 test signs was presented in Figure 1. The colors of the various sign elements were as close to the correct color as could be obtained. The background for all the signs was sky blue.

The laboratory scenarios called for the slides to be projected in a sequence consistent with the simulated trip. The slides were projected on a built-in wall screen in the laboratory by using rear-projection techniques. Viewing conditions and legibility of the signs shown to the subjects were controlled to approximate the average legibility requirements of signs on freeways. The design, placement, and display of the test signs along the route were selected such that large differences between the amount of information on each sign were not placed on consecutive locations. One practice slide was provided at the start of the trip to acquaint the subjects with the laboratory testing procedure. Map slides, similar to Figure 2 but showing the present location of the trip, were alternated between the 22 test signs so that the subjects, it was hoped, would know the information needed to navigate along the route. These slides were presented for 2 s and then advanced to the next test slide.

The subjects were asked to select the sign panel

Figure 2. Map of hypothetical city and route used in sign-reading study.



that provided information to the final destination from the set of panels on the sign bridge. It was assumed, and stated in the laboratory, that the sign panels would be placed immediately overhead of the corresponding freeway lane to drive in. The lane (or sign) number selected was given in the slides for each panel. At the bottom of each sign panel, a number representing the position of that panel on the bridge was printed on the slide.

Some subjects may have been confused in a few cases where the signing sequence (from the left) did not correspond to the lane assignments. For example, the first sign from the left may have been over lane 2 and the associated sign panel number would have been 2. To aid the subjects, the relative positions of sign panels over specific lanes were consistently maintained throughout the study.

Measurements

Estimates of subject reading times of the signs were obtained from electronically timed measurements of the time the slide became visible (human-operator input) until the time each subject activated his or her recorder unit. A student responder was used in the laboratory study for the subjects to indicate which sign panel contained the information that he or she needed to follow his or her route. There were five buttons located on this responder that corresponded to the five possible positions for the sign panels. Under each sign panel was a number, and this number was then entered into the responder system by each subject if that panel contained the required information. A recorder was at a master console at the back of the room recording the number entered by each subject. One of five numbered buttons could be selected, with the correct choice varying with each test sign. Subjects were asked to respond as soon as they were confident of their lane-assignment answer by pushing the corresponding numbered button on their data-recording unit. The accuracy of each response was also recorded for each subject. A maximum of five subjects could be tested at one time.

The subjects' average reaction time to a zero-level information sign was developed such that this reaction time could be subtracted from the overall response times so that the reading time could be estimated. The zero-level information slides were slides that had a distinct red background with the message "Push Button No. 1" on them. The subjects were shown one of the slides prior to testing, were informed of its purpose, and permitted to practice responding to it one time. The subjects were told that four of these slides would be randomly distributed through their trip and to respond to it accordingly. From these signs it was determined that an average subject population reaction time of 1.0 s existed. This time was subtracted from all measured response times to determine sign reading times.

Test Sequence

The 6- and 4-s display times were tested in the laboratory during March 1978. The sequence of projection times began with 6 s and the subjects drove the trip not knowing that, after a 10-min break, the trip would be redriven by using the same set of signs but displayed at 4 s. This procedure did result in some learning effect and improvement in response skill due to the previous experience. This was expected since the repeat test was conceptualized as a simulation scenario of semi-familiar urban freeway motorists who are experienced with the types and locations of decisions required.

The 2.5-s rate was a test to see how the same subjects would perform under anticipated and expected high-stress levels. This study was conducted one year after the previous tests. Some 8 improvements to the original 22 signs were made to improve route following and were used during all three (2.5-, 4-, and 6-s) display-time tests. Some of these signs were creating biased results due to the location of the sign, the complexity of the sign, or the location of the sign panel on the bridge.

RESULTS

The results of the 6-, 4-, and 2.5-s display times are presented in the following sections. The results show that the faster the display time, the faster the subjects responded. The results also show that, in general, the greater the information load, the slower the reading time. It is also important to note that the faster the display rate of the greater information load on a sign, the lower the percentage of correct responses. Most anomalies in the results to follow can be explained by either the simplicity or complexity of determining the correct sign panel (and lane) of a particular sign as tested in the laboratory. The subjects were told to select the sign that presented the information to their destination and then respond. They were not required to respond in a set length of time; therefore, the traditional speed-accuracy trade-off did not take place. However, after the test slide was removed from the subjects, it took longer to respond; thus, the response would be less accurate.

Display Time of 6 Seconds

A summary of the results for the 6-s display time, as presented in Figures 3 and 4, shows that the median (50th percentile) reading time was 2.9 s and the 85th percentile was 4.6 s. The average percentage of correct responses is 75 percent for 84 usable subjects. Some of the subjects did not answer within a reasonable time compared with the other subjects. Therefore, not to bias the time data, we proceeded to the next situation. Those subjects' responses we did not take into consideration when determining the reaction time or the accuracy of the responses.

There are some important trends to be noted that result from this study. As the amount of information units per sign panel (and total on the sign structure) increased, increased reading times and decreased accuracy levels generally were the result. These inverse trends are interrelated as the following comparisons show. The average values of the 50th percentile reading times and 85 percentile accuracy levels for all 2-panel signs are 2.2 s and 89 percent, respectively. On the other hand, the average values of the 50th percentile reading times and 85th percentile accuracy levels for all 5-panel signs are 3.3 s and 70 percent, respectively. Assuming that the 80th percentile correct response is selected as the minimum acceptable value, then four of the five 2-panel signs would be acceptable, whereas only one of the five 5-panel signs would be acceptable. The 6-unit 3- and 5-panel signs were selected very accurately and with unusual speed as a result of the location during the trip scenario. The location of the test signs were randomly selected, which resulted in a certain amount of bias in the results. This was the cause of the very poor accuracy rate of the 6-unit 4-panel sign. These were 3 of the 8 changes made in the signs before the start of the second study, as described earlier.

Figure 3. Reading times and percentage of correct responses for 2- and 3-panel signs at 6-s display times.

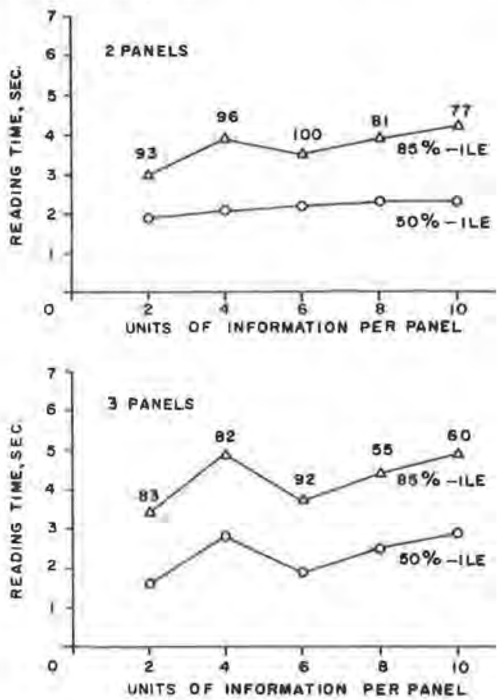
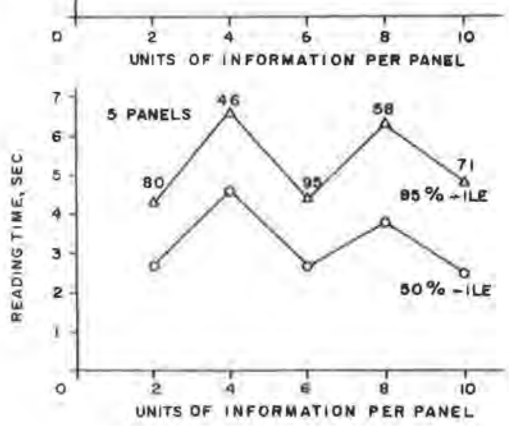


Figure 4. Reading times and percentage of correct responses for 4- and 5-panel signs at 6-s display times.



Display Time of 4 Seconds

The 4-s display test was a repeat of the same 22 signs used in the 6-s study. As noted previously, a break of about 10 min separated the two simulated trips. The subjects were given no advance clues that the second study was going to be a repeat of the first run. Some learning effects and skills improvement were expected. The reason for the repeat laboratory test was that it might more readily simulate a semi-familiar motorist who has driven the facility in the recent past.

A summary of the 4-s display test is presented in Figures 5 and 6. The median (50th percentile) reading time was determined to be 2.0 s, the average 2.3 s, and the 85th percentile 3.5 s. A mean percentage of correct response of 78 percent was obtained for 84 usable subjects. This is a 3 percent increase above the initial run and illustrates subject improvement due to learning and experience.

The inverse relation between reading time and accuracy continued with the 4-s display experiment. For example, the average values of the reading times for all 2- and 5-panel signs were 1.7 and 2.3 s, respectively. That is, reading times increased with increasing information load. The respective accuracy levels, on the other hand, decreased from 89 to 71 percent. Again, by using 80 percent as a minimum acceptable accuracy level, all 5 of the 2-panel signs performed acceptably. Only two of the five 5-panel signs had acceptable accuracy levels.

Display Time of 2.5 Seconds

The 2.5-s display-time laboratory was conducted one year after the previous two studies. Of a total of 70 subjects, 67 usable subject responses were evaluated. Some improvements to the design of 8 of the initial 22 test signs were made in addition to rearranging the test sequence for several of the modified test signs to improve the logic of the signing sequence. As will be shown later, these modifications produced significant improvements in route-selection accuracy and clouded the aggregate accuracy results.

A summary of the 2.5-s display test results are presented in Figures 7 and 8. The median (50th percentile) reading time was calculated to be 1.7 s, the mean 1.8 s, and the 85th percentile 2.8 s. An average percentage of correct response of 78 percent was determined for 67 usable subjects.

The inverse relation between reading time and accuracy level continued to be observed in this subsequent experiment. The average of the 50th percentile reading times was determined to be 1.7 s for all 2-panel signs and 1.9 s for all 5-panel signs. The average percentage of correct responses for all 2-panel signs had accuracy levels above 80 percent. However, only two of the five 5-panel signs performed acceptably.

Figure 5. Reading times and percentage of correct responses for 2- and 3-panel signs at 4-s display times.

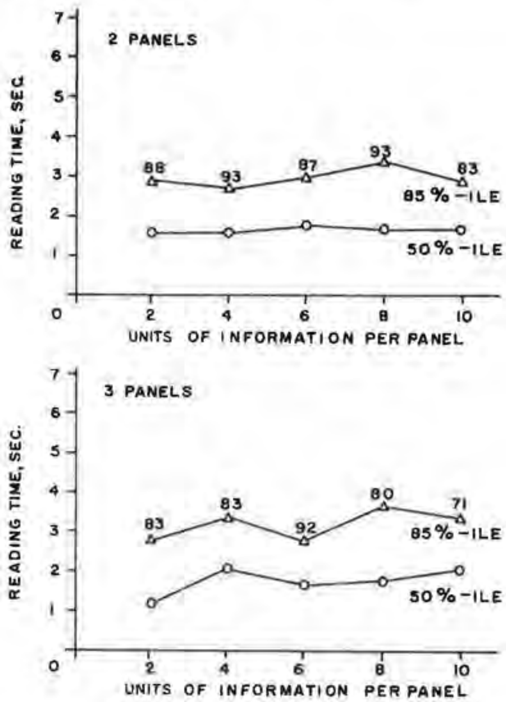
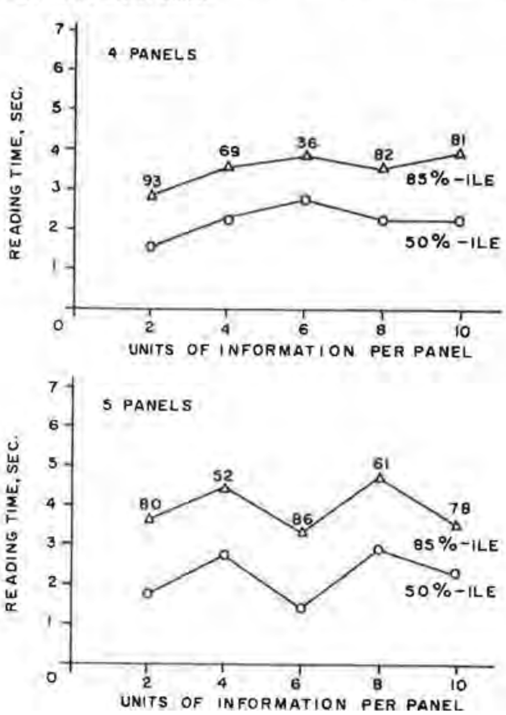


Figure 6. Reading times and percentage of correct responses for 4- and 5-panel signs at 4-s display times.



DISCUSSION OF RESULTS

A discussion of the results of the three display-time experiments follows. Comparisons are made from among the accuracy and reading-time results. Useful research findings are drawn from these comparisons and analyses.

Accuracy

The ability of the laboratory subjects to select the correct sign panel was found to depend on several variables; namely, total units of information on the sign, display time, and experience. Sign modifications also were found to impact accuracy results.

Information

A summary of the average percentages of correct response results by the sign-information test variables--number of sign panels per sign structure and information units per test panel--is presented in Table 1. At the outset, the sign-modification impacts between the 6- and 2.5-s display rates should be noted from Table 1. Of the 14 test signs not modified, 11 showed reductions in accuracy levels, 1 was unchanged, and 3 experienced slight accuracy increases. The mean percentage of correct responses of the data set dropped 5 percentage points on average, from 82 to 77 percent. Of the 8 signs that were modified, all 8 showed increases in the percentage of correct responses. The accuracy levels of the 8 modified signs increased 13 percentage points from 64 to 77 percent. Although there was no objective originally to suboptimize the sign designs, these findings do show that suboptimal sign designs can be significantly improved.

If it is assumed that the 6-, 4-, and 2.5-s display-test results represent samples of existing sign designs, reading requirements, and representative driver experiences for design-evaluation purposes, then the results of the 66 tests (3 display rates by 22 test signs) may be pooled to analyze combined accuracy results. The following analyses are conducted under this assumption.

The pooled accuracy results of Table 1 suggest that 6 units of information/panel are about optimum, which recognizes that 2 units are not a practical value. This conclusion is drawn from a consideration of the average accuracy levels of the 2-, 4-, 6-, 8-, and 10-unit signs in Table 1 (i.e., 86, 78, 79, 70, and 73 percent, respectively). It can also be determined from Table 1 that the average percentage of correct responses decreased with an increasing number of panels and with total information load I, where I is the product of number of panels P, and by the average number of units of information per panel B, or $I = P \times B$. The average percentage of correct responses for 2-, 3-, 4-, and 5-panel signs in Table 1 is 87, 78, 73, and 71 percent, respectively. The average percentage of correct responses for I-levels of 8, 12, 16, 24, and 40 units of information is calculated from the average of two cells for each I-level to be 91, 91, 79, 51, and 65 percent, respectively.

An analysis of the 66 individual data points from the three display-time experiments further reveals the reduction in accuracy rates with increasing total information levels on a sign. From Table 1 it can be determined that all 21 test signs that have I-levels of 12 units or less had accuracy levels of 80 percent correct or better. Again, 80 percent correct response is assumed to be the minimum acceptable level per test for this laboratory. These results are reflected by the upper curve in Figure 9. This curve shows the percentage of all data points (i, p) that have i ≤ I, which also have accuracy levels p ≥ 85 percent. Ninety percent (27 of 30) of all samples that have I-levels for 18 units or less had accuracy levels of 80 percent or more. Only 78 percent (28 of 36) performed acceptably. Over the complete data set, 41 of 66 (or 62 percent) of the signs were acceptable, as the upper curve in Figure 9 depicts at the upper bound I-level of 50 units.

Figure 7. Reading times and percentage of correct responses for 2- and 3-panel signs at 2.5-s display times.

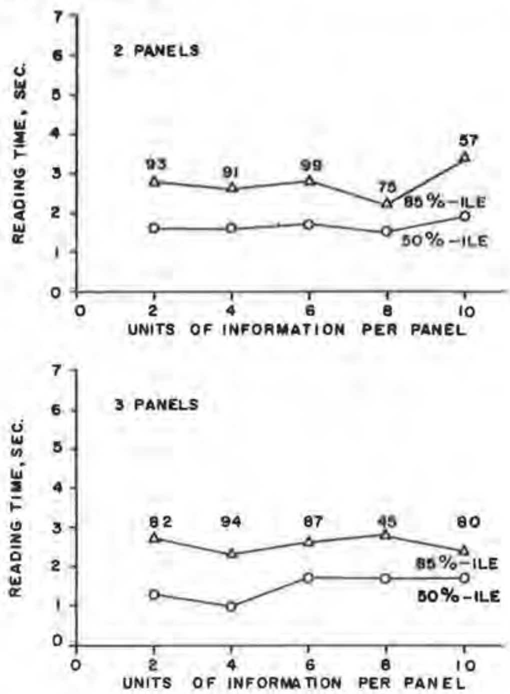
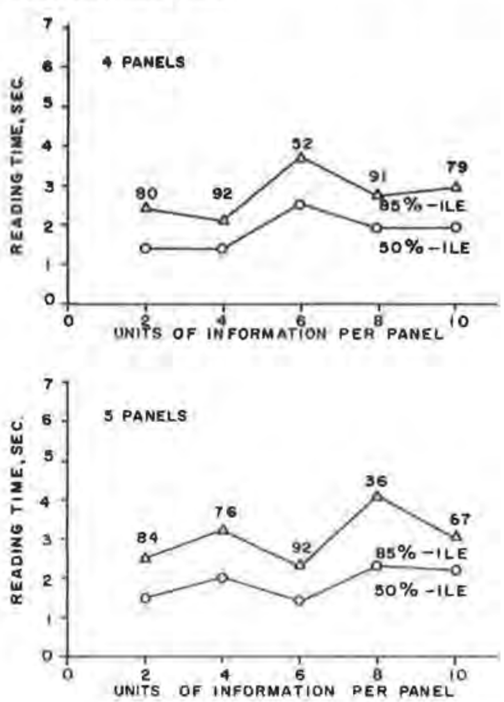


Figure 8. Reading times and percentage of correct responses for 4- and 5-panel signs at 2.5-s display times.



The average percentage of the signs performing acceptably (i.e., 80 percent correct response), based on the laboratory results, is given at the bottom of Figure 9 for three intervals of information load. In the interval of 1-15 units, 100 percent of the signs performed acceptably. In the interval from 15 to 30 units, 51 percent of the signs were acceptable. In the interval from 31 to 50 units, only 33 percent of the test signs were found to be acceptable.

Another sign-design parameter that seems to affect accuracy levels to some extent is the ratio of the number of panels P, divided by the average information unit rate per panel B, or R = P/B. If one analyzes the 4- and 2.5-s display-time results in Table 1, it will be determined that in 8 of 14 paired comparisons that exist at similar total information levels up to, but not including, 20-unit I-levels, that in only 1 of 6 cases did the percentage of correct responses increase as the ratio R decreased for a given I-level. This one case was at an I-level of 16 units with 2.5-s display time. However, in the 9 cases where paired comparisons were possible from I-levels of 20 bits or more, no trend is evident; 4 cases rose with decreasing R values and 4 cases dropped. Again, it is concluded that somewhere in the vicinity of 15-20 units of information (I) per sign structure is the maximum amount of information to be presented. Above this level there are just too many choices (panels) or too much clutter per sign panel for efficient decisionmaking to occur.

Display Time

A comparison of the 14 test signs not changed between the 6- and 2.5-s display-time experiments showed that this significant reduction in display time resulted in a moderate drop in route selection accuracy from 82 to 77 percent. It should be noted, however, that most of the signs that were not modified tended to be the smaller, less-complex signs.

Experience

The results of the 6- and 4-s display-time experi-

Table 1. Summary of average percentage of correct responses by number of panels and information per panel.

Information per Panel (units)	Display Rate (s)	Number of Panels per Sign 2	3	4	5
2	6	93	83	94	80
	4	88	83	93	80
	2.5	83	82	80	84
	Mean	91	83	89	81
4	6	96	82	63	46
	4	93	83	69	52
	2.5	91	94[a]	92[a]	76[a]
	Mean	93	86	75	58
6	6	100	92	33	95
	4	87	92	36	86
	2.5	99	92	52[a]	92
	Mean	95	90	40	91
8	6	81	55	76	58
	4	93	80	82	61
	2.5	75	45	91[a]	36
	Mean	83	60	83	52
10	6	77	60	82, 65	71, 70
	4	83	71	88, 75	83, 73
	2.5	57	80[a]	81, 78	55[a], 79[a]
	Mean	72	70	78	72

[a]Modified before 2.5-s laboratory.

Figure 9. Acceptance levels related to total information on sign.

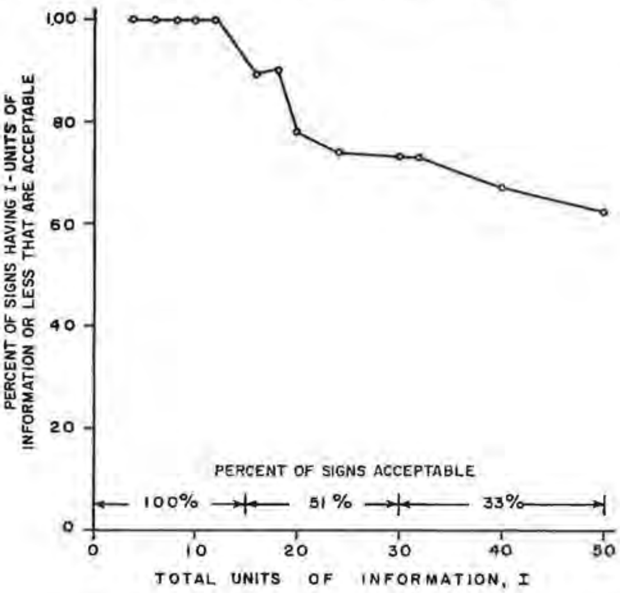


Table 2. Desirable and minimum reading times for overhead freeway guide signs (in seconds).

Units of Information per Panel	Design and Operating Conditions	Number of Sign Panels for Overhead Sign Structure: 2	3	4	5
2	Desirable	3.1	3.5	3.9	4.4
	Minimum	2.7	2.7	3.0	3.3
4	Desirable	3.6	4.2	5.0	5.7
	Minimum	2.7	3.2	3.7	4.2
6	Desirable	3.8	4.5	-	-
	Minimum	2.8	3.4	-	-
8	Desirable	3.9	-	-	-
	Minimum	2.9	-	-	-
10	Desirable	4.0	-	-	-
	Minimum	3.0	-	-	-

ments demonstrate how driver familiarity and experience yield improved driver performance. The mean percentage of correct responses increased from 75 to 78 percent, even though the average display time was reduced 33 percent. A total of 14 of the 22 test signs showed increases in the percentage of correct responses, whereas only 5 showed decreases.

Reading Time

The time the subjects took to read the signs depended not only on the sign-design parameters but also on how much time was available to perform the task. This was to be expected as normal behavior for the subjects. A brief review of the averages of the 85th pecentile reading times for each display rate illustrates this point, as follows (note: DT = display time, RT = reading time):

DT (s)	85th Percentile RT (s)	Ratio (DT÷RT)
6	4.6	1.30
4	3.5	1.14
2.5	2.8	0.89

A plot of these data shows that a 3.0-s display time would have produced a 3.0-s 85th percentile reading time, or a ratio of display time to reading time of 1.00 for the 85th percentile driver. Thus, it would appear that the 4-s display time would represent a test condition that is pressurized but yet provides minimum acceptable conditions. Since the 4-s display 85th percentile reading times were 75 percent of the 6-s times, the 6-s display time represents what may be reasonably considered to be a desirable set of operating conditions.

Linear regression analyses were performed to develop equations for estimating the reading times. The advantage of this approach is that smoothed estimates of each test sign can be estimated based on trends and characteristics of the complete study. Estimated desirable and minimum reading times based on these analyses are presented in Table 2. Minimum reading rates were assigned to be 75 percent of the desirable values subject to a 2.7-s minimum. Sign structures that have a total of more than 20 units of information on them are not recommended and usually do not exist in the field.

CONCLUSIONS AND RECOMMENDATIONS

The results of this detailed laboratory study of the urban freeway guide sign reading task form the basis for the following conclusions and recommendations. It is apparent that route-selection accuracy decreases as the number of route choices (and related sign panels) increase. It is also clear that the information content of a large sign structure should not exceed 6 units of information/panel and at no time should the total information loading exceed 20 units of information (absolute maximum); however, a more desirable maximum information loading would be 16 units/sign bridge. The time required to read a sign also increases with the number of route choices available and total information on the signs, as presented in Table 2.

The sign designs given in the table below represent what are recommended as desirable and maximum acceptable design parameters for overhead freeway guide signing in urban areas (note: * = undesirable design; sign spreading, removal of redundant concurrent routing, or other appropriate techniques should be examined):

No. of Route Alternatives (panels)	Maximum Units of Information on Sign: Condition	Bits
2	Desirable	12
	Maximum	16
3	Desirable	16
	Maximum	18
4	Desirable	18
	Maximum	20
5	Desirable	-*
	Maximum	20

The desirable and minimum design parameters were developed based on reading times determined during the laboratory studies and personal comments from the subjects based on heavily loaded signs. We felt that a cut-off of approximately 4.0 s to read any sign was critical for safe handling of a vehicle along urban freeways.

Any sign that does not provide desirable design conditions, with respect to the number of panels and the level of information in each panel, should have a sign layout that optimizes all other sign-design criteria. Minimization of costs should not be the only one controlling consideration for the minimum condition designs. All signs that do not meet minimum conditions should not only be redesigned, but the route structure should be redesigned to eliminate concurrent routes, unnecessary exits, etc.

REFERENCES

1. T.W. Forbes and R.S. Holmes. Legibility Distances of Highway Destination Signs in Relation to Letter Height. Proc., HRB, Vol. 19, 1939, pp. 321-335.
2. O.W. Richards. Vision at Levels of Night Road Illumination. HRB, Bull. 56, 1952, pp. 36-65.
3. A.L. Straub and T.M. Allen. Sign Brightness in Relation to Position, Distance, and Reflectorization. HRB, Bull. 146, 1956, pp. 13-44.
4. T.M. Allen. Night Legibility Distance of Highway Signs. HRB, Bull. 191, 1958, pp. 33-40.
5. A.W. Roberts. A Study of Lateral Sign Placement. Connecticut Highway Department, Hartford, 1968.
6. D.L. Ivey, E.K. Lehtipuu, and J.W. Button. Rainfall and Visibility--The View From Behind the Wheel. Journal of Safety Research, Vol. 7, 1975, pp. 156-169.
7. C.J. Messer, W.R. Stockton, J.M. Mounce, D.A. Andersen, and J.M. Turner. A Study of Physical Design Requirements for Motorists Information Matrix Signs. Texas Transportation Institute, Texas A&M Univ., College Station, unpublished rept., Research Foundation Project 3112, 1976.
8. G.F. King. Some Effects of Lateral Sign Placement. HRB, Highway Research Record 325, 1970, pp. 15-27.
9. N. Abramson. Information Theory and Coding. McGraw-Hill, New York, 1963.
10. R.R. Mourant, T.H. Rockwell, and N.J. Rackoff. Driver's Eye Movements and Visual Workload. Systems Research Group, Department of Industrial Engineering, Ohio State Univ., Columbus, Jan. 1969.
11. V.D. Bhise and T.H. Rockwell. Toward the Development of a Methodology for Evaluating Highway Signs Based on Driver Information Acquisition. HRB, Highway Research Record 440, 1973, pp. 38-51.

Publication of this paper sponsored by Committee on User Information Systems.

Use of Computerized Roadway-Information System in Safety Analyses

MARK A. FLAK AND JAMES C. BARBARESSO

The investigation of safety problems at a site requires that accident patterns be identified to aid in defining whether a safety problem exists and the level of the safety problem. Traditionally, safety engineers have identified accident patterns based on a combination of a visual inspection of collision diagrams and a site review of the study location. This approach, although enabling the engineer to identify accident patterns at a site, fails to define whether the magnitude of the patterns is sufficient to warrant the critical review of specific patterns and to recommend countermeasures to alleviate them. Rather, these decisions are made based on the engineering judgment of the safety engineer. A procedure that describes a mathematical approach used to identify safety problems and recommends countermeasures is described. By using a comprehensive roadway-information data base, this procedure is being used in Michigan by the Oakland County Road Commission in its highway risk management program. Similar roadway sites are selected based on geometric and operational parameters provided in the data base. These data, in combination with an accident data base for the defined roadway system, permit the analysis and comparison of accident characteristics of a study site to other similar sites. These analyses are also used in the identification of favorable countermeasures by comparing accident characteristics between similar sites with and without a particular countermeasure. These findings can also be used to develop accident-reduction factors for specific countermeasures.

Highway-safety professionals have long recognized the need for an organized approach for the correction of safety problems. This objective led the Federal Highway Administration (FHWA) to establish several programs to assist state and local agencies in improving safety on highways under their jurisdiction. These programs served to guide, assist, and encourage highway agencies to upgrade and maintain their highway system based on highway-safety objectives. The programs ranged from funds provided to identify and correct hazardous locations to the earmarking of funds for specific categorical programs and for the development of procedures to efficiently and effectively maintain a highway-safety improvement program (HSIP).

The HSIP was described in Federal Highway Program Manual (FHPM) 6-8-2-1. The FHPM described a systematic procedure for organizing a safety-improvement program. FHPM 6-8-2-1 was superceded in 1979 by FHPM 8-2-3. FHPM 8-2-3 recommends that processes for planning, implementing, and evaluating highway-safety projects be instituted on a statewide basis. It is planned that each state develop and implement, on a continuing basis, a HSIP that has the overall objective of reducing the number and severity of accidents and decreasing the potential for accidents on all highways.

A major component of the HSIP is the planning component. The planning component consists of the collection and maintenance of traffic, highway, and accident data; the identification of hazardous locations; the application of engineering studies; and the development of safety projects and their implementation. Prior to the development of safety projects, individual site investigations are required. The investigation requires that the site and its roadway, traffic, and accident characteristics be reviewed and that safety problems be identified in order to select favorable countermeasures. A review of the cost-effectiveness, safety impacts, and other significant factors for each countermeasure results in the development of a safety project.

To aid in the development and maintenance of the HSIP and highway risk management program for Oakland County, Michigan, the Oakland County Road Commission developed a comprehensive roadway-information system. This system uses computerized roadway, roadside, and accident data bases and a statistical computer package to assist the Road Commission in its safety analyses. The system not only permits the maintenance of highway, traffic, and accident data but allows a variety of safety analyses to develop and maintain the county's safety program. One specific tool developed from this system is intended to aid in the definition of safety problems and deficiencies and to assist in the countermeasure-selection activity. The purpose of this paper is to describe this analytical tool, its development, and its application for use in the county's highway-safety and highway risk management programs.

BACKGROUND

As part of the Road Commission's highway risk management program, Road Commission officials sought to develop a computerized roadway-information system that could be related to an existing computerized accident data base [monitored by the Traffic Improvement Association (TIA) of Oakland County]. This system was intended to permit accident information to be merged with roadway-information system characteristics. This system would then allow for the analysis of accident characteristics based on the roadway and roadside features established in the roadway-information system. The services of Goodell-Grivas, Inc., were obtained to develop the computerized roadway-information system and its analysis capabilities.

Development of Data Bases

Initially, roadway and roadside-obstacle inventories were developed from photolog records. Parameters recorded in the roadway inventory included the following: main-street name, cross-street name (where applicable), beginning and end points of roadway characteristics, intersection type (where applicable), number of through lanes, number of turn lanes, surface type, curb type, shoulder type, shoulder width, approach width, and median width, type, and presence of curbing. Other parameters that were also recorded included direction of travel that inventory is proceeding, date that roadway characteristics are initiated or updated, land use characteristics, speed limit (posted or unposted), horizontal geometrics (degree of curve, direction of outside of curve, radius of curvature, and rate of superelevation) in direction of travel, vertical geometrics (grade direction and percentage of grade) in direction of travel, passing-zone markings, date of initial paving for a section of roadway, average daily traffic (ADT), and stopping-sight distance.

Each section of roadway was divided into 0.1-mile segments for analysis purposes. Based on previous research and study efforts (1,2), this segment length was determined to be reliable to adequately define a section of roadway based on its roadway characteristics.

The roadside-obstacle inventory was also developed from photolog records. This inventory included the following parameters: main-street name, cross-street name, distance and direction of object from a reference cross street, side of street that object is situated, presence or lack of curbing, horizontal geometrics (degree of curve, direction of outside of curve, radius of curvature, and rate of superelevation) in direction of travel, vertical geometrics (grade direction and percentage of grade) in direction of travel, rigidity of fixed object (assigned by Road Commission for each obstacle type), ADT, speed limit, distance object is located from roadway (pavement) edge, roadway type (primary or local), obstacle type, and priority factor (based on formula derived by Road Commission staff).

A common location-reference system was applied to both inventory data bases in order to allow a merge of the roadway, roadside-obstacle, and accident characteristics. The reference system used was based on the Michigan Accident Location Index (MALI), part of a statewide reference system that involves the designation of a primary road (PR) number for each roadway or roadway section and a system of mile points used to locate the reference distance along the roadway. This reference system was selected due to its availability, accuracy, and current use in the accident data available to the Road Commission.

The accident data base currently existed as a part of the statewide MALI system. These data records are periodically sent to TIA, which is responsible for the local use and needs of the accident data. The information supplied on each accident record includes the following: PR number for street in which the accident occurred, mile point along street in which the accident occurred, accident report number, date and day of accident, time of accident, functional classification of route in which the accident occurred, weather conditions during accident, lighting conditions during accident, road surface conditions during accident, road defects (if any), traffic controls in the accident area, highway area type, roadway alignment (general), accident location (general description of roadway type), accident type, accident where (e.g., at intersection, at driveway, not at intersection), accident how (accident situation, e.g., head-on, rear-end, etc.), and accident severity. The accident record also included information on the following characteristics: drinking-related accident; police agency performing investigation; contributing circumstances; police enforcement action; driver age, residence, sex, and degree of injury; vehicle information; driver or pedestrian intent; hazardous actions; presence of visual obstructions (if any); object hit (if object-related accident); vehicle impact codes; and other special information.

Once the individual data bases were developed (i.e., roadway, roadside, and accident), a merged roadway-accident-roadside information system was developed. The purpose of this system was to provide a single data base that contains the roadway, accident, and roadside characteristics along all sections of county-maintained roads. This merged system would allow the detailed analysis capabilities for all merged roadway, accident, and roadside characteristics. By using a common reference system for each data base, the merge system was easily generated.

Data-Analysis Capabilities

To permit desirable data-analysis capabilities with the comprehensive roadway-information system developed for the Oakland County Road Commission, the use of a statistical computer package was employed. The statistical package selected was in use in the Oakland County Computer Center and, as such, was readily available.

The selected computer package was the Statistical Package for the Social Sciences (SPSS). SPSS is specifically designed for analysis of social-science data. Its use is readily intended for analysis of the data bases provided in the developed inventories. SPSS affords the user the capabilities of sorting, ranking, duplicating, and searching various options, numerous data transformations, and various statistical and numerical analyses.

Although most of the available SPSS capabilities are employed in the SPSS version available to the Road Commission, the most common subroutines planned for use consist of the following:

1. Frequencies,
2. Crosstabulations,

3. Condescriptive
4. Regression,
5. Breakdown,
6. Analysis of variation (ANOVA),
7. Report,
8. NPAR tests, and
9. Various bi-variate correlation analysis subprograms.

Other subroutines are available that can provide further detailed statistical analyses of the data bases. Their uses are documented in the SPSS manuals (3,4).

SYSTEM USE IN SAFETY PROBLEM ANALYSIS

One of the planned uses of the computerized roadway-information and analysis system developed for the Oakland County Road Commission is in the identification of safety problems and countermeasures for identified hazardous locations or features. The use of this system for these purposes is based on the premise that a site or feature has previously been identified as hazardous and requires identification of safety problems.

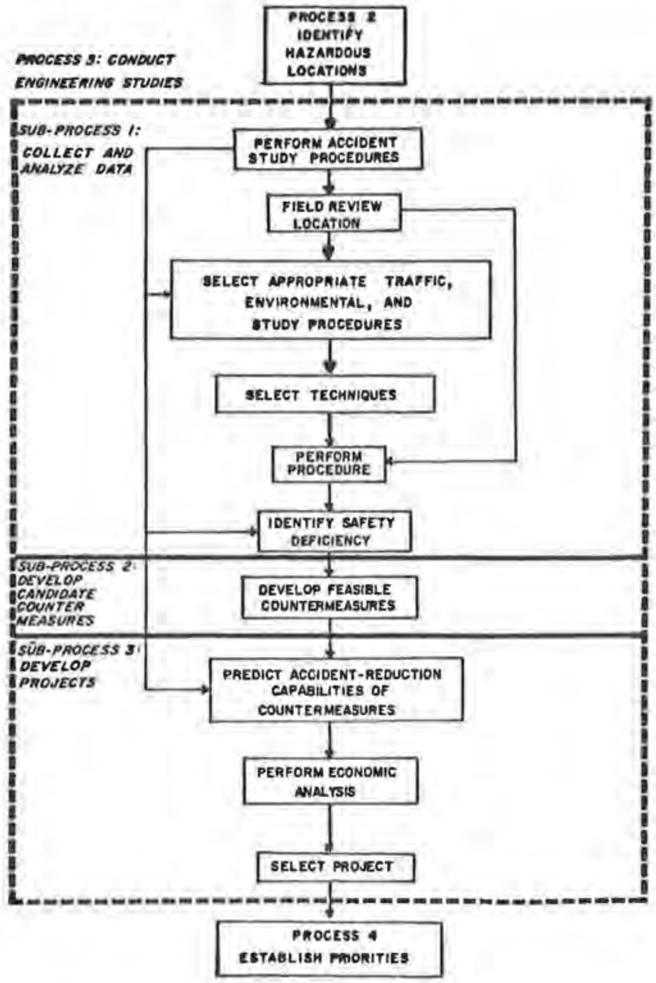
The basis for determining safety problems at a location is by identifying patterns in accident characteristics. Accident characteristics that recur at a location provide information used to aid in identifying safety problems. For example, the occurrence of a pattern of wet-weather accidents may indicate a slippery roadway surface as a possible safety problem. Other accident information or engineering studies may verify that a slippery roadway surface did or did not cause a safety problem. Initially, however, a safety problem was indicated by the occurrence of an accident pattern related to wet-weather conditions. Accident patterns, therefore, are the primary means used to identify safety problems.

Once the safety problems are defined, accident patterns are used to suggest feasible countermeasures to reduce or alleviate the defined safety problems. Cost-effectiveness studies and the results of other considerations are used to select a safety project for a site. A flow chart of the safety engineering study process is shown in Figure 1.

In the safety engineering study process displayed in Figure 1, the review of accident characteristics is vital in various steps of the process. For instance, the accident data are used as input in the field review; the selection of traffic, environmental, and special studies; and the identification of safety deficiencies. As a result, the need for accurate identification of accident patterns is critical to the safety study of a location. For accident-analysis purposes, two approaches to identify an accident pattern are typically available: (a) cluster analysis and (b) expected-value analysis.

The cluster-analysis method involves the visual inspection of accident-collision diagrams or summaries to observe the clustering or grouping of specific accident characteristics. Patterns are normally identified by an abnormal occurrence of a specific characteristic in relation to other accident characteristics that occur at a site. For instance, at a spot location with one rear-end accident, four run-off-road accidents, and two sideswipe accidents, the run-off-road accidents could be defined as a pattern. In some cases, the sideswipe accidents may also comprise a pattern. The identification of accident patterns in these cases is primarily based on the engineering judgment of the safety engineer. However, in everyday use, the accident activity that will comprise a pattern may vary from site to site and agency to agency, depending on the safety engineer's past experience in accident analysis.

Figure 1. Safety engineering investigation procedure.



A more reliable means of accident-pattern identification employs the statistical analysis of accident information. One means of statistical analysis is expected-value analysis. Expected values for specific accident characteristics are determined based on a statistical review of similar type sites. The expected values are used to define overrepresentations or critical values of the accident patterns for a study site. A comparison of the accident characteristics at the study site with other similar sites is made to determine accident patterns at a site. Assuming that the accident occurrences are normally distributed, accident characteristics below the critical value will not represent a pattern; however, values at or above the critical value will identify an accident pattern. This type of analysis defines critical values for accident characteristics at all types of locations.

Expected-Value Analysis

The expected-value analysis is a systematic mathematical procedure for identifying abnormal accident characteristics. To use this approach, accident data for similar sites (geometrics, volume, traffic control, etc.) are obtained and the average number of specific accident characteristics is determined. To account for variability or fluctuations in the data between sites, the use of the variance of the frequency of the specific accident characteristic is made to establish a range for an expected value. Assuming that accidents are normally distributed, the range of expected values can be defined as follows (5):

$$EV = \bar{X} \pm Ks \quad (1)$$

where

- EV = expected range of frequency of an accident characteristic,
- $\bar{X}$ = average frequency of an accident characteristic (e.g., accidents per number of sites),
- K = factor that relates selected level of confidence that a site will have a specific accident frequency, and
- s = standard deviation of accident frequencies.

Average and standard deviations may be obtained by standard statistical relations:

$$X = \Sigma FX/n \quad (2)$$

and

$$S = \sqrt{[\Sigma F (X - \bar{X})]^2/(n-1)} \quad (3)$$

where

- $\bar{X}$ = average frequency of an accident characteristic (accidents per number of sites),
- F = number of sites with a given frequency of a selected accident characteristic,
- X = accident frequency of nth site,
- n = total number of sites, and
- S = standard deviation of frequency of accident characteristic.

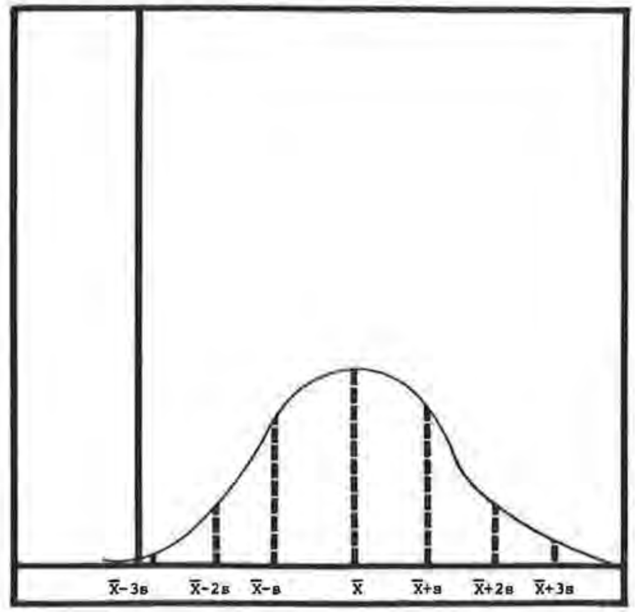
The K-value is based on the desired level of confidence that a site will have accident characteristics in a defined range. Values of K and their respective confidence levels are shown in the table below (note: 50.0 is the average value):

K	Confidence Level (%)
0.00	50.0
1.00	68.3
1.96	95.0
3.00	99.7

The expected range values for any given section with similar characteristics increase as the confidence level increases. Figure 2 displays this relation.

The above values are used in defining a range of expected values for an accident characteristic of X ± Ks. For accident-analysis purposes, safety analysts are typically concerned with the expected values along the right-hand portion of the normal curve. In the past, many safety engineers have used average values as the critical level used to define an accident pattern. This situation would allow for a 50 percent level of confidence that an accident characteristic would occur above the critical value. However, to increase the reliability that the expected value is defining an accident pattern, a greater level of confidence is desired, as shown in Figure 2. From this figure, it is shown that all accident frequencies to the right of the critical or expected value would represent an overrepresentation, i.e., an accident pattern.

Figure 2. Relation of confidence level in a normal distribution.



An example of the use of expected-value analysis is shown below. Accident-type characteristics for 14 similar sites reveal that the average annual frequency for the sites are as follows: 2.05 left-turn accidents, 3.56 rear-end accidents, 1.52 right-angle accidents, 1.32 sideswipe accidents, and 0.50 pedestrian accidents. Accident-type characteristics for the similar study site were as follows: 1.42 left-turn accidents, 2.48 rear-end accidents, 3.52 right-angle accidents, 0.78 sideswipe accidents, and 0.00 pedestrian accidents. By using expected-value analysis at the 50 percent level of confidence (average values), the right-angle accidents result in an overrepresented condition or accident pattern. The other accident types are below the defined critical level and do not represent accident patterns.

Expected-Value Analysis by Using the Comprehensive Roadway-Information System

The concept of expected-value analysis is currently used by the Oakland County Road Commission to assist in the accident analysis and countermeasure-selection phases of its highway-safety and highway risk management programs. By using the comprehensive roadway-information system, similar sites can be identified. Site characteristics used in defining similar sites include the following:

1. Number of traffic lanes,
2. Intersection type (where applicable),
3. Roadway surface (paved versus unpaved),
4. Shoulder width (ranges),
5. Horizontal geometrics (tangent, 1°-3° curve, 4°-6° curve, etc.),
6. Vertical geometrics (flat, 1 percent grade, 2-3 percent grade, 4-6 percent grade, etc.),
7. Curb type (curbed versus uncurbed),
8. Presence of median (median type and width),
9. Land use,
10. Speed limit, and
11. Volume--ADT (ranges).

In some cases, to acquire a large enough sample of similar sites it is required that some of the roadway parameters be relaxed to allow a wider range for a specific parameter and permit a greater number of sites to fit the criteria. Also, in some cases certain parameters may be considered insignificant for a specific analysis and may be removed from the criteria for that specific analysis. The site characteristics of the study location dictate the values of the parameters to be used in the expected-value analysis. For example, if a rural two-lane curved section of highway is identified as the study location, the site characteristics for this site will set the parameter values.

The SPSS computer package is used to identify similar sites, their accident characteristics, and the standard statistics (mean, standard deviation, minimum and maximum values, number of cases, etc.) associated with the accident characteristics. Initially, the data bases are searched to select from the roadway inventory only those records with similar parameters, as specified by the user. For example, to select similar sites for rural two-lane curved sections of highway, the computer command may read as follows:

```
SELECT IF [(ROADLANE EQ 2) AND (LANDUSE EQ 2) AND
   (HORIZGEO GTE 002)].
```

It should be noted that in SPSS, all alpha values require recoding to a numeric character. The value of 2 for the LANDUSE parameter actually refers to a 2=rural condition.

Once the data file has selected only those roadway records that have the specific roadway characteristics, a statistical subroutine is applied to the subfile to define the key statistics for the required accident characteristics. For example, the SPSS command used for obtaining descriptive statistics is CONDESCRIPTIVE--VARACC01 TO VARACC12. This command will provide the descriptive statistics for those accident variables listed above (i.e., VARACC01 TO VARACC12).

The output of this command can supply all or some of the following statistics and numbers: mean, standard error, standard deviation, variance, kurtosis, skewness, range, minimum, and maximum. These statistics are used in the formula for the expected value to derive the critical level of accident characteristics for the study site.

The value of each accident characteristic for the study site is then compared with the critical values defined for the similar sites. By using the expected-value concept, accident patterns are identified where the study-site frequencies are at or above the critical values. Values below the critical numbers are not defined as a critical pattern.

To adequately use the expected-value analysis, two considerations should be observed. First, in order to develop significant critical values, it is necessary to select an adequate sample size of similar sites. Sample-size formulas are available in various sources (6,7). The sample size is a key determinant of the level of significance of the findings. Where a sufficient sample size is not available, parameters may be relaxed or removed where their impact is not considered significant, or a lower level of confidence for the findings must be assumed. On a county level, the sample size will typically be insufficient. However, the relaxation or removal of limiting parameters can significantly increase the sample sizes to acceptable levels.

A second consideration concerns the level of confidence chosen for the analysis. The effect of a high-confidence level will result in a short list of critical accident patterns. Less-stringent confidence requirements will result in a greater list of accident patterns. The length of the list is dependent on the safety engineer's experience and judgment. If the safety engineer desires to resolve the most critical safety problems, more stringent requirements should be made. However, if the objective is to resolve all levels of safety problems, lower requirements should be specified. With recent reductions in highway-safety budgets, a more stringent level may typically be used. By resolving the critical safety problems, more cost-effective solutions can be developed, thereby resulting in more efficient use of available safety funds.

OTHER SAFETY USES OF COMPUTERIZED ROADWAY-INFORMATION SYSTEM

In addition to the use of the computerized roadway-information system as a tool for identifying safety problems at a site, the system can also be used in selecting favorable countermeasures or safety projects for a site. The procedure involves the identification of accident characteristics for a group of similar sites. The analysis of a group of sites with field conditions similar to those of the study site, except for one altered condition, will provide a record of accident statistics for the field situation with the altered condition. For example, the safety impacts of paving the shoulder of a two-lane roadway can be predicted. A comparison of the accident characteristics for the two situations will identify whether the countermeasure has a positive or negative safety impact. In this way, accident-reduction factors can be developed. It can be assumed that the difference in accident characteristics represents the effect that a specific countermeasure (altered condition) will have on the field situation. These analyses can be made for several countermeasures to assist in determining the most favorable project based on highway-safety objectives.

SUMMARY AND CONCLUSIONS

The development of the computerized roadway-information system for the Oakland County Road Commission provides a major advantage to the Road Commission's highway risk management and highway-safety programs. A significant increase in the roadway, roadside, and accident study and analysis capabilities occurred following the development of the system. As increased use of the system occurs, greater developments and innovations in its applications will result.

This paper discussed two current uses of the roadway-information system. They are (a) the identification of safety problems and (b) the selection of countermeasures. Some deficiencies do occur with the data base due to the limited number of similar sites typically afforded a county area. However, with the relaxation of constraints, this problem can be dealt with effectively. The roadway-information system does, however, allow the Road Commission to study and review the accident activity and their characteristics for the area under its jurisdiction. In this way, the county can more closely monitor and dictate the safety needs of its highways. The result is an increase in safety and an improved highway risk management program, thereby permitting more effective use of the highway tax dollar in the Oakland County area.

REFERENCES

1. Analysis of Urban Arterial Road and Accident Experience. FHWA, Draft Rept. DOT-FH-11-9500, June 1981.
2. Methods for Evaluating Highway Safety Improvements. NCHRP, Rept. 162, 1975, p. 5.

3. N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent. SPSS--Statistical Package for the Social Sciences, 2nd ed. McGraw-Hill, New York, 1975.
4. C.H. Hull and N.H. Nie. SPSS Update--New Procedures and Facilities for Releases 7 and 8. McGraw-Hill, New York, 1979.
5. Highway Safety Engineering Studies. FHWA, Rept. DOT-FH-11-9685, June 1981.
6. Accident Research Manual. FHWA, Rept. RD-80/016, Feb. 1981.
7. Elementary Sampling for Traffic Engineers. Eno Foundation for Highway Traffic Control, Westport, CT, 1962.

Publication of this paper sponsored by Committee on User Information Systems.

Four Approaches to Instruction in Occupant-Restraint Use

A. JAMES McKNIGHT AND KENARD McPHERSON

The results of four test programs for increasing teenage occupant-restraint use are presented. Each program contained an informational component while three programs provided additional learning experiences—a testimonial, operation of a vehicle, and use of a safety-restraint convincer. Conclusions show that the programs are a promising way to educate teenagers about using restraints.

Although the use of occupant restraints represents the single most valuable way of reducing traffic injuries and fatalities, use continues to be very low. The use rate for drivers in the United States is only about 10 percent while the rate for young people (ages 16-19) is even lower. Since young drivers are overrepresented in the number of traffic accidents, it is extremely important to encourage the use of occupant restraints in this segment of the population.

The National Highway Traffic Safety Administration funded a study for the development, implementation, and evaluation of several supplementary driver-education programs to be taught subsequent to the standard driver-education curriculum. One of these programs deals with occupant restraints.

The main objective of the restraint program is to teach teenage drivers to use safety restraints and encourage their passengers to do the same. Other objectives include teaching the students the value of safety belts in reducing injuries and fatalities as well as the risks associated with nonuse. In addition, the course encourages favorable attitudes toward restraints, including the belief that restraints are valuable and that the safety of passengers is the driver's responsibility.

NATURE OF PROGRAM

To attain these goals and objectives, four individual driver-restraint programs were developed. Each program contained an informational component while three of the programs provided additional learning experiences, which included a testimonial, operation of a vehicle, and use of a safety-restraint convincer. A brief description of each program is presented below:

1. Information only--The information program consists of 3 h of classroom instruction. No behind-the-wheel or other learning experience is provided. The classroom activities are, however, supported by a film. The information contained in the student materials and the film is directed toward cognitive and additudinal aspects of safety restraints. This program is designed to provide students with factual information about restraints and to increase their perception of the risks associated with nonuse.
2. Testimonial--The testimonial program includes the information contained in the previously described program. In addition, it provides an audiovisual presentation that consists of a testimonial in which an age peer describes an accident, the nature and extent of any injuries, and the disabilities that resulted from the crash.
3. Vehicle--The vehicle program adds to the information program the experience of riding in a vehicle, both restrained and unrestrained, through a series of emergency maneuvers. The maneuvers were selected to show the effect of restraint use on ability to control the vehicle in an emergency.
4. Convincer--The convincer program combines with the information program the use of a device designed to demonstrate the forces experienced in a crash. A sled with a car seat and safety-restraint system is mounted on an inclined plane at approximately a 45° incline. The sled is raised to the top of the incline and allowed to slide freely to the bottom. Persons, properly restrained, ride the sled and can feel the forces exhibited in a simulated crash.

METHODS

A before-and-after design was employed to evaluate each of the four programs. The programs were administered at four high schools in St. Louis, Missouri. Approximately 100 students were available at each school, each school administering only one program. The use of four different schools was necessary to be able to determine the effects of each program on actual restraint use. If more than one program had been given at each school, there would have been no way of knowing, as students arrived and left in their cars, which students received which program.

The measures employed to evaluate the program were as follows:

1. Knowledge test--A paper-and-pencil test that contained items on the facts of restraint use, the risks of injury, and the effects of nonuse on occupants;
2. Attitude test--A multiple-choice measure that presented scaled opinions concerning the use of restraints; and
3. Use of restraints--Observations were made on students' use of safety restraints while coming to

school and when going home.

All subjects were administered a knowledge and attitude test prior to commencing the program. Baseline performance was observed for three consecutive days prior to the beginning of the program at each school. Postprogram knowledge and attitude tests were administered on completing the course while postprogram performance was observed over the same three days of the week as the baseline period. Each subject took different forms of the knowledge test on preprogram and postprogram administration. Both forms were used equally often as pretest and posttest in order that differences in forms would not bias precourse and postcourse comparisons. A single form of the attitude measures was used.

In the schools giving the vehicle and convincer program, all three measures were given on a follow-up basis one month after postprogram administration. In the case of the knowledge test, the form given as a preprogram measure was repeated.

RESULTS

Results from the study of the restraint program will be discussed in terms of (a) actual use of restraints, (b) knowledge about restraints, and (c) attitudes toward restraints.

Use of Restraints

Results obtained from monitoring use of restraints are shown in Table 1. It is evident that wide preprogram and postprogram differences existed among the students who received the four programs. These differences prevent any meaningful comparisons being made across the four programs.

Each of the four programs produced a gain in use. Gains for the information, testimonial, and vehicle programs were statistically significant ($p < 0.05$). Significance was assessed through a one-tail test of correlated means. The gain for the convincer group was well within chance variation ($p = 0.26$).

As noted earlier, follow-up comparisons could be made only for the vehicle and convincer programs. The students in the vehicle program maintained the substantial gain obtained earlier. The gain for the convincer group continued to be nonsignificant.

Knowledge

Results obtained from the administration of knowledge tests appear in Table 2. Although students in all four programs were given the same information presentation, the information and testimonial groups showed the largest percentage gain--42 and 32 percent, respectively. Both gains were highly significant. The knowledge gain for the vehicle group, although statistically significant, was considerably smaller. The convincer group failed to show a significant information gain.

Although the information component of all four programs was the same, the conditions under which it was delivered differed among the groups. For the information and testimonial programs, it was presented in a classroom situation with opportunity for interaction. The information presentations for the vehicle and convincer programs were, on the other hand, given to all students collectively in an assembly hall. There was no interaction and no way of making sure students were paying attention. These conditions may explain the small amount of gain.

The follow-up knowledge administration evidenced a somewhat lower but still statistically significant gain for the vehicle program. Results for the convincer program were the same as they were immediately after the program.

The failure of the convincer group to show a significant knowledge gain may help explain the lack of a significant gain in restraint use. It appears that what little gain in restraint use occurred for the convincer group came as a result of the experience in the convincer and not the information that was presented.

Attitudes

Results obtained from administration of the attitude measures to students in the four programs appear in Table 3. All groups evidenced statistically significant attitude pregains and postgains. Attitude gains for the information and testimonial groups paralleled the knowledge gains. This is not surprising, considering the relation between knowledge and attitudes. The attitude gains for the vehicle and convincer groups surpassed knowledge gains and seem to indicate that experiences in the vehicle or convincer also influenced attitudes. The results of follow-up measures suggest that the attitude changes experienced by students in the vehicle and convincer programs tend to endure.

SUMMARY

The overall results indicate that all four programs are capable of having a beneficial effect. The

Table 1. Restraint program results of use measures (percentage using restraint).

Program	Pre-program	Post-program	Preprogram/ Postprogram Difference	Follow-Up	Preprogram/ Follow-Up Difference
Information	3.3	8.5	+5.2[a]	-	-
Testimonial	4.1	6.7	+2.6[a]	-	-
Vehicle	13.5	26.7	+13.2[a]	27.7	+14.2[a]
Convincer	9.0	13.2	+4.2	11.8	+2.8

[a] $p < 0.05$.

Table 2. Restraint program results of knowledge measures—mean scores.

Program	Pre-program	Post-program	Preprogram/ Postprogram Difference	Follow-Up	Preprogram/ Follow-Up Difference
Information	8.5	12.1	+3.6[a]	-	-
Testimonial	9.5	12.5	+3.0[a]	-	-
Vehicle	9.8	11.2	+1.4[a]	10.8	+1.0[a]
Convincer	9.7	10.3	+0.6	10.3	+0.6

[a] $p < 0.05$.

Table 3. Restraint program results of attitude measures—mean scores.

Program	Pre-program	Post-program	Preprogram/ Postprogram Difference	Follow-Up	Preprogram/ Follow-Up Difference
Information	12.2	16.5	+4.3[a]	-	-
Testimonial	13.1	16.7	+3.6[a]	-	-
Vehicle	11.9	15.3	+3.4[a]	15.2	+3.3[a]
Convincer	12.4	14.1	+1.7[a]	14.5	+2.1[a]

[a] $p < 0.05$.

information, testimonial, and vehicle programs produced significant gains in knowledge about, attitudes toward, and use of restraints. How long these gains were sustained could be determined only for the vehicle program. However, the fact that gains realized through this program appeared to endure is encouraging.

The vehicle program appeared to produce the most substantial gains in restraint use. However, it would be dangerous to make comparisons. The fact that the program use rate was highest among students who received the vehicle program may be an indication that they were a more responsive group than those who received the other programs.

The effectiveness of the convincer program is difficult to evaluate. The failure to obtain any significant gains in use is certainly discouraging. However, this failure is accounted for at least in part by (a) failure of the information component of the program to communicate effectively and (b) large day-to-day variation in prevailing restraint use.

From the results obtained, the following conclusions may be offered:

1. It is possible to influence the use of safety restraints among teenage drivers by means of an in-school program;
2. Communication of factual information about restraints and the risks associated with failure to use them are necessary elements of any program; and
3. More research is needed to determine whether any additional benefit is derived from experiencing the consequences of nonuse through operation of a vehicle, a ride in a convincer, or the testimony of someone who has been injured in a crash.

Publication of this paper sponsored by Task Force on Occupant-Restraint Systems.

Contributory Negligence in Promotion of Safety Belt Use

ROBERT N. GREEN AND GILBERT S. SHARPE

Contributory negligence, or contributory fault, can be described as unreasonable behavior on the part of an individual by which he or she contributes to injuries caused him or her by another's negligent act. Historically, under common law, once contributory negligence on the part of a plaintiff is established by a defendant in a personal-injury action, this serves as a complete bar to the plaintiff's claim. Even in jurisdictions without seat belt legislation, the common law over the past two decades has been increasingly recognizing that the failure to wear seat belts constitutes contributory negligence. It appears that if the common law continues to develop by itself, the seat belt defense will be increasingly recognized by the courts in the assessment of contributory negligence. If the seat belt defense is to be recognized by law, such statutes should be broad rather than restrictive to provide just penalties for contributory negligence.

Many studies in laboratories, as well as on-site motor vehicle crash investigations, have shown that modification of collision forces to prevent human injuries requires occupant restraint. Canada, the United States, the United Kingdom, and many Western world countries have mandated the availability of three-point restraint systems during the past 10-15 years. The effectiveness of seat belts is clearly established from the many studies on the subject. The disadvantages of wearing seat belts appear to be negligible.

Data were recently collected on seat belt legislation and its effectiveness in many countries of the Western world (1). A summary of the major findings is cited below:

1. Countries that have enacted seat belt laws seem to have evolved to a state where mandatory seat belt legislation was considered acceptable by the majority of the public prior to actual enactment. Where this is not the case, the law has either been repealed, has no penalty associated with it, or is not rigorously enforced by the police.
2. Seat belt laws enacted by various countries usually pertain to the driver and front-seat passenger only. Also, the laws are generally applicable to passenger cars and vans.
3. Most countries with seat belt laws have penalties associated with the legislation. In some cases the amount of the fine has a substantive upper limit ($200-$300). However, where statistics are available, it has been shown that the average fine is usually less than $10. Some countries have penalties for noncompliance that include imprisonment.
4. All countries allow exemptions from seat belt legislation. Exemption generally applies to passengers of a particular age or size, passengers with certain medical conditions, and drivers of commercial vehicles.
5. All countries studied have regulations regarding the installation of seat belts in both new and old cars. Most countries specify that the three-point inertial-retractor-type belt be installed.
6. Public information and education programs have been used to some extent by all countries that have enacted seat belt legislation. However, it was found that while these programs may be of value in

changing the attitude of motorists concerning the safety and effectiveness of seat belts, they do not result in any appreciable behavioral change regarding the wearing of seat belts.

7. Most countries that have enacted seat belt legislation expended considerable effort to publicize the law prior to its enactment. After the law became effective, some countries provided a grace period of at least one month before they began to enforce it.

8. Studies in virtually all countries revealed that the seat belt use rate rises from 200 to 300 percent immediately after the seat belt law becomes effective. The rate subsequently drops as much as 10-20 percentage points and then rises to some plateau, depending on the amount of attention and enforcement provided by government officials.

9. The results of attitudinal studies in countries with seat belt laws reveal that 60-80 percent of the people interviewed prior to enactment of the law indicated that they were in favor of mandatory seat belt use. However, the use rate was so much lower that it bore no relation to the results of the attitudinal studies.

10. Police officials are reluctant to enforce laws that are not supported by the general public; therefore, in those countries where the law is unpopular, enforcement is very low.

11. In almost all countries it was found that the seat belt law was not enforced independently of other traffic infractions. It is almost always enforced only as an ancillary action in connection with some other traffic violation.

12. Enforcement of seat belt laws appears to be essential to a high seat belt use rate. In several countries it was determined that the use rate was directly related to the level of enforcement, with high use rates usually associated with stringent enforcement. However, in some cases it did appear that the people's cultural propensities for being highly law-abiding obviated the need for stringent enforcement.

13. In several countries with seat belt laws it was found that the courts have ruled that insurance compensation should be reduced for victims who were not wearing seat belts at the time of the collision. In order to support such a ruling, it is necessary to have the crash investigated by experts, who must then testify that the injuries sustained would have been less if the victim had been wearing a seat belt. The amount of reduced compensation has been set as high as 50 percent in several European countries.

Whether a given jurisdiction has chosen to follow the Australian example and mandate the use of the available restraint system (as has happened in much of Canada in the past few years) or to continue a policy of voluntary use (as in the United States), the increasing application of personal-injury tort law will undoubtedly influence future legislation. The premise of contributory negligence can be stated that where a situation or activity involves a foreseeable risk of accident, there is a common law duty to exercise reasonable care to guard against it. If failure to exercise reasonable care to avoid a foreseeable risk of an accident is the proximate cause of injury to another, legal liability follows unless contributory negligence on the part of the injured party was also a proximate cause. Therefore, contributory negligence or contributory fault can be described as unreasonable behavior on the part of an individual by which he or she contributes to injuries caused him or her by another's negligent act.

LEGAL EVOLUTION OF APPORTIONMENT

Historically, under common law, once contributory negligence on the part of a plaintiff is established by a defendant in a personal-injury action, this serves as a complete bar to the plaintiff's claim. In fact, this remains the law today in a number of American states. Perhaps this explains the reluctance that American courts have demonstrated to find failure to wear seat belts as contributory negligence, as this would permit negligent defendants to escape liability in total. This harsh judicial rule was subsequently altered by means of so-called apportionment legislation. Ontario enacted the first Canadian statute in 1924 and the other provinces followed over the next few years. Ultimately, this legislation spread to the United Kingdom and Australia and has, over the last few years, been appearing in the United States.

The Ontario Negligence Act (R.S.O. C296 S.4, 1970) provision is as follows:

> In any action for damages which is founded upon the fault of negligence of the defendant, if fault or negligence is found on the part of the plaintiff that contributed to the damages, the court shall apportion the damages in proportion to the degree of fault or negligence found against the parties respectively.

Under the statute, in a personal-injury action the finder of fact attempts to determine the extent to which the plaintiff's conduct contributed to his or her damages caused through the negligent conduct of the defendant.

Turning now to seat belts, the common law standard of the reasonable man in the role of a driver or passenger in a motor vehicle must be established. Until the 1950s, when seat belts were generally not available, any argument that failure to wear seat belts constituted contributory negligence would have been moot. As seat belts became generally available, first as an option and ultimately as a requirement incumbent on manufacturers, and as evidence came to light of a significant indication that the use of seat belts constitutes a major factor in saving lives and in preventing serious injuries in motor vehicle collisions, the notion of what was reasonable behavior changed. The courts have taken a similar view in imposing contributory negligence where workmen fail to wear safety goggles and safety ropes. The inconvenience of using the available seat belt is slight, and so the reasonable person would use it in the face of an unreasonable risk, even if it offered only a small amount of protection.

In one of the leading Canadian decisions that deals with the seat belt defense, Mr. Justice Munroe of the British Columbia Supreme Court reduced an award by 25 percent because of a plaintiff's negligence in failing to wear a seat belt. In stating what is now, in our view, the legal position in Canada, Mr. Justice Munroe said, "A person must use reasonable care and take proper precautions for his own safety, and such precautions include the use of an available seat belt" [Yaun v. Farstad, 66 D.L.R. (2d) 295 (B.C.) 302, 1967].

CONCEPT OF RISK

A year after the British Columbia Supreme Court decision, Mr. Justice Dubinsky of the Nova Scotia Supreme Court [McDonnell v. Kaiser, 68 D.L.R. (2d) 104 (N.S.), 1968] did not invoke the seat belt defense to mitigate damages because he expressed doubts about the general effectiveness of seat belts and concluded that "the effectiveness of seat belts is still in the realm of speculation and controversy." This attitude to disallow the seat belt defense was raised again in 1974 [Reineke v. Weisburger, 46 D.R.L. (3d) 239 (S.Q.B.), 1974], which cited a number of negative and uninformed arguments, such as no need to wear the belt within city limits and suggesting that even had the plaintiff worn her seat belt, she may have been seriously injured by another occupant tossing about the car. Mr. Justice Sirois rejected the established logic of Yuan v. Fairstad because no Canadian authorities had been quoted in that judgment. Mr. Justice Sirois further resurrected the following emotional suggestion: "A person driving down the highway on his proper side of the road is entitled to assume that other persons using the highway will obey the laws of the road still appeals to me and it is not negligence not to strap oneself in the seat like a dummy, a robot or an astronaut." These emotional arguments and others have arisen from a 1967 Hastings Law Journal article (2), which has been categorized as undistinguished and incomplete by Mr. Justice Allen Linden (3). Any doubt that Mr. Justice Dubinsky and others might have had over a decade ago surely can no longer exist in view of recent overwhelming evidence as to the effectiveness of belts.

In a 1974 Ontario Supreme Court decision [Smith v. Blackburn, R.T.R. 533; 2 Lloyd's Rep 229 (Q.B.D.) 1974], where the seat belt defense was not recognized, there was the suggestion that it was necessary for the individual plaintiff to actually foresee the exact risk that arose. In most judgments in the past decade in Canada, the court has found it only necessary that the general risk be foreseeable by a reasonable person (Drage v. Smith, R.T.R. 1 at 5F, 1975).

In regard to risk, 1 out of every 10 cars in Ontario is involved in a crash each year (according to a pamphlet by the Insurance Bureau of Canada, You and Your Car Insurance). One-half of the population of Ontario will be injured in a motor vehicle collision at some time during their lifetime (Ontario Legislative Debates, Dec. 2, 1975, p. 1185). This considerable risk is similar throughout Canada and the United States. Although earlier cases seem to have been decided against defendants because of a failure to prove the effectiveness of seat belts or the degree of risk involved, the currently available body of evidence permits the expert witness to statistically establish the concept of considerable risk and to readily establish the effectiveness of seat belts in preventing death and severe injury.

ROLE OF EXPERT WITNESS

Bearing in mind the above considerations, failure to wear a seat belt per se does not entitle a defendant to employ the seat belt defense as contributory negligence. The onus rests with the defendant to establish that, had the available seat belt been used, the injuries sustained either would have been avoided entirely or would have been of less severity. Thus, the causal connection between failure to use the belts and the resulting injury must be established. Many cases have turned on findings of fact as to whether failure to employ belts contributed to the plaintiff's injuries. Courts have placed much reliance on the use of expert evidence. Indeed, in some jurisdictions, medical and engineering experts are developing in this field. The challenging job for the expert witness, usually a physician with a special interest and expertise in motor vehicle trauma, is to prove causation in order to establish contributory negligence. His or her testimony is sometimes augmented by that of an engineering expert.

One of Canada's recognized leading experts, Carl Shiels, Research Engineer from the University of Saskatchewan, has been most helpful as an expert to the courts in many cases. However, the medical aspects of expert testimony may require confirmation by a licensed physician and expert, as pointed out by Mr. Justice McPherson in his judgment in the case of Ohlheiser v. Cummings [1978, Sask Q.B. (not yet reported, case study available from authors)], where he asked if he should accept Mr. Shiels as qualified to say how much force would break the plaintiff's leg. "Without disrespect to him, therefore, I cannot accept Mr. Shiels' opinion that Mrs. Ohlheiser's leg might not have been broken if she had been restrained." Despite this ruling, the court did accept Shiels' opinion that the injuries to the plaintiffs would have been less severe if they had worn their seat belts. He, therefore, found contributory negligence.

More cases have failed to establish contributory negligence on the point of causation than for any other reason. When causation is not proven, the cases are seldom appealed, as it would mean overturning a factual ruling rather than a point of law.

In some cases an expert is able to establish causation of injury for some vehicle occupants but not for others. In some cases of multiple injuries to one occupant, testimony can clearly establish causation for some injuries and be unclear in the remaining injuries. Where another passenger who was wearing a seat belt suffers no injury, the courts are often more willing to find causation. In some instances, the courts should be willing to conclude from the very nature of the injuries that they would have been prevented (Toperoff v. Mor, R.T.R. 419 at 421F, 1973).

Historically, tort law has played an important role in deterring certain types of conduct. Although some contend that tort law has no business establishing new standards of care without legislative initiative, this has always been the case. For example, tort law insisted on reasonable speed on the highway long before statutory speed limits were set. Under the umbrella of the reasonable-man test, tort law has fostered safety by taxi companies, public transit systems, and even the medical profession. The courts have also held as contributory negligence the failure to use safety devices such as a safety rope and safety goggles (3).

In Canada, most of the apportionment statutes speak of fault that causes "damage or loss", as it is not alleged that the plaintiff's negligence caused the collision but rather that it caused or contributed to his or her injury. If the courts uniformly adopted the view that once the causal connection is established between failure to wear seat belts and the injuries suffered by the plaintiff, such failure constitutes contributory negligence and thus would offset the amount against the overall reward; this could go a long way toward educating the public and encouraging them to use belts. Until now the offset through failure to use belts, even where it has been established that such failure was the major cause of the injury suffered by a plaintiff, has amounted to, at most, about 35 percent. The range appears to be between 5 and 35 percent, with the average offset being about 25 percent. In a recent case [Ulveland v. Marini, 4 C.C.L.T. 102 (B.C.S.C.), 1977], the plaintiff wore the available lap belt but failed to employ the shoulder harness. In judgment, the 25 percent contributory negligence offset was quoted as reasonable and fair where no belt was worn, and a 15 percent offset was awarded where the shoulder harness had not been used.

Although an offset as low as 1 percent has been found in seat belt contributory negligence [Plitchie

v. MacNeil, 1979, N.S.S.C. (not yet reported; case study available from authors)], a recent judgment in the British Columbia Supreme Court [Needham v. Harron Brothers, 1980, B.C.S.C. (not yet reported; case study available from authors)] decided that the failure to wear the seat belt must be treated as a small contribution and fixed this small contribution at 15 percent. Except in jurisdictions where the ancient common law traditions of contributory negligence apply (that is, where contributory negligence acts as a complete bar to a claim by the plaintiff), in jurisdictions where the seat belt defense is recognized the better approach would be to permit an offset against the overall damages found against the defendant according to the percentage causation established by the defendant from failure to wear the belts. This is certainly the spirit behind the provisions of the Ontario Negligence Act and of similar apportionment legislation elsewhere.

INFLUENCE OF MANDATORY STATUTE ON COURT DECISIONS

What role do statutes that mandate the use of seat belts play in all of this? In these instances, the courts often seize on the statutory duty as establishing the standard of care expected of the actor and, where that standard is breached and the other elements required to prove negligence are established, liability is imposed on the defendant. Indeed, breach of a statutory duty can serve as grounds for contributory negligence on the part of a plaintiff.

In the motor vehicle situation, it seems obvious that safety statutes such as traffic rules are designed for the broad purposes of preventing collisions in which plaintiffs are just as likely to be hurt as defendants. Therefore, it can be argued that violation of these statutory standards serve as evidence of negligence. Thus, the fact that an individual made a left-hand turn at an intersection where such a turn was prohibited may serve as evidence of negligence, even though without such a statutory requirement, at common law an obligation not to make that turn at that intesection did not exist.

When we consider the use of seat belts, a somewhat different situation exists. As we have said, the use of seat belts over the past few decades has been increasingly recognized as an important deterrent to prevent serious or fatal injury. Therefore, there has been growing recognition at common law that a reasonable driver or passenger in an automobile with available seat belts would use those belts. Thus, at common law, failure to use the belts would be evidence of negligence. We are reluctant to apply the term prima facie evidence of negligence as some courts have done because this may not appear reasonable. The issue of whether the evidence presented by the defendant on the matter as to whether the causal link between the failure to use the belts and the injury suffered was established usually remains one for the fact finder.

In Jackson v. Miller [25 D.L.R. (3d) 161 O.H.C., 1971], Mr. Justice Osler determined from the nonexpert evidence that plaintiff Jackson had been ejected from the vehicle and concluded that this was prima facie evidence of contributory negligence. "I have no doubt that if Jackson had remained with the car, he would not have suffered the injuries he received." He further found that, as a matter of fact, the plaintiff's injuries resulted from coming into contact with the ground after being ejected from the car. Also, as a matter of law, he found the injuries to Jackson were contributed to by his own negligence. The mechanism of injury and the protective role of seat belts might be established adequately for some courts to find prima facie contributory negligence but, in the great majority of recent cases, failure to establish the causal link was the result of poverty or absence of expert testimony.

UPDATING OF REASONABLE BEHAVIOR IN EYES OF THE COURT

It appears from recent court decisions that plaintiffs are not excused from using the seat belt because they hold certain views as to its effectiveness [Gagnon v. Beaulin, 1 W.W.R. 702 (B.C.S.C.), 1977], or because they were not advised by the driver to use the seat belt [Beaver v. Crowe, 49 D.L.R. (3d) 114 (N.S.S.C.), 1974], or they were unaware of the availability of the seat belt in the vehicle in which they were riding (Jackson v. Miller).

It seems from the recent British Columbia Supreme Court decision [Arnie v. Adams, 1980, B.C.S.C. (not yet reported; case study available from authors)] that, where there is a duty established to wear belts, in the absence of belts in the vehicle, there is a duty to have them installed.

In a leading motor vehicle negligence case [Sterling Trusts Corp. v. Postma, 48 D.L.R. (2d) (S.C.C.), 1964], the Supreme Court of Canada held that a breach of the Ontario Highway Traffic Act shifted the burden onto the defendant to disprove negligence. Although this principle has since been challenged, another motor vehicle negligence case in the Ontario Court of Appeal [Queensway Tank Lines Ltd. v. Moise (1969), Ont. Reports 1970, Vol. 1, 535 (O.C.A.)] heard judgment by Mr. Justice Mackay that suggested that a breach of the Highway Traffic Act is prima facie evidence of negligence unless the defendant can show by evidence no fault or want of care on his or her part. Applying this rule to the failure on the part of the plaintiff to use the available seat belt as mandated in the Highway Traffic Act, should the burden of justifying the nonuse of the available restraint system fall on the plaintiff? In Ohlheiser v. Cummings, Mr. Justice MacPherson addressed this question in his judgment as follows: "When we all pay for one another's hospital and medical care and other loss through taxes or insurance, we each have a right to say to a driver and to a passenger: 'Fasten your seat belts in my interest if not in your own. If you don't fasten them, then you may have to pay part of your loss if you are hurt.' What we have here is not a new interference with private rights but the creation of a new public duty in the automobile age."

The imposition of a statutory duty to use seat belts only confirms what has been developing at common law. That is, so-called reasonable drivers and passengers do use seat belts. With these statutes, they must use them. Although such statutes will no doubt encourage greater numbers of people to wear belts, which depends on the extent to which the seat belt statutes are enforced, the growing evidence and education of the public as to the effectiveness of belts in preventing serious injury and death and the growing adoption by the courts of the contributory negligence defense (where persons have failed to wear belts and such failure has contributed to the injuries) were having a similar effect in any event.

One could question whether the existence of a statutory duty by itself is sufficient to permit a court to find contributory negligence in circumstances where the common law might not require it. For example, consider the individual sitting in his or her car at the side of the road with the motor off who is not wearing a seat belt. The statute may require that it be worn. At common law, the reasonable person might not wear it in that situation. The car is subsequently struck by another vehicle.

The expert evidence establishes that, had the person been wearing the belt as required by the statute, the injuries would have been less severe. At common law, the reasonable person in this position may not have buckled up because he or she may not have foreseen the possibility of being involved in an automobile collision in those circumstances. If and when a case of this sort comes to court, the matter of whether the seat belt statute by itself serves as evidence of contributory negligence will then be determined.

In fact, few cases that involve failure to wear seat belts ever come to court because of the general practice that most motor vehicle collision claims are settled out of court. At the present time in Canada, where expert opinion indicates that failure to wear a seat belt contributed to the injury, lawyers are becoming sufficiently sophisticated in this area to advise their clients of the offset that courts would likely award. However, because of the significant effect that seat belts have on preventing serious injuries and death, conviction under a statute for failure to wear a seat belt could have a similar effect on a person's insurance premiums as do a number of speeding convictions.

CHILD-RESTRAINT SYSTEMS

There is also the aspect of statutory standards that govern the use of child-restraint systems. With an increasing emphasis on the rights of children and their access to independent legal representation, where parents fail to take appropriate measures to ensure that their children are adequately placed in a restraint assembly, and these children suffer injuries directly caused by such failure, an action will lie against their parents for negligence. Where a parent with a young child improperly places the child in a system or does not put the child in one at all and is involved in a collision, even though the collision may be entirely the fault of someone else, the defendant will have available the seat belt defense to the extent that the injuries suffered by the child may be directly attributed to the improper use or the failure to use the child-restraint system. That amount that is offset against the global award may be claimed by the child against his or her parent. Here the standard of the reasonable parent will likely be determined by the manufacturers' standards for the use of the child-restraint system set out in one type of legislation and by legislation that mandates the use of such a system.

We would strongly recommend the adoption of uniform seat belt legislation that standardizes the types of restraint systems required to be installed by manufacturers and also mandates their use. Such legislation should make it clear, however, that the existence of the seat belt defense does not constitute a complete bar to a plaintiff's claim, thus granting immunity to a negligent defendant. Rather, we would suggest that language similar to the Ontario Negligence Act be adopted, which indicates that the apportionment will be in direct relation to the causation demonstrated through expert evidence. In those states where contributory negligence constitutes an absolute bar to recovery, apportionment statutes should be enacted.

CONCLUSIONS

In summary, the effectiveness of seat belts in reducing or preventing injury in motor vehicle collisions has been well established. However, it has been demonstrated that, regardless of educational campaigns designed to increase the public's awareness of the usefulness of seat belts, public acceptance has evolved very slowly. Where mandatory seat belt legislation is in effect, there is generally a sharp increase in use. However, to improve and sustain use in regions where mandatory belt laws apply, enforcement must be seen to be ever present.

The concept of contributory negligence has of late shown a greater acceptability in motor vehicle collisions that involve plaintiffs who are not wearing their belts during a collision. Contributory negligence is conduct on the part of the plaintiff that falls below the standard to which he or she is required to conform for his or her own protection. Further, as a result of such conduct, the plaintiff must have suffered harm. Thus, although defendants may be found negligent in violating their duty of care to the plaintiff and would otherwise have been held liable for the full extent of the proven damages, plaintiffs may either be denied recovery altogether or their damages may be reduced in proportion to the extent of their own negligence in not wearing a seat belt on the principle that their own conduct disentitles them to fully succeed.

In a motor vehicle action where this concept is raised, the judge or jury must conclude both that the seat belts are an established, desirable safety feature of the vehicle and that in the collision the available restraint system would have reduced or prevented the injury. It will be easier to establish the first element, the desirability of the use of the belt, where seat belt legislation is in effect.

Even in jurisdictions without seat belt legislation, the common law over the past two decades has been increasingly recognizing that failure to wear seat belts constitutes contributory negligence. Seat belt studies should make it clear that there is no limit on the amount that may be offset against the overall damages awarded. That is, contributory negligence for failure to wear seat belts should permit an offset against the global award to the full extent that expert evidence establishes that failure to wear the belts contributed to the injuries suffered, with no ceiling on this amount.

The causal relation between the violation of the statute (or the failure to wear the belt at common law) and the harm to the plaintiff must still be established. It therefore becomes necessary to analyze the kinematics of the collision, relate them to the described injuries, and predict the modification of the injury patterns that likely would have prevailed if the available restraint system had been properly employed.

Mr. Justice Linden, at the time Professor of Tort Law at Osgoode Hall Law School, urged the judiciary to consider the following (3):

> Tort law should do what it can to encourage the use of seat belts. It has at its command the machinery for this purpose. If it were held that the failure to wear belts amounted to contributory negligence, it might help to educate the public to their undoubted effectiveness. Although the deterrent role of tort law has diminished in importance since the rise of liability insurance, the threat of a finding of contributory negligence may still have some force. The unbuckled plaintiff cannot merely shrug his shoulders and say his insurance company will pay for his negligence; it is money out of his own pocket if he neglects to strap himself in. Moreover, by adopting the seat belt defence, our courts may act as a catalyst to our sluggish legislatures. By moving into this field, perhaps tort law can stimulate more comprehensive legislative treatment, something that would be prefer-

able to the piece-meal approach of the common law.

We submit that if the common law continues to develop by itself, the seat belt defense will be increasingly recognized by the courts in the assessment of contributory negligence. If the seat belt defense is to be recognized by law, such statutes should be broad rather than restrictive to provide just penalty for unreasonable behavior on the part of an individual by which he or she contributes to injuries caused him or her by someone else's negligent act.

ACKNOWLEDGMENT

Acknowledgment is due Harry Munro, law student, University of Western Ontario, for his extensive literature search and advice in support of this paper.

REFERENCES

1. F.G. Fisher. Effectiveness of Safety Belt Use Laws, Final Report. National Highway Traffic Safety Administration, U.S. Department of Transportation, May 1980. NTIS: PB80-209901.
2. J.M. Kleist. Seat Belt Defence--An Exercise in Sophistry. Hasting Law Journal, Vol. 18, March 1967, pp. 613-622.
3. A.M. Linden. Torts--Seat Belts and Contributory Negligence. Canadian Bar Review, Vol. 49, Sept. 1971, pp. 475-485.

Publication of this paper sponsored by Task Force on Occupant Restraint Systems.

Impact of Legislation and Public Information and Education on Child Passenger Safety

K.W. HEATHINGTON, JOHN W. PHILPOT, AND RANDY L. PERRY

The State of Tennessee passed legislation in 1977 (effective January 1, 1978) requiring that children under four years of age who are traveling in motor vehicles, with certain exceptions, be restrained in child-restraint devices (CRDs). A large-scale public information and education (PI&E) program was established that concluded with an analysis of the impact on child passenger safety of the legislation and PI&E program. The PI&E program involved two intensity levels of application: (a) a higher-intensity level, called the comprehensive plan (CP), and (b) a lower-intensity level, called the basic state plan (BSP). At the end of a two-year period, the CRD use rate was increased 103 percent over the baseline rate based on statewide estimates. The CP, when applied to target areas during the operational period of this research, was significantly more effective in increasing CRD use than the BSP. The expected number of deaths was reduced by 10 over a three-year period. There was a strong correlation between individuals using seat belts and individuals protecting their children by placing them in CRDs. Characteristics of nonusers of CRDs were identified through various statistical analyses. A nonuser is (a) less likely to be wearing a seat belt, (b) more likely to have a lower education-attainment level, (c) more likely to have more passengers in the vehicle, (d) more likely to be transporting older children (under four years of age), (e) less likely to be the parent of a child, (f) likely to be in a lower income bracket, and (g) less likely to own the vehicle.

The State of Tennessee passed legislation in 1977 requiring that children under four years of age who are traveling in a motor vehicle, with certain exceptions, be restrained in child-restraint devices (CRDs). The legislation became effective January 1, 1978. As a result of this legislation, a large-scale public information and education (PI&E) program was established in the state that concluded with an analysis of the impact on child passenger safety of the legislation and the PI&E program. The State of Tennessee, by passing an active child-restraint law, provided a unique research situation in the United States. Until Tennessee passed the restraint law in 1977, which required that children under four be protected in most moving vehicles, no state had any type of passenger-restraint law for any age group.

The research reported on in this paper was designed to investigate the effect of the Child Passenger Safety Program on the reduction of fatalities and injuries to children under four years of age in Tennessee for a two- to three-year period after the law and PI&E programs were implemented. [This paper is one portion of the larger research effort of the Child Passenger Safety Program (1-12).] Study areas were selected and procedures were developed to collect data on CRD use. The data-collection instruments were designed to record information from both observations of CRD use and interviews with parents. The information collected included characteristics of children under four years of age as well as characteristics of their parents.

BACKGROUND

Target Areas

The target areas chosen for this research were representative of both urban and nonurban areas in Tennessee. The five major metropolitan areas of the state were selected for the urban sampling; i.e., Memphis, Nashville, Knoxville, Chattanooga, and the Tri-Cities area of Johnson City, Kingsport, and Bristol. Three nonurban areas, one in each of the geographical divisions of the state, were chosen to represent the more rural population. The nonurban target area was made up of merged data from Dyersburg, Columbia, and Morristown. The term more rural is used because the three areas in which the sampling occurred may not be considered rural by most standards, although the population that surrounds each town within an approximate 30-mile radius is largely rural. Each of the nonurban areas chosen, however, has towns within the 30-mile radius that have more than 5000 persons in population. The east Tennessee area has three towns within 30 miles that have more than 5000 residents, the middle Tennessee area has two, and the west Tennessee area has one.

An average of five sites was chosen within each urban area to collect data. The nonurban areas had one or two sites each. Shopping areas, regional and local, were selected as the sites to collect a large percentage of the data because of the large volume of traffic composed of parents who stopped with small children. The selected sites represented a variety of types of shopping areas that attracted a broad range of shoppers from low to high socioeconomic and educational levels. The locations of the urban and nonurban areas chosen as study areas for this research are illustrated in Figure 1.

Tennessee Characteristics

The State of Tennessee is divided naturally into three geographical divisions--the mountains and valleys of east Tennessee, the basins and rolling hills of middle Tennessee, and the flat lowlands of west Tennessee. In 1970, the population was 3 923 687, with approximately 59 percent of the people residing in nonrural settings. Nonrural is defined as any place of 2500 or more inhabitants. The population estimate in 1977, just before the beginning of the research, was 4 299 000. The urban populations of the areas under study are shown in Table 1 (13), and the populations of the nonurban areas used in the study are shown in Table 2 (13). In 1975 Tennessee had 81 272 miles of highways and streets, 12 308 miles of which were classified as urban. There were 2 725 569 registered motor vehicles in 1975, which included more than 2 000 000 automobiles. A total of 2 434 206 persons had valid driver's licenses in Tennessee in 1975. There were 32 926 million vehicle miles driven in Tennessee in 1975. The estimation for 1978 by the Tennessee Department of Transportation was 37 500 million vehicle miles. The estimate of the total number of children under four years of age, as of January 1, 1978, was 251 132.

Figure 1. Location of target areas for data collection.

Table 1. Population, selected Tennessee urban areas.

Area	Population, 1970	Estimated Population, 1975
Memphis	623 530	661 319
Memphis (SMSA)	834 006	873 300
Memphis SMSA (TN)[a]	750 015	767 000
Nashville (metro)[b]	448 003	451 200
Nashville (SMSA)	669 144	753 100
Knoxville	174 586	183 383
Knoxville (SMSA)	409 409	436 100
Chattanooga	119 082	165 282
Chattanooga (SMSA)	370 016	393 000
Chattanooga SMSA (TN)[a]	260 567	294 000
Tri-Cities	85 772	99 365
Tri-Cities (SMSA)	312 876	339 000
Tri-Cities SMSA (TN)[a]	292 808	313 000

Note: SMSA = standard metropolitan statistical area.
[a]Includes only population inside Tennessee. [b]Includes all of Davidson County.

SAMPLING AND PI&E IMPLEMENTATION PLAN

The sampling and PI&E implementation plan for this research is shown in Figure 2. CRD use data were collected before the effective date of the law and every six months after the effective date of the law, for a two-year period. The data collected

Table 2. Population, selected Tennessee nonurban areas.

Area	Population, 1970		Estimated Population, 1975	
	City	County	City	County
Dyersburg	14 523		14 694	
Dyer County		30 427		31 727
Columbia	21 471		22 124	
Maury County		43 376		45 879
Morristown	20 318		20 655	
Hamblen County		38 696		43 405
Total	56 312	112 499	57 473	121 011

Figure 2. Data collection and PI& E implementation plan.

Target Area	Oct. 77–Jan. 78	Jan. 78–July 78	July 78–Jan. 79	Jan. 79–July 79	July 79–Jan. 80	Jan. 80–July 80	July 80–Oct. 80
Memphis		BSP	CP + LP	CP + LP	CP + LP	CP + LP	Analysis and Report Preparation
	BLD	SAS	SAS	SAS	SAS		
Nashville		CP	CP	CP	CP	CP	
	BLD	SAS	SAS	SAS	SAS		
Knoxville		BSP	CP	CP	CP	CP	
	BLD	SAS	SAS	SAS	SAS		
Chattanooga		BSP	BSP	CP + LP	CP + LP	CP + LP	
	BLD	SAS	SAS	SAS	SAS		
Tri-Cities		BSP	BSP	CP	CP	CP	
	BLD	SAS	SAS	SAS	SAS		
Nonurban Dyersburg Columbia Morristown		BSP	BSP	CP	CP	CP	
	BLD	SAS	SAS	SAS	SAS		

Legend: BLD = Baseline Data; SAS = Semiannual Survey; BSP = Basic State Plan; CP = Comprehensive Plan (includes BSP); LP = Loaner Program prior to January 1, 1978, provided the baseline data for comparison purposes throughout the program. As can be seen from Figure 2, baseline data were collected in each of the target areas. In addition, a semiannual survey was conducted at each location.

A comprehensive plan (CP) and a basic state plan (BSP) for PI&E were initiated for six-month intervals at different target areas. The use of these two plans in different target areas was to provide analyses relative to the cost-effectiveness of comprehensive PI&E programs. Two target areas, Chattanooga and Memphis, were selected for implementation of loaner programs. The loaner programs provided a mechanism whereby low-income families could receive CRDs without having to purchase them at regular retail prices.

The BSP was designed to distribute brochures informing parents of children under the age of four of the law and how they could protect their children. Stand-up posters for offices were designed and distributed with the brochures. Distribution was made to hospitals, doctors' offices, clinics, and other strategic places where parents with small children visited frequently.

The CP included using a mass media approach to inform the general public about the law and the need for passenger protection. Public service announcements, news spots, and talk shows on television and radio were used. Newspapers were encouraged to run feature stories and to cover events such as press conferences. Newspaper editorials were also effective public information sources. Billboards were also used as part of the CP. As can be seen from Figure 2, the CP was initially implemented in Nashville on January 1, 1978. The overall master plan shown in Figure 2 called for the number of target areas receiving the CP treatment to be increased during each six-month interval until all target areas were included.

A loaner program, designed to provide CRDs to selected citizens who could not afford them, was implemented at one target area beginning six months after the effective date of the law and PI&E program. A second target area received the loaner program six months after the first loaner program was initiated.

Evaluation Limitations

This study was designed to evaluate the impact of legislation and a PI&E program promoting child passenger safety. Measures used for evaluation were CRD use and the change in the number of fatalities and serious injuries among children under four years of age. One limitation recognized early in the project was the inability to restrict the PI&E treatment rigidly to a given target area called for by the study design. Nashville was designated as the target area to receive the CP initially. An evaluation of the effectiveness of the CP was based on the premise that Nashville could be compared with all other urban target areas that had received only the BSP. Minor leakage of CP information intended only for the Nashville target area was reported. Urban areas other than Nashville had some programs promoting CRD use that were not a part of the project design. These programs likely had an influence on CRD use rates beyond that which the BSP might have had. However, from normal observations, it did not appear that the leakage was sufficient to greatly influence the analysis.

There was another limitation that might have affected CRD use. Nashville, for instance, is the state capital; therefore, there is a larger number of government employees than in other areas of Tennessee. Nashville area residents are more likely than residents of other urban areas to be aware of new legislation because of local publicity. Thus, the use rates in Nashville could have been increased slightly by this difference in characteristics from other areas in the state.

Great care was taken to minimize any leakage of information from the CP into areas in which it had not been introduced. All activities of the program were controlled carefully to minimize any possible effects of external variables.

Data Requirements for CRD Use

Relatively large samples were required to make competent assessments of changes in CRD use. The sample-size calculations were predicated on the need to detect any substantial increases in the use rates of CRDs at critical points in the implementation plan. For example, in tests contrasting urban target areas within a particular operational period, would the pooled CP results give higher CRD use rates than the pooled BSP results? Or, when comparing across time periods for a given target area, would the BSP engender higher CRD use rates than those observed for the baseline? And would the CP result in higher CRD use rates when compared with the BSP across time? A number of such scenarios were investigated.

The appropriate statistical technique to determine if an observed increase is significant is a test of proportions. Figure 3 illustrates the hypothesis that was to be tested.

The next step was to specify the conditions under which the tests would be made. All calculations assumed the following:

1. $\alpha = \beta = 0.10$;
2. The baseline CRD use rate would be low, on the order of 5 to 10 percent;
3. The Δp increase in the CRD use rate induced by the BSP would be about 0.03 for the first six-month period and about 0.01 for the next six-month period;
4. The Δp increase resulting from the CP would be at least 0.05 for each six-month period; and
5. All effects would be additive.

In addition, when comparing across time periods, sample sizes per interval were initially assumed equal, and the following equation was used:

$$n = (4z^2/d^2)(p_1q_1 + p_2q_2) \tag{1}$$

where $d = p_2 - p_1$. Figure 4 shows curves for determining sample size by using the above formula for the stated conditions.

To illustrate the calculations for sample size, consider one of the more important comparisons, where the CRD use rates in Knoxville in operational periods 1 and 2 were to be tested to determine if the CP generated a significant increase in CRD use when compared with the BSP results. Let p_1 = expected CRD use rate for Knoxville in the first operational period (BSP) and p_2 = expected CRD use rate for Knoxville in the second operational period (CP). The sample sizes required for various combinations of p_1 and p_2 can be read from Figure 4 or, more precisely, calculated from Equation 1. The results are summarized in the array below:

p_1	0.05	0.05	0.06	0.10	0.08
p_2	0.11	0.10	0.11	0.16	0.13
n	266	363	406	410	491

The first combination (p_1 = 0.05, p_2 = 0.11, n = 266) also is illustrated in Figure 4. It can be

Figure 3. Basis for the statistical test.

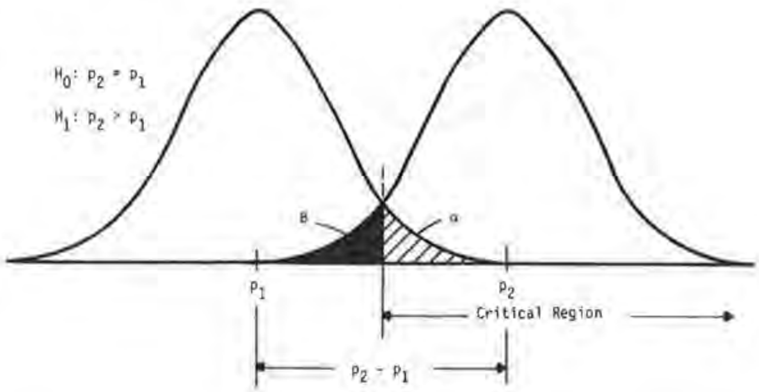


Figure 4. Sample size requirements for CRD use.

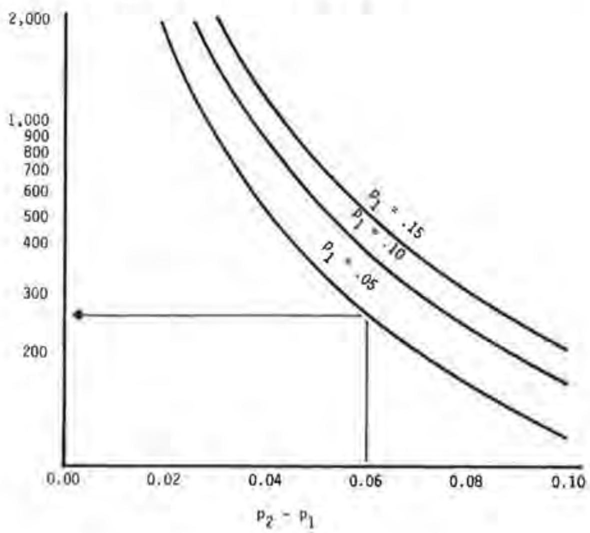
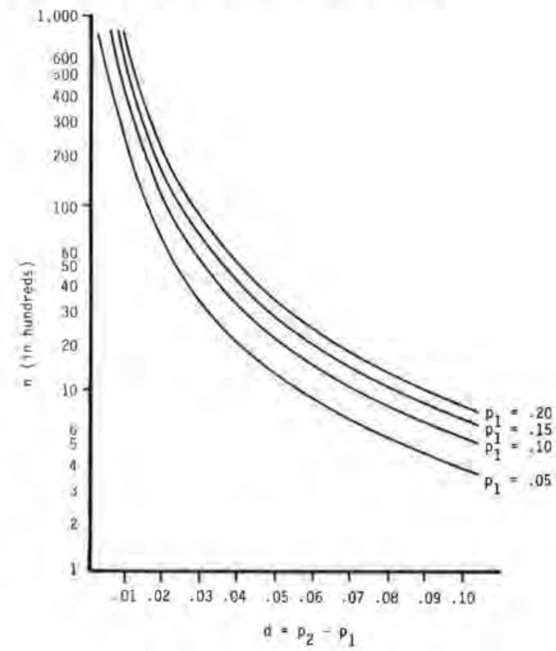


Figure 5. Sample size requirements for seat belt use.



seen that for most of the situations envisioned for this test, a sample size on the order of 400 was indicated.

A number of other test scenarios were investigated in a similar manner. No one sampling plan could satisfy both the budget constraints and the requirements for precision. The compromise solution was to obtain 800 observations for each target area in the baseline period, 400 observations for each target area per BSP operational period, and 500 observations for each target area per CP operational period.

The table below shows a comparison of design sample sizes versus the actual sample sizes:

Item	Design	Actual (avg)
Baseline	800	869
BSP	400	560
CP	500	630

It can be seen that, in general, the design sample sizes were exceeded by a comfortable margin.

Data Requirements for Seat Belt Use

A family of curves was developed for sample-size selection for seat belt use (Figure 5) for both urban and nonurban areas by using the same procedures for CRD use sample-size selection. The sample size for each category was based on a power of test of 0.99 and an alpha level of 0.01. The sample size required for each urban target area was 3800 observations. This sample size was chosen by assuming a 15 percent use rate during the baseline data-collection period and a 19 percent use rate during the first operational data-collection period. The sample size required for nonurban areas was 2913 observations--971 at each nonurban location. The assumption of 5 percent seat belt use by drivers at the nonurban locations was made after preliminary review of one day's data collection at one nonurban location. It was assumed that there would be about the same increase (3 percent) in seat belt use by drivers as the assumed increase in CRD use because of the fact that the PI&E program had an underlying message for all occupants and not just child passengers. It was anticipated that this sample size would be greatly exceeded if all drivers were observed while collecting the sample size for CRD use.

Death and Injury Data Requirements

To determine the death and injury rates resulting from motor vehicle accidents it was necessary to depend on accident records from accident investigation files. Based on the records of previous years, approximately 1000 accident injuries to children under four years of age were expected each year. Fewer than half of these were investigated by the Tennessee Highway Patrol. To compare death and injury rates by CRD use category, the Tennessee Department of Safety provided accident records of child occupants under four years of age involved in vehicle accidents; the specific data came from a supplemental accident data-collection form used by the Tennessee Highway Patrol.

Levels of Data Collection

Three tiers of levels of CRD use data were collected in this project. The tier 1 level was designed to record observed information in a matter of seconds as vehicles passed an observation point. Tier 1 data were not recorded on vehicles specifically exempted by the law. Data were gathered only if those eligible vehicles had at least one child estimated by the observer to be under four years of age. This level of data collection was performed primarily at entrances to parking areas at shopping centers, although a few observations were made at public health centers, pediatricians' offices or clinics, and children's hospitals. The information recorded on the tier 1 instrument included the disposition of the child or children in the vehicle (e.g., restrained or unrestrained), the use of the seat belt by the driver of the vehicle, and the license number of the vehicle for identification purposes.

The tier 2 level of data collection was designed as a combination observation, personal interview, and self-administered questionnaire. This instrument was used to gather specific information about the child, parent or guardian, vehicle, and CRD, if one were present in the vehicle or if one were owned but not present. The self-administered portion of the questionnaire was used to collect demographic data on the parent or guardian. The personal interview took approximately 30-60 s, and the self-administered part took about 60-90 s. Tier 2 level respondents were a subset of those observed at the tier 1 level where only an estimate of the child's age was made. Therefore, the first question at the tier 2 level was the age of the child.

The tier 3 level of data collection involved a questionnaire given to the parent or guardian to be filled out at a later date and mailed to the Transportation Center at the University of Tennessee. (This presentation does not report on any of the tier 3 level of information.)

RESULTS OF ANALYSIS

Baseline Data Results

A sizable number of vehicles (68 884) were observed during the baseline data collection in order to have a sufficient number of vehicles with small children. Only 9.1 percent of the vehicles observed had small children as passengers. The drivers of vehicles with children under four years of age as passengers were the parents of the children in 87.2 percent of the cases observed. The majority of the drivers with children under four were females (53.6 percent). There were more female (58.9 percent) than male drivers observed with children using CRDs. Of the types of CRDs owned, almost 93 percent were car seats and infant carriers. The car seat and infant carrier types account for 79.4 and 13.5 percent, respectively.

The composite use for all target areas combined for the baseline period was 9.2 percent (see Table 3). This percentage is an average of the percentages of use for urban and nonurban areas. A weighted average for just urban areas showed 11.8 percent use compared with 6.5 percent use for nonurban areas.

Table 3. Statewide CRD use rate estimates.

Target Area	Baseline	Operational Period First	Second	Third	Fourth
Nashville	14.0	22.1	19.0	19.1	24.6
Memphis	10.9	13.5[a]	16.5	22.6	18.9
Knoxville	12.8	20.4[a]	22.3	21.5	26.9
Chattanooga	10.9	16.5[a]	9.2[a]	15.0	23.7
Tri-Cities	10.7	17.9[a]	15.1[a]	19.9[a]	20.6
Urban average[b]	11.8	18.3	17.0	20.0	22.9
Nonurban average	6.5	12.5[a]	9.7[a]	13.0	14.5
Statewide estimates[c]	9.2	15.4	13.4	16.5	18.7

[a] BSP.
[b] Weight according to sample size.
[c] Estimates = (1/2) (nonurban + urban average).

A contingency table analysis of use for the baseline period revealed that there is a significant relationship between CRD use and the age of the child. The use rate was highest for the youngest children and lowest for three-year-olds. The number of children under four years of age in the vehicle had a significant bearing on use rates. The use rate was greatest when two or more children under four were present in the vehicle. Other significant relationships between selected variables and use included family income, marital status, number of adult passengers in vehicles, employment status of respondent, employment status of the couple, educational status of respondent, educational status of the respondent's mate, and educational status of the couple.

Seat belt use of all drivers observed during the baseline data-collection period was only 7.0 percent. The percentage of seat belt use by drivers with small children was even smaller at 4.1 percent.

There was a significant relationship between the driver's decision to use seat belts and the driver's decision to place a child in a CRD. It was discovered that, of those drivers who used seat belts themselves, 44 percent placed their children in CRDs. Of those drivers not using seat belts themselves, only 7.3 percent had their children restrained in CRDs.

Data Results of Operational Periods

CRD Use

Data were collected at each of the target areas each six months for a total of four data-collection periods after the law went into effect January 1, 1978. Thus, data were compiled over a period of two years after the law went into effect. Table 3 provides a summary of the statewide CRD use rate estimates for the baseline and four operational data-collection periods. It can be seen from Table 3 that there was generally an increase in CRD use rates for both urban and nonurban areas. Figure 6 shows the urban and nonurban CRD average use rates for the baseline and four operational periods.

It is interesting to note from Figure 6 that the second observational period showed a decrease in CRD use rates from the first operational period. The reasons will not be discussed in detail in this paper, but it was believed that this decrease was attributable to the enforcement program. The general public seemed to perceive that there would be strong enforcement at the beginning of the implementation of the law. However, the plan was to provide a strong educational program for the general public and then begin a strong enforcement program at a later date. When the enforcement did not occur at the beginning, the public most likely thought that enforcement was not going to occur. However, when it was substantiated that the general public did understand the legislation and was knowledgeable of its requirements, a strong enforcement program was emphasized. This began toward the end of the second operational period. From that point, there was a continual increase in CRD use rates.

Figure 7 is helpful in illustrating CRD use rates by target area and PI&E plan. This figure reveals that in every contrast indicated by arrow pairs, each BSP CRD use rate is significantly higher than its corresponding baseline value. In turn, each CP CRD use rate is significantly higher than its corresponding BSP value. It also shows that Nashville's four periods of the CP led to significantly higher rates compared with its baseline measure.

In summary, significant increases in CRD use were observed at each target area by the end of the two-year program. CRD use statewide was improved by more than 100 percent over the baseline rate of 9.2 percent. Comparisons across time indicated that the BSP resulted in a significant increase over the baseline and the CP generated a significant increase over the BSP rates. Comparisons within time periods tended to confirm these PI&E results. Nonetheless, the increases were relatively small in absolute terms, and, therefore, the cost-effectiveness of the programs becomes more important.

Comparisons with Another Study

The Child Passenger Safety Program measurements of CRD use were independently verified by another agency that was monitoring CRD use rates in two of the urban target areas. The Insurance Institute for Highway Safety investigated CRD use not only in Knoxville and Nashville, Tennessee, but also in Lexington and Louisville, Kentucky. By using a different methodology than the Child Passenger Safety Program, the Insurance Institute for Highway Safety collected data on cars exiting from shopping centers (14). CRD use rates were measured three times--August 1977 (pre-law), April 1978, and May 1980.

The Insurance Institute for Highway Safety was interested both in a comparison of pre-law and post-law CRD use in Tennessee and in a comparison of Kentucky and Tennessee rates, the most important difference being that Kentucky had neither a law nor an extensive PI&E program.

The CRD use rates obtained by the Child Passenger Safety Program for Knoxville and Nashville are superimposed on the Insurance Institute for Highway Safety results in Figure 8. The general trend for the Tennessee target areas, as measured by the two agencies, is the same: a low baseline rate, a moderate but significant increase shortly after implementation of the law, and a substantial increase in CRD use rates by the end of the study. The levels of CRD use for Tennessee, measured by the two agencies, are comparable as well.

The pattern for Tennessee, however, is in sharp contrast to that observed for Kentucky. While the two states had roughly the same initial CRD use experience, by the end of the study Kentucky's use rate had regressed to 14 percent, near its 1977 level, while Tennessee's use rates had climbed to nearly 30 percent. If Kentucky is a valid control, then the final result is vivid evidence of the long-term impact of the PI&E and law enforcement combination.

Figure 6. CRD use rates.

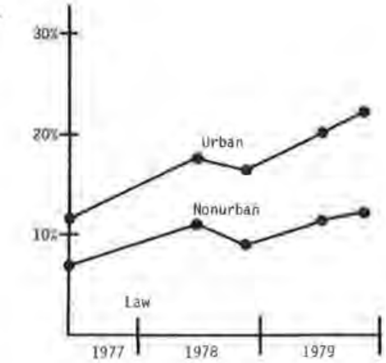
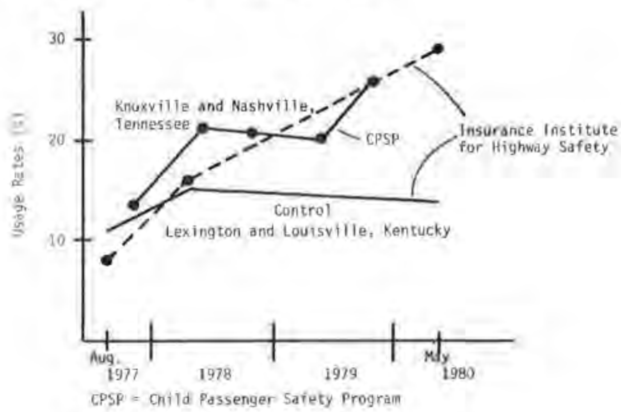


Figure 7. Comparison of use rate changes.

	Baseline	BSP (average)	CP (average)
Nashville	14.0		21.2
Memphis	10.9	13.5	19.3
Knoxville	12.8	20.4	23.6
Chattanooga	10.9	12.9	19.4
Tri-Cities	10.7	17.6	20.6
Urban Average	11.8	16.1	21.0
Nonurban Average	6.5	11.1	13.8
Statewide Estimates	9.2	13.6	17.4

Figure 8. Observed CRD use rates by agency.



Seat Belt Use

While the use of seat belts for all drivers declined overall after the initial measurement (baseline period), the use rate increased for those drivers with small children. There was less use in nonurban areas than in urban areas. Nonurban areas had 2.2 percent use of seat belts by all drivers, while the urban areas had 7.0 percent use for the operational periods.

The change in seat belt use rates for drivers with small children went from 4.5 percent to 7.1 percent in urban areas while in nonurban areas use shifted from 2.7 percent to 2.8 percent. Generally, the subset of drivers with small children had lower initial use than the set of all drivers.

The relationship between drivers' use of seat belts and restraint of children in CRDs, which proved to be significant for baseline data, was also significant when tested by using operational frequencies for the variables. The percentage of drivers observed as seat belt users and who also had children in CRDs increased between the two time periods from 44 percent during the baseline period to 55.3 percent in the operational periods. On the other hand, those drivers who did not use seat belts used CRDs for the children with them at only a 19.7 percent rate; but this percentage is a tremendous improvement over the 7.3 percent observed for the baseline period.

Babes-in-Arms Clause

The legislation in Tennessee permitted an older person to hold a child rather than placing the child in a CRD while traveling in an automobile. There are numerous data that indicate that this is a very serious and dangerous manner in which to transport children in an automobile. Many thought that parents and guardians in Tennessee would use this exemption within the law to circumvent using CRDs; however, the proportion of children under four being held by older passengers in the vehicle did not change significantly after implementation of the legislation and the PI&E program. Twenty-six percent of children in the baseline period and 22.6 percent of children in the operational periods were held by older passengers. Thus, the expectations of circumventing the law by holding children did not occur. (The clause permitting children to be held by older passengers was later removed.)

Accident Analysis

A supplemental accident data-collection form was developed for the Tennessee Highway Patrol and used in the investigation of accidents involving children under four years of age. The Tennessee Highway Patrol normally investigates accidents on state highways and would not normally be involved in the investigation of an accident on a road that was not under state jurisdiction. Thus, the total accidents analyzed and reported on here are less than the total that occur in the state.

Data taken from these supplemental forms for a three-year period (1978-1980) are reported in Table 4. It is seen from Table 4 that 350 observations were made on accidents in which CRDs were used. In addition, 964 observations were made on accidents in which CRDs were not used. The data were broken down into accidents in which there were no injuries to children, those in which there were injuries, and those in which fatalities occurred. It is also seen from Table 4 that an injury rate was calculated for CRD nonusers. Some 47.1 percent of CRD nonusers involved in accidents sustained no injuries; however, 49.5 percent of CRD nonusers involved in accidents did sustain injuries and 3.4 percent resulted in fatalities. It is immediately apparent that CRD users have two major advantages over the unrestrained if they are involved in automobile accidents: (a) CRD users are more likely to escape without injury and (b) CRD users have almost no risk of fatality. Other data also indicate that CRD users have less risk in each injury category (i.e., minor injury and major injury).

If one assumes that these same injury rates for CRD nonusers would apply to CRD users involved in accidents if they did not use CRDs, one can develop an expected number of no injuries, injuries, and fatalities for the CRD user category, as shown in the table below:

Type of Injury	Expected for CRD Users	Observed	Savings for Three-Year Period (expected-observed)
No injury	165	232	-67
Injury	173	116	+57
Fatality	12	2	+10

Thus, based on the table, one could expect 165 no-injury accidents for a three-year period, but 232 were observed. Thus, there was an increase of 67 noninjuries through the use of CRDs. In addition, if the CRD users had not used CRDs, one would have expected 173 injuries, but 116 were observed. Thus, there were 57 injuries that did not occur because of the use of CRDs. Even more important, one would have expected 12 fatalities to occur in this three-year period; however, only two occurred. Thus, there was a saving of 10 fatalities through the use of CRDs in this three-year period.

The total number of deaths and injuries of children under four in the state did not decrease under the Child Passenger Safety Program but remained in the same range as before implementation of the law. However, the total number of deaths fluctuates (i.e., ≈ 10 to 30) from year to year without significant changes in population, motor vehicle miles traveled, number of drivers, or other variables that tend to influence the number. Thus, the use of the actual number of death reductions as a sole measure of effectiveness is not really appropriate. It is more useful to investigate individual accidents to determine the death and injury rates for CRD nonusers, and then to apply these rates to CRD users to estimate the saving that most likely occurred with the use of CRDs.

CONCLUSIONS AND RECOMMENDATIONS

The evaluation of the PI&E program involved the measurement of the effectivensss of two intensity levels of application. The highest-intensity level, the CP, was applied in progression to specific target areas during the study. The lower-intensity level, the BSP, was used statewide for the entire period after implementation of the law. An evaluation was made of the two intensity levels by comparing the target areas having the CP with target areas having only the BSP.

There were factors that may have influenced the CRD use rates that were practically uncontrollable. These factors included the leakage of information, which only CP target areas were to receive, to other urban areas and independent programs in the urban areas, both of which were outside the control of this study. However, it is believed that the impact of this leakage was minor and, if completely eliminated, would give even more importance to the CP.

The results of this study should have application for similar situations in other states. The implications of the results of this study are included in the following summary of the major conclusions and recommendations:

1. The rate of CRD use was significantly increased in Tennessee after implementation of a law, a PI&E program promoting child passenger safety, and a law enforcement campaign. This conclusion is based on evaluation after two years of operation, assuming no seasonal variation. The final CRD use rate was 103 percent higher than the baseline rate, based on statewide estimates.

Every state should develop methods to increase child passenger protection. Legislation requiring the use of CRDs by small children should be one of the more important methods developed.

2. The CP, when applied to target areas during

Table 4. Automobile accidents for children under four years of age.

Type of Injury	CRD Users 1978	1979	1980	Total	CRD Nonusers 1978	1979	1980	Total	Injury Rate (%)
No injury	78	72	82	232	175	110	169	454	47.1
Injury	34	38	44	116	148	163	166	477	49.5
Fatal	-	-	2	2	3	17	13	33	3.4
Total	112	110	128	350	326	290	348	964	

the operational period of this research, was significantly more effective in increasing CRD use than the BSP. The actual size of the difference was partially masked by the bleeding of information into BSP areas. The CP is also substantially more costly than the BSP.

The decision of whether or not to use a CP as defined in this project should also be based on economic considerations. A lower-intensity plan, such as the BSP in this study, has a relatively low cost. Since the CP campaign had a definite impact, low-cost mass media programs should be considered. Any PI&E program should be coordinated with a law enforcement campaign.

3. There were no reductions in overall fatalities or serious injuries during the Child Passenger Safety Program for children under four; however, the children in CRDs had significantly more protection than those that were not in CRDs. Of the 35 individual deaths investigated in a three-year period, 33 involved children without CRDs. By this measure, use of CRDs prevented at least 57 injuries and 10 fatalities during the three years. This estimate is a minimum because the child deaths and injuries investigated are a subsample of the total child deaths and injuries of the state.

Since the frequencies of fatalities are low, this should not be used as a measure of effectiveness of this type of safety program. The best measure of effectiveness of the program is to apply the injury and fatality rates of unrestrained children to the group using CRDs.

4. The proportion of children under four being held by older passengers in the vehicle did not change significantly after the implementation of legislation and the PI&E program; 26.0 percent of children in the baseline period and 22.6 percent of the children in the operational periods were held by older passengers. However, holding children while traveling in an automobile is very dangerous.

Legislation should not permit an older passenger to hold a child while traveling in an automobile. (The Tennessee law's babes-in-arms clause has been rescinded.)

5. There was no increase in seat belt use by all drivers observed between the baseline and operational measurement periods. However, when a subset of drivers who had small children with them was measured, there was a significant increase in seat belt use. Drivers who are users of seat belts tend to protect their children by placing them in CRDs; i.e., there was a significant relationship between the drivers' decision to use seat belts and their decision to place their children in CRDs.

To increase the use rates of both seat belts and CRDs and thus to decrease deaths and injuries, a passenger-restraint use law for all vehicle occupants should be passed and strictly enforced.

6. The variables that best distinguish between users and nonusers of CRDs were identified. By using these variables as descriptors, a nonuser is (a) less likely to be wearing a seat belt, (b) more likely to have a lower educational-attainment level, (c) more likely to have more passengers in the vehicle, (d) more likely to be transporting older children (under four years of age), (e) less likely to be the parent of the child, (f) more likely to be in a lower income bracket, and (g) less likely to own the vehicle.

The major focus of PI&E campaigns should be directed toward the specific target groups that fit characteristics of the nonuser.

ACKNOWLEDGMENT

The research reported in this paper was sponsored in part by the National Highway Traffic Safety Administration and by the Tennessee Governor's Highway Safety Program. This presentation is only a portion of the larger research effort of the Child Passenger Safety Program at the Transportation Center, University of Tennessee. Many researchers, including Jo Lynn Cunningham, Clyde A. Pentz, Virginia Kraemer Redford, William A. Goodwin, E. Christy Hughes, Carol J. Culler, Pamela Moss Munz, Linda S. Geiss, Kevin C. Trent, Mark Lo, Dewey A. Wyrick, Dianne B. Sontag, and Julie S. Howard, have contributed to the overall research program and to the information contained in this presentation.

REFERENCES

1. C.J. Culler and J.L. Cunningham. Compliance with the Child Passenger Protection Law: Effects of a Loaner Program for Low-Income Mothers. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT HS-805 801, Oct. 1980.
2. J.L. Cunningham, C.J. Culler, and C.A. Pentz. Child Restraint Device Loaner Programs. National Highway Traffic Safety Administration, U.S. Department of Transportation, June 1981.
3. J.L. Cunningham, E.C. Hughes, J.W. Philpot, and C.A. Pentz. Parents' Knowledge, Attitudes, and Behavior About Child Passenger Safety. National Highway Traffic Safety Administration, U.S. Department of Transportation, May 1981.
4. K.W. Heathington. The Tennessee Child Passenger Safety Program. National Highway Traffic Safety Administration, U.S. Department of Transportation, Aug. 1981.
5. K.W. Heathington and J.W. Philpot. The Proper Use of Child Restraint Devices in Tennessee Saves Lifes and Reduces Injuries. Transportation Center, Univ. of Tennessee, Knoxville, Feb. 1981.
6. J.S. Howard, D.B. Sontag, K.W. Heathington, and M. Lo. Judicial Perspectives on Child Passenger Protection Legislation. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT HS-805 803, Aug. 1980.
7. P.M. Munz. Development of a Child Passenger Safety Component for Driver Education Programs. National Highway Traffic Safety Administration, U.S. Department of Transportation, June 1981.
8. R.L. Perry, K.W. Heathington, J.W. Philpot, C.A. Pentz, and M. Lo. The Impact of a Child Restraint Law and a Public Information and Education Program on Child Passenger Safety in Tennessee. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT HS-805 640, Oct. 1980.

9. J.W. Philpot, K.W. Heathington, D.B. Sontag, C.J. Culler, and J.L. Cunningham. Use of Telephone Surveys to Determine Awareness of Tennessee's Child Passenger Protection Law. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT HS-805 804, Sept. 1980.
10. D.B. Sontag, K.W. Heathington, and M. Lo. Enforcement of the Child Passenger Protection Law. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT HS-805 802, Oct. 1980.
11. D.B. Sontag, K.W. Heathington, E.C. Hughes, C.J. Culler, and L.S. Geiss. Organizational Networks for Promoting Child Passenger Safety. National Highway Traffic Safety Administration, U.S. Department of Transportation, Rept. DOT HS-805 806, Oct. 1980.
12. D.A. Wyrick, D.B. Sontag, and K.W. Heathington. Development of Materials and Public Relations Efforts to Promote Child Passenger Safety. National Highway Traffic Safety Administration, U.S. Department of Transportation, Oct. 1980.
13. Tennessee Statistical Abstract, 1977. Center for Business and Economic Research, Univ. of Tennessee, Knoxville, 1977.
14. Insurance Institute for Highway Safety. Status Report, Vol. 15, No. 10, June 25, 1980.

Publication of this paper sponsored by Task Force on Occupant Restraint Systems.

Automobile-Restraint Controversy: Analysis and Recommendations

DOUGLAS B. BRITTAIN AND YOSEF SHEFFI

Some of the costs and benefits of motor vehicle passenger-safety systems and policies, including passive seat belts, air bags, and a mandatory seat-belt-use law, are analyzed. This paper argues that since the last alternative is significantly more cost effective than the first two, the federal government should have offered it as an option to states instead of abolishing the passive restraint requirement.

In October 1981 (1), the Reagan Administration abolished the most significant regulatory action of the National Highway Traffic Safety Administration (NHTSA) by eliminating the passive restraint requirements from Federal Motor Vehicle Safety Standard (FMVSS) 208 (2). This regulation required cars manufactured after September 1982 to be equipped with passive restraints for front-seat occupants. Unlike conventional seat belts, passive restraint systems require no action by either driver or passenger in order to be activated, thus providing automatic protection to automobile occupants in almost every accident. These systems are aimed at the heart of the problem with current seat belt systems--the low use rate. Only about 10 percent of U.S. drivers wear their seat belts.

Manufacturers are planning to meet the pending requirement by using one of two systems: passive belts or air bags. Passive belts are standard seat belts that are automatically buckled up as the occupant enters the vehicle while air bags consist of large pillows that inflate in case of an accident, thus restraining the occupants' movement. A more detailed technical description of these systems is provided later in this paper.

The analysis offered in this paper shows that the recent action by the Reagan Administration can be easily justified on the basis of cost-effectiveness considerations. In order to be justified, though, it has to be compared with an alternative course of action--the implementation and enforcement of a mandatory seat-belt-use law. A comparison of this alternative with the passive belt and air bag solutions suggests that the federal government should have complemented its recent action by instituting a seat-belt-use law.

The federal safety concern is motivated by the high number of automotive accidents, which cause more than 50 000 deaths and 2 million injuries every year in the United States (3,4). The regulation under consideration concerns roughly 60 percent of the fatalities (or about 30 000 deaths) that are automobile occupants; the others include mainly pedestrians, bicyclists, motorcyclists, and streetcar occupants. All the alternative courses of action mentioned above are designed to save as many of these lives as possible as well as reduce the number and severity of injuries and ease the economic hardship associated with automotive accidents. This paper chooses the number of automobile-occupant fatalities as the key criterion in evaluating the effectiveness and costs of the three aforementioned alternatives. This criterion should, however, be treated only as a measure of effectiveness; the reduction in the number and severity of injuries may be a much stronger impact of these courses of action.

The paper is organized as follows. First, the passive restraint provisions set forth in FMVSS 208 are summarized, some background on the functioning of seat belts is presented, and the concept of effective use rate that is used in the analysis that follows is described. Second, analyses of the three alternatives are presented, i.e., passive belts, air bags, and mandatory belt use. This presentation includes a technical description as well as a discussion of effectiveness, costs, and other considerations. Third, the three alternatives are compared with each other in terms of several measures and, finally, the last sections summarize and conclude the paper with an outline of an implementation strategy.

The paper does not report new experimental results but rather tries to shed light on the controversy through a comparative evaluation of the major alternatives by using existing data. [An abridgment of this paper is given elsewhere (5).]

BACKGROUND

This section provides background information for the analysis described later in this paper. First, a review of the passive restraint requirements set forth in FMVSS 208 is presented. This is followed by a short description of how restraint systems work by using conventional seat belts as an example. Finally, the concepts of effectiveness and use rate are discussed; these are major factors in evaluating alternative restraint systems.

Regulations

The first large-scale federal regulation of the automobile industry was initiated in 1966 with the passage of the National Traffic and Motor Vehicle Safety Act. In this Act, many automotive safety requirements were established, most of which were introduced on 1968 model-year vehicles.

Most of the more recent safety regulations have been devised by NHTSA, an agency of the U.S. Department of Transportation (DOT) established by Congress in 1970. Although NHTSA has been involved in several areas over the years, its most significant regulatory action was the promulgation of the passive restraint law, officially referred to as FMVSS 208.

The passive restraint provisions of FMVSS 208 were originally scheduled to take effect as early as 1973, but the effective date has been postponed several times. As the law stood until October 1981, large and intermediate-sized cars built after September 1, 1982, were required to be fitted with passive restraints, as were small cars beginning in September 1983. In October 1981, however, the Reagan Administration abolished this law, citing the expected use of passive belts (rather than air bags) by manufacturers as one of the reasons.

Currently, all passenger cars sold in the United States are equipped with a three-point seat belt (one that incorporates both a lap and a diagonal shoulder belt) for the driver and front-seat occupant as well as with lap belts for rear-seat occupants. FMVSS 208 incorporated a number of provisions dealing with vehicle-occupant protection in several types of accidents. For front-seat occupants, passive protection (measured by injury criteria for head, chest, and femur) should have been provided in a frontal (±30°) barrier-crash test at 30 mph. In addition, passive protection should have been provided for a 20-mph lateral-impact test (within specified head and chest injury criteria) and for a 30-mph rollover test (during which the occupants must remain inside the car). A vehicle did not have to be subjected to the last two types of tests if active or passive lap belts (or three-point belts) were provided in conjunction with the passive restraint system. Note that this requirement virtually guaranteed that all manufacturers would have chosen the latter approach, typically by offering active lap belts.

How Seat Belts Work

In a frontal crash against a solid object at 30 mph, the vehicle comes to a complete halt in about 0.1 s, during which time the front of the car is crushed approximately 20 in (6). In such a crash, an unrestrained front-seat occupant continues forward at nearly 30 mph, leaving his or her (quickly decelerating) seat behind him or her. By the time he or she arrives at the dashboard, however, the vehicle is almost stopped. The result is the so-called second impact, in which the occupant must decelerate from close to 30 to 0 mph in the space of about 1 in. The resulting forces can cause massive injuries. The alternative outcome, in which the occupant is thrown through the windshield, can be equally disastrous.

If the occupant is wearing a three-point seat belt, he or she is subjected to much lower forces. As the seat decelerates, the occupant decelerates along with it (the so-called ride-down effect). In addition, belt stretch allows an additional 8 in or so of forward movement. Even when the car's post-impact recoil is taken into account, this still allows 12-16 in for the occupant to come to a halt, which is much more desirable than the 1-in deceleration space available without belts. Also note that with the use of the belts the head is prevented from contacting any unforgiving surface. Instead, the head's deceleration is partly absorbed by the neck movement without resulting in serious injuries (6).

In addition to providing protection in frontal collisions, a properly worn belt reduces injuries in practically all other collisions as well (7). Not only does a seat belt decrease the decelerative forces on the body, but it also dramatically reduces the possibility of occupant ejection. This is particularly important for accidents that involve rollovers.

The effectiveness of seat belts has been demonstrated and reported in several European and American research projects. Grime (6) analyzes many of these studies, observing that the estimated reduction in serious injuries varies from 45 to 70 percent, depending on the type and severity of the accidents investigated. One study found that for accidents equivalent to barrier crashes at approximately 35 mph, the probability of fatal or serious injury was six times greater without a seat belt. Also, the various studies showed that head injuries were reduced by half and ejections are practically eliminated with belt use.

Restraint-System Effectiveness

In discussing the effectiveness of any restraint system one has to distinguish between two components: (a) the effectiveness of the system given that it is activated and ready and (b) the activation rate or the rate of vehicles on the roads in which the system is active. These two components are discussed in turn in this section.

The first component of effectiveness may be termed technical effectiveness (or conditional effectiveness) to emphasize the fact that it measures reductions in death and serious injury given that vehicle occupants enter an accident with the restraint system in an active mode. (This means that if one considers a seat belt system, it is buckled and properly tightened and if one looks at an airbag system, it is armed and ready for use.) The technical effectiveness depends on three major factors: accident-type effectiveness, accident-type frequency, and system reliability. These factors are explained below.

1. Accident-type effectiveness--The various restraint systems are clearly not equally effective in all types of accidents. The accident-type effectiveness is the fraction of cases where a restraint system would prevent death or serious injury for an accident of a particular type. Thus, accident effectiveness of 80 percent in a frontal crash at 35 mph means that, on average, four out of five occupants who use the system are likely to escape death or serious injury. As another example, note that the effectiveness of any restraint system at a frontal accident at a speed of 60 mph or more is about 0.
2. Accident-type frequency--Different types of accidents (with different accident-type effectiveness measures) occur at different frequencies. In order to get a summary measure applicable to all types of accidents, one has to compute the average

accident-type effectiveness weighted by the accident frequency.

3. System reliability--The probability of a restraint-system malfunction could also be included, and the effectiveness measure described in 2 above could be reduced to account for it. This paper assumes that the reliability of all the systems considered is very high; thus, this factor can be ignored in the comparative evaluations.

In order to compare the effectiveness of the various policy options with regard to restraint systems, it is important to recognize that any system has to be in working order and operative in order to function. As shown later, the activation rate of the passive restraint systems is not 100 percent as argued by some proponents but significantly lower than that due to deliberate deactivation (mainly of passive belt systems) and low maintenance level (mainly of the air-bag systems). The low use rate of current seat belts was, of course, the motivation for FMVSS 208. Thus, one can define an activation rate as the proportion of vehicles on the road with active and operative restraint systems.

The product of the two above mentioned components--conditional effectiveness and activation rate--is a possible measure of effectiveness of a given automotive restraint policy option. Note, though, that these two components are not independent, as it is reasonable to assume that people who do not use their restraint system are more accident-prone than people who do (this effect is discussed in more detail later). The above measure of effectiveness should thus be scaled down to account for this selectivity bias. In this paper, the estimated activation rate for each system is modified to give a virtual activation rate that is always lower than the activation rate itself, thus accounting for this bias. The measure of effectiveness is thus the product of the technical effectiveness and the virtual activation rate.

As an example, one can calculate the effectiveness of the standard seat belt system. The conditional effectiveness of this system has been estimated by various sources to be around 70 percent. In 1976, then Secretary of Transportation William T. Coleman used a seat belt conditional effectiveness of 60 percent (and a lap belt conditional effectiveness of 40 percent) in evaluating various restraint alternatives (8). Thus, one can assume that if activated, the standard seat belt may be effective in, say, between 45 and 65 percent of the cases as a conservative range. The use rate of seat belts in the United States is only about 10 percent; thus, the combined measure of effectiveness is expected to be in the range of 4.5-6.5 percent. If one believes, now, that drivers who wear seat belts are more safety conscious and thus less accident-prone than the rest of the population, the use rate can be adjusted from 10 percent to, say, 9 percent. Multiplying this virtual use rate by the technical effectiveness, one gets a range of 4-6 percent. This range represents the effectiveness of the standard seat belt without mandatory activation regulation.

The measure of effectiveness described in this section and all its components are used in the analysis described in the next sections.

ANALYSIS OF ALTERNATIVES

This section presents safety, cost, and implementation-related aspects of the three major alternative approaches to occupant protection: passive belts, air bags, and (active) seat-belt-use laws. As mentioned above, the first two represent the current technology that manufacturers were planning to use for compliance with government regulations. The third is an alternative that has been implemented in many locations outside of the United States.

Note that these three alternatives are not the only methods of protecting automobile occupants or for increasing restraint use. Several other restraint designs exist, and the systems discussed here show potential for significant technical improvements. The focus of this paper, however, is on currently available options.

Passive Belts

Passive or automatic belts require no buckling or other passenger action to make them operative. The most familiar example of this is the Volkswagen VWRA passive shoulder belt system available on most of its models sold in the United States. A diagonal belt is attached to the upper rear of the door and runs down across the front of the seat back to a take-up reel positioned between the front seats. As the door is opened, the belt unrolls from its reel and moves forward away from the seat, thereby allowing the driver or front-seat passenger room to enter the vehicle. As the door is closed, the belt returns to restrain the seated individual. The spring-loaded emergency-lockup reel keeps the belt snug so that no manual adjustment of the belt is necessary. In addition, an emergency release of the belt is provided so that the occupant may still escape the vehicle in the event that the door becomes inoperative in a collision.

Although a single diagonal belt can effectively restrain the torso in a frontal crash, there is little to keep the front-seat occupant from sliding down under the belt, a phenomenon known as submarining. A lap belt is effective in preventing submarining, but the Volkswagen system does not incorporate a lap belt. Instead, a padded knee bolster is positioned under the dashboard and runs the width of the vehicle. Since this bolster prevents substantial forward movement of the knees, the possibility of submarining is practically eliminated.

Door activation of passive belt systems, as employed by Volkswagen, is an easy system of belt activation, but there are other possibilities as well. There now exists automatic belt systems that use electric motors to correctly position the belts during vehicle operation while still allowing sufficient belt-to-seat clearance for easy ingress and egress (9). Also, some manufacturers have designed three-point passive belts that need no knee bolster in order to be effective.

It should be noted here that the current belt and bolster design as employed by Volkswagen and other manufacturers seem to be less effective than the lap belt in four aspects:

1. The bolster can be depressed only 1-2 in while a belt can stretch and allow 4-8 in of body movement;
2. Bolsters must allow for normal vehicle operation and, unlike snug belts, therefore subject knees to a loading impact;
3. Bolsters transmit the impact to the knees, which are relatively fragile in comparison with the pelvis; and
4. Bolsters are of no use in other than frontal impacts; i.e., they do not provide protection on rollover or do not prevent ejections.

The remainder of this section discusses the effectiveness of the passive belt system and its costs.

Effectiveness

As mentioned previously, the effectiveness of a restraint device is the product of its conditional effectiveness and its virtual activation, where both components are in turn a function of several factors. In evaluating the effectiveness of a passive three-point belt, it can be assumed that its conditional effectiveness would be similar to that of the standard three-point belt. The particular Volkswagen design may have a lower technical effectiveness due to the above mentioned disadvantages of the knee bolster and the possibility of the door opening, which would leave the occupant unrestrained. Thus, one may use a conditional effectiveness of 45-65 percent for a three-point system, and, maybe, 40-50 percent for a Volkswagen-like design.

In order to estimate the activation rate of a passive belt, one can look at cars that offer such a system, e.g., Volkswagen Rabbit and Chevrolet Chevette, which use a similar design. Field data show a 75 percent use rate (i.e., 25 percent of the drivers disconnect the system). (This seemingly high deactivation rate may be rooted in the ease with which the system can be disconnected; it is, in fact, as easy to disconnect as to debuckle the standard seat belt.) This activation rate may be a high estimate, though, since drivers of these models may be more safety conscious than the rest of the population. This is evident from the high activation rate of the same model cars with standard seat belts, which is higher than 30 percent, or more than three times the national average. Thus, the activation rate in the general population will probably be lower than 70 percent. A low bound for the estimated activation rate for the entire population may be obtained from the ratio 75/30, which depicts the activation rate with passive belts and without it for the Rabbit and Chevette models. When applied to the general population (with its current 10 percent activation rate), this bound translates to an activation rate of about 25 percent. This figure is undoubtedly low, since people may get used to the passive belts as time goes on. A reasonable estimate of the activation rate lies probably between the two bounds in the range of 45-65 percent.

As mentioned earlier, the people who actually will be using the system and not deactivating it may be more safety conscious than the people who purposely are deactivating the system. To account for this, the above mentioned activation rate of 45-65 percent may be adjusted to a range of, say, 40-60 percent. The overall effectiveness may thus be estimated at approximately 20-40 percent.

Costs

When Volkswagen introduced the passive belt as an option in 1975, it was priced at $30. Current estimates range between $30 and $150, depending on the complexity of the system (10). This is in addition to the $100 that is the approximate cost of the currently required active belts. A figure of $50 is used in this paper as a working hypothesis for the incremental cost of the passive belts.

Air Bags

Air-bag restraining systems are somewhat more complicated than passive belts. They consist of large inflatable pillows, usually made of nylon, that inflate when the car hits a solid object at a speed greater than 12 or 15 mph. The driver's air bag is stored folded in the steering-wheel hub and the bag for the other front-seat passenger or passengers is stored in the lower right end of the dashboard. Air-bag deployment is electrically triggered by sensors located in the front bumper.

Immediately on receiving the signal from the crash sensor the air bags begin to inflate, typically by burning sodium azide, which produces nitrogen to fill the bag (11). On the order of 40 thousandths of a second later, the air bags check forward movement by restraining the occupants' heads and torsos. As with seat belts, the occupants are allowed several inches of forward movement in order to moderate their deceleration rates but, with the bag system, the forces are distributed over a much larger area of the body. Submarining is prevented by knee bolsters or by two-part air bags that have an additional inflatable chamber that controls forward movement of the knees.

Deflation of the air bags begins almost immediately due to the use of porous bag material or bags with holes in them (10). Within 0.5 s or so after initial impact, the automobile occupants are seated with the deflated air bags lying in their laps. Note that air bags are not reusable and must therefore be replaced after each collision severe enough to cause inflation. The remainder of this section discusses the effectiveness, costs, and other factors associated with the use of air bags.

Effectiveness

In a frontal impact, the technical effectiveness of air bags should be equal to that of a seat belt and may be even slightly better due to broader impact distribution with the bag. The bags were found as effective as belts in controlled experiments, but limited evidence suggests inferior performance in actual accidents. As a working hypothesis it is assumed that the conditional effectiveness of air bags is similar to that of the standard three-point belt.

Air bags, however, unlike belts, work only in frontal impacts. No protection is offered in side impacts, rollovers, or other nonfrontal impacts. It should be noted that fewer than 57 percent of occupant injuries are the result of frontal or near frontal (±45°) impacts (12). Furthermore, air bags would not deploy in low-speed collisions, which cause a substantial number of injuries. Thus, it is clear that air bags may not provide protection in more than half of the injury-producing collisions. This problem can be partly remedied with a lap belt, and such belts probably would be included with most air bags. The lap belt (which should be manually activated) would eliminate injuries in most types of accidents where air bags do not offer protection. Thus, the air bag-lap belt combination may be as technically effective as the standard belt, with a rating of 45-65 percent.

Another factor, though, may still detract from the air bag technical effectiveness, and this is their lack of performance in multiple-impact accidents. Due to the rapid deflation of air bags, protection will not be offered during a secondary impact. A recent British study (7) noted that a third of the occupants in the accidents studied were involved in multiple-impact accidents or rollovers. In these types of accidents it is important that the restraint system operate continuously. The problem with the air bag can, again, be partly remedied with a lap belt, and fully solved with a standard three-point system that could be supplied in addition to the air bag.

Assuming that such a system would not be supplied with air bags (and if it would, its activation rate would be negligible), the estimate of the technical effectiveness should be lower than the technical effectiveness of the standard seat belt. This paper, however, uses a rate of 45-65 percent for this system. This rate is similar to the technical effectiveness of the standard belt and is, thus, a conservative estimate (from the point of view of this paper).

The activation rate of an air bag-lap belt system is not 100 percent as some proponents of this system believe. In the first place, the lap belt requires active buckling, and field observations suggest that less than 10 percent of the lap belts are used in air-bag-equipped cars. There is little reason to believe that this figure will grow in the future. Furthermore, in light of the high replacement costs of air bags (this point is discussed later in this section), it is expected that a large number of them will never be replaced once initially used. For this reason, the air bag's conditional effectiveness would be limited to the first frontal impact (of a multiple-impact collision) of the first accident that the car would experience.

In order to compute the activation rate of the air-bag system, one should distinguish between those accidents where it would inflate (which are less than 70 percent of the injury-causing accidents) and those where it would not. In the first category of accidents, the activation rate measures the fraction of the population who will drive with an operative system, which may be as high as 90-95 percent (especially if replacing used bags will be covered by insurance companies). Because between 15-20 percent of these accidents may involve multiple collisions and rollovers, where the air bag would not be operational after the fist impact, this rate can be assumed to be in the range of 80-90 percent. This rate can be further reduced to account for the selectivity bias in the replacement of used air bags to a range of say, 75-90 percent. For the (injury-causing) accidents where the air bag would not inflate, protection is provided by the lap belt that, as mentioned above, is likely to be activated in no more than 10 percent of the cases.

Thus, for about 70 percent of the accidents the conditional effectiveness is 45-65 percent and the activation rate is 75-90 percent. For the other 30 percent of the accidents, the lap belt (with technical effectiveness of 30-40 percent) is the only protection and, even so, its activation rate is only about 10 percent. One can thus assume a technical effectiveness of 35-50 percent and an activation rate of 60-75 percent. The resulting effectiveness rate is on the order of 20-40 percent. This figure is in line with the range of estimates included in the 1972 report by the Ad Hoc Committee on the Cumulative Regulatory Effects on the Cost of Automotive Transportation (RECAT), which stated that pessimistic estimates predict air-bag benefits to be equal to a 33 percent seat-belt-use rate (10). The optimistic estimate included in the same report is of effectiveness equal to 80 percent that of a seat belt, a figure that seems a little high in light of the calculations above.

Costs

As with other air-bag-related issues, the issue of exactly how expensive air bags would be is open to great dispute, with NHTSA, consumer groups, and the insurance companies generally claiming relatively low costs and the automobile manufacturers predicting relatively high costs. Back in 1976, then Secretary of Transportation Coleman observed that estimated costs of air-bag systems ranged from $70 to $520 (13). (Note that these are estimates of the incremental cost of an air-bag system over the standard three-point belt.) The optional air-bag system being sold by General Motors at that time was a loss-leader at $315 (14).

Air-bag costs, like everything else, have increased with inflation. In 1978, NHTSA estimated the cost of air-bag systems in full production to be $200. Ford has recently recomputed its retail-price estimate for air bags and predicts that air bags would cost new car buyers $575 in 1982 for sales of 787 000 units annually. If most consumers prefer passive belts and only 200 000 air-bag-equipped cars were sold annually, then the retail price would rise to $825 (15).

The prices discussed so far are for air bags as original equipment. The costs of replacing air bags are expected to be considerably more, as is typical with the automobile repair business. The insurance coverage of some of these costs would, of course, be borne by consumers in the form of higher premiums.

In light of the above figures, it seems conservative to estimate the (incremental) costs of a car air-bag system at $500/car and, taking into account the fact that many cars would have to be equipped with the systems more than once, the actual life-cycle cost estimate may be twice that amount. For analytical purposes, this paper uses a working hypothesis of $600/car (borne either directly or in part in the form of higher insurance premiums).

Other Considerations

Many of the problems of the early air-bag systems have been solved. Two issues, however, still cloud the expected benefits of this system: their performance with respect to out-of-position children and the use of sodium azide as a gas generant.

The first issue is that children seated too far forward on the front seat can be subjected to unnecessarily high deceleration rates. This is because a child who is sitting on the front edge of the seat (so his or her feet can dangle over the edge) is pushed back into the seat (into the correct seating position) by the rapidly inflating air bag. Such a child has approximately 10 in less in which to decelerate. Children who are sitting more out of position, such as lying back with their feet up on the dashboard, can in some instances actually be hurt more by the inflating air bag than by the collision itself. Out-of-position adults present less of a problem due to their greater weight.

The problem of using sodium azide is not necessarily inherent to air bags, but the technology chosen by all American manufacturers is to inflate the air bags with nitrogen created during the combustion of sodium azide. Unfortunately, sodium azide is a class-B poison, a known mutagen, a suspected carcinogen, and becomes explosive when it contacts various metals. Although most car owners would have no contact with sodium azide, the use of this substance presents some real hazards to scrapyard workers.

The analysis presented later in this paper does not account for these issues; it is assumed that they will be solved in the future. Of course, the resulting air-bag costs may be even higher than $600/car.

Belt-Use Law

The effectiveness of automobile-occupant restraints has been recognized by many countries. Unlike the United States, more than 20 of these countries have chosen to enact a law that requires automobile drivers and passengers to wear their safety belts.

Most belt laws now in effect require front-seat occupants to have their seat belts buckled while their cars are in motion. Punishment for noncompliance can range from nothing to $300 or more and may also include a jail sentence, although the average punishment is a fine of approximately $10 (16). The degree of enforcement also varies widely, but in almost all cases no citations are issued unless the persons involved have already been cited for some other traffic-law infringement (8).

The remainder of this section discusses the effectiveness, costs, and constitutionality of a belt-use law.

Effectiveness

The technical effectiveness of the standard three-point seat belt was discussed earlier, where a range of effectiveness of 45-65 percent was suggested as a working hypothesis. The activation rate of the seat belt under such a law is not easy to estimate due to noncompliance problems. Foreign experience in countries where belt use is required by law suggests a use rate of 70-80 percent as typical (8,10), with considerable variation between countries. This number, however, has to be adjusted for selectivity bias that, in the case of this alternative, may be stronger than in the previous cases. The 20-30 percent of drivers who choose to violate the law may be significantly more prone to be involved in accidents than the average driver (17). If the accident likelihood of this group is, say, 25 percent higher than the population average, then the virtual activation rate is likely to be in the range of 55-70 percent instead of 70-80 percent.

Costs

A compulsory belt-use law would not add to the price of a new car but would instead cost state and local government additional money for enforcement. In the past, DOT estimated that enforcement of such a law would cost approximately $5 million annually (8). This figure may now be adjusted to $6 or $7 million to account for inflation; thus, a figure of $7 million is used in this paper.

Constitutionality

If belt-use laws were enacted, it seems probable that their constitutionality would be tested. Motorcycle-helmet-use laws have generally been found constitutional in the past, and thus it appears likely that belt-use laws would be upheld as well. The argument that belt wearing is a matter of personal safety and therefore not subject to state regulations is expected to be used, based on the Ninth and Fourteenth Amendments. The argument used by the courts in the case of the motorcycle-helmet-use law was that this action affects others in numerous aspects and therefore is subject to state regulation.

Even with the absence of belt-use laws in this country, some courts have held that failure to fasten an available seat belt constitutes negligence (e.g., Mount v. McClellan, 91 Ill App 2d 914). In most cases, neglecting to wear seat belts is not considered negligence that contributes to an accident but negligence that contributes to injuries. Consequently, the question of belt use is typically only one that governs the extent of compensation for injuries and does not block the compensation itself. In a few cases, the nonuse of belts has not been allowed to have an effect on the extent of the compensation (e.g., Britton v. Doehring, 286 Ala 498). If nationwide belt-use laws were to take effect, though, one can expect more rulings of the former type, which should increase the use rate. This would make it practically impossible to collect damages for injuries that a seat belt could have prevented, although the nonwearing of a belt would probably not be considered criminally negligent unless failure to use the belt was felt to be the proximate cause of the accident. In many foreign countries with compulsory belt wearing, courts have issued similar rulings (16).

It should be noted that although more than 110 mandatory belt-use bills have been introduced and defeated in 32 states between 1972 and 1977, some forms of the law currently do exist. The Federal Highway Administration, for instance, requires all truck drivers engaged in interstate commerce to wear belts. Tennessee requires all children under four years of age to be properly restrained and California requires belt use in all driver-education vehicles. In addition, state employees, officials, and policemen must wear safety belts while on duty in many states (18). Thus, mandatory seat belt use has at least some important precedents in the United States at this time.

COMPARISON AND ANALYSIS

This section compares the three alternatives presented in the last section along three dimensions: effectiveness, cost-effectiveness, and personal freedom.

Effectiveness

In order to create a more vivid measure of restraint-system effectiveness, one can use the expected number of lives saved by each of the alternative systems. As mentioned in the first part of this paper, annual fatalities of car occupants amount to about 30 000. Before any of the alternative approaches to occupant restraint are evaluated, though, it should be recognized that the current arrangement, which includes mainly active belts, has some positive effectiveness. In fact, it was estimated at approximately 5 percent. The number of annual fatalities used in comparing the various alternatives should thus be adjusted to 31 500.

Table 1 presents the conditional effectiveness, activation rate, and the expected number of saved lives for each system. The table presents both a likely (or midrange) estimate for each figure shown and the range itself.

The number of lives saved shown in the table was computed as the product of the potential number of fatalities (31 500), the activation rate, and the technical effectiveness. Note that the ranges of estimated number of saved lives shown in Table 1 overlap with each other and, even though the mandatory belt-use alternative seems to be the best in terms of the midrange estimate, it does not dominate the other alternatives in absolute terms.

Note also that the table displays an approximate number of lives saved annually with all cars on the road equipped with the system. Since either passive restraint alternative would be introduced only gradually, an immediate compulsory belt-use law would mean a large increment of lives saved in the early years. If one assumes 10 years as the period in which the number of passive restraint-equipped

Table 1. Effectiveness of alternative approaches.

Item	System		
	Passive Belts[a]	Air Bags[b]	Seat Belt Law
Technical effectiveness (%)			
Midrange	55	42.5	55
Range	45-65	35-50	45-65
Activation rate (%)			
Midrange	55	67.5	60
Range	45-65	60-75	55-70
No. of lives saved annually			
Midrange	9350	9040	10 395
Range	6380-13 310	6615-11 810	7800-14 330

[a]Includes passive lap belts. [b]Includes active lap belts.

Table 2. Cost-effectiveness of alternative approaches.

Item	System		
	Passive Belts[a]	Air Bags[b]	Seat Belt Law
Incremental cost per new car ($)	50	600	-
Incremental total annual cost ($000 000s)	500	6000	7
Incremental no. of lives saved			
Midrange	8030	7540	8895
Range	4880-11 810	5115-10 310	6300-12 830
Cost per life saved ($)			
Midrange	62 300	796 000	790
Range	42 300-102 500	582 000-1 173 000	560-1100

[a]Includes passive lap belts. [b]Includes active lap belts.

cars would approach 100 percent and assuming linear growth to 100 percent, it can be calculated that the belt-use law would save more than 55 000 more lives than air bags or passive belts in the first decade.

Cost-Effectiveness

Due to the difficulty in quantifying the costs of loss of life and limb, one of the most widespread techniques of evaluating investment in safety is their cost-effectiveness. In other words, safety investments are compared with other safety investments in terms of expected number of lives saved per dollar invested. Alternatively, one can calculate the investment costs per life saved, as is done in Table 2. This table assumes annual new car sales of 10 000 000 and the cost figures presented earlier for each alternative approach. The calculations of cost per life saved require some explanation. As mentioned, the dollar figures attached to passive belts and air bags represent the incremental cost of these systems over the standard three-point seat belt. In order to be consistent, one has to compute the incremental number of lives saved by these systems. Since the estimate of the number of lives saved annually with the current system is approximately 1500, this figure should be subtracted from the estimated number of lives saved shown in Table 1. (In other words, the calculations are based on midrange estimates of 8030 lives saved by passive belts, 7540 by air bags, and 8895 by a seat belt law.) The cost per life saved shown in Table 2 is calculated by dividing the total incremental annual costs by the expected incremental number of lives saved. The range shown in Table 2 corresponds to the range of estimates of number of lives saved annually depicted in Table 1.

The figures that represent the incremental cost per added life saved for each of the alternatives in Table 2 speak for themselves. There are several orders-of-magnitude difference between the compulsory seat-belt-use alternative and the others. Any potential inaccuracies in the assumptions made in the course of developing these figures are dwarfed in comparison to the magnitude of the difference in the cost-effectiveness of the alternatives. The compulsory seat-belt-use alternative is almost a hundred times more cost effective than passive belts and more than a thousand times more effective than air bags. (Note that we do not present any incremental analysis of the cost-effectiveness between the alternatives presented since, if based on the midrange figures, it would show a negative benefit for any of the passive restraint alternatives over the seat-belt-law approach.)

The only possible drawback of such a law may be the implied loss of personal choice involved. This point is discussed next.

Personal Freedom

Any discussion of a mandatory safety-belt-use law invariably generates a debate on the question of personal freedom or the right to choose whether to buckle up or not. In these arguments, the situation under a mandatory seat-belt-use law (where a person can choose to either buckle up or risk a fine) is compared with the current situation (the do-nothing alternative) where a driver does not risk a fine for not using the seat belt.

This paper argues, however, for an alternative to the status quo, and the law should thus be compared with other alternatives such as the passive restraint option. Under a passive restraint course of action similar to the provisions of FMVSS 208, all consumers would be required to purchase passive restraints with their new cars. No alternative will be available.

Under a mandatory seat-belt-use law, consumers would be given the choice of whether to buckle up or pay the extra money for passive restraint and thereby avoid the bother of buckling up (one can also choose to risk a fine and not use any restraint system). (Admittedly, the extra money is a little more than if all cars were equipped but, then again, no consumer will be forced to buy it.) It is highly likely that manufacturers would not have to be regulated to provide such an option, as under a mandatory seat-belt-use law market forces would encourage it. With mandatory seat belt use, the choice of buying or not buying a passive restraint option will be reduced to a cost versus inconvenience trade-off, similar in principle to the trade-off between manual and automatic transmission, manual or remote-control side mirror, and many other options.

In conclusion, it seems that a well-written nationwide belt-use law would actually provide more personal freedom and choice opportunity than the situation that would have arisen under FMVSS 208.

CONCLUSIONS

As evident from the analysis presented in this paper, mandating the use of seat belts (or any other restraint system) should be preferred to requiring manufacturers to equip cars with passive restraints as required in FMVSS 208. The Reagan Administration has recently dropped the passive restraint requirements of FMVSS 208 but did not go all the way and institute a mandatory belt-use law.

By its action, the Administration missed a unique opportunity to implement a belt-use law without a lengthy repeal process and furor from consumer groups. Instead of abolishing the regulations, the Administration should have exempted states that passed such laws from the passive restraint provisions of FMVSS 208. States then could have chosen between passing a mandatory seat-belt-use law or forcing consumers to pay more for their new cars. It is reasonable to assume that most states would have chosen the former route.

This course of action should be considered if and when the passive restraint requirements are considered again. Without the carrot (or exemptions), the federal government does not have an effective means to get states to pass and enforce such laws. The stick approach of withholding highway funding aid was proven ineffective and unpopular in the case of the 55-mph speed limit and is not likely to work in the case of a mandatory seat-belt-use law.

Last, two notes are in order. First, the Administration may trade off an issue that is not directly related. Thus, for example, states that pass a mandatory seat-belt-use law may be exempted from the 55-mph speed limit. This course of action may be less elegant than the aforementioned one but still possible and very effective. Second, it may be interesting to know that, starting in the summer of 1982, all drivers in England will have to wear seat belts.

REFERENCES

1. Federal Motor Vehicle Safety Standards: Occupant Crash Protection. Federal Register, Vol. 46, No. 209, Oct. 29, 1981.
2. Federal Motor Vehicle Safety Standard 208: Occupant Crash Protection. Code of Federal Regulations, Title 49, Section 571.208, Oct. 1, 1980, pp. 313-325.
3. I.G. Black. Going After the Real Cause. Automotive Industries, Vol. 161, No. 3, March 1981, p. 7.
4. B. Carraro. Motor-Vehicle Deaths Edge Upward in 1980. Traffic Safety, Vol. 81, No. 3, May/June 1981, pp. 20-25, 30.
5. Y. Sheffi and D. Brittain. Motor Vehicle Safety: Passive Restraints Versus Mandatory Seat Belt Wearing. ITE Journal, Feb. 1982, pp. 26-29.
6. G. Grime. The Protection Afforded by Seat Belts. Department of the Environment and Department of Transport, United Kingdom, Supplementary Rept. 449, 1979. NTIS: PB80-127186.
7. C.A. Hobbs. The Effectiveness of Seat Belts in Reducing Injuries to Car Occupants. Department of the Environment and Department of Transport, United Kingdom, 1978. NTIS: PB-300334.
8. A.A. Altshuler, J.P. Womack, and J.R. Pucher. The Urban Transportation System: Politics and Policy Innovation. MIT Press, Cambridge, MA, 1979.
9. L. Givens (ed.). Occupant Protection: An Integrated Approach. Automotive Engineering, Vol. 85, Oct. 1977, pp. 44-47.
10. Safety Effectiveness Evaluation of the National Highway Traffic Safety Administration's Rulemaking Process, Volume 2--Case History of Federal Motor Vehicle Safety Standard 208: Occupant Crash Protection. National Transportation Safety Board, U.S. Department of Transportation, Sept. 1979. NTIS: PB80-114085.
11. A.D. Little, Inc. Identification of Approaches for the Control of Health, Environmental, and Safety Hazards Associated with Air Bag Use and Disposal. U.S. Department of Transportation, Final Rept., Aug. 1979. NTIS: PB80-129109.
12. R.L. Braun. Toward Safer Motor Vehicles. IEEE Spectrum, Vol. 14, Nov. 1977, pp. 81-86.
13. E.L. Bolton. Hill Unit Hits Coleman on Air Bag. Automotive Industries, Vol. 155, Nov. 15, 1976, p. 9.
14. L.J. Carter. Auto Safety: Coleman to Act on Controversial Air Bag Issue. Science, Vol. 193, Sept. 24, 1976, pp. 1219-1222.
15. P.D. Hodges. Air Bag Battle Erupts on Hill. Automotive Industries, Vol. 159, Sept. 1979, pp. 21-23.
16. F.G. Fisher, Jr. Effectiveness of Safety Belt Usage Laws. U.S. Department of Transportation, Final Rept., May 1980. NTIS: PB80-209901.
17. A.F. Williams and B. O'Neill. Seat Belt Laws: Implications for Occupant Protection. The Engineering Index Annual, 1979, p. 548.
18. W.G. Wilson. Occupant Restraint Legislation Handbook: A Guide for Proponents. U.S. Department of Transportation, Feb. 1979. NTIS: PB-298461.

Publication of this paper sponsored by Task Force on Occupant-Restraint Systems.