10 RESEARCH IN ACCIDENT PREVENTION

THE READINGS:

INCREASING ATTENTION is being paid in accident research to the evaluation of the efficacy of preventive measures. But this is a substantially new trend. Relatively few accident countermeasures—some in use for decades—have yet been subjected to adequate scientific scrutiny. As a result, it is usually impossible to specify the return achieved in accident prevention per dollar spent and how the efficiency of that expenditure can be increased.

This is a serious deficiency in light of the substantial evidence that intuition, "common sense," and traditional assumptions constitute an inadequate and even erroneous basis for understanding the causation of accidents and for planning and evaluating attempts at their prevention. For example, Garrett has pointed out that

> until the last decade, it was widely accepted as a "fact" that when an automobile accident occurred, occupants who were thrown from the car were safer than those who remained inside. Indeed the connotation of the phrase "thrown clear" was that the ejectee who survived this experience would otherwise have been killed. Although such incidents have in fact been documented, a 1954 Automotive Crash Injury Research (ACIR) report[1] showed that they were the exception rather than the rule . . . that door opening was both a frequent and a hazardous event: In injury-producing automobile accidents, about 44 percent of the cars had one or more front doors opened, and, contrary to general opinion, occupants who were hurled through these doors were often "thrown clear" to eternity—not to safety.[2]

Among other widely accepted assumptions now known to be incorrect was the premise that heavy intoxication is an accident preventive. Thus, as recently as 1956 an authoritative text stated that

> It is the slightly intoxicated driver who characteristically demonstrates impairment of judgment more than impairment in sensory functions or psychomotor responses who is the real threat. . . . The "cockeyed drinker" constitutes neither a pedestrian nor a driving problem. Most of these individuals are either too drunk to drive or to walk and hence sleep it off. . . .[3]*

Similar evidence of the danger of basing programs merely on common sense is exemplified by the results of Barmack and Payne's evaluation of the Smith System of driver training and by McMonagle's documentation of increases in accidents after the installation of traffic control devices at some locations (see below). It is likely that many additional examples will be found when other preventive measures are studied objectively.

Despite such strong evidence against basing programs on unsupported presumptions, programs of this kind continue to be introduced, often with dogmatic public assurances as to their efficacy. Although conceivably justifiable as a stopgap until adequate evidence can be marshaled, this process has tended to delay the necessary fact-finding during much of the past half century, a period in which, in the highway accident field alone, more than 1.4 million drivers and pedestrians were killed in the United States. The introduction of unevaluated measures contrasts sharply with practice in other public health areas. There, before measures of prevention are *permitted* to be used, it is customary to document not only their pertinence, efficacy, and cost but their safety as well. The chlorination, filtration, and

* For research evidence to the contrary, see McCarroll and Haddon, Chap. 3, Haddon *et al.*, Chap. 4, and references 4 and 5.

fluoridation of public water supplies, the pasteurization of milk, and the Salk and Sabin vaccines have been among the many measures thus evaluated. Parallel procedures are employed in the evaluation of new drugs and most new surgical and other clinical procedures. Until similar evaluation can be achieved with accident countermeasures, there is little basis for believing that we are preventing as many accidents as our available resources permit, or that many measures are preventing rather than causing accidents.

In essence, accident prevention measures attempt to interfere with the sequences of events that culminate in damage to animate or inanimate structure. This interference is of three broad types. The first type attempts to prevent the potentially harmful chemical or physical forces from reaching the body or other structure to be protected. The second type attempts so to modify their interaction with that structure that damage is reduced or prevented. The third type attempts, through emergency, early and late clinical, and other care to lessen the long-range consequences of damage not prevented by measures of the first two types. Since research is concerned with measures of each of these kinds, we shall consider them in order.

Measures Directed at Factors Leading to Accidents

Measures designed to prevent potentially harmful chemical or physical energy from reaching a susceptible structure may attempt, in the order of preference, (1) to prevent the marshaling of the hazardous energy per se; (2) to prevent or modify its release; (3) to separate it and the susceptible structure in time or space; and (4) to interpose a barrier that blocks the energy from reaching the structure to be protected.[6, 7]

The prevention of damage from a nuclear device can illustrate these four initial levels of prevention. First, its manufacture might be prevented. Second, its use might be prevented. Third, the nuclear device and the persons and structures to be protected might be separated by a safe distance. Fourth, blast, thermal, and radiation shelters might be employed. It is both theoretically and practically useful to analyze in similar terms all measures which seek to prevent potentially harmful energy from reaching susceptible structures. This applies not only to chemical, mechanical, electrical, and thermal energy but also to such newer hazards as ionizing radiation and magnetic energy dangerous to biological and other systems.[6, 7]

All of these levels of *pre*-accident prevention are exemplified by measures long in use, such as lessening the manufacture of nitroglycerine in favor of safer explosives, restrictions on the discharge of firearms, the separation in space and time of pedestrian and vehicular traffic streams, and the use of insulation on electrical and thermal devices. Nonetheless, the failure to analyze accidents in such terms has made it difficult to determine the exact portions of the causal sequences that offer the greatest possibilities for research and prevention. Neither accident research nor accident prevention will have come of age until this is done skillfully and as a matter of course.

There is little well-designed research on the prevention of the marshaling of given forms of energy in hazardous amounts—for example, by prohibiting the manufacture of fireworks or by limiting the conditions under which vehicles are set in motion or electricity is generated at hazardous voltages and frequencies.† Similarly, there is little research on the modification or elimination of environmental components that favor traumatic interference with normal bodily energy exchange,[6, 7] as in the case of refrigerators that are so designed as to permit the entrapment and suffocation of young children.

BEHAVIOR OF YOUNG CHILDREN UNDER CONDITIONS SIMULATING ENTRAPMENT IN REFRIGERATORS

—Katherine Bain, M.D., Marion L. Faegre, B.A., Robert S. Wyly, B.S.

The following paper is an already classic study of human behavior in relation to a specific man-made environmental hazard. The study was one result of public pressure and Congressional concern which had led in 1954 to attempts to require the manufacturers of refrigerators to modify their models so that entrapped children would be able to escape.

The hearing on the proposed legislation,‡ which was opposed by refrigerator manufacturers and the Secretary of Commerce, led to "a request that the National Bureau of Standards [of the U.S. Dept. of Commerce] work with the refrigerator manufacturing industry . . . to develop performance criteria for evaluating safety release devices." The resulting investigation by an exceptionally competent research group for the first time placed the problem in qualified hands, a development which might have taken place much earlier had the subject involved a more classic public health area. As we shall note below, however, the manufacturing standard subsequently promulgated by the Secretary of Commerce was sufficiently inconsistent with the results of this work to make it likely that deaths would continue to occur. This illustrates a common problem in accident prevention: the failure to apply research results properly.

† One of the factors that make electrocution more likely with currents of some types is their ability to render those who make contact with them physically incapable of breaking loose. Reference 8 describes research to determine characteristics of electric currents that permit the breaking of contact.

‡ The record of this hearing,[9] like that of hearings before the Roberts Committee (see below), illustrates (1) the contending forces which influence the public safety; (2) the tendency of many concerned with accidents to predicate their control on completely unevaluated presumptions; and (3) the common tendency, when dangerous products are discussed, to emphasize public education, local police action, and measures other than product modification or elimination. No mention was made during the hearing of the fact that there had apparently been no well-documented instance of the *widespread* elimination through public education and local police action of any type of accident due substantially to the characteristics of a dangerous product. (This is still the case in 1964.) In addition, no mention was made of the fact that such accidents have often been successfully eliminated through the redesign of equipment, as in the development and compulsory use of the automatic railroad coupler and the airbrake.[10]

See reference 11 for an earlier and somewhat different report of this work.

EACH YEAR a number of young children perish as a result of entrapment in iceboxes, refrigerators and freezers. As numbers go, these are few compared with accidental deaths from other causes, but the thought of even a small number of helpless children suffocating needlessly is so appalling as to have created widespread interest in the problem.

Because of increasing public awareness of the fatalities resulting from entrapment of young children in refrigerators during a period of several years preceding 1956, manufacturers, engineers, governmental bodies and others became interested in methods for alleviating this hazard. Congressional hearings on proposed legislation resulted, in 1955, in a request that the National Bureau of Standards (NBS) work with the refrigerator manufacturing industry represented by the National Electrical Manufacturers Association (NEMA) to develop performance criteria for evaluating safety release devices. Considerable progress was made toward this objective by mid-1956, at which time Congress passed an Act which required "certain safety devices on household refrigerators shipped in interstate commerce" that would allow the doors of such refrigerators to be opened easily from the inside. It further required the development of standards for such release devices. Such standards were published in the Federal Register of August 1, 1957. They require that all devices meet at least one of three specified performance requirements, and specify in some detail tests for the purpose of determining compliance with these requirements.

Consideration of the problem by NBS and NEMA made clear that it was not only an engineering problem, but also a problem in child behavior and so the aid of the Children's Bureau was enlisted.

In developing performance criteria for release devices, it was necessary to correlate the mechanical forces required to keep refrigerator doors securely closed against their gaskets and the forces young children are able to exert when seeking to escape from entrapment. Because no data were available on this point, late in 1955 the Children's Bureau and the NBS conducted tests on children in nursery schools in an attempt to gain this information. In this preliminary experiment, some 60 children between the ages of 2 and 5 years were tested in an experimental enclosure, which simulated a refrigerator only with respect to inside dimensions. The enclosure was camouflaged to represent a gay red "Santa Claus chimney," with a window and door. The children were urged to use, and were rewarded for using, their utmost strength in competitive pushing against the door, from both sitting and standing positions. These tests indicated that a significant proportion of the young children tested failed to exert forces in excess of 10 pounds. However, practical manufacturing considerations make it hard to design for assembly-line production a release device which will respond to a direct push of this magnitude on a refrigerator door and which, at the same time, will permit the refrigerator door to seal so as to allow the refrigerator to perform satisfactorily its primary function—food preservation.

A further investigation was therefore undertaken during the summer of 1956 to provide additional information on the force efforts of children, as well as information on child behavior in general with respect to release devices currently obtainable, the experiment being carried out under conditions simulating actual entrapment as closely as possible. No studies had previously been made under such conditions, insofar as could be determined.

From death certificates and newspaper accounts of refrigerator deaths, a few facts are known and some assumptions can be made. The age range, for all practical purposes, is 2 through 9 years, with the peak

[Reprinted, with permission, from *Pediatrics*, *22*:4, Part I, pp. 628-647, 1958.
The summary, 13 tables, and all photographs have been omitted.]

between 3 and 6. Males far outnumber females. Children enter refrigerators singly or in groups. In one instance four, and in another case five, children died together, while a number of situations are recorded in which two children were fatally entrapped together. Some of the refrigerators are abandoned in dumps but many are in homes, only temporarily in disuse (as in empty apartments), or are in the process of defrosting. Some children probably get into refrigerators as into a playhouse, some probably are hiding from companions, a few are shut in by playmates.

Part I. Behavior Study

In designing an experiment to simulate as closely as possible the real situation precautions had to be taken to protect the experimental subjects. If a real refrigerator were used or the nature of the experiment disclosed, children's interest in exploring refrigerators might be aroused. But more important still, entrapment in an enclosed dark space is a fear-provoking experience. If it had not been for the dearth of information and the important use to be made of the results, the originators of the plan would not even have considered subjecting children to fear-provoking conditions.

In an effort to make the entrapment bearable, not only by the children but by the experimenters, a time limit was proposed. On the advice of consultants to the experiment—a child psychiatrist, child psychologists and pediatricians—a time limit of 3 minutes was set as the maximum time that a child might safely and excusably be allowed to cry.

Especial care was taken to see that both before- and after-test experiences were pleasant and that the children left in a cheerful, relaxed frame of mind.

Setting for the Tests

The ideal environment for such an experiment, it seemed clear, would approach what children are used to in the home or at play. However, practical problems and the time schedule agreed on for the study precluded the possibility of conducting the experiment in children's homes or at nursery schools. These considerations dictated the choice finally made—that of a former residence on an estate now a part of the NBS grounds. Trees, shrubs and spreading lawns, together with a large terrace, contributed much to the environment. Two very large first-floor rooms, with the adjoining tiled terrace, were used for an office, reception quarters and testing space.

In the office-reception room, toys, crayons, coloring books and puzzles were provided for the children, also magazines for parents to read while their children were taking part in the tests.

Test Equipment and Facilities

Test enclosure and recording equipment: The plywood test enclosure resembled a child's playhouse, with door, roof and chimney. The inside dimensions ($40 \times 18 \times 25$ in.) were based on the measurements of a number of currently available household refrigerators of 8 to 11 ft capacity, and represent, approximately, the maximum inside dimensions excluding the space occupied by the freezing unit. A safety-glass panel formed the ceiling of the enclosure so that motion pictures of the child could be taken from above. A 16-mm motion-picture camera and illumination equipment designed for infrared photography were housed under the roof. Forced ventilation provided for the child's comfort while in the enclosure.

Several identical doors were constructed into which different release devices were inserted, thus saving time when changing from one release mechanism to another. A snooperscope, which replaces an infrared image with one of ordinary visible light, was used behind the enclosure for observing the children. Under the low intensity of infrared illumination used, the children were in what seemed to them total darkness. Microphones and tape recorders picked up sounds the children made, comments of the observer, and time and force readings during the tests.

Conventional release devices: The release devices furnished by the household refrigerator manufacturing industry were of two broad classes of construction. One general type included devices which responded to the application of a force to the inside of the door panel, the most effective area being near the latch edge of the door (D 1, D 2). The second general type included devices which responded to manipulation by hand (D 3). Obviously escape by means of the latter type of device depended not only on a child's finding and having the strength to use a releasing mechanism that covered a limited area, but also on his familiarity with operating similar mechanisms.

Specially designed escape devices: Observations of the children's exploratory search of the interior of the playhouse led to the investigators' hunch that what the children were seeking might be a doorknob—something familiar to most of them and one of the things they are eager to manipulate and conquer as soon as they can walk.

To test this possibility, a doorknob linked to the latching mechanism was made (D 4). In order to help children find it in the dark, a circular plastic rim containing luminescent material was attached. This knob triggered the door's latching mechanism if turned slightly in either clockwise or counter clockwise direction or if it was pushed in or pulled out.

When it became obvious—early in the test program—that some children were going to remain very quiet and move about very little, another experimental device was engineered (D 5). A push-open door linked with a floor panel riding on ball bearings enabled a child to release himself with as little as one-half the effort required without the movable floor panel. This door could be opened by a very slight forward, backward, or sideways movement of the floor panel such as would result if the child pushed on the door or on any wall of the enclosure. A child weighing between 30 and 50 pounds would not have to push very strenuously to produce the small force (6 to 12 pounds, depending on point of application) that would open the door for him.

Device D 6 was very similar to device D 5 except that it was designed to indicate the maximum force efforts of the child and to permit his release only at the discretion of the observer.

Test Plan

Test subjects: Tests were set up for ages 2, 3, 4, and 5 years, with equal numbers of children at each age, and of each sex. Insofar as practicable each of two experiments handled equal numbers of children of each age. Most of the test subjects were obtained as the result of a letter distributed among the NBS staff, broadly describing the nature of the study and inviting their co-operation. The response was immediate and generous. Personal acquaintances and neighbors brought in a scattering of other subjects of the required ages.

The 201 children tested came from 157 families.

Records: A record card was kept on each child. Parents filled in their own names, years spent in school and occupation, as well as the children's names, sex and birth dates. Each child's height, weight, date of test, and his reactions before and after the test were recorded.

Sounds made by a child while in the test enclosure, comments on his behavior made by the observer, also time and force readings spoken into microphones during the testing were recorded on tape and later transferred to record forms.

A moving-picture record of each test had been planned. Unfortunately, the required infrared film was not obtainable in sufficient quantity and, in consequence, the test behavior of only the first 42 and last 48 subjects was recorded on film.

Test Procedure

A feature of the original plan that was

adhered to in practically all cases was that each child should be accompanied to the test by one or both of his parents, or by a close relative. In a few instances where this rule was not observed, children were brought to the test by their nursery-school teacher, a substitute mother figure with whom they felt comfortably at home.

After greeting the parent and child, the experimenter who was to handle the child during the test made friends with the child. Once children's attention was diverted to the toys, most of them played happily until it was time for them to go with the experimenter to play ball on the terrace. Parents remained in the office while each child was taken individually to the test room.

Separation of a child from his parents sometimes required finesse, especially in the case of younger children. Once the child was left in the hands of the experimenter, she could almost invariably establish an easy relationship, and get him to go willingly with her to the terrace and from there to the test room. In only three cases were children so unco-operative that attempts to test them had to be abandoned.

After playing on the terrace, the experimenter weighed and measured the child in the test room, and then led him gradually down the room toward the test enclosure. Older children were often curious about the enclosure from the time they entered the room, but the attention of the younger children usually had to be drawn to it. When the experimenter considered the time was right, she signalled the engineer behind the scene to start the color cartoon which lured the children into the "playhouse."

The sound track with its music and Donald Duck chatter drew most of the children into the playhouse. In some cases the experimenter bent down to look at the movie and told the child what was going on in order to interest him in going inside.

As soon as a child became absorbed in the movie, the experimenter told him she was going back to the office, or otherwise tried to convey to him the idea that he was being left alone. At this point an observer, who could see the child from behind the playhouse in a mirror placed high on the wall, closed the door of the enclosure. Simultaneously the cartoon stopped, the playhouse became dark, a shutter closed the screen, sound- and force-recording equipment began to function, and the overhead camera began taking infrared pictures. Now that the child was shut into an enclosure from which most outside sound was excluded, the observer and engineers in charge of observing and recording behind the booth could speak all pertinent information into microphones, to be preserved on tape. This information included comments on the child's behavior, as seen through the snooperscope, time readings at 10-second intervals, and, in one series of tests, the force output as registered on a gage and observed through a telescope. The child's vocalizations were also recorded on tape.

If a child did not release himself, the observer determined from his behavior when to let him out. This determination was formed on the basis of the amount of effort he was exerting and the degree to which he appeared to be disturbed.

Outside, the experimenter, who had started away or given that impression when the child entered, was there to comfort him the moment he emerged. The cartoon was at once continued on the outside screen and the child was invited to watch it or to fetch his parents to watch it with him and to see the playhouse. This proved to be an effective way to calm him, take his mind off his experience in the playhouse, and help him carry away a happy impression of his visit.

Findings

Success in Escaping: A child's success in escaping depended on at least three factors: the device with which he was dealing, his age and size, and his behavior. Greatest success was achieved with device D 2, which required a releasing force of 12 pounds directed against the door panel near the latch edge (or more, if applied elsewhere), device D 4 equipped with a knob to be turned, pushed or pulled, and device D 5 with a movable floor panel. Two-thirds to three-fourths of the children taking part in tests making use of these devices let themselves out. In tests in which device D 6 was used, the child could not escape, but he was considered successful if he exerted a maximum force in excess of 15 pounds. Success with each of the six devices was affected by age and height and weight. Boys and girls got out with equal ease. Other factors not determined in these tests may have influenced success, such as the child's intelligence or the socio-economic group from which he came. The only measure of the latter factor obtained in the study was the combined years of education of the parents. A higher rate of success was associated with fewer years of education of the parents.

Behavior in the Test Enclosure—General Response to Entrapment: Particularly striking was the wide range in behavior shown by the children in response to the entrapment situation, varying from complete inactivity to violent panic. Three major behavior patterns were observed:

1. Inaction, with no effort to escape or only slight effort (24%).

These were the children who stood or sat patiently, apparently waiting to be let out or for the movie to come on. Some made slight exploratory movements, gently touching the door or walls. A few knocked politely, saying "Please let me out" or "I'm ready to come out now." A few were almost motionless. Some cried gently, others made no sound and apparently were unconcerned. More than one child sat quietly for 15 minutes. Age was not a factor in determining which children would show this passive behavior. A few more girls than boys were in this group.

2. Purposeful effort to escape, without violence (39%).

These children went to work, usually immediately, to find a way out—pushing, feeling, or trying to manipulate a knob or device, if present. Some were quiet, others cried or called out, but at the same time made direct efforts to let themselves out. This kind of behavior, equally characteristic of boys and girls, increased with age.

3. Violent action, with or without purposeful effort to escape (37%).

These were the children who kicked, banged, jumped up and down, threw themselves against the door, or exhibited anger. Many of these directed their violence toward escape, but some panicked to the degree that no purposeful effort was apparent. This type of behavior decreased with age and was somewhat more characteristic of boys.

This marked variation in behavior obviously influenced success in getting out. The four passive children who escaped did so because the slight shift of their weight on the movable floor (device D 5) opened the door for them.

Greatest success (86% of Group 2 above) was achieved by those who went about the job purposefully, but most of these were also the older children. Panic, anger and violence interfered with success, so that the group with this type of behavior (Group 3 above), most of whom were also younger, achieved success in only 32% of the cases.

Behavior in the Enclosure—Specific Acts: In early consideration of the problem of release devices it had been presumed that children would push, yet not all did so. Considering only the four devices (D 1, D 2, D 5, and D 6) in connection with which pushing was appropriate and when no release gadget was present to attract the child's attention, 67 (61%) of the 110 children pushed to some degree. About the same number knocked, banged, slapped or kicked the door or walls. When confronted with a gadget which could be grasped (D 3 and D 4), 18% pulled it; 9% pushed it; and 40% made turning motions.

Hand movements of some of the children were particularly noticeable. About one-fourth of the children put their hands to their mouths or faces. A small number made curious twisting, twining and picking movements of the fingers or clenching and clasping of the hands. Very few sucked the thumb or fingers and none masturbated. Wringing the hands, as an adult does, was observed in several children.

With sound recordings available on all children it was possible to determine the vocal response they made to entrapment. Some children were silent, with a range from only 6% of the 2-year-olds to 50% of the 5-year-olds. About a quarter of the children screamed, the younger children more often showing this behavior than the older ones. Many of the children called for help.

Although as they entered the playhouse they had been told the experimenter was leaving, some called to her to be let out. But most of the children called "Mommy" or "Mother" even though she was well out of earshot. Only six children called "Father" or "Daddy" (5 girls and 1 boy) although about one-third of the children who called for help had been accompanied to the test by father alone or by both parents.

Duration of Test: Time in the enclosure was short for most children. One-fourth got out by themselves or were released in less than 10 seconds and three-fourths either got out by themselves or were released in less than 3 minutes. One-half of those who released themselves did so in less than 10 seconds. If a child became panicky and seemed much upset, he was immediately released. If his disturbance seemed mild or moderate the limit of 3 minutes agreed upon through psychiatric and psychologic consultation was adhered to. A few inactive children remained in the enclosure for relatively long periods, six in this group staying over 10 minutes.

Behavior on Entering and on Leaving the Enclosure: There was little resistance to the test situation, but the small number reluctant to leave their parents (17%) or resistant to entering the playhouse (13%) contributed more heavily to the group characterized by violent action in the playhouse than did the co-operative children.

Upon escape or release the experimenter was right there to comfort the child if necessary and to show the movie on the outside screen for the child's entertainment. One-third of the children emerged unruffled; about half were upset but could be comforted or distracted by the movie; and a small group (11%) emerged from the enclosure upset. These were the younger children and also those who had shown a violent reaction in the playhouse.

Force Exerted by the Children: Device D 6, designed to indicate the horizontal force exerted by children (Table XIV) no matter where the force was applied, was used to test 31 children. This force was found to range up to a maximum of 29 pounds. The average by age group ranged from 10.4 pounds for 3-year-olds to 20.6 pounds for 5-year-olds. The average for 2-year-olds was 12.8 pounds. A study of the data did not reveal why the 3-year-olds exerted less force, on the average, than did the 2-year-olds.

TABLE XIV.—MAXIMUM HORIZONTAL FORCES BY POUNDS, EXERTED BY CHILDREN (ACCORDING TO AGE) ON DEVICE D6*

	2 Years	*3 Years*	*4 Years*	*5 Years*
	14.5	12.3	17.3	16.7
	14.5	6.8	23.4	19.0
	19.0	9.0	11.2	16.7
	16.7	14.5	15.6	22.3
	14.5	16.7	10.0	29.0
	0.0	5.6	12.3	10.0
	10.0	7.9	19.0	22.3
	13.4	—	10.5	29.0
Av.	12.8	10.4	14.9	20.6

* Designed to measure horizontal force without permitting self release.

Though direct measurements of forces exerted with other devices were not made, the fact that some children released themselves with devices set at specific thresholds indicated the minimum force they exerted at the moment release occurred. Device D 1 required a force of at least 18 pounds to effect release, and D 2 a force of at least 12 pounds. Using the results for children who released themselves in tests using these devices, and combining these results with those for children exerting various forces measured by D 6, Figure 10 is derived.* From this it can be seen that 25 to 30% of the children taking part in tests in which horizontal force was an appropriate effort exerted in excess of 18 pounds and 65% exerted in excess of 12 pounds.

Discussion

No experiment could be designed to reproduce all the conditions actually present in a naturally occurring entrapment. The kind of child who gets entrapped in a refrigerator is unknown. Is he bold and aggressive and may he therefore be expected to be active in releasing himself? Or does he often seek solitude? What kind of child is lured in, or shut in, by companions? The best the experimenters could do was to take from a volunteer population a sample chosen with the aid of competent statistical advice and containing children of ages known to be susceptible to this type of tragedy. Obviously no test would have been considered that involved depriving a child of oxygen, so it is possible only to speculate as to a child's activity in an atmosphere in which oxygen depletion was taking place.

Probably the greatest differences between the experimental situation and real life lay in the fact that the children knew people were not far off, and in the protection provided against psychologic trauma. The experimenters who handled the children were warm, friendly people, fond of and widely experienced with children, who first developed rapport with each child to be tested. Having got a child's confidence they were unable to bear his fear reaction long (in contrast to the observer who, having developed no such relationship, could be more objective). His cries brought a quick appeal by the experimenter for his release. How a child would ultimately have solved his problem was therefore often not determined. Left in the enclosure longer some children might have quieted down and released themselves. Success, therefore, is possibly underestimated.

The wide range of behavior of the children was especially interesting in light of the purpose of these experiments—to provide data for the development of performance standards for release devices. If one assumes that behavior of these children was generally typical of those entrapped by chance, then a significant number will probably not release themselves by the use of any currently practical device requiring purposeful physical effort. A device utilizing their purposeless movements would increase the escapes. The movable floor panel device was developed for this purpose, and showed some promise in this connection in the few tests in which it was used. However, problems in its manufacture and sanitary care, and the impracticability of making its releasing features accessible from all spaces in which a child might become entrapped, appear to place serious limitations on it. In addition, were it sensitive enough to be effective for passive children, it might interfere with normal use of the refrigerator.

The association of lower education level of parents with degree of success might possibly be explained on the grounds that parents with no more than grade or high school education may give their children more opportunity to play independently than parents with college or advanced degrees.

* There are at least three possible sources of error in the observed force values for device D 6. First, friction, weight, weight distribution on the floor panel, and direction of applied force may have affected the directly observed force values. Second, the person who observed the force gage probably was unable to read with great precision the values indicated by the sometimes-fluctuating needle. Third, the calibrated accuracy of the force gage affected the results.

Computations based on the calibration data taken for device D 6, loaded centrally with a weight of 34 pounds on the floor panel, showed that the maximum deviation of applied force from that indicated by the calibration curve used in the preparation of Figure 10 and Table XIV was —1.7 and +1.5 pounds for rearwardly directed forces (as in the case of a child pushing on the door) and sidewardly directed forces, respectively. The corresponding average deviation was —1.2 and +0.4 pounds. Available information indicated that the calibrated accuracy of the force gages used in the calibration of device D 6 and in the tests was within ±0.2 pound. It is estimated that the person who watched the force gage while children were in the enclosure was able to read the maximum forces indicated by the fluctuating needle within ±0.5 pound.

The turning movements made by many children when taking part in tests utilizing a device that could be grasped, suggested that the household doorknob was familiar enough to all children to be useful. This proved a valid assumption.

Fundamental to the establishment of performance requirements relating to devices that release by pushing was knowledge of children's pushing behavior. Would a child push? Where? With how much force? Not all children pushed; some directed their efforts to walls rather than to the door, some to the hinge rather than the latch side of the door. Others let themselves out accidently when they leaned against the door or when in their violent activity they bumped against it. The size of the space, the size of the child and chance all played a part here.

Part II. Follow-up Study

Though the experiment produced data which can be expected to save lives, the investigators would still have been uncertain of their justification if the subjects had been harmed. The complacency with which most of the children took the testing, and the ease with which those who became upset could be comforted, reassured the directors of the experiment that it was permissible and justifiable. Nevertheless, more objective data on the aftereffects, if any, seemed desirable to round out the experiment. Deep-seated anxieties in the children could not be uncovered without extensive psychologic testing, and even if such deviations were found their relation to the experiments could not be determined, since previous personality studies of the children had not

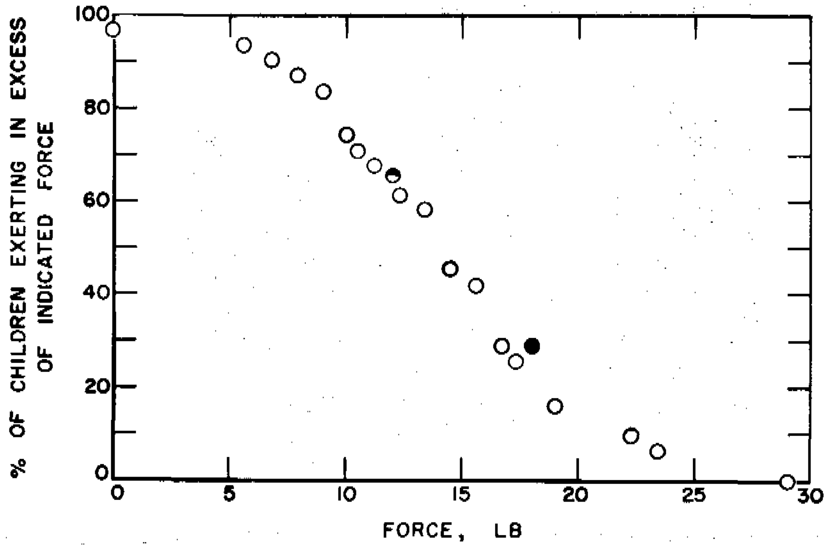


FIG. 10. Horizontal components of forces exerted by children, ages 2 through 5 years.

Device	*Number Tested*	*Number Who Released Themselves*	*Symbol*
1	31	9	●
2	32	21	◓
6	31*	0	○

* All released by observer.

been done. The only type of study which seemed practical was interviews with the mothers, who were in a position to observe any evidence of changes in behavior following the tests. Also, the mothers were able to give a first-hand account of how the child reacted after the test, how much he talked about it, and what residual effect was apparent on the surface.

Interview Methods

About 8 months after the test, follow-up interviews were held with the mothers of 96 of the children who had been test subjects. This sample, of almost half the children originally tested, was drawn so that it consisted of equal numbers of boys and girls of each age from 2 through 5 years, with an equal number handled by each of the two experimenters.

In order to avoid the suggestion that harm might have been done the children, the approach to the parents was made on the grounds that when the tests were held in the summer of 1956, time did not permit getting background material on the children's health, personality and daily routines, such as eating and sleeping.

Mothers without exception responded cordially to the request. The interviews which were conducted in the homes by one interviewer, a member of the original staff, lasted from 30 to 45 minutes. Times of day were chosen to suit the mothers' convenience, often at children's nap-time or when they were in kindergarten or nursery school. Only 34 of the children were present during the interview, the others being in the room only momentarily or not at all.

Studies of young children who have undergone traumatic experiences—bombing, separation from parents, and operative procedures—have shown that regressive behavior in patterns of sleeping, eating and toileting is not unusual. The interview started in relatively structured fashion with questions about appetite, taking the bottle, sleeping, thumb sucking and wetting. The questions were phrased to describe his behavior, any change in recent months and, if any change, the approximate date. Thus any behavior changes in these areas could be related to the time of the tests without asking directly for such information or implying that there might be an association.

The interview then moved more freely into a discussion of the child's usual emotional reactions, i.e., was he easy-going, fearful, etc. This line of inquiry, though of no practical value in answering the questions of the study, led easily into comments by the mother on how the child seemed to feel about the test, did he remember it, had he talked about it, had he seemed upset. Other possible upsetting experiences and their timing were also sought, such as hospitalization, new sibling, moves, or separation from parents.

The first part of the interview, then, yielded data on regressive behavior, the second part some impressions of effect of the tests as noted by the mothers.

Findings

No child went back to taking a bottle after the test, although several had a history of reversion briefly before the test by trying out a new sibling's juice or milk bottle.

No unfavorable change in the sleeping behavior of any children, such as restlessness or crying out, was reported for a period of 4 months after the test, with the exception of one child whose family moved from an apartment to the country soon after the test, and who then began ("rarely") to cry out in his sleep because, he said, he feared animals in the country. He wanted his door open and his light on.

Of the 37 children who sucked their thumbs or fingers or, more rarely, a blanket, all had done so since babyhood. Their age distribution was: 10, 2 years; 13, 3 years; 8, 4 years; 6, 5 years.

Of the children who were bedwetters none reverted to this behavior anywhere near the time of the test with the possible exception of one 5-year-old whose mother could not recall just when it began.

Many of the 2-year-olds were not yet

talking at the time they took the test. Only four of them ever talked about it later, two of them indicating they did not like the dark room. The other 20 never mentioned the experience, as far as the mother knew.

Children 3, 4, and 5 years old were for the most part quite voluble about their experience, their comments reflecting their pride in success, their enjoyment of the attention they got from the experimenters who played with them, or their puzzlement over *why* they were shut in.

The attitudes of the children, as recalled by their mothers, 8 months after the test, ranged through more or less casual acceptance, remarks about not liking "that kind of playhouse," thinking it was "a lark, and bragging about it," showing great pride in having earned money ($5 a child was given to recompense parents for transportation expenses, baby sitting, etc.), curiosity about the whole thing, and resentment at "a dirty trick." A good many children thought they got out by themselves when actually they did not; their believed success made them feel exultant.

Illustrative Case Reports

Case 76: A 2-year-old girl (3 years at time of interview) frequently refers to the test, though the parents have never brought up the subject, her last mention of it being to her grandparents whom she visited recently. This mother did not realize what the test would entail, that is, being locked in, and would not do it again. She thinks, however, it did not prove to be a bad experience for the child because she had opportunity to watch her mother being locked in and getting out.

Case 41: A 4-year-old boy, the youngest of his family, who was very hard to separate from his mother, did not make a very great physical effort to get out of the enclosure, but did a great deal of calling to "Woman!" (seven times) to let him out. He seemed somewhat disturbed for a few days afterward, but accepted his mother's explanation of the test. Recently, while on the NBS grounds, he said, "I don't want to see the movie again," although no one had mentioned the experiment. This is a family which has been having a struggle to meet the expenses of the mother's illness. The boy showed no other overt evidence of being disturbed, no changes in routine behavior patterns save a "wolf dream" approximately 6 months later.

Case 125: A 4-year-old boy whose mother reported he "came home with a happy expression as if he'd been to a party" had been taken to the test with his younger sister by his father. He is next to the youngest in a family of five children; neither he nor his 3-year-old sister showed ill effects, the little girl's only reaction being that "something was put over on her." The family background is serene, the parents appearing to be stable, mature people.

Case 19: A boy who was 3 years of age (summer, 1956) is described by his mother as being somewhat dominated by his 5-year-old sister. At the time of the test he went into the test house readily, but made no effort to get out beyond calling "Mommy" a great many times. He was released because of crying at 3 minutes, 20 seconds. While the experimenter was cheering him up, she found a tiny baby rabbit in the tall grass, and this so entranced him that it seems to be all he recalls. His mother speaks of his having "enjoyed the whole experience."

Case 20: The sister of Case 19, 5 years of age, who went into the house quickly because she thought it was a dollhouse, got out by herself in 2 minutes, 19 seconds, showing little or no concern. She has made no mention of the experience in recent months. The parents have never brought up the subject.

Discussion

In designing the experiment the staff was deeply concerned to avoid exposing the children to a damaging experience. They took great care that no child should become too upset and that all children should have a pleasant experience after the test. Apparently the precautions taken were adequate, for the children in the sample followed up 8 months later showed no reversion in behavior, and the parents, when given an opportunity to discuss reactions, produced no evidence of trauma or concern over aftereffects.

Other factors may also have contributed to the apparently low level of anxiety.

The fact that almost all the parents tried out being shut into the enclosure seemed, as was intended, to absorb some of the children's concern. To have father or mother double up and squeeze into the small space was a source of merriment.

A number of mothers reported that their children talked the test over among themselves. This kind of "shared experience," the opportunity to talk freely with others who had been through the same thing, was reported in "Operation Schoolhouse" to help sufferers to integrate the disaster, because their feelings did not have to be repressed. Under incomparably milder stress, quite a number of children in this experiment demonstrated what may be similar tension-relieving behavior: three children in one family (2, 3, and 5 years old) talked about their experience in terms of the toys they had played with; a 4-year-old boy and his 3-year-old sister talked about the test, and he told his playmates about it.

An interesting aspect of the experiment, which seemed to be reflected in the children's behavior, was that a child's parents were not involved in his "trouble," and so did not become excited and upset. In many real-life experiences in which a child is hurt or hospitalized, the parents are much distressed, something a child may sense even though the parents try to conceal their feelings.

In the present experiment the parents did not come on the scene until the child had recovered his equilibrium or was on the way to doing so, and they were invariably calm and unalarmed. In the Vicksburg study it was remarked that "much of a child's later behavior around the topic of the tornado was almost wholly parent-determined." Our impression is that the usually nonchalant behavior of the parents of the children in the NBS study may well have been related to the casual way in which the children seem to have reacted to the experience.

* * *

This paper illustrates many points that we have noted in previous chapters: (1) the necessity, in some cases, of using human experimental subjects; (2) the problems which the use of such subjects entails, particularly in relation to injury and the ethical issues involved; (3) the importance of studying accidents in relation to the environments in which they occur; (4) the consideration, in mature accident research, of many classes of variables and the use of correspondingly varied methods of data collection; (5) the use of subjects as similar as possible, particularly with respect to age and sex, to those whose accidents are of interest; (6) the problems inherent in extrapolating laboratory findings to the real world in the absence of epidemiological and related studies which would serve as the basis for estimating their relevance; and (7) the inherent variability of biological material and the consequent inappropriateness of using average values to describe its characteristics. This research, by documenting the substantial error in the assumption of the Secretary of Commerce that "it is doubtful that a panicky child would find or use such a release device,"[9] demonstrates again the danger of basing conclusions on unsupported presumptions.

The standard promulgated by the Secretary of Commerce [12-14] based on this work specified, among other provisions, that household refrigerators must

> meet at least one of three performance requirements: (1) a force not greater than 15 pounds exerted near the latch edge of the door from the inside will release the door; (2) a turning moment of not more than 5 in.-lb. applied to a knob similar to a conventional doorknob will release the door; or (3) an automatic device that permits the door to be opened in accordance with requirement (1) whenever interior spaces are created which might permit a small child to enter. . . .[14]

The specification of a limit of 15 lb. in the first of these requirements is remarkable in view of the data upon which the standard is stated to have been based. For example, as summarized in Table 14, only three of the 15 two- and three-year-olds exerted a force of more than 14.5 lb. on device D-6, and only nine exerted more than 10 lb. Even among the four-year-olds only half exerted more than 12.5 lb. of force, indicating that even for this group a device requiring "a force not greater than 15 pounds" would often be tragically inappropriate.* This unfortunate discrepancy between the standard and the evidence upon which it is stated to be based[14] may have resulted from the tendency, often seen in the work of engineers unfamiliar with biological materials, to use average values rather than values based upon a careful consideration of the actual distributions of the data from which such averages are calculated. (See McFarland's comments, Chap. 2.) Neither the foregoing report nor its predecessor[11] provides data in support of requirement (2), above, and it is to be hoped that it was not similarly based.

Finally, it is important that, although many of the same considerations apply, and although children have been trapped in such other devices, the standards promulgated specifically excluded deep freezers[13] and did not apply to dryers, washers, and similar appliances. This well illustrates the tendency to approach accident prevention problems on a piecemeal basis and the failure to identify and apply general principles.

As we have noted above, most attempts at accident prevention, and hence most evaluations of their efficacy, have taken for granted the presence of hazardous physical and chemical agents in the environment and have been concerned with preventing injurious interactions with them. In the case of transport accidents, it is usually assumed that people and objects will be placed in motion, thus acquiring what the physicist terms "kinetic energy," and that the fundamental problem is avoiding decelerations of such abruptness that the resultant forces are not sustainable without damage (see De Haven, Chap. 9, and reference 6).

Measures directed at preventing such injurious interactions have been most frequently based on attempts to modify behavior, even where there is insufficient scientific evidence as to the relative importance of other initiating causes—for example, mechanical failures and medical events. This behavioral emphasis is well

* The data strongly support the conclusion that the standard should have specified "a force not greater than 5 pounds," since 30/31 of the subjects referred to in this table could then have escaped. In this connection it would be interesting to determine the percentages of currently manufactured refrigerators that can be opened by such force and the force required to open refrigerators in which children are now being entrapped. It would also be very useful to know the factors that lead trapped children *to attempt* to escape, a point not adequately dealt with in the foregoing work.

represented in the selections that follow, although growing attention to other classes of variables, especially in connection with motor vehicle crash design, is also evident (see below).

AN EXPERIMENTAL FIELD TEST OF THE SMITH-CUMMINGS-SHERMAN DRIVER TRAINING SYSTEM

—*Donald E. Payne, Ph.D., Joseph E. Barmack, Ph.D.*

This work and the discussion of its merits in the two letters that follow it illustrate (1) the extent to which a seemingly reasonable accident countermeasure can come into wide use before it has been completely evaluated; (2) the differences between common-sense assumptions concerning efficacy and research evidence; (3) several of the methodological and other problems involved in the proper evaluation of a preventive measure directed at behavioral modification; (4) some of the problems of "communication between the researchers and the program people" (see Imhoff, below); and (5) the importance of continuous study of the effectiveness of countermeasures in parallel with standard practice in approaching the prevention of disease.

SINCE 1957, a training program for professional drivers developed by H. L. Smith, J. J. Cummings, and R. A. Sherman has been offered to motor fleets throughout the United States. For several years Mr. H. L. Smith was sponsored by the Ford Motor Company to promulgate this program.

The principles of the training system emphasize two main points: (1) Developing systematic search habits to detect potential driving hazards and (2) Using driving strategies to dispose of potential hazards before they become critical.

The students in the program ordinarily spend a week learning and practicing the system under supervision. During the week they receive practice in rating drivers (using a special 12-item rating scale), in demonstrating correct driving habits, and in instructing other drivers. Students are given a variety of visual aids for classroom instruction of drivers. The visual aids include movies, film strips and pamphlets.

The Smith-Cummings-Sherman training system has considerable appeal. The principles are logical, the objectives of the training are clear, the methods well organized. Training aids are available. Any bugs in the system presumably have been worked out through widespread use in a large number of fleets.

All that was lacking was concrete evidence of effectiveness. No one had collected the necessary data from those fleets which were using the training system. Therefore, a survey was planned for this purpose. A list of the fleets which had sent personnel to learn the system was provided by Smith. Questionnaires were sent to 49 fleets, of which 35 returned partially or fully completed questionnaires. The results of the survey have been published elsewhere.

The results of the survey indicated that there were small improvements in accident rate in most of the fleets. The enthusiasm of the fleet safety supervisors outran the dem-

Reprinted, with permission, from *Traffic Safety Research Review,* 7:1:10-14, 1963, published by the National Safety Council.

onstrated results of the training system. These findings, while not conclusive, were sufficiently positive to justify an experimental evaluation of the effectiveness of the system.

The Hayes-McLean Field Test: To conduct a thorough field test of the training system, we needed a fleet of substantial size with comprehensive accident and mileage records, and willing to modify operating procedures to meet the requirements of the experiment. The McLean Trucking Company of Winston-Salem, N. C., generously provided the necessary arrangements through its Hayes Freight Lines division.

Hayes Freight Lines is a common carrier based in Indianapolis, Ind. The fleet operates through a wide area of the midwestern United States. The Interstate Commerce Commission granted McLean temporary authority to manage and operate Hayes in July 1958. Hayes became a division of the McLean Trucking Company in December 1959.

Method

When the field test began in May 1960, there were 131 over-the-road drivers of tractor-semitrailers on the Hayes roster. There was a range in seniority of the drivers from less than a month to slightly over 30 years.

The general plan for the field test was as follows:

1. The drivers were to be divided into two groups, matched for seniority and accident history.
2. One of the groups was to be trained, using the Smith-Cummings-Sherman training system.
3. Following the training there was to be a 15-month waiting period. Individual records of mileage and accidents were to be collected for each driver during this period. To minimize outside influences, no major changes were to be made in the fleet's safety program or accident recording procedures.
4. At the end of the waiting period the records of the two groups of drivers were to be compared. If the training program was effective, the trained drivers should have better accident records than the untrained drivers.

Selecting and Preparing the Trainers: Two Hayes drivers were chosen to be trainers. Both men were highly rated by management and accepted by the other drivers. One had slightly more than 11 years of driving experience, the other approximately 23 years.

The trainers attended a regular one-week course given by Harold Smith. Following this, they received an additional two weeks of practice in rating and training drivers in one of the subdivisions of Hayes Freight Lines, operating out of a different terminal from the field test groups. The two weeks of special practice was supervised by an experienced driver trainer nominated by Smith.

Between May and July 1960, the trainers attempted to conduct two training sessions with each of the 60 drivers who were to be trained. The training program had to be terminated the end of July, at which time 33 of the drivers had received one training session, 27 had received two.

A follow-up training program was arranged nine months later. It consisted of a single training session for each driver in the trained group, during the period from April 22, 1961 to May 16, 1961. The follow-up training was conducted by the professional driver trainer who had supervised the preparation of trainers I and II. He was able to retrain 51 of the 60 drivers.

No further training was given to any of the drivers. At no time did any of the drivers in the untrained group receive direct training from any of the trainers. Three drivers who were not members of either the trained or untrained groups were trained during the follow-up period.

We could not exclude the possibility that some of the trained drivers might pass along tips or suggestions to buddies in the untrained group. We assume that if this did happen, it was unsystematic.

Classification of Accidents: The Smith-Cummings-Sherman training system does not aim to prevent all accidents. A driver could carefully follow all five principles of the system and still have an accident if his trailer came unhitched while travelling on a road, or if his unit were run into while properly parked. Therefore, we classified all the accidents among Hayes drivers into one of two classes as follows:

1. Avoidable accidents: These are accidents which a driver should be able to avoid by carefully following the principles of the Smith-Cummings-Sherman system. Examples from the Hayes records are: (1) Hayes unit started to pass another vehicle; the other vehicle slowed down and the Hayes unit struck it in the rear, or (2) Trailer of Hayes unit was out of line; the driver rode too close to a line of parked cars and sideswiped them.

2. Unavoidable accidents: These are accidents which a driver could not avoid, even if he carefully followed the principles of the system. Examples from the Hayes records are: (1) Someone moved the Hayes unit while the driver was in a truck stop; left side of top front of trailer was damaged, or (2) Hayes driver was attempting to back tractor under a semitrailer; the semitrailer dropped too low and rolled backward, cracking the dollies.

It is important to keep in mind that "avoidable" and "unavoidable" accidents are not necessarily "preventable" and "not preventable." Some accidents classified as unavoidable might be considered preventable; for instance, the accident listed above in which the semitrailer's dollies were damaged. Avoidable and unavoidable refer only to *whether or not the accident could have been avoided by following the principles of the Smith-Cummings-Sherman training system.*

To evaluate the effectiveness of the training system, primary emphasis should be placed upon that class of accidents which the system can reasonably be expected to affect.

Reliability of the Accident Classification: Since classification is a matter of judgment, it is possible that an accident might be classified differently by different judges. If the classification is to be useful, it must be reliable. This means that there must be a high degree of agreement among judges regarding what accidents are avoidable and what accidents are unavoidable. In a study of the rating scale used in the training system, we reported on the agreement between judges when classifying accidents as avoidable or unavoidable.

Four judges—Smith and three members of the research staff—independently classified a total of 157 consecutive accidents as avoidable or unavoidable. All four judges agreed on the classification of 62.5 per cent of the accidents. Three of the four judges agreed on the classification of 92.4 per cent of the accidents. All of these accidents had occurred during the 22-month period prior to the start of the field test.

As a further check on the classification system, the first 12 accidents which occurred after the beginning of the training period were classified by three judges. All three judges agreed on the classification of 11 out of the 12 accidents.

The classification appeared to be reliable. Therefore, in analyzing the results of the training, separate comparisons were made for the avoidable and unavoidable classes of accidents.

Mileage Rates: When mileage data were computed for the drivers, it was found that the trained drivers had driven somewhat more than the untrained drivers. The trained drivers had a total combined mileage of 4,837,000 miles; the untrained drivers had a total combined mileage of 4,576,000 miles. Because of this difference, we could not directly compare the number of accidents among the trained and untrained drivers. Instead, we used their mileage rates, i.e., accidents per 100,000 miles.

Results

The accident rates for the trained and untrained drivers during the 15 months following the training program are shown in Table 1.

None of the differences in Table 1 were statistically significant. In other words, the differences in accident rates between the trained and untrained drivers could be explained simply by random fluctuation.

TABLE 1.—COMPARISON OF ACCIDENT RATES: TRAINED AND UNTRAINED DRIVERS

ACCIDENT CLASS	TRAINED DRIVERS ($N = 59$)*	UNTRAINED DRIVERS ($N = 60$)
	Accidents per 100,000 miles	
Avoidable	.43	.61
Unavoidable	.54	.52
Total	.97	1.13

Note: The statistical test of significance of differences in accident rates was developed by Dr. Herbert H. Jacobs and others. It is analogous to the comparison of two frequencies from binomial populations, given by Hald. The assumptions and method of computation were presented by Irby & Jacobs in an earlier issue of the *Research Review*.

For the comparisons presented in this table, the results are:
Avoidable accidents, trained versus untrained drivers: Z = 1.06; P = ns.
Unavoidable accidents, trained versus untrained drivers: Z = .23; P = ns.
Total accidents, trained versus untrained drivers: Z = .68; P = ns.

* Mileage data were not available for one driver in the trained group. Theoretically, if his rate were very low, if might increase the differences between the groups to the point of significance. However, we conducted a special calculation to see if his presence could change the results. According to this calculation, even if he had driven 200,000 miles without an accident during the 15-month period (highest actual recorded mileage was 128,000) it would not have affected the outcome (e.g., for avoidable accidents, Z is increased to 1.19, which is still far short of generally accepted levels of significance).

The trained and untrained drivers were compared also on accident costs, as shown in Table 2.

None of the differences in Table 2 were statistically significant.

The comparisons of the trained and untrained drivers failed to provide significant support for the training system on either count. Nevertheless, comparisons involving avoidable accidents (the class which the training program would be expected to reduce) favored the trained drivers. The trained group had fewer accidents in total, had a lower avoidable accident mileage rate, and had a lower median cost per accident for the avoidable class.

On the other hand, the trained group did *worse* than the untrained group on all comparisons involving *unavoidable* accidents. The trained group had more unavoidable accidents in total, had a higher unavoidable accident mileage rate, and had a higher median cost per accident for the unavoidable class.

It might be argued that these differences, though not statistically significant, seem to lend at least impressionistic support to the training system. All differences in avoidable accidents favored the trained group. We disagree with this argument because it ignores the results of the statistical test in favor of an impressionistic judgment. Statistical tests are designed precisely to protect us from errors we would otherwise make by relying upon our impressions. By the same logic, one could conclude that the training increases a driver's vulnerability to costly unavoidable accidents. In both cases, the statistical tests show that the apparent differences are, in fact, well within the range of random fluctuation. We must conclude that the Smith-Cummings-Sherman system had no statistically significant effect on accident vulnerability or costs.

Impact of Different Trainers: Two different trainers took part in the training program. We could not overlook the possibility that these trainers might differ in what they taught the drivers or how they taught them. To detect any differences which might exist, we compared the accident rates of the three subgroups of trained drivers. The first group contained drivers who had been trained only by trainer 1 (either one or two sessions), the second group contained drivers who had

TABLE 2—COMPARISON OF MEDIAN ACCIDENT COSTS: TRAINED AND UNTRAINED DRIVERS

Accident class	TRAINED DRIVERS ($N = 60$)	UNTRAINED DRIVERS ($N = 60$)
	Median accident cost in dollars	
Avoidable	110	156
Unavoidable	98	75
Total	110	96

Note: The statistical test employed for determining the significance of differences in accident costs was the Median Test, as described in Walker and Lev, using the Chi-square corrected for continuity.

For the comparisons presented in this table, the results are:
Avoidable accidents, trained versus untrained drivers: Chi-square = .70; df = 1; P = ns.
Unavoidable accidents, trained versus untrained drivers: Chi-square = .02; df = 1; P = ns.
Total accidents, trained versus untrained drivers: Chi-square = .01; df = 1; P = ns.

been trained by both trainers (each driver had two sessions, one with each trainer), and the third group contained drivers who had been trained only by trainer II (either one or two sessions). The results are shown in Table 3.

TABLE 3.—COMPARISON OF ACCIDENT RATES FOR DRIVERS TRAINED BY DIFFERENT TRAINERS

Accident class	*Only by trainer I* (N=25)*	*Drivers trained by both trainers* (N—14)	*Only by trainer II* (N=20)	*Untrained Drivers* (N=60)
Avoidable	.19	.47	.79	.61
Unavoidable	.41	.62	.64	.52
Total	.60	1.09	1.43	1.13

Note: The statistical test employed for accident rate comparisons was described earlier (see footnote to Table 1).

For the comparisons presented in this table, the results are:

Accident class	*Trainer I versus both*	*Trainer I versus trainer II*	*Trainer I versus untrained*
(Statistic: Z referrable to unit normal curve)			
Avoidable	1.04	2.04†	2.64‡
Unavoidable	.58	.67	.41
Total	1.28	2.13†	2.18†

Accident class	*Trainer II versus both*	*Trainer II versus untrained*	*Both versus untrained*
(Statistic: Z referrable to unit normal curve)			
Avoidable	1.07	.50	.40
Unavoidable	.18	.29	.20
Total	.61	.69	.02

* Mileage data not available for one driver in this group.
† $P < .05$
‡ $P < .01$

The avoidable accident rate and the total accident rate for the drivers trained by trainer I were significantly lower than the corresponding rates for the drivers trained by trainer II and the untrained drivers. The odds were less than 1 in 20, that the lower rates among the drivers trained by trainer I were due to random fluctuation.

The size of this difference came as a surprise to us, even though we were prepared for possible differences between the effects of the two trainers. The differences in avoidable accident rate were quite large—e.g., drivers trained only by trainer II had an avoidable accident rate four times higher than drivers trained only by trainer I. Despite the breaking up of the trained drivers into smaller groups, several of the differences were statistically significant.

We believe that the issue the foregoing data raise is worthy of study. To identify the characteristics of effective trainers may be equally as important in fleet safety as to identify the characteristics of effective training systems.

TABLE 4.—COMPARISON OF SUBGROUPS OF TRAINED DRIVERS

Characteristics	*Only by trainer I* (N — 25)	*Drivers trained by both trainers* (N = 14)	*Only by trainer II* (N = 20)
Average age (in years)	39.0	41.5	37.2
Average seniority (in years)	6.8	7.8	4.8
Average mileage	86,000	91,000	70,000
Average number of accidents prior to training	1.3	1.1	1.1

Note: The statistical test employed for estimating the significance of differences was the t-test recommended for use when the population variances are unknown but presumed unequal and the samples are small.

For comparisons presented in this table, the results are:

	Trainer I versus trainer II	*Trainer I versus both* "t"	*Trainer II versus both*
Average age	.64	.84	1.47
Average seniority	.99	.48	1.19
Average mileage	1.95	.70	2.23*

* $P < .05$

The accident data could not meet the requirements of the t-test, hence were compared by means of the less restrictive Chi-square. None of the Chi-squares were significant.

Because the differences in accident rate between the groups of drivers might have been due to other characteristics, Table 4 was prepared, showing the average age, seniority, mileage, and the accidents they had prior to the Smith training.

Table 4 shows that the drivers trained by trainer I were slightly older, had slightly more seniority, had accumulated slightly more mileage, and had experienced slightly more accidents than the drivers trained by trainer II. None of the differences is statistically significant, except the mileage difference between drivers trained by trainer II and those trained by both trainers ($t = 2.23$, $n = 32$, $p = .05 > .02$). Furthermore, these differences cannot explain the post-training differences in accident rate because the drivers trained by both trainers (who were intermediate in accident rate) were older, more experienced, and accumulated more miles than *either* of the other groups.

What Conclusions Can Be Drawn From the Field Test? Certain of the results obtained from the field test hold special interest for research. However, the major conclusions of a practical nature may be stated as follows:

1. Effectiveness of the Smith-Cummings-Sherman training system—as a system—in preventing certain types of accidents by experienced professional drivers, was not demonstrated unequivocally. Neither accident rates nor accident costs differed significantly between the trained and untrained drivers.
2. It is possible that the merits of the system might be demonstrated more easily and might produce more convincing results with beginners rather than with professional drivers.
3. Effectiveness cannot be evaluated independently of the trainers. It is possible that the Smith-Cummings-Sherman system may produce useful results with some trainers. If some trainers are more effective than others, it is important to identify who will be an effective trainer.
4. One important practical question is still unanswered. Should the Smith-Cummings-Sherman training system be recommended for fleet use? A blanket answer is not possible. The field test results indicate that the fleet safety director who uses the training system may or may not get significant accident reductions, depending upon who does the training. Because of this uncertainty, the final decision must be an individual one.

Research and the Practitioner

To the editor:

Three cheers for Doctor Goldstein for his article "Whither Accident Research?" in the March 1963 issue of *Traffic Safety*. And particularly for one of his parting observations concerning the "need for communication between the researchers and the program people." We could not agree more.

Program people, among them fleet safety directors, must cope with the vehicle accident problem on a day-to-day basis and keep management and the drivers sold on the idea that their safety programs make sense. These people have more problems than an umbrella has rain drops in a storm and deserve better of the research fraternity who should be trying to diminish their problems instead of adding to them, and who should be giving them new and better tools to work with instead of breaking up what few imperfect tools they have.

There seems to be a trend in accident research today to shoot down the old and honorable concepts that have been the mainstay of successful fleet accident prevention programs for many years.

The type of research we are talking about goes something like this:

> For many years fleet supervisors have been using Idea X in their accident prevention programs. We don't think that Idea X is as good as the fleet supervisors think it is. Accordingly we set up a research project to find out. Sure enough, we did find out that Idea X is not as good as they think it is.

Now the researchers do not say that Idea X is bad. They may even admit with ill concealed reluctance that Idea X does some good. Their main pitch is that Idea X is not perfect and that were it not for research, fleet safety supervisors would never become aware of this great fact.

Recently what is known as the Smith System was checked somewhat along these

[Reprinted from *Traffic Safety Research Review*, 7:3:2-3, 1963, published by the National Safety Council.]

lines by Payne and Barmack (*Research Review*, March, 1963). Now the Smith System has been around for a long time and has been accepted by a very wide range of practicing fleet safety supervisors. For the driver trainer and the driver supervisor, the Smith System offers a welcome method of organizing the driving problem and teaching drivers how to cope with the hazards of every day traffic.

Justification for this research is interesting. It was reported that questionnaires had been sent out to a list of 49 fleets who had used the Smith System. The list, incidentally, was supplied by Harold Smith himself. The replies to this questionnaire indicated that there were small improvements in the accident rates in most of the fleets that had used the Smith System. Now this finding alone should have been sufficient to prompt the abandonment of this research project and moving on to some more important problem. After all, there is no such thing as a small improvement. The prevention of just one accident is a substantial achievement. But there was an additional curious fact revealed by the survey. It was this—that ". . . the enthusiasm of the fleet safety supervisors outran the demonstrated results of the training system." This is a problem? Payne and Barmack do not state how they measured the enthusiasm of the fleet safety supervisors, but apparently by some yardstick there was such a great disparity between the enthusiasm of the supervisors and the results as to justify "an experimental evaluation of the effectiveness of the system."

Armed with this rather flimsy reason, Payne and Barmack proceeded with their research project. And what did they find? They found that the effectiveness of the Smith System was not "demonstrated unequivocally"—that the fleet safety director who uses the training system may or may not get significant results—and that the final decision as to whether the system should be used will have to be an individual one.

The mountain has labored and brought forth a mouse. But a rather mischievous mouse. The effect for accident prevention will be to promulgate a certain degree of reluctance to employ the Smith System—a system which the researchers admit has effected small improvements in the accident rates of fleets reporting in their original survey.

The point is that there are no perfect tools in the field of accident prevention. It is a job of finding and using the least imperfect of the tools available. Having some tools to work with is better than having none at all.

Chris Imhoff
Motor Transportation Department
National Safety Council

In Reply

To the editor:

Chris Imhoff, in his letter to the *Research Review*, is unhappy with the negative results that turned up in our evaluations of the Smith-Cummings-Sherman System. So were we. But we also know the value of negative results and welcome the opportunity of pointing out why.

Our original purpose was to determine whether specific behind-the-wheel behavior could be related to specific types of accident experience, or to accident experience generally. We wanted to know whether it was more important, for example, to "aim high in steering," or to "keep your eyes moving," or to "leave yourself an out."

We approached the system with high hopes. The basic ideas of the system made sense. It was the best and most integrated of the training systems that we knew about. It had been available to fleets for several years. It was reported to have been used with a large number of professional drivers, for whom comprehensive driving and accident records had been kept (estimated to be about 50,000 drivers). A driver performance rating scale had been developed.

Here, it seemed, was a system in which specific performance behind-the-wheel could be related to accidents. With such information available, we could find out what behavior to concentrate upon, thereby making the training more effective. That was our goal. Our general program was as follows:

1. Establish contact with fleets which had

used the training system, to secure training program information and accident records for detailed analysis, and to solicit their research cooperation.

2. Determine the reliability and validity of the driver performance rating scale. (This had not been done before.)

3. Initiate a program of data collection with participating fleets to obtain a large enough number of cases to identify statistically significant differences in accident frequencies for specific performance items.

Upon contacting the list of fleets provided by Smith, we found that many of them had only partially adopted the system, had mixed it in with other programs, or had not extended the prescribed training to the drivers. It also became clear that the training and accident records which were available were inadequate for a detailed study of relationships between rated driver performance and accident experience. There were just not enough drivers with adequate training and records to permit the original study. This was our first disappointment. But we found from this disappointment that you can't do a study as refined as the one we intended, without extensive long range preparations with many fleets. This phase of the research program was reported by Payne and Prince in the *Research Review*.

However, there were other characteristics of the system which could be studied, so we proceeded to do so.

Our evaluation of the reliability and validity of the driver performance rating scale provided a second disappointment. In a special field test, drivers were rated by Smith and by four other raters trained by Smith. Two of the raters were truck drivers, each with more than 10 years of experience, selected by Smith; the other two were professional members of the research staff. Coefficients of reliability among the raters ranged from +.18 to +.51—far below minimum acceptable standards for assessment of individual performance. In both the field test and in additional data collected from three other fleets (involving 1,000 drivers) there was no relationship between ratings and accident experience.

In spite of the disappointment, we learned something positive from this experience—that useful predictions about accidents from performance ratings of experienced professional drivers probably cannot be achieved. Performance rating scales have little if any value in driver screening, though they may have some use as incentives in driver training. Their value in a training program, however, remains to be demonstrated. This phase of the research program was reported by Barmack, Fabrizio, Payne, and Prince in the *Research Review*.

Despite the inadequacy of the rating scale, the training itself might be effective. It was clear that no convincing assessment could be made from the available raw fleet data, so our next step was to conduct a controlled experimental field test of the training system under fleet operating conditions. The McLean Trucking Co. generously provided the opportunity through its Hayes Freight Lines Division.

The experimental field test required a year and a half to complete. It included 120 drivers, for whom monthly accident and mileage records were carefully compiled and analyzed. The effect of the training program on these experienced professional drivers was small. As it turned out, statistical tests showed that the trained and untrained drivers did not differ significantly on accident rates or costs.

This finding was also disappointing, but here, too, we learned something. There seem to be important differences between trainers—in other words, those who teach may be as important as what is taught. We carefully pointed out that the system might produce more convincing results if it were tested on beginners, rather than on professional drivers. Both of these clues—that is, the importance of the teacher, and of the type of pupil—are worth following up with additional research. The experimental field test was also reported in the *Research Review*.

No small part of our disappointment arose because we had the unhappy task of reporting evidence which did not support some of the beliefs and claims made by

Smith, Cummings, and Sherman, who were most cooperative throughout the course of our study.

On the other hand, accidents are a major health problem. Their costs—in deaths, injuries, and dollars—are all too familiar to the transportation industry. If progress is to be made in controlling and preventing accidents, we know of no alternative to rigorous, objective testing of the programs offered to reduce accidents. All of us who are interested in the problem should be prepared to accept our disappointments, and to learn from them.

The impressive progress which has been made in controlling and preventing disease did not result because medical research was content to rely upon untested treatments. It resulted from continuing efforts to find out which remedies do or do not work, and to find new ones which work even better.

If Mr. Imhoff has an alternative to this approach, we should be happy to have him share it with us.

JOSEPH E. BARMACK, PH.D.
DONALD E. PAYNE, PH.D.

THE EFFECTS OF DRIVER IMPROVEMENT ACTIONS ON DRIVING BEHAVIOR

—B. J. Campbell, Ph.D.

The inadequacy of evidence as to the efficacy of certain widely accepted and expensive accident prevention measures is well indicated by this pioneering work, which provided the first scientific evaluation of measures that have been in use for decades. In view of the hazards of basing countermeasures on common sense, it is noteworthy that in this instance the measures were shown to have some favorable effect.

THE PRACTICE of withdrawing driver licenses upon conviction for certain violations of the motor vehicle law is almost universal in North America. The law in most jurisdictions requires revocation of the driving privilege upon conviction for any of several specific offenses, and permits suspension for others. Moreover, in most jurisdictions the license may be taken for an accumulation of offenses even though none warranted action by itself. In recent years increasing amounts of attention have been given to the latter group of drivers whose total record is serious though no single offense is particularly flagrant. These driver improvement programs have in common the fundamental assumption that drivers who violate motor vehicle laws are dangerous and must be dealt with.

Driver improvement programs can justify their existence only to the extent that they accomplish their fundamental purpose—that of bringing about desirable changes in the "behind-the-wheel" behavior of drivers with whom they deal. McFarland has mentioned the lack of studies concerning the socio-legal control of drivers and has stated that there is need for carefully controlled research in this area. The lack of clear-cut evidence showing that driver improvement programs produce desirable changes in drivers indicates a necessity for investigating the fundamental effectiveness of such programs before proceeding too far with

Reprinted, with permission, from *Traffic Safety Research Review*, 3:3:19-31, 1959, published by the National Safety Council.

elaborate refinements which carry the implicit *assumption* that they are effective. For this reason, one of the major undertakings of this study was an investigation of changes in driver behavior subsequent to departmental action.

The organization of this study is more easily presented if driver improvement action is thought of as involving (1) a process of selecting drivers in need of attention, and (2) a policy of taking one of several actions toward each such driver. It is basic that driver improvement programs need to contact the drivers most likely to cause or be involved in motor vehicle accidents and through appropriate action to reduce this tendency. At first glance it would seem that departments would therefore select for action those drivers who have the most accidents. The actual and universal fact is, however, that action is initiated toward most drivers not because they have been involved in accidents, but because they have been convicted for violations of the motor vehicle law (most of which had nothing to do with an accident). This policy is pursued on the assumption that drivers who violate the law are more likely to cause or be involved in accidents than drivers who do not violate the law. Kelly stated this fundamental assumption very well when he referred to a violation as "a symptom of an accident to come." Of course, many jurisdictions have programs under which accident repeaters are designated for some kind of attention, but, by far, the majority of candidates for driver improvement are selected on the basis of violation records.

After the process of selecting the driver comes the problem of deciding the type of action to be taken toward those selected for driver improvement. One of the most common actions is that of suspending the driving privilege, which presumably reduces the immediate hazard by removing the dangerous driver from the road and reduces the future hazard by making him drive more carefully in order to avoid losing his license again. Probation is another common device used in driver improvement which, presumably, produces desired changes by making him understand that a breach of probation will result in suspension. Advisory letters are used even more frequently than the preceding methods and are thought to have persuasive value in improving drivers.

The basic task of this study is to evaluate the subsequent record of drivers selected on the basis of violations and dealt with in certain ways including suspension, probation, etc. The subsequent record might be expected to show a decrease in accidents and violations, and the decrease should be of sufficient magnitude to be attributable to departmental action. Moreover, the influence of this action might be shown through changes in the type of violation committed subsequently and through the time lapse between departmental action and the next offense.

Design of the Study

In New Jersey, where the data for this study were gathered, a point system is in operation. Each motor vehicle violation carries a certain number of demerit points, and when a driver accumulates a certain number of points, some driver improvement action is taken toward him. Drivers receiving such action constitute the experimental group, and changes in their driving records after action are compared with changes in the record of control drivers whose records warranted action, but who received *none*.

The necessity for a control group would seem to rule out the possibility of an adequate study, because the usual policy in driver improvement is to take action toward *all* drivers at a given level. Thus, ideally, there would *not* be any drivers who warranted but did not receive action. It is found, however, that in several jurisdictions the work load is so heavy that only part of the cases at a given level are processed. The usual way of handling the case surplus is to review all cases, act on the most serious ones, and route the less serious back to file. The study could not be carried out in such a jurisdiction because the control group would consist of less serious cases and would not be comparable to the experimental group. It is fortunate for this study that the case load in the New Jersey point system exceeds the number that can be processed, and, even more important, that the method of deciding *which* cases at 12 points will receive action does not involve reviewing the cases. Because of the existence of an adequate control group, New Jersey was selected as the jurisdiction in which to carry out this research.

In order to obtain a valid comparison, the experimental and control groups must be similar in all pertinent respects except receipt of action. There are two common methods of equating groups to meet these requirements. The first involves matching each driver in the experimental group with a control driver of the same age, sex, occupation, driving experience, accident and violation record, etc. The matching process is tedious and is feasible only with small samples of drivers and large amounts of time. The second method depends upon dealing with an entire population of drivers having certain characteristics (in this case a bad driving record) and assigning samples of the population to the experimental or control group in a random manner. If the assignment is random, it may be assumed that the groups differ only by chance in all respects except the variable under observation. The latter method was used in this study.

Collecting the Data: The information on which this study is based was taken from a case file containing records of drivers who reached 12 points and received departmental action, and a master file containing records of all drivers with one or more points who have not received action. Approximately 8,000 folders from the case file were screened and 2,400 were eliminated because the driver had received some other driver improvement action in addition to that under the point system (such as suspensions imposed by magistrates and revocations imposed under mandatory provisions of the law). The 5,600 remaining cases are records of drivers whose only contact with driver improvement was through the point system. In addition to these, the records of approximately 8,000 drivers were obtained from the master file. These records lay, without distinguishing marks, among nearly one million abstracts of conviction. To obtain these cases (drivers who warranted but did not receive action), 183 of the 265 file drawers were searched. The term search is used in a literal sense because in order to locate the cases each had to be examined for point total and dates of violation. The information was gathered by more than 35 workers who devoted nights and weekends to this project over a ten-week period.

Assumptions: In this study two basic assumptions are submitted to test:

1. It is proper to select drivers for departmental action on the basis of an accumulation of violations, because such drivers also have more accidents than average.

2. Departmental action is more effective than no action in bringing about desirable changes in the driver's behavior.

The first assumption is tested by observing the numerical relationship between certain convictions and reported accidents. The second assumption is tested by observing changes in the frequency and nature of subsequent violations committed by drivers who received action as compared to those who did not. Originally, it was planned to evaluate changes in the frequency and type of accidents as well, but a systematic error was discovered in the original information which introduced a bias and, most unfortunately, required discarding the accident data.

Accident-Violation Contingency

The first of the two basic assumptions is that drivers with violations on their records are more likely to have accidents than those with no violations. Previous studies have suggested a contingency between accidents and violations, but it is not clear whether they included controls for any of the factors which may cause spurious correlation. The first of these factors pertains to the kind of violations entering into the correlation. It often happens that a violation is charged as an outgrowth of an accident investigation.

This results in the fact that any large group of drivers, each having several accidents, will have some violations on their record, if for no other reason than because of the charges filed at the scene of the accident. Such a correlation is of little value to driver improvement, because (1) the violation has no predictive value since it was charged after the accident, and (2) violations charged in accidents are only a small fraction of the total arrests which constitute the basis for driver improvement action. (In North Carolina only about ten percent of state highway patrol arrests are connected with accidents.) For these reasons it is probably more useful to observe the contingency between non-accident violations[1] and accidents. The question then becomes, "Do drivers who commit violations not connected with accidents also, on other occasions, tend to have accidents?"

The second factor that tends to cause spurious correlation between accidents and violations is that of driving experience or age. Young drivers, who have been licensed only a short while, have not had time, as a group, to becomes involved in very many accidents *or* very many violations. On the other hand, older drivers, or drivers with many years experience, have had ample time to be involved in numerous accidents and violations. Extreme variation in age or experience may therefore cause an artificially inflated correlation. One way of checking the degree of this artificial correlation is to compare the accident-violation contingency of two groups of drivers, one having a wide range with respect to age and experience, and the other being more restricted in age and experience.

Figure 1 shows the contingency between accidents and violations taking into account the two factors mentioned. Only non-accident violations are used, and a group more restricted in age and driving experience is included as a comparison with the variant total sample. The information from which the graphs were made is shown in Table 1. It is obvious that there is a stable and substantial contingency between non-accident violations and accidents when dealing with the average record of a large group of drivers. The slope of the graph is not greatly different even when age and experience are restricted. At least one source of artificial correlation remains uncontrolled in that drivers with high exposure (high annual mileage) may have both more accidents and more violations than those with low exposure, and this would tend to make the slope steeper than if exposure were controlled.

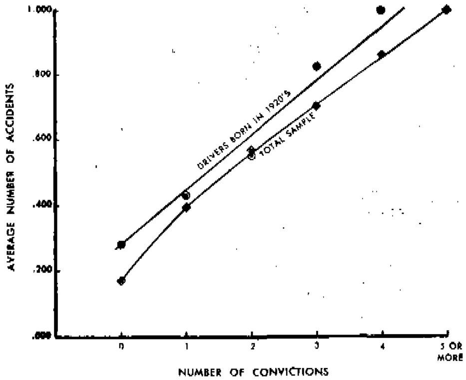


FIG. 1.

The information used to show the accident-violation contingency was obtained by analyzing the records of more than 40,000 North Carolina drivers. The analyses reported in the remainder of the chapter are based on the research conducted in New Jersey.

EFFECTS OF THE ADVISORY NOTICE

Ordinarily an advisory notice or letter is the first contact between the department and the driver who is beginning to build up a violation record. In New Jersey, for example, drivers who reach eight to 11 points were sent an advisory notice containing a request for their cooperation, a statement of their present point total, and a schedule of points. The letter in New Jersey, as in other jurisdictions, is sent in the hope that it will influence the recipient to drive more carefully and lawfully.

Little is known about the ability of a

[1] A non-accident violation is the typical situation in which the driver is apprehended and convicted for a moving traffic offense which did not occur in connection with an accident.

TABLE 1.—AVERAGE ACCIDENTS FOR DRIVERS WITH VARIOUS NUMBERS OF VIOLATIONS

Number of non-accident violations	*Drivers in total sample*	*Average accidents per driver*	*Drivers born in 1920's licensed 1947-50*	*Average accidents per driver*
0	29,984	.167	5,357	.279
1	5,921	.391	400	.435
2	2,221	.560	146	.548
3	1,042	.699	57	.825
4	595	.857	40	1.075
5	704	1.001		
Total	40,647		6,000	

letter to control a driver's behavior, but the topic is most pertinent and at least one prior study has been carried out. The Texas Department of Public Safety recently completed two studies, one concerning advisory letters and the other concerning re-examinations. In the first study, 100 drivers eligible for an advisory letter were selected from each of the four largest cities in Texas. Such a letter was sent to half the drivers from each city and was not sent to the other half. In the second study the same procedure was used but the letter was sent to drivers with a more severe record and required that they be re-examined. The records of these 800 drivers were kept under observation for approximately two years. The results of the study are presented in such a way as to permit computation of the percentage of drivers in each group who had subsequent accidents and violations. In each study the results are in the expected direction, but in only one is the difference large enough to be statistically significant. In the study of re-examinations, significantly fewer drivers who were re-examined had subsequent violations than drivers not sent examination letters. It is not clear, of course, whether the examination itself or mere contact with the department accounts for the differential improvement. Results derived from the Texas study are presented in Table 2.

The Texas study is commendable despite the fact that the results are less than startling. All the trends are in the desired direction, and indeed, if the same proportions held true in a larger sample, all the results would be statistically significant. The important thing about the Texas study is that the groups were set up in advance. This simple step eliminates the intricate and frustrating problems of sampling that were encountered in the New Jersey study to be reported. By establishing the control group in advance and setting the time period over which records are observed, most of the difficulties of after-the-fact research are eliminated. More will be said of this kind of study design at the conclusion of the chapter.

The New Jersey study is similar to the Texas study in some respects and dissimilar in others. It is similar in that drivers who were and were not sent letters are compared in terms of subsequent violations. It is different in that the subsequent time period in the New Jersey study is not necessarily the same for each group. The practice of sending advisory letters in New Jersey was in effect for only eight months at which time personnel problems forced abandonment of the service. The letter group is drawn from those drivers whose record reached eight to 11 points during the period that letters were sent. The no-letter group is drawn from those drivers whose record reached eight to 11 points during the four months immediately before or immediately after the period during which the letters were sent. The average time at which the letters became due was about the same in each group. Since this is true, such things as fluctuation in enforcement stringency should affect both groups approximately equally. The New Jersey research, then, hinges on the comparison between two fairly similar groups, one sent letters and the other not.

TABLE 2.—TEXAS DEPARTMENT OF PUBLIC SAFETY STUDY OF ACCIDENTS AND VIOLATIONS AFTER TWO TYPES OF LETTERS

	Drivers sent advisory letter	*Drivers not sent advisory letter*	*Drivers sent examination letter*	*Drivers not sent examination letter*
Number of drivers	200	200	200	200
Proportion with further violations	.580	.625	.525	.690
Proportion with further accidents	.200	.265	.175	.250

Drivers who were sent letters may be divided into two categories, one consisting of those whose record subsequent to the letter continued to be so severe that further action had to be taken and the other consisting of those whose subsequent record improved to the extent that no further action was required. In terms of point accumulation, this means that the first and by far the smaller group reached eight and went on to 12 or more points, and the other, probably constituting 80 per cent or more of those sent letters, reached eight but did not accumulate as many as 12. Each of these two groups is important to the over-all study, and they should be combined and analyzed in their proper proportion. Unfortunately, however, they must be analyzed separately because of a sampling problem inherent in the after-the-fact nature of the research. The first set of comparisons concerns those drivers who reached eight points and did not go on to reach 12 points. These are presented in the most detailed form, because they are the most significant and concern the largest class of drivers. Since all these drivers showed enough improvement to avoid further trouble, whether sent a letter or not, the only way that the influence of the letter can be shown is through an even greater improvement in the subsequent record of the letter group than in that of the no-letter group. Such a comparison puts the letter to a most severe test.

Comparability of Groups Prior to Action: It is important that the letter and no-letter groups be compared with respect to their record up to the time of reaching the letter level. If the groups are substantially different *before* reaching letter level it would be impossible to know whether subsequent differences were attributable to the effects of the letter or to differences inherent in the groups prior to the time the letter level was reached. One way of comparing the letter and no-letter groups is to calculate the average number of points they had when they reached the letter level. Depending on the particular combination of violations that a driver commits, he could be sent a letter at eight, nine, ten, or eleven points. Two speeding convictions, for example, give a driver eight points and he would be sent a letter. If, however, he committed a speeding violation and then a stop sign violation, his point total would be seven and no action would be taken. If his next violation were speeding, he would then have eleven points and would be sent a letter. If the letter and no-letter groups are comparable, the average points at the time of reaching the letter level should be the same.

The 2,880[2] drivers in the letter group are compared to the 3,814 drivers in the no-letter group. At the time they reached the letter level the letter group had an average of 8.84 points, and the no-letter group had an average of 8.76 points. As is shown in Table 3, the averages differ by less than one-tenth of a point. Not only are the point totals similar, but the times at which the groups reached the letter level are very nearly the same. The month in which the average driver in the letter group was sent a letter was November, 1955, and the month in which the average driver in the no-letter group reached the letter level was January, 1956.

Another indication of the similarity of the letter and no-letter groups is the average number of points per violation committed prior to reaching the letter level. If points are scaled according to seriousness, the number of points per violation should give some indication of the severity of the violations committed. Table 3 shows that drivers sent letters were assessed an average of 3.84 points per violation, and drivers not sent letters received an average of 3.71 points per violation. These estimates were obtained by dividing the average number of points by the average number of violations. Obtaining the

[2] The reader will note throughout the article that the sample size of a given group varies from one analysis to the next. There are two reasons for this variation. First, some punched cards were discovered to be partly incorrect and were therefore eliminated from some analyses. Second, some cards became worn, jammed in the machine, and were destroyed as the analysis proceeded. There is no reason, however, to assume that any systematic error is introduced by these slight variations in sample size.

TABLE 3.—SUBSEQUENT VIOLATIONS OF DRIVERS SENT LETTERS AND DRIVERS NOT SENT LETTERS

	Number of drivers	*Average points before letter*	*Average length post letter period*	*Proportion with subsequent violations*	*Points per violation before letter*	*Points per violation after letter*	*Lapse between letter and next violation for drivers with subsequent violations*
Drivers sent letters at 8 points	2,880	8.84	22	.117	3.84	3.71	M = 10.0 s = 6.11 N = 317
Drivers not sent letters at 8 points	3,814	8.76	20	.192	3.71	3.60	M = 8.7 s = 6.5 N = 720
				CR = 8.33 $p < .01$			CR = 3.03 $p < .01$

ratio of points to violations in this manner is not algebraically correct, but the particular way in which the information was gathered and arrayed on the punched cards makes it impossible to arrive at the ratio properly. This estimate is presented for what it may be worth.

The final comparison of prior records of the two groups concerns the *type* of violations committed. Chi square is used to test differences in the frequencies. The table also shows each violation frequency as a percent of the total. Table 4 shows that the letter and no-letter groups differ significantly, even before the letter, in terms of the frequency breakdown of the violations. The frequency of speeding violations, for example, is considerably more than expected in the letter group and considerably less than expected in the no-letter group. It is interesting to find

TABLE 4.—FREQUENCY OF VARIOUS TYPES OF VIOLATIONS PRIOR TO REACHING LETTER LEVEL

	DRIVERS SENT LETTERS			DRIVERS NOT SENT LETTERS	
Violation	*Observed frequency*	*Observed per cent*	*Expected per cent*	*Observed per cent*	*Observed frequency*
License or restriction	13	0.4	1.5	2.2	85
Equipment	4	0.1	0.4	0.6	23
Traffic control	208	7.2	7.9	8.4	319
Passing	40	1.4	1.6	1.8	68
Lane travel	41	1.4	1.5	1.6	60
Others	22	0.8	0.9	1.1	40
Careless driving	170	5.9	7.3	8.4	317
Speeding	2,263	78.3	73.6	70.1	2,655
Reckless driving	55	1.9	2.1	2.2	85
Leaving scene	73	2.5	3.1	3.6	137
Total	2,889	99.9	99.9	100.0	3,789

$\chi^2 > 85.0$ df = 9 $p < .01$

that the two groups show significant[3] differences in the type of violation committed, even though the average points for each was very nearly the same. The number of observations makes the significance test very powerful, and relatively small absolute differences approach statistical significance. Considering all the comparisons, it seems that the groups are reasonably alike in terms of their need for driver improvement even though certain significant statistical differences appear. The nature of the differences is not such as to suggest that one or the other group would show a superior record subsequently.

Differences Subsequent to Action: If advisory letters produce any desired effects, fewer drivers sent letters should show subsequent violations than those not sent letters. Table 3 shows that only 12 per cent of drivers sent letters had further violations whereas 19 per cent of those not sent letters had at least one. Naturally, these values are partly dependent on the time period over which the violations were recorded. If one group was measured over a one-year period and the other over a two-year period, the latter probably would show more violations simply because a greater time period was involved. Since the present study was not a controlled experiment, set up in advance, the time period for each group is not the same though fortunately the averages are nearly equal. The average time over which the letter group was observed was 22 months as compared with 20 months for the no-letter group. The signif-

[3] $\chi^2 = 87.6$ ($p < .01$).

icance[4] of the superiority of the subsequent record of the letter group is underscored in view of the fact that relatively fewer violated despite a somewhat longer time period. Among drivers who reached eight but not 12 points, it appears that the letter reduces by almost half the per cent of drivers who would otherwise commit an additional violation. Note that the reduction in violations in the control group is rather marked. This is probably partly regression to the mean (a rather low mean—only about seven per cent of New Jersey drivers violate each year) and partly a result of other deterring factors.

Despite the influence of the letter and other deterring factors, 12 per cent of drivers sent letters and 19 per cent of drivers not sent letters show further violations. The letter appears, however, to influence even those that violate subsequently by delaying the onset of the violation. Considering only the 400 drivers who committed violations after being sent a letter and the 700 who committed violations but did not receive such a letter, the average lapse between the time of reaching letter level and the next violation was computed. For those sent a letter, the lapse is 10.0 months and for those not sent a letter, the lapse is 8.7 months. The lapse following the letter is significantly[5] greater than in the cases in which no letter was sent.

Differences Independent of Action: Though the letter is the variable operating differentially, there are other deterring forces operating in both groups. Drivers in both groups had to pay fines, perhaps endure censure, etc. If this is true, drivers not sent letters should show improvement as time, fines and censure go on. Two comparisons suggest that this is true. As was mentioned earlier, drivers not sent letters had an average of 3.71 points per violation, but later, though no letter was sent, the average number of points per violation was found to be 3.60. If the average number of points received for each violation decreases, then it must be true that high-point-value-violations (more serious ones) were committed relatively less often later than they were earlier. This hypothesis can be tested by studying the no-letter group with respect to the type of violation the group committed as their first offense and the type committed subsequent to reaching the letter level (ordinarily their third or fourth violation). Table 5 shows that speeding, reckless driving, and leaving the scene of an accident are committed relatively more frequently as first violations than they are as later violations. Naturally, if the percentage of these decreases as time goes on, something else has to increase since total violations must still add up to 100 per cent. It is found that the increase is distributed over several violations generally of a less serious nature, and the changes are very significant.[6] Very similar changes were found in the letter group by comparing the first violations committed with the ones committed immediately after the letter was sent.[7] It is evident that (1) the type of violations committed changes significantly as the driver

TABLE 5.—FREQUENCY OF VARIOUS TYPES OF VIOLATIONS AMONG DRIVERS NOT SENT LETTERS

	FIRST VIOLATION			THIRD VIOLATION	
Violation	*Observed frequency*	*Observed per cent*	*Expected per cent*	*Observed per cent*	*Observed frequency*
License or restriction	85	2.2	2.3	2.7	20
Equipment	23	0.6	0.9	2.2	16
Traffic control	319	8.4	9.6	15.7	117
Passing	68	1.8	2.4	5.7	42
Lane travel	60	1.6	1.9	3.4	25
Others	40	1.1	1.3	2.8	21
Careless driving	317	8.4	9.2	13.3	99
Speeding	2,655	70.1	67.3	53.4	397
Reckless driving	85	2.2	2.0	0.7	5
Leaving scene	137	3.6	3.0	0.1	1
Total	3,789	100.0	99.9	100.0	743

$\chi^2 > 190.0$ df = 9 $p < .01$

[4] The critical ratio of the difference between proportions is 8.33 ($p < .01$).

[5] Critical ratio equals 3.03 ($p < .01$).

[6] Chi Square equals 190.

[7] Note that in this and succeeding tables the assumptions of independence inherent in the use of Chi Square are not met. It is likely however that the resulting error results in an *underestimate* of the true significance (See Appendix 1).

continues to violate, (2) the nature of the change does not appear to be related to receipt or non-receipt of the advisory letter, and (3) changes are in the direction of less serious violations. *These* changes seem logical, but some others are not so readily explained. It is found that the letter and no-letter groups differ significantly in terms of the type of violations they committed first (see Table 4), each undergoes significant changes as the violation sequence develops, and finally, after reaching the letter level, they are more similar than they were earlier. The meaning of these latter changes is not clear, but they are interesting to note.

Drivers Who Eventually Reached Suspension Level: The foregoing comparisons pertained to drivers who reached the letter level and, though they may have received *some* additional points, did not reach suspension level. Most drivers who accumulate enough points to warrant a letter fall into this category, but a few (usually not more than 10-20 per cent) continue to accumulate points until they reach suspension level. The records of such drivers were subjected to the same analyses as those just reported. In a sense such drivers are, by definition, those who are not deterred by the various forces of society, therefore, it should not be surprising to find that receipt or non-receipt of an advisory letter has no appreciable influence on the nature or severity of their subsequent records. The only significant finding is that drivers who were sent a letter and went on to suspension violated sooner than drivers not sent a letter who went on to suspension. Advisory letters show appreciable effects on the 80-90 per cent of drivers who reach letter level but do not become real problems, but show little influence on the small proportion who continue to violate. It is gratifying to find that advisory letters are the most effective with the largest class of drivers and that action at suspension level has desirable effects on those few drivers who are apparently insensitive to other deterrents.

Summary of Findings: Comparison is made between two groups of drivers, all of whom reached eight points and none of whom went on to reach suspension level (12 points). One group was sent advisory letters at eight points and the other was not. After establishing comparability of the groups with respect to prior record, it is shown that fewer drivers sent letters have subsequent violations than those not sent letters. The letter is shown to influence even those who *do* violate subsequently in that the lapse between the letter and the next violation is longer than the comparable period in the no-letter group. Both the letter and the no-letter group showed significant changes in the type of violation committed third in series compared with those committed first. The changes are in the direction of less serious violations and do not appear to be influenced by receipt or non-receipt of an advisory letter. The smaller group of drivers who go on to suspension level was, of necessity, analyzed separately and does not appear to be influenced by the letter (at least not desirably influenced).

Discussion of Results: It is clear that the advisory letter induces desirable changes in the behavior of many of its recipients. Some readers may be disappointed to find that letters are no more effective than to reduce 19 per cent violations to 12 per cent, but perhaps further consideration will change that somewhat. By extrapolation of the New Jersey results onto a larger scale it is possible to see that small differences in effectiveness of a technique affect large numbers of drivers on a continent-wide basis. Suppose that there are 85 million drivers in the United States and Canada and that letters are sent in all jurisdictions in the proportion found typical in point system jurisdictions. This would mean that about 850,000 letters per year would be sent. Since other evidence indicates that only about 15 per cent of drivers go on to suspension level after receiving a letter, it would seem that 85 per cent or 722,500 drivers would fall into the category of drivers studied in New Jersey. If the New Jersey results are applied to this number, it would be supposed that 19 per cent of the total would violate subsequently. Nineteen per cent of 722,500 amount to 137,275 drivers. If 12 per cent would violate after receiving the letter, the number would be 86,700. The difference between these two numbers is more than 50,000 which indicates the number of drivers that would be deterred each year by full use of the advisory letter. To carry this fanciful example one step further, suppose that through refinement of the letter it was found that 11 per cent violated subsequently instead of the present 12 per cent. The one per cent difference would involve more than 7,000 drivers each year, which means that adoption of such an improved letter for three years would affect more than 20,000 drivers. Everyone realizes that extrapolation of such figures is fraught with error, and it is not the intention in this illustration to estimate the number of drivers influenced by the letter. It is important, however, to emphasize the fact that small increases of driver improvement practices will influence large numbers of drivers each year and when such improvements are carried over a period of years the results are staggering to the imagination. There is little validity to arguments that small scale refinements in driver improvement are not worthwhile.

In addition to preventing violations, there is evidence that the advisory letter tends to delay the onset of the violation in those who do violate. A most interesting study could be carried out which would consist of sending a group of drivers a friendly, semi-congratulatory letter about six months after the advisory letter, provided that there had been no further violation. Such a letter might bolster the waning effects of the advisory letter and prevent the onset of the violation even further. If proper controls were established and the results were carefully documented, it is possible that a powerful and worthwhile means of influencing drivers would emerge.

Another interesting finding is that the type of violation committed undergoes significant change as the driver continues to violate. This finding also indicates that when the forces of society fail to control the behavior of a driver, they do not fail completely. Such drivers continue to violate, but they apparently commit less serious violations (if the reader grants that violations with lower point values are generally less serious). From this initial study it is not possible to know whether changes in violation type over a period of time are the result of changes in individual drivers or changes in the constituency of the group. It is possible that changes in the type of violation committed first as compared to those committed third exist only because many drivers present in the earlier group are absent from the later (because they did not continue to violate). Thus, drivers who have three violations may, as a group, commit the same type third as they did first, but, whereas they constitute the entire group with three violations, they are only a small part of the group with one violation and may be "drowned out" by the drivers whose one violation is different in nature. These interesting side lights could be explored in further research.

The study of advisory letters in New Jersey produced several satisfying and provocative findings. In addition to the specific results, the findings indicate that the next step is to carry out a thorough study of the effects of various types of letters, sent at various levels. Such research should furnish means of obtaining the best use of letters. It is quite possible that one type of letter is distinctly better than another, and it is also possible that letters are more effective at one point level than another. It is obviously to the administrator's advantage to find the level at which the effectiveness of the letter is maximum. Of course, he must consider both the expense involved and the degree of influence on the driver. In order to carry out this type of investigation, it would be necessary to establish groups and introduce controls in advance. After-the-fact research like the New Jersey study is a satisfactory prelude for controlled experimentation but is no substitute for it. Controlled experimentation takes more time than after-the-fact research, and when time means lost lives it is tempting to take the shortest course. Indeed, in the present study it was necessary to carry out the various projects after-the-fact despite a two-year schedule. It may be disturbing to contemplate, but it is true that a five-year program of controlled experimentation will yield more usable information than could be obtained from a series of shorter, after-the-fact researches.

Conclusions and Recommendations: The advisory letter appears to prevent violations in many drivers, delay violations in some and have no effect on others. It influences more drivers than it fails to influence, however, and should be used universally as a part of driver improvement. Advisory letters should be studied further by systematic experimentation in order to improve the contents of the letter and determine the best level at which letters should be sent. Also worth study is the idea of a follow-up letter which is, in effect, a commendation. Such a program of study would permit improvement of the effectiveness of advisory letters, and even slight improvement will affect additional thousands of drivers on a continent-wide basis.

Effects of Action at Suspension Level

When a driver's record reaches a certain level of severity, the department ordinarily initiates action which leads to suspension of the driver's license or takes other appropriate action. In New Jersey, drivers who accumulate 12 points in three years are sent a notice stating the department's intent to suspend. The notice informs the driver that he may request a hearing, and if he does so the suspension is stayed pending the outcome of the hearing. If the driver does not request a hearing, suspension goes into effect shortly thereafter. Many drivers request a hearing, and after the hearing their license may be suspended or they may be placed on formal or informal probation. Presumably, each action has its own ability to improve the driver's behavior, and each driver has his own individual characteristics which make him more responsive to some kinds of action than others.

As was the case with advisory letters, little research has been conducted to determine the influence of departmental action. With the exception of the Texas study cited earlier in the article, most previous studies of departmental action have been carried out after-the-fact. A study carried out by the California Motor Vehicles Department compares drivers before and after an interview. A very old study done in Wichita, Kansas, compares the subsequent records of drivers receiving various actions. Both studies are difficult to interpret because they lack a control group. The latter, however, has a number of groups whose members received one of several kinds of action (arrests, warnings, traffic violator schools, etc.). The Wichita study showed that fewer drivers violate subsequent to attending a traffic violator school than subsequent to an arrest or warning but does not deal with comparability of the groups prior to action.

In the present investigation, seven groups are studied including six experimental groups and a control group. These groups, separately and in combination, are compared with respect to frequency and type of violation committed before and after departmental action. The groups are defined in the following outline:

I. Drivers Sent Hearing Notice At 12 Points (experimental group-total 4,600 drivers).
   - A. Ignored notice
     1. received short suspension (default-short suspension group includes 475 drivers).
     2. received long suspension (default-long suspension group includes 625 drivers).
   - B. Attended hearing
     1. were suspended
        - a. short suspension (hearing-short suspension group includes 1,750 drivers).
        - b. long suspension (hearing-long suspension group includes 675 drivers).
     2. were placed on probation
        - a. formal probation (formal probation group includes 825 drivers).

b. informal probation (informal probation group includes 825 drivers).

II. Drivers Not Sent Hearing Notice At 12 Points (control group-total 3,100 drivers).

A. Have subsequently received action (case file sample-total 1,175 drivers).

B. Have not subsequently received action (master file sample-total 1,950 drivers).

There were a number of sampling problems connected with obtaining an adequate control group and they are discussed in Appendix One.

Comparability of Groups Prior to Action: In order to be able to interpret post-action differences between groups, it is necessary to have some idea of their pre-action similarity. The first indication of group comparability is the average number of points each had when they reached suspension level. The combined experimental group had an average of 13.20 points when action was taken and the combined control group had an average of 13.04 points when they became eligible for action. The difference is less than two-tenths of a point with the control group having slightly the better record. Though such a slight difference has little practical significance, the difference is statistically significant[8] due to the extremely large samples. Since, however, the control group shows a slightly better prior record than the experimental group, the difference is in the least objectionable direction, because in order to show subsequent superiority the experimental group must improve even more than the control group.

When the experimental groups are broken down according to type of action received, a number of interesting differences is seen. Table 6 shows that the average points at action level vary from 12.71 per driver in the informal probation group to 13.74 for drivers in the hearing-long suspension group. Virtually all of the measures are significantly different from the control group and from each other. It seems obvious that the number of points on a record is at least one of the factors determining the type of action taken. Drivers given long suspensions at the hearing had an average of 13.74 points and those given short suspensions had an average of only 13.18 points. These two figures differ to a highly significant degree. The same trend holds true for those who defaulted and were given short or long suspensions. It thus appears that, from a large population of drivers, those with more points tend to get longer suspensions. To carry the speculation further, it appears that drivers placed on informal probation may have received that action because as a group they had fewer points (12.71). The group placed on formal probation, however, had an average as high as some of the groups receiving suspension. Formal probation represents a compromise between the severity of their record and extreme need for the license and is granted at the discretion of the hearing officer.

The differences in the prior point level of the groups probably do not bias the data, because the subsequent violation record does not appear to be related to the specific point level reached before action was taken. Several groups, separately and in combination, were divided according to the point level at which action was taken. It was found that drivers received the various actions at 12, 13, 14, 15, or 16 points. There were no trends observed in the proportion violating subsequently. Those suspended a given length of time at 16 points violated no more or less often than those suspended the same length of time at 12 points. Another comparison that is relevant to comparability of the groups is the average date that they reached action level and this information is also in Table 6. The month in which the average case in each group reached action level falls within the eight month period beginning January, 1956. This means that gradual changes in such factors as enforcement policy should affect the groups to a similar extent. Each month listed in the table is an average, and the drivers within each group are scattered about the average. Note that both of the

[8] Critical ratio 5.52.

TABLE 6.—AVERAGE POINTS AT THE TIME OF ACTION FOR SEVERAL EXPERIMENTAL GROUPS AND THE CONTROL GROUP

	COMBINED EXPERIMENTAL	CONTROL	DEFAULT *Short susp.*	DEFAULT *Long susp.*	HEARING *Short susp.*	HEARING *Long susp.*	PROBATION *Formal*	PROBATION *Informal*
Average points	13.20	13.04	12.73	13.56	13.18	13.74	13.56	12.71
Standard deviation	1.29	1.13	1.03	1.30	1.23	1.41	1.39	1.22
Average month case reached action level	——	June 56	Jan. 56	Apr. 56	Jan. 56	Apr. 56	Aug. 56	Mar. 56
Number drivers	4,516	2,760	487	599	1,735	651	230	814

short suspension groups have an earlier average date than either of the long suspension groups, reflecting perhaps a recent tendency to impose longer suspensions than has been the practice in the past.

A final check on the similarity of the groups prior to action is made by comparing the types of violations each group committed prior to action. Table 7 shows the data for a comparison of frequencies, and violation type is significantly[9] different for the two groups. Strangely, the significance is due to one violation, that being equipment (muffler violations mostly), and without it the groups would not have differed significantly. Why the experimental group should have fewer than expected muffler violations is not known, but at any rate the over-all difference between violation types does not seem to be great or to show meaningful trends.

[9] Chi Square greater than 57.

TABLE 7.—VIOLATION TYPE BEFORE ACTION—EXPERIMENTAL AND CONTROL GROUPS

Violation	COMBINED EXPERIMENTAL GROUP *Observed frequency*	COMBINED EXPERIMENTAL GROUP *Observed per cent*	*Expected per cent*	CONTROL GROUP *Observed per cent*	CONTROL GROUP *Observed frequency*
License or restriction	195	4.2	4.4	4.5	143
Equipment	18	0.4	1.0	2.0	63
Traffic control	511	11.1	11.5	12.1	381
Passing	143	3.1	3.2	3.3	104
Lane travel	79	1.7	1.8	1.9	59
Others	109	2.4	2.5	2.8	88
Careless driving	540	11.7	12.1	12.7	401
Speeding	2,754	59.8	58.5	56.6	1,780
Reckless driving	184	4.0	3.5	2.7	86
Leaving scene	73	1.6	1.5	1.3	41
Total	4,606	100.0	100.0	99.9	3,146

$\chi^2 > 60.0$ df = 9 $p < .01$

The prior records of the groups are dissimilar in some respects. Apparently one reason for their difference is that action was determined on the basis of the prior record. The size of the pre-action differences is small, however, and the nature of the differences does not suggest a predisposition of one group or the other to show a certain kind of subsequent record; therefore, the groups are thought to be sufficiently similar to permit interpretation of subsequent differences as being a product of the action received.

Differences Subsequent to Action: There are many large and reliable differences among the subsequent violation records of groups receiving various actions from the department. Table 8 lists several comparisons, and many more may be computed from the information therein. The second row of figures shows the proportion of drivers in each group who violated subsequent to action. These figures, however, cannot be compared directly, because they are based on substantially different periods of observation. Note, for example, that the observation period for group three (the default-long suspension group) is less than ten months, and the period for group four is more than 15. While the proportion of drivers violating is almost the same (.216 against .207), it is obvious that a given proportion represents worse driving if a shorter time period is involved. In view of the variation in time periods, it is necessary to correct the original proportions to account for this variation.

The observation period for each group is determined by computing the month in which the average case reached action level and finding the length of time between that month and September, 1957. September, 1957 is the month in which information was gathered and, since entries on the record subsequent to that time are not included, that month is the cut-off point of the study.

Table 6 showed that the average case in the hearing-long suspension group was acted on in April, 1956. These drivers, then, could have committed a violation subsequent to action any time from April, 1956 to September, 1957, a period of about 16¾ months. This is not, however, the figure indicated as the period of observation in Table 8. The reason for this is that the drivers involved actually were not able to drive all of that time, because for at least part of the time their license was suspended. To obtain the final figure for the period of observation, the length of time without a license is subtracted from the first figure. For this particular group the length of time without a license was 6.8 months (this figure includes the actual suspension plus any delay in reinstatement due to difficulties in securing insurance, etc.). Subtracting 6.8 from the original period gives 9.87 months (shown in Table 8), which was the period during which the drivers could drive legally. These computations bear the implicit assumption that driving while under suspension is negligible.

Before comparison is made the original proportions are adjusted by increasing or decreasing the value to a 12-month period. Thus, if 30 per cent of the drivers violate in 15 months, it is assumed that two per cent violate each month and that the 12-month figure would be 24 per cent. It is difficult to decide which constitutes the least objectionable alternate, that of having varying periods of observation or that of altering the proportions to correct for time. Though the time-corrected proportions are the basis of the comparisons listed, the original proportions are included and may be compared. The significance of most of the comparisons is the same regardless of which of the sets of data are used, but there are a few important differences.

The most striking features of the data as analyzed are the large and reliable differences between the control group, the combined suspension groups, and the combined probation groups. The control group shows 37 per cent with further violations, the combined suspension groups 22 per cent, and the combined probation

TABLE 8.—PROPORTIONS OF DRIVERS WITH VIOLATIONS SUBSEQUENT TO VARIOUS KINDS OF ACTION

<table>
<tr><th></th><th>CONTROL</th><th colspan="2">DEFAULT</th><th colspan="2">HEARING</th><th colspan="2">PROBATION</th></tr>
<tr><th></th><th></th><th>Short susp.</th><th>Long susp.</th><th>Short susp.</th><th>Long susp.</th><th>Formal</th><th>Informal</th></tr>
<tr><td>Number drivers</td><td>2,769</td><td>489</td><td>628</td><td>1,752</td><td>694</td><td>232</td><td>826</td></tr>
<tr><td>Proportion with subsequent violations</td><td>.451</td><td>.182</td><td>.216</td><td>.207</td><td>.291</td><td>.116</td><td>.145</td></tr>
<tr><td>Average length post-action period</td><td>14.80</td><td>11.19</td><td>9.83</td><td>15.13</td><td>9.87</td><td>12.36</td><td>17.39</td></tr>
<tr><td>Proportion with subsequent violations-time corrected to 1 year period</td><td>.366</td><td>.195</td><td>.264</td><td>.164</td><td>.354</td><td>.113</td><td>.100</td></tr>
<tr><td></td><td></td><td colspan="2">.234</td><td colspan="2">.218</td><td colspan="2">.103</td></tr>
<tr><td></td><td></td><td colspan="4">.223</td><td></td><td></td></tr>
<tr><td>Group number</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr>
</table>

COMPARISONS

Groups	*Critical ratios*	*Groups*	*Critical ratios*
1 vs 3	4.64*	4 vs 6	2.04
1 vs 5	0.54	6 vs 7	0.58
1 vs 6	7.78*	1 vs 2-3-4-5	4.07*
2 vs 3	2.71	2-3-4-5 vs 6-7	8.63*
2 vs 4	1.61	2-4 vs 3-5	10.00*
3 vs 5	3.53*	2-3 vs 4-5	1.07
4 vs 5	10.27*		

* $P < .01$

groups 10 per cent. These differences are statistically significant and are considerable in size. Drivers not receiving action show a much higher rate of violations than do drivers receiving some kind of driver improvement action. In addition to these findings, the usefulness of probation as a driver improvement tool is confirmed. This is shown by the fact that significantly fewer drivers violate after probation than after suspension. It is impossible from these results, however, to know whether the superior showing of drivers on probation is due to inherent properties of that driver improvement method or the fact that the hearing officer is a good judge of the type of driver that will respond to such handling.

The results also show that more drivers violate subsequent to a long suspension than to a short suspension. On the basis of available information it is impossible to know whether these results are due to the length of suspension itself or to the fact that the worst drivers are given the longest suspensions, and after reinstatement they continue to be the worst drivers. Another comparison indicates that there is not a significant difference between the proportions violating after a suspension by default or a suspension after a hearing. This is hardly a confirmation of the wide belief that it is better to suspend a driver after talking to him than to suspend him by mail. No firm conclusion can be reached, however, because it happens that the prior record of the group requesting the hearing was slightly worse than those defaulting (taking average points as the criterion), and the proportion violating subsequently was slightly greater. There is no indication that subsequent record is dependent on number of prior points, but it is not within the power of this study to settle this important question. A similar set of circumstances is encountered in comparing formal and informal probation. The difference between the proportion violating after formal or informal probation is not significant; however, those placed on formal probation had worse prior records (higher average points) and may, therefore, have shown more improvement, even though the results do not indicate that possibility.

As in the analysis of advisory letters, there is a possibility that receipt of action has some influence even on drivers who violate subsequently. Drivers who violated subsequent to action in the combined experimental groups were compared to control group drivers with subsequent violations with respect to the lapse between receiving action (or reaching action level) and the next violation. It was found that experimental group drivers who violated subsequently waited an average of 7.6 months before violating. Control drivers who violated waited an average of 6.3 months. The difference was significant.[10] Note that the lapse for both groups is shorter than the comparable lapse at the eight point level, though the size of the difference is the same.

Changes in Type of Violation: Significant changes occur in the type of violation committed as the sequence of violations develops. Changes occur regardless of the action received and, in fact, occur *whether or not* action was received. The basic nature of the change seems to be that drivers who have several violations tend to commit less severe violations as time goes on. An indication of this decrease in severity can be obtained by comparing the points per violation early in the violation history with the points per violation later. Table 9 shows that there is a decrease in points per violation in all cases which presumably indicates a decrease in severity since the worst violations carry the most points. Remember that the values of points per violation were obtained in an algebraically incorrect manner, and one should not expect too much from them.

The indicated decrease in severity is confirmed by an analysis of the types of violations committed early and late in the violation history of the driver. Speeding and reckless driving constitute a much lower

[10] Critical ratio greater than 4.0.

TABLE 9.—POINTS PER VIOLATION BEFORE AND AFTER ACTION FOR SEVERAL GROUPS

	CONTROL	DEFAULT		HEARING		PROBATION	
		Short susp.	*Long susp.*	*Short susp.*	*Long susp.*	*Formal*	*Informal*
Points per violation before action	3.68	3.86	3.74	3.77	3.82	3.72	4.54
Points per violation after action	3.50	3.53	3.44	3.35	3.28	3.58	3.26

than expected percentage of the total *after* action. Other violations have to replace those that decrease—the total still has to add 100 per cent—and increases are observed in such things as violations of traffic controls and careless driving violations. Table 10 shows changes in frequencies for the probation and suspension groups, and these changes are typical of those in the other groups.

SUMMARY OF FINDINGS

The total experimental group was found to differ slightly from the control group in number of points prior to action, but it was the experimental group which had the more severe prior record, which makes their subsequent superiority even more significant. When the experimental group was broken down according to type of action taken, the groups were found to differ from one another in a significant degree indicating that the kind of action taken was partly dependent on the point score at that time. Drivers given long suspensions, for example, had more points as a group than drivers placed on probation. The comparability of groups prior to action was further examined by calculating the month that the average case reached action level, and all groups fell within the period January-August, 1956. Also, the frequency of various violations committed prior to action by each group was similar except for a curious, unexplainable difference in equipment violations.

Many very significant changes in the record followed the receipt of driver improvement action. Far fewer drivers placed under suspension violated subsequently than did drivers in the control group. Also, drivers placed on probation showed a better subsequent record than those suspended. Fewer drivers given short

TABLE 10.—FREQUENCY OF VARIOUS TYPES OF VIOLATIONS BEFORE AND AFTER ACTION

	COMBINED EXPERIMENTAL GROUP BEFORE ACTION (1ST & 2ND VIOLATION)			COMBINED PROBATION GROUP AFTER ACTION	
Violation	*Observed frequency*	*Observed percent*	*Expected percent*	*Observed percent*	*Observed frequency*
Traffic control	1,005	11.4	11.6	19.2	29
Other	728	8.3	8.3	7.9	12
Careless driving	1,104	12.6	12.7	17.2	26
Speeding	5,566	63.4	63.2	51.7	78
Reckless driving	372	4.2	4.2	4.0	6
Total	8,775	99.9	100.0	100.0	151

$\chi^2 = 13.5$ $df = 4$ $p < .01$

	COMBINED EXPERIMENTAL GROUP BEFORE ACTION (1ST & 2ND VIOLATION)			COMBINED SHORT SUSPENSION GROUP AFTER ACTION	
Violation	*Observed frequency*	*Observed percent*	*Expected percent*	*Observed percent*	*Observed frequency*
Traffic control	1,005	11.4	11.4	11.3	17
Other	728	8.3	8.4	16.0	24
Careless driving	1,104	12.6	12.7	20.0	30
Speeding	5,566	63.4	63.2	49.3	74
Reckless driving	372	4.2	4.2	3.3	5
Total	8,775	99.9	99.9	99.9	150

$\chi^2 > 21.0$ $df = 4$ $p < .01$

suspension violated subsequently than was the case with those given long suspension. No differences were found in the subsequent records of drivers given formal or informal probation, but the former had a worse prior record. Similarly, drivers suspended by default were not different from those suspended after hearing, but the latter had a more severe prior record. There was no indication, however, that severity of prior record, as such, influenced subsequent violation record.

After-the-Fact Research and Its Alternate: It has been mentioned that after-the-fact research is adequate for showing gross changes, but it can only be regarded as a prelude to controlled experimentation. Some of the problems encountered in the New Jersey study would be avoided in a study in which the groups were set up in advance. Such experimentation takes time, and the very urgency of time often results in resorting to after-the-fact research. Objections may be voiced to "experimenting with human lives," but these objections can be met by proceeding carefully. No driver need be callously handled for the sake of experimentation. Even if such were necessary, it would seem that a clearer conscience would be justified than is the case now upon considering that many drivers die each year because such experimentation has not taken place and thereby has not resulted in the refinement of techniques to the degree of maximum ability to influence dangerous drivers. The task of carrying out such experimentation will no doubt be difficult, time-consuming, and expensive, but the fruits of the study could help to guide the development of driver improvement with fact instead of speculation. After-the-fact research has quite satisfactorily served a purpose in showing the grosser influences of driver improvement, but now it is necessary to deal more precisely with the elements of causation, and this can be done only through controlled experimentation.

Conclusions and Recommendations: It is most obvious that driver improvement justifies its existence by clear demonstration of its ability to influence drivers' behavior. Further investigation is recommended for the purpose of determining methods of refining techniques. In the meantime, the use of standard driver improvement practices is recommended.

Appendix 1

The New Jersey study is quite complex in terms of sampling procedures and adjustments of the control group but is fairly simple and straightforward in terms of statistical methods. The experimental design consists of comparing various groups with respect to post-action violation records after having ascertained the degree of their pre-action similarity. The dependent variables include proportion of drivers with post-action violations, length of time between departmental action and the next violation, and percentage breakdown of violation types.

Statistical Methods: The post-action records of various groups are compared through use of the critical ratio of the difference between proportions, the critical ratio of the difference between means. Comparisons of proportions always involved independent proportions and the formula shown in McNemar was used. The correction for continuity was not included in the computations because the proportions were not extreme and the Ns were large. When the test is the critical ratio, the null hypothesis is rejected when p is less than .01 for a two-tailed test.

Chi Square was used in several places to test the null hypothesis that violation types occurred with equal frequency in two situations. Most of the uses of the test in this study fail to meet one of the assumptions of Chi Square—independence. While it is true that the violation categories are independent, in several cases the groups are *not* independent because two separate observations are made on the same individuals. The situation arose because the analysis of violation types was not planned in advance (this promising line of investigation was noted rather by chance), and the data were not gathered so as to permit an adequate analysis. It is felt, however, that the statistic obtained is a *conservative* estimate of the true significance, and the results are therefore included. The results are felt to be conservative, because it seems almost certain that the variance of the occurrences *within* a sample is less than that between samples, and the distribution of Chi Square is based on such between-sample variance.

Adequacy of the Control Group: It is important that the operations be described by which the control group came into existence, because unless it can be assumed that the control group constitutes an adequate basis for comparison, the study is limited. In New Jersey, all drivers who accumulate 12 points in three years are supposed to receive action, but the fact is that some do not. The factor that causes this situation is the hearing officers' backlog. During periods of normal functioning the records of all drivers who reach 12 points are removed from the master file and designated for action. After a few weeks of such procedure, however, a surplus of cases is built up, and hearings are scheduled far in advance. This happens because the daily rate at which cases reach the 12 point level is greater than the daily rate of disposition of cases. When the surplus grows to a certain size the point system is "turned off" until the surplus is dissipated. (Note: Since release of this study this practice has been eliminated.) During the period of non-operation, no cases reaching 12 points are removed and designated for action, but rather all are left in the master file. Once the surplus is dissipated, the normal procedure of removing cases resumes.

Drivers whose records reached 12 points during the time the point system was "turned off" are not designated for action and have "escaped" action at least temporarily. Such drivers are still subject to action, however, and if they commit a further violation, they will probably be designated for action. Such a case would be handled no differently from the ordinary 12 point case except that the point total would be higher (perhaps 16 or more). On infrequent occasions it happens that a driver is passed over twice. That is, his case reaches 12 points, he is passed over, commits another violation and is passed over again.

The control group, then, includes three types of drivers, all of whom were originally passed over when they reached 12 points:

1. Drivers who reached 12 points, were passed over, and have committed no further violations (such drivers have not received departmental action).

2. Drivers who reached 12 points, were passed over, committed another violation, and were passed over again (such drivers have not received departmental action).

3. Drivers who reached 12 points, were passed over, committed another violation, and were designated for driver improvement action.

To be representative, the control group must include each of the above three types of drivers in the proportions in which they occur in the population. This poses a difficulty because the proportion in which they occur cannot be rationally determined. Considering, however, the low probability of violations and the phenomenon of regression to the mean, it can be deduced that Group One is the largest, Group Three the next, and Group Two the smallest.

The actual records of drivers in the three control sub-groups are scattered among two files. The first, containing most of the control group cases, is the master file. Because of the periods of non-operation, the master file contains the records of drivers who have reached 12 points, have been passed over, and have not received departmental action. Thus, Groups One and Two from above are in the master file. The second file is the case file, which contains records of all drivers who have received departmental action under the point system. Most of the records in this file belong in the experimental group, but a few are records of drivers who reached

12 points, were passed over, committed another violation and eventually received action. These control group cases are those in Group Three above.

The control group must include drivers who did violate as well as those who did not violate subsequent to the time action was due. If, for example, only case file drivers were used as the control group, the subsequent records would look too "bad" because 100 per cent of these drivers had subsequent violations. If only the master file were used, the control group would look too "good" because most of these drivers did *not* have subsequent violations. The obvious solution to the problem is to compose the control group of *all* drivers who were originally passed over, regardless of the file in which their records are stored, because in this manner the groups would be combined in their natural proportion. Unfortunately, this solution would require a complete search of both files, a task beyond the resources of this study. It was possible to search the entire case file and thereby to obtain the entire population of control drivers in that file, because it contained only about 10,000 current cases, but the master file, containing nearly one million cases, could be only partially searched. The partial search of the master file included 183 of 265 file drawers. Each of these randomly selected drawers was searched completely, and they constituted 69 per cent of the total file.

Unweighted combination of *all* case file components of the control group with only 69 per cent of the master file components would bias the results by making the control group appear to be worse than it actually is. In order to obtain the best estimate of the control group, two assumptions are made to compensate for these sampling difficulties. First, it is assumed that the drivers passed over are scattered evenly throughout the master file and that, therefore, searching 69 per cent of the files yielded 69 per cent of the total control drivers. (Since the drawers were selected randomly, and the records are filed alphabetically, it is assumed that the sample is representative of the total.) Second, it is assumed that the best estimate of the proportion of drivers violating subsequent to the time they reached action level (and were passed over) would be obtained by combining the master file sample (69 per cent of the population) with 69 per cent of the case file population.

The question of whether or not indicated differences can be assumed to reflect true differences hinges on the question of whether or not the control group, as constituted, actually represents an adequate base against which the experimental groups may be compared. Despite extensive sampling problems, it seems likely that the control group is usable, though not as precisely determined as is desirable. Fortunately, the magnitude of post-action differences between experimental and control groups is large, which indicates that differences would still have been significant had the control group shown a considerably better record than it did. This is mentioned because the apparent severity of the control group's record is dependent on the characteristics of the drivers who escaped action and are in the unsearched part of the master file. If the number obtained in the 69 per cent of the master file grossly under-estimates the number in the total file, the control group is better than it appears to be, thus the differences between it and the experimental group are less. This seems unlikely to be the case, however, due to the fact that *most* of the file was searched.

Treatment of the Data: Since data collection began shortly after September 1, 1957, only violations occurring before that time were considered. For each group, the month in which action was taken toward the average case was computed. The difference between this value and September 1, 1957, was the average number of months of "post-action" driving time. Of course, drivers in some groups were not able to drive all this time because of being under suspension. The average length of time drivers in each group were without their license had to be subtracted from the average post-action period. The value used was not simply the suspension time, but the *actual* time the driver was without his license. For example, in some cases a suspension of three months results in six months without a license because the driver may not be prompt in applying for reinstatement, or he may have trouble filing proof of financial responsibility. The final value, the average length of post-action driving time, was obtained by computing the difference between group means rather than the mean of individual differences.

The actual proportions of drivers violating subsequent to action was corrected to equate for length of post-action driving period. This correction was made by assuming equal proportions violating per unit time. This, of course, is not valid over wide ranges. If ten per cent violate in the year following action, it is not proper to assume that *all* would violate within ten years, because other evidence indicates that most drivers would not violate regardless of the time period. Most cases, however, were sampled over periods of 11 to 15 months with the extreme being only nine to 17 months; therefore, a time correction was made. It was assumed that if 30 per cent of the drivers violated in 15 months, then 24 per cent would have violated in one year. Such a correction is a poor substitute for experimental control of the subsequent driving period, but it is felt that this correction is a lesser evil than unequal time periods.

Finally, it is necessary to discuss one part of the advisory letter research. The data show that approximately 12 per cent of drivers sent letters have further violations, and 19 per cent of drivers not sent letters have further violations. Remember, however, that these are drivers who had a post-action violation but did not reach 12 points. Though it seems far-fetched, the data as presented do not rule out the possibility that the advisory letter makes many drivers *so much worse* that they reach 12 points (and consequently are not in the advisory letter sample). If advisory letters make drivers worse, then it should be true that a greater proportion of drivers with 12 or more points have been sent letters than those with eight but less than 12. Such a test was made, and it was found that the opposite was true to a statistically significant degree, *thus ruling out the possibility that advisory letters make drivers more likely to violate.* More drivers in the eight to 12 point range were sent letters than those in the 12 or more point range. Thus, the effects of the letter remain as indicated—desirable.

Campbell's work illustrates a high degree of methodological sophistication in the face of many practical problems. Particularly noteworthy is his plea for appropriate study to provide the basis for maximizing the effectiveness of control measures. As he implies, the unnecessarily prolonged continuance of the present situation, in which "many drivers die each year because such experimentation has not taken place," raises moral issues which should greatly concern all whose responsibility is accident prevention. The point is applicable to accidents of all types.[6, 15, 16]

Any modification of human behavior in order to prevent or reduce accidents involves two measures of its efficacy. Not only must the validity of behavior change be measured in terms of a reduction of accidents but the degree to which the behavior has in fact been changed must also be evaluated instead of being assumed. The suspension of drivers' licenses—that is, an administrative attempt to alter behavior by restricting the driving of certain individuals—is a case in point. Quite aside from the question of whether the restriction of driving by suspension of licenses results in a reduction of accidents is the point that the act of suspension may not in fact restrict driving. There apparently has been no scientifically adequate study of the extent to which drivers whose licenses are suspended or otherwise restricted ignore such official actions, although it is believed on the basis of spot checks by state police and other informal evidence that in at least some areas between 5 and 10 percent of drivers are operating without valid licenses.[17] Whatever the actual situation, the present ignorance of its dimensions indicates not only the inadequacy of knowledge of the revelance and efficacy of long-used control measures but also the hazard in assuming that administrative actions necessarily produce changes in driving behavior.

In addition, what is known leads to the conclusion that samples of drivers derived for research purposes from motor vehicle agency records may be substantially unrepresentative of the driving population because the population includes an unknown percentage of drivers who are unlicensed. Furthermore, the possibility that the unlicensed driver may have different exposure, accident, and violation characteristics from other drivers, and that accidents in which he is involved may be under-reported, must always be considered. This possibility is borne out, for example, by evidence that (1) although only "about 5 percent of [Texas] motorists are unlicensed, . . . 20 percent of drivers killed in accidents either never had a license or were driving with an expired, suspended, or revoked license [and] the percentage is even higher among apprehended violators";[17] (2) in Oklahoma "about 25 percent of motorists involved in repeated accidents and serious traffic violations lack valid licenses," in contrast with the estimated 10 percent of motorists in the same state who are unlicensed; and (3) 15 percent of Swedish drivers arrested for drunken driving were unlicensed (1948-49), and their social characteristics tended to differ from those of licensed drivers arrested for the same offense.[18] These facts further emphasize the importance of studying the real world and not merely secondary sources assumed to describe its characteristics.

THE USE OF "CONTROL GROUPS" IN HIGHWAY ACCIDENT RESEARCH

—Joel W. Novak, Robert P. Shumate

Care in the selection of control or comparison groups is particularly important in the evaluation of the efficacy of accident countermeasures, since the use of inappropriately chosen controls can lead to results that seem to justify ineffective countermeasures that may in turn delay the development and application of more effective approaches. The following selection illustrates how a choice of inappropriate controls might have led to erroneous conclusions concerning the effects of enforcement. It illustrates also the hazards in basing conclusions on secondary data sources in the absence of firsthand knowledge of the many factors which may influence the results.

EVERY YEAR LARGE AMOUNTS of time, money and effort are spent on police traffic enforcement. It is reasonable to question whether such efforts are well placed. Experimental researches into the effects of police enforcement on traffic behavior are costly and often are impeded by difficult methodological problems. Thus, most of the data available relating those two variables are anecdotal in nature. There are a few exceptions.

Shumate, using rather rigorous analytical methods, has concluded that police traffic enforcement produces no observable effect on the average speed of automobiles in a traffic stream. However, he did note that the variability of speeds about those averages was significantly affected.

Specifically, the results indicate that, as the level of enforcement on a highway was increased, the variability of speeds observed in the traffic stream was significantly decreased. Concomitant with this, it was observed that the frequency of fatal and personal-injury accidents was also significantly reduced.

One might reasonably hypothesize from this that some functional relationship is shared by the two variables, speed variability and traffic-accident rate. Further study *might* reveal that as two vehicles move in and out of close proximity to one another (i.e., overtaking), some "extra" hazard comes into being. It is this hazard which is mitigated by police enforcement through its effect on speed variability. The reader is reminded that those are merely hypotheses or conjectures suggested by Shumate's investigation. Although the hypotheses certainly merit further study, there is not yet direct evidence to support them. It was decided that before further investigation of the dynamics underlying accident-generation is undertaken, additional support of the general hypothesis relating police enforcement and accident reduction should be sought. The need for further support of that hypothesis seems particularly acute since a certain conflict appears in the research literature regarding that relationship. Michaels has concluded that no evidence could be found to support the hypothesis that traffic enforcement is related to a reduction in highway accident rate, while Shumate's results indicate that the two are directly and significantly related.

This conflict stems solely from the differences in the analytical methods employed by the two investigators, not from any difference in the data which were studied, since the same set of data was used by both investigators.

Such instances of conflict are not uncommon, and although they are sometimes disturbing to the casual reader, they are not always avoidable. The selection of a particular method for analysis should take into consideration numerous factors. Such questions as how the data were collected, what the data represent, and what questions the analyst wishes to answer with those data must be considered. If investigators differ on what are appropriate answers to those questions (and it is reasonable to expect this) it should not be surprising that we find conflicts in research results appearing in the literature.

In addition, the analysis and interpretation of highway accident data is made difficult by the presence of large, unexplained and unpredictable variations in the accident characteristics of highways under study and the lack of analytical devices for controlling those variations. The most prominent of those sources of unexplained variations can be described in time series. That is, it can be observed that the accident characteristics of a set of highways will change in time for unknown reasons and in ways which are not yet completely predictable. It can be suggested that the weather, the economy, etc. change

Reprinted, with permission, from *Traffic Safety Research Review*, 5:2:20-24, 1961, published by the National Safety Council.

in time, and hence, highway usage is affected; but precise knowledge of how those mechanisms work and the extent to which they work are not yet available. This imposes rather severe limitations on the utility of simple before-and-after studies attempting to relate an experimentally introduced variable such as police enforcement to such complexly determined data as highway accident rates.

In the past "control groups" have been used to estimate the uncontrollable variations in an experimental area, and in essence provided the basis for removing the effect that those variations brought to the data.

Whether or not such devices are valid depends on the extent to which time variations (if that is our interest) in the control group are representative of the time variations in an experimental group. It would seem that when membership in either group is determined by some random device, there would be no problem. However, when one group is selected to represent another, the selection process deserves a great deal of scrutiny.

The conflicting results that are reported by Michaels and Shumate are, in part, due to the employment of different techniques for handling the uncontrollable variations in time. Both investigators used control groups, but to different ends.

The present study was originally undertaken in an effort to resolve the conflict suggested by a comparison of the Shumate and Michaels studies and, in the process, to seek some further insight into the effectiveness of control groups in the analysis of highway accident data.

In addition, there will be included here a discussion of the statistical and experimental aspects of the analysis as it develops.

The Field Study Plan

In October 1958, the Wisconsin State Patrol placed approximately 110 miles of state trunk highway in the southern part of that state under concentrated and systematic enforcement patrol. Prior to that time enforcement on that highway had been minimal; no state personnel had been assigned to the area. Police enforcement until that time had been restricted to actions taken by country police and state police officers moving from one area to another.

Accident statistics are available for those highway segments for the yearly periods, October '56—September '57, October '57—September '58, October '58—September '59, the latter period being that one during which enforcement was applied to the "experimental" roads for the first time.

Also available were accident statistics for highways which have not yet been subjected to concentrated and systematic enforcement patrol. Accident data from the latter group of highways are used for control group information. The proportional fluctuations observed on those highways from year to year were used to derive the expected number of accidents for the set of "experimental" highways. Statistical tests were then made between the number of accidents expected and those actually observed on the experimental segments to determine if the differences noted for the yearly periods, October '57—September '58 and October '58—September '59, could be explained in terms of sampling error. Interpretations of the data were then made to determine if the use of control groups so constructed was appropriate for testing the effect of police enforcement on highway accident rates.

The "Experimental" Segments: In this report a highway *segment* is defined as a length of highway which:

1. Is located within a specifiable country,
2. Is located within a specifiable township (sub-county classification), and which is
3. Identifiable by federal and/or state highway numbers. It is a unique and easily identified length of highway, and because it incorporates into its identification the jurisdictions under which it falls, accident data for individual highway segments are easily obtained. With those features the segment is particularly useful as a unit of study. Also, it is easily seen where several segments can be combined as a single unit of study.

Twenty-one such continuous segments constituting approximately 110 miles of state trunk highway were subjected to concentrated and systematic enforcement patrol in the southern part of Wisconsin for the first time beginning October, 1958. As indicated, enforcement on those roads had been minimal prior to that time.

All of the segments selected had relatively high accident densities and carried relatively high volumes of both truck and passenger cars representing local and through traffic. For the most part, the highways have two lanes and are of relatively high design standard. The entire system is rural in nature. Speed limits for the most part were maximums for rural highways, 65 m.p.h. during the day and 55 m.p.h. at night.

The Amount of Enforcement: The method employed for assignment of enforcement personnel is referred to as "line patrol assignment." In the case at hand, each enforcement unit was assigned to a 15-mile sector. Units were not permitted to leave their assigned sectors except while in the pursuit of violators, to take violators to the nearest stations for booking, or for extreme emergencies. Patrol methods were generally varied, utilizing both moving patrol and off-street patrol. Both marked and unmarked enforcement units were used, with marked units predominating in a five-to-one ratio.

There are two common measures of traffic enforcement. One is the number of enforcement units present on a highway network during any given continuous time period. In this case one enforcement unit was assigned to each 15-miles of highway length during the 18 peak hours of traffic each day.

The second measure is generally referred to as the enforcement index, and is calculated from the equation

$$EI = \frac{C}{AF + API}$$

where,

EI = Enforcement Index.

C = The number of convictions for moving hazardous violations.

AF = The number of fatal accidents.

API = The number of personal-injury accidents.

With data supplied by the Wisconsin State Patrol, the enforcement index during the October '58—September '59 period for the set of "experimental" roads was calculated to be:

$$EI = \frac{6626}{10 + 108}$$

or

$$EI = 56.17$$

As indicated, prior to October, 1958, enforcement activity was minimal on the experimental highway segments. In terms of the enforcement index, we will assume that there were zero convictions (C); the resulting enforcement index (EI) is computed to be zero.

The "Control" Segments: Concentrated highway enforcement patrol was initiated in the state of Wisconsin in 1955. Since that time, the amount of highway mileage which receives enforcement activity has steadily grown. However, there still remains a large part of the state highway system which has never been subjected to concentrated and systematic enforcement patrol. There are, in fact, several thousand individual highway segments which remain unpatrolled. From those, two control groups were selected—a "random control" and a "matched control."

The Random Control Group: A random sample of 21 segments was drawn constituting approximately 95 miles of state trunk highway. Since the sampling technique was random, we can expect that this control group represents an unbiased picture of unpatrolled highway segments in terms of geographical and physical characteristics. Further, we can expect that the accident statistics which describe this sample of segments provide unbiased estimates of the population parameters which are characteristic of the universe of highway segments which have not been subjected to enforcement activity.

The Matched Control Group: A "matched control" sample was also drawn from the universe of segments which have not received concentrated and systematic enforcement patrol. It was intended that this set of highway segments should reveal as nearly the same set characteristics as the experimental group as possible. The matched sample was comprised of 21 continuous highway segments of approximately 125 miles in length. They are situated on the western perimeter of Madison, Wisconsin, whereas the experimental group of segments is found on the northern and eastern edges of that city. The physical characteristics of the two groups of highways are roughly comparable and, although the "matched control" sample carries less traffic, it can still be considered a relatively high-density highway system. Speed limits are the same for both the experimental and matched control segments.

The Accident Data: The state of Wisconsin requires by law that the drivers of automobiles involved in accidents which have incurred damages exceeding $100 file an accident report. The data used in this study are from reports filed for accidents which occurred on the experimental and control segments.

It is recognized that this reporting system is not always perfectly reliable; all cases are not always reported, and those reported are sometimes not complete or are in error. However, experience indicates that the statistics for the limited class of accidents involving fatalities and personal injuries are rather reliable. With this in mind, only accidents involving fatalities or personal injuries on the highway systems under study were considered.

Results and Discussion

With accident data from the experimental and control groups collected, the total frequency of fatal and personal-injury accidents was computed for the three groups for the three yearly periods, October '56—September '57, October '57—September '58, and October '58—September '59. Inspection reveals yearly fluctuations of accident rates in all groups; and, in general, the experimental and control groups of segments follow the same pattern.

Specifically, we notice that the number of fatal and personal-injury accidents was lower during 1957-58 than it had been during the previous year, and then increased in 1958-59.

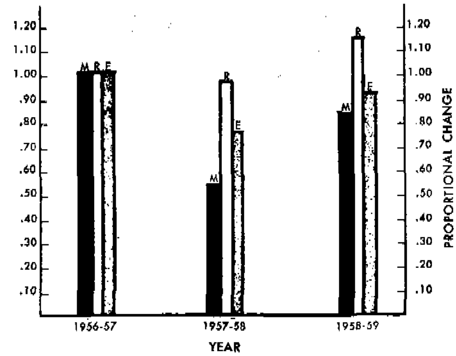


FIG. 1. Proportional changes in accident rate for 1957-58 and 1958-59 for the "matched control," "random control," and "experimental" groups, using 1956-57 as the base year.

Figure I shows the proportional change in accident frequency for the 1957-58 and 1958-59 periods for the two control groups and the experimental group. In each case the yearly period 1956-57 was used as the base year and the proportional yearly fluctuation was determined relative to the accident rates for that year. The principal questions to be answered through this investigation are:

1. Were the observed proportional changes in the frequency of fatal and personal-injury accidents the same for the experimental set of highway segments and control sets of highway segments in 1958-59, when concentrated and systematic enforcement activity was applied only to the experimental set of highway segments; and,

2. Were the observed proportional changes in the frequency of fatal and personal-injury accidents the same for both experimental and control groups in 1957-58, when neither was subjected to concentrated and systematic enforcement activity?

Table I shows the total number of accidents reported for both of the control groups for the three yearly periods, the proportional changes noted in 1957-58, and the propor-

TABLE I

	1956-57	1957-58	1958-59
FREQUENCY OF FATAL AND PERSONAL-INJURY ACCIDENTS			
Random control	40	39	46
Matched control	79	43	67
PROPORTIONAL YEARLY CHANGE IN FREQUENCY OF FATAL AND PERSONAL-INJURY ACCIDENTS WITH 1956-57 AS THE BASE YEAR			
Random control		.975	1.150
Matched control		.544	.848

tional changes noted in 1958-59, using 1956-57 as a base year.

Table II shows the total number of fatal and personal-injury accidents reported on the experimental segments for the three yearly periods and the expected number of accidents which would have occurred had this set of segments followed precisely the same proportional fluctuations for the three-year period as random control and matched control areas. Again 1956-57 was used as the base year.

Four chi-square (χ^2) tests were made comparing the observed with the expected frequency of fatal and personal-injury accidents

TABLE II

	1956-57	1957-58	1958-59
Observed frequency of fatal and personal-injury accidents.	127	96	118
Expected frequency of fatal and personal-injury accidents using proportional change in random control area to establish expectancies.		123.83 123.83	146.05 146.05
χ^2 resulting from comparison of observed and expected frequencies using random control area to establish expectancies.		$\chi^2 =$ 6.25*	$\chi^2 =$ 5.19*
Expected frequency of fatal and personal-injury accidents using the proportional change in matched control area to establish expectancies.		69.09	107.70
χ^2 resulting from comparison of observed and expected frequencies using matched control area to establish expectancies.		$\chi^2 =$ 10.48*	$\chi^2 =$ .99

* χ^2 Significant at .05 level.

in the experimental group for the yearly periods 1957-58 and 1958-59.

The first χ^2 test was made by comparing the observed with the expected frequency of accidents for the 1958-59 yearly period using the proportional changes observed in the control area to establish the expectancies. The observed χ^2 statistic was 5.19. According to tables of the distribution of χ^2 with one d.f., we can reject the hypothesis that these two frequencies are equal at the .05 level of significance. In effect, this means that there are less than five chances in 100 of our *rejecting* the hypothesis that the observed and the expected frequencies are equal *when that hypothesis is actually true.*

Looking back at the observed and expected frequencies of accidents in the experimental segments for 1958-59 then ($fo = 118$, $fe = 146.05$), we notice that the accident rate in those segments *did not* increase as much proportionally as that noted in the random control group. By virtue of our field study plan, one might be led to deduce that this phenomenon is explainable through differences in the enforcement activity which was applied to the two sets of highway segments. It is at this point that lack of complete control over the experimental situation prevents fullest interpretation of these data.

It can be argued that the use of proportional fluctuations observed in the random control sample is not valid for developing the expected frequency of accidents in the experimental group, since the latter should not be expected to have the same accident characteristics as a random sample; in fact, the experimental segments were selected for enforcement patrol because of their unusual accident characteristics. In terms of an experimental interpretation this suggests that, for one reason or another, one can not reasonably expect the proportional changes observed in the random control group to be representative of the proportional change which could be expected in the experimental group. If this were in fact the case, the differences between the observed and expected frequencies of fatal and personal-injury accidents would prove to be statistically significant irrespective of any outside influences such as police enforcement.

A preliminary test of that hypothesis can be made by comparing the differences between the observed and expected accident frequencies on the experimental segments for the yearly period 1957-58, during which time neither group was subjected to concentrated and systematic enforcement activity. Reviewing the data for that period we notice that 96 fatal and personal-injury accidents were reported on the experimental segments, whereas 123.83 fatal and personal-injury accidents were to be expected using the proportional changes noted in random control area. A χ^2 test was made comparing those two frequencies and the resulting statistic was computed to be $\chi^2 = 6.25$. With one d.f. the hypothesis that the observed and expected frequencies of fatal and personal-injury accidents are the same during the 1957-58 period is rejected at the .05 level of significance. Rejection of this hypothesis makes questionable our initial deduction that enforcement is responsible for some reduction in accident rate, and suggests that either police enforcement leads to no observable change in accident rate, or that the use of yearly fluctuations of fatal and personal-injury accidents on a group of randomly selected highway segments is not appropriate for establishing expectancies in a relatively high-density accident area.

A χ^2 test was then made comparing the observed with expected frequencies of fatal and personal-injury accidents in the experimental group in 1958-59, using the proportional changes noted in the matched control group to establish the expectancies. A χ^2 of .99 was computed and the hypothesis that the two frequencies were the same could not be rejected. The χ^2 test comparing the observed and expected frequencies of fatal and personal-injury accidents in 1957-58 was computed to be 10.48. This was sufficient to reject the hypothesis of equality at the .05 level of significance.

Interpretation of the collected results now begins to assume somewhat more meaning. Although there have been no statistical tests which clearly relate a reduction in accident rate to police enforcement, there is mounting evidence that the methods employed in the selection of the random control and matched control groups are not appropriate for controlling yearly accident variations observed in the experimental segments. With this information, the validity of any analytical method which is dependent upon the control groups so selected is questionable. We are left then with a situation wherein the effect of any experimentally introduced variable such as police enforcement, if it exists, is concealed by year-to-year fluctuations in accident rate.

This condition should not be entirely surprising. Crowther and Shumate, in describing a sampling design for fixed-point speed measurements, state that the simple two-period, two-sample, before-and-after experiment does not meet acceptable statistical standards. At the locations they observed it was noted that average speeds tended to shift over time both with and without the presence of experimentally introduced variables. Van Til also reports that this data showed before-and-after studies to be "inconclusive as far as assigning the cause of a change in accident rate is concerned since there may be many other changes in traffic or roadway which contribute to the change."

In conclusion we may state the general findings of this study suggest that:

1. The use of a random control group is inappropriate for determining the frequency of fatal and personal-injury accidents which can be expected on a set of high-accident-density highways.

2. With current notions as to what constitutes a matched control group, those data are also inappropriate for determining the frequency of fatal and personal-injury accidents which should be expected on a set of experimental highway segments.

3. Before-and-after studies do not appear to be sensitive enough to detect the effect of that level of police enforcement which was employed in this study.

The evaluation of any preventive measure requires a very carefully developed experimental design with full provision for unmanipulated controls for comparison with the manipulated groups. Although measures introduced without consideration for scientific evaluation can sometimes be appropriately studied after they have been instituted—as Campbell's work well demonstrates—frequently the timing and location of such measures preclude scientific evaluation because they eliminate appropriate controls for comparison purposes. Thus, for example, if the research issue is the efficacy of police patrols as employed on a given type of road, and if the investigator is presented with a situation in which all such roads are being so patrolled, it may be impossible for him to set up an adequate study.

Because this and many other aspects of experimental design,[19] data collection, and research logistics should be considered in advance, the research investigator must be involved as early as possible in the planning of the application of any measure which is later to be evaluated. In the illustration just cited, for example, he should be permitted to determine in advance the sections of the given type of road to be patrolled, the ways in which this is to be done and documented, and the timing to be followed. In addition, the design should provide for the observation of the accident rates of the roads or drivers of interest before, during, and after the introduction of the measure being evaluated (see Barmack and Payne, below, and reference 20).

THE LACKLAND ACCIDENT COUNTERMEASURE EXPERIMENT

—Joseph E. Barmack, Ph.D., Donald E. Payne, Ph.D.

The work that follows is a well-designed evaluation of the efficacy of a countermeasure designed to reduce accidents through the modification of behavior. It illustrates (1) the use of the specific results of well-designed accident causation research (see Barmack and Payne, Chap. 8) as the basis for the design of a prevention program of correspondingly specific focus; (2) the design, by exceptionally competent research workers, of a practical control program and their evaluation of its efficacy; (3) the importance of obtaining the full cooperation of the pertinent authorities; (4) the necessity of obtaining suitable baseline data in advance of the initiation of the measure to be studied; (5) the careful weighing of the possibility that extraneous factors contributed to the results obtained; (6) the fact, previously demonstrated by Irby and Jacobs,[20] that there may be a substantial lag between the initiation of an accident countermeasure and the onset of its effects; (7) the obtaining of a substantial reduction in accident incidence at relatively low cost; (8) the possibility that the continuing high tolls of off-duty military accidents[21] might be greatly reduced; and (9) the possibility that similarly specific and well-designed measures applied to civilian groups—for example, to high school and college students and other young adults—might be comparably productive.

THIS STUDY was part of a research program to develop and evaluate measures for the prevention of personal injury accidents among servicemen driving privately-owned motor vehicles. The first phase of the research was to identify those events occurring at the time of the accidents to which causal influence might be imputed. Biographical histories of the accident drivers were also collected. Comparable data on the driving habits and personal histories of a randomly selected group of controls were obtained.

Roughly two-thirds (64.5 percent) of the accident drivers had been drinking prior to the accident, most of them heavily; that is, six or more drinks. Among the controls the comparable figure was 5.6 percent. This figure includes the proportion of times a driver was behind a wheel after drinking to the same criterion for any part of an hour, and for multiples of hours of exposure. Of the drinking accident drivers, 60 percent had 6 or more drinks. Only 30 percent of the drinking controls reported drinking as heavily. Compared with not-drinking-accident drivers and control drivers the drinking accident drivers had significantly higher incidences of broken childhood homes, currently disrupted home life, legal and disciplinary conflict, problem drinking parents and personal problem drinking. The distributions of these characteristics showed substantial overlap among the three groups.

The problem of altering the drinking-driving pattern is a formidable one. For many young airmen even reckless driving following drinking has group tolerance. The existence of a sizable minority of problem drinkers in the accident sample suggested that posters and slogans probably would not be successful. A third (33.7 percent) of the drinking drivers involved in accidents met the criteria of problem drinking; that is, one or more previous convictions for offenses involving drinking, other than the drinking associated with the current accident.

Driver testing and license control are popular approaches, but there is no test currently available which would permit useful prediction of those who will have accidents. Biographical correlates of accidents also have low predictive effectiveness.

Since the injury-accident group was drawn primarily from among those airmen whose personal histories contained a variety of socially and emotionally traumatic experiences, an alternative to screening might be counseling. Unfortunately, there is no convincing evidence that counseling can prevent accidents. In addition, the cost of such a program would make it impractical. Some form of counseling or group psychotherapy might prove feasible if limited to extreme cases, for example, chronic violators or accident repeaters. This approach, for instance, is currently being studied by W.A. Tillmann. However, 73.3 percent of the accident drivers interviewed had experienced a lost-time automobile accident for the first time while in the service. Consequently, selective counseling (after the occurrence of an accident) would not reach the majority of airmen who are having accidents.

To sum up, neither screening nor counseling appears practical as a countermeasure for the class of accidents and drivers under consideration. Therefore, a special countermeasure was designed and instituted at Lackland AFB for preliminary evaluation.

The primary aim of the countermeasure was to undercut the favorable image that many young adults have toward "tanking up and taking off" in a car. Rather than view such behavior as an act of personal courage and daring or as a tolerable peccadillo, it was held up as an example of disturbed or "sick" behavior.

Statements to this effect, however, are not enough. It was believed necessary to back up this view with psychiatric and administrative support in order for the view to have impact.

The theme was first outlined by General Stillman, Base Commandant, in a memo-

[Reprinted, with permission, from *Highway Research Board Proceedings*, *40*:513-522, 1961, published by the National Research Council, Washington, D.C.]

randum announcing the start of the experimental countermeasure program November 3, 1958. Follow-up publicity included a series of articles in the base newspaper and special material prepared for commander's call meetings.

Administrative Support

Any airman who became involved in a lost-time injury accident while driving a privately-owned motor vehicle was, in effect, "tagged" for special attention. By having the accident, he exposed himself to a thorough review of his value to the service. The accident was not in itself considered evidence of ineffective behavior, but it set in motion a screening procedure to uncover evidence of ineffectiveness. The airman's record was investigated by his squadron commander, who requested special record reviews by the base provost marshal, the base surgeon, the base ground safety office, the airman's duty officer, etc.

When the review was completed, the squadron commander could recommend the airman for discharge if evidence of ineffectiveness was uncovered, or if not, he could refer the airman to the base psychiatrist for further evaluation. For the year in which the countermeasure was in operation, no one met the criteria in the regulations for discharge for inaptitude or unsuitability.

Psychiatric Support

Psychiatric treatment of the airmen drivers involved in lost-time injury accidents was important for two reasons: (a) As an overt indication of the earnestness of the point of view expressed in the educational program, that is, reckless driving after drinking is a sign of maladjustment; and (b) to provide psychiatric assistance to those airmen who might need it.

The circumstance of referral put the base psychiatrist in a somewhat unconventional role, a role which might be seen as threatening or punitive. It should be noted, however, that punitive action—if any were to be taken—preceded referral. There has been a growing awareness that psychiatric appraisal should be extended to persons involved in a variety of antisocial acts. For many but by no means all classes of accident drivers, this approach would be equally pertinent.

During the countermeasure experiment, the base psychiatrist had three alternate courses of action open to him. Each driver referred to the psychiatrist was evaluated as thoroughly as, in the judgment of the psychiatrist, the circumstances warranted. Following evaluation the psychiatrist could take the following actions:

1. Return the airman to duty without further recommendation.
2. Offer psychiatric assistance if the airman seemed in need of it.
3. Recommend discharge on medical grounds if appropriate.

The Educational Program

Eighteen items were prepared for dissemination via commander's call meetings, bulletin boards, and the base newspaper. Of the 18 items, one gave the details of the program. Three comprised a factual series on social drinking, alcoholism, and alcohol and the road. Three others were of a general appeal type. Eleven were part of a series called "snatches of conversation with an accident victim from the notes of an accident investigator"—these were abstracted from case histories obtained at other bases, with the identity of the victim carefully disguised. An example of one of these is the following:

> I have been court-martialled a couple of times. Once it was for AWOL. I got tired of the NCO riding me, so I took off. It was only a couple of days—they said eight. I got a summary court for that.
>
> I got married once when I got drunked up. It didn't work out; I don't know where she is now.
>
> Now it looks like this marriage is going to break up. She thinks I'm running around.
>
> I was upset that day, because my wife went off to stay with some people. We had a fight. I went to a friend's house. We killed a fifth between four o'clock and seven o'clock. Maybe I had ten or twelve shots.
>
> I was going back to the base. I was thinking about my wife. I saw the light changing from green to amber, but I thought I could make it.

> They say I hit a car coming through the intersection. They took me to the hospital, but I don't remember any of it. I woke up with a terrible headache.
> I've been in accidents before, but never one where I got hurt.
> I guess I should have been paying more attention.

One of the general appeal articles, titled "The Thin Line," began:

> A new car, a straight and empty road, so let her go. Who is there who doesn't get a charge out of seeing the speedometer edging up to 80, 90, or more, maybe? It's normal to feel a thrill as the car surges forward. But would it still be normal behavior to drive at such speeds in rain, or on an icy road at night, or on a busy highway, or through crowded downtown streets? Where do we draw the line between what is normal behavior and what is emotionally disturbed behavior?

The article then went on to point out that the normal person is one who cares not only for his own welfare, but also for the welfare of others. The tendency of young people to take larger risks than more experienced people was singled out for special attention—as abnormal behavior when the risk is almost certain to result in serious harm to someone.

The article then reported the story of a young airman who, after a beer-drinking party with some buddies, ran off a serpentine mountain road at 120 mph. The article pointed out that alcohol did not cause the accident—but that it weakened the airman's shaky controls over emotionally disturbed behavior. It was pointed out that, at Lackland, this airman would be referred to a psychiatrist to determine the nature of his emotional disturbance.

The article concluded by urging Lackland drivers to stay on the right side of the thin line between normal and disturbed behavior. Airmen were urged: "If you must drink, don't drive. If you must drive, don't drink. If you must do both, drive as if your life depended on it. It does."

The countermeasure was installed at Lackland Air Force Base for the period of one year, from November 3, 1958, to November 2, 1959. The end of the countermeasure experiment was announced by the base commander, and publicized in the base newspaper.

RESULTS

To evaluate the impact of the countermeasure, accident data were collected for a 2-year period:

1. A control period—the year preceding the installation of the countermeasure (November 3, 1957 to November 2, 1958); and

2. The experimental period—the year during which the countermeasure was in operation[2] (November 3, 1958, to November 2, 1959).

The Lackland test of this countermeasure was restricted to lost-time accidents reportable on the AF Form 122. The criterion was injury resulting in more than 24 hours time lost from duty. This class of accidents is more reliably reported than the property damage type. It is also the class of accidents with which the services are most directly concerned.

The total number of lost-time accidents for each month during the control and experimental periods was recorded by the base ground safety office.

The measure of accident reduction should include some index of exposure. There are no completely satisfactory measures of exposure available for a large population operating private motor vehicles. As a compromise, the air training command uses a rate based on the number of accidents per 100,000 man-days of exposure. This rate adjusts for variations in manpower and days of the month. The same exposure index was used in the study. There is no assurance, of course, that true driving exposure corresponds entirely to the number of personnel stationed on any base at one time.

[2] A third year of data collection was included in the original research plan. It was hoped to determine whether the base accident rate would climb to previous levels after the countermeasure was terminated. Unfortunately, support for the project terminated before this could be accomplished.

The cumulative accident rates for Lackland Air Force Base during the control and experimental periods are shown in Figure 1. The cumulative accident rates represent successive monthly averages of the rates which accumulated in the experimental and control years. The rate during the countermeasure period did not begin to diverge from the rate during the control period until after the third month of the experiment. The difference became increasingly large thereafter until June. There was an increase in rate for the next two months and a decline toward the end. By the end of the experiment there had been one-half as many accidents during the countermeasure period as during the preceding year. There were 40 lost-time injury accidents during the control period, but only 19 during the countermeasure experiment, a 52.5-percent decline.

The fact that accidents began accumulating at approximately the same rate during the first three months of the countermeasure period as they had during the equivalent period of the preceding year suggests that there was a lag between the time the countermeasure was first announced and the time it began to exert a detectable effect upon the base personnel. A similar lag was found by Irby and Jacobs in a study which tested a patrol intensification countermeasure at a military base.

The lag between application and detectable effect could be accounted for in two ways. It is possible that the spread of information about the countermeasure is an accretive process which necessitates repeated announcements (for example, the series of articles in the base newspaper and in squadron meetings), and word-of-mouth communication until a substantial majority of the population becomes aware of the countermeasure. Equally tenable is the hypothesis that the information about the countermeasure spreads rapidly but has no effect until concrete actions have been taken with specific cases. The present data do not permit a test of these hypotheses, though further research would seem justified.

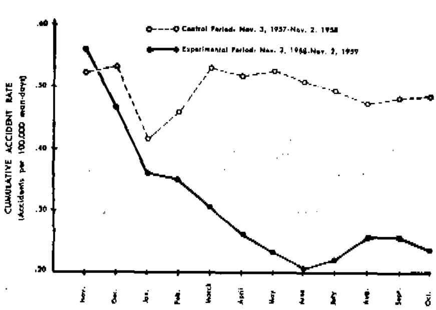


FIG. 1. Cumulative accident rate during the experimental and control years.

A statistical test of the difference in accident rate indicated that the rate during the experimental countermeasure period was significantly lower than the rate during the control period; the probability that the drop in accident rate was due merely to chance is less than one in a hundred. This value was obtained by a method analogous to the comparison of two frequencies from binomial populations given in Hald.

The trend in accident rates for the experimental and control periods (Fig. 1) represents the most conservative statement of the differences because: (a) they include the data of the first three months (that is, the lag time for the countermeasure to take effect); and (b) they are limited to the accident rate criterion. When other criteria are used, a somewhat more favorable difference is obtained.

Table 1 shows the differences for the two periods when other criteria are used, such as "military personnel injured" or "drivers injured." The percentage drop during the experimental period ranged from 50 to 60 percent. All of the differences were significant.

Having established that the differences are not due to chance, the next question was whether the extra-chance differences were due to the countermeasure or to other factors. Since the study was conducted

TABLE 1.—ACCIDENT EXPERIENCE CRITERIA FOR CONTROL AND COUNTERMEASURE PERIODS

CRITERION*	PERIOD *Control*	PERIOD *Countermeasure*	DIFF. %	p†
Number:				
Accidents	40	19	—52.5	<0.01
Total injured	54	24	—55.0	<0.001
Drivers injured	33	13	—60.0	<0.01
Rate (per 100,000 man-days):				
Accidents	0.48	0.24	—50.0	<0.001
Total injured	0.65	0.30	—53.8	<0.001
Drivers injured	0.40	0.16	—60.0	<0.001

* The criteria are not independent. Accidents and total injured (phi coefficient = +0.84), accidents and drivers injured (phi coefficient = +0.75), and total injured and drivers injured (phi coefficient = +0.75) are all related.

† Significance levels for number criteria estimated by χ^2 with theoretical frequencies equal for each period. Significance levels for rate criteria estimated by rate-method described previously.

in a natural environment it was not possible to control all relevant factors; it is possible, however, to examine whether any of them could have produced the accident reduction.

Accident Trends: The incidence of accidental injury varies from year to year in ways which cannot as yet be satisfactorily explained or predicted. Consequently, if the countermeasure experiment took place in a time or region characterized by a declining accident rate, the results might be credited to unidentified factors responsible for these general shifts. For the experimental period in question, the Lackland experience ran counter to national, state and local city trends. Figures published by the National Safety Council indicate that, for the nation, 1959 as compared with 1958 was a year of increase. The motor vehicle death rate also increased in Texas. Records of the San Antonio police department indicate that personal injury accidents within its jurisdiction increased from 184 per month to 198 per month.

The accident experience of the civilian drivers in the vicinity of Lackland might be different from the experience of military personnel for reasons quite independent of the countermeasure. Fortunately for this study there are two air training command bases in San Antonio—Lackland and Randolph. The lost-time injury accident experience of each base is shown in Table 2. At Lackland, where the countermeasure was applied, accidental injuries declined; at Randolph, where the countermeasure was not applied, accidental injuries increased, although the increase was not statistically significant.

The trend data warrant the conclusion that the decline in accidental injuries at Lackland cannot be attributed to unidentified causes affecting national, state or local accident experience.

Civilian Police Activities: State police patrol activities contain two primary classifications for recording violations: warnings and citations. State police records indicate that during the control period an average of 1,612 citations and 3,455 warnings were issued each month; during the experimental period an average of 1,667 citations and 4,385 warnings were issued each month.

Since the state police issued more warnings and citations during the countermeasure period than they had during the control period, the possibility must be considered that the decline in accidents noted among Lackland personnel might have been the result of an intensified state police patrol campaign. There are two lines of evidence which suggest that the state police campaign was unrelated to the Lackland experience, as follows:

1. If the increase in the number of

TABLE 2.—COMPARISON OF LACKLAND AND RANDOLPH AFB

DETERMINATION	BASE	PERIOD *Control*	PERIOD *Countermeasure*	DIFF. (%)	p*
Total accidents	Lackland	40	19	—52.5	<0.001
	Randolph	17	19	+11.8	0.42
Total injured	Lackland	54	24	—55.0	<0.001
	Randolph	22	23	+ 4.5	0.48
Drivers injured	Lackland	33	13	—60.0	<0.001
	Randolph	17	20	+17.6	0.32

* The significance of the difference between control period and countermeasure period was calculated using the rate-difference method described earlier. The exposure base was: control period—Lackland 8,274,973 man-days, Randolph 1,568,694 man-days; countermeasure period—Lackland 8,033,433 man-days, Randolph 1,333,688 man-days.

citations and warnings affected Lackland's experience, it should also have affected civilian accident experience—yet both violations and accidents increased. If increased state police activity were playing a major role, then as violations-issued increased, accidents should have decreased.

2. It might be argued that the state police activities affected Lackland personnel differently from civilian personnel. This would be the case, for instance, if the state police concentrated upon military personnel. However, there are two facts which suggest that this was not the case. First, if the state police had concentrated on military personnel then both Randolph and Lackland accidents should have declined. Randolph accidents did not decline. In the second place, courtesy reports (special reports prepared by the state police and sent to the base—including the names of all Lackland personnel who are apprehended committing off-base moving violations) sent by the state police to Lackland indicate that the average number of Lackland personnel who were cited for moving violations off-base declined (that is, from an average of 55 per month during the control period to an average of 51 per month during the countermeasure period).

The weight of evidence indicates that the decrease in accidents among Lackland personnel cannot be attributed to state police activities.

The records of the San Antonio police department were similar. Although the number of citations issued increased from 7,052 per month during the control period to 7,374 per month during the countermeasure period, the average number of injury-producing accidents per month also increased.

In short, the reduction in accidents which occurred during the countermeasure period cannot be accounted for by local law enforcement activities of the state or city police.

Lackland Military Police Activities: There is one activity which might affect the off-base accident experience at Lackland Air Force Base, that is, the on-base activities of the air police.

Quantitative information on air police activities were not available for the period before January 1958. Consequently, to provide comparable periods for analysis, control period data were limited to the January-October span of 1958 and experimental period data were limited to the same time span for 1959.

Shortly after the initiation of the countermeasure, a new provost marshal was appointed. He had not been present at the staff briefing in which it was requested that no major changes in operating procedure be initiated. The new provost marshal expanded the number of on-base patrols and stepped up their activity. This intensified campaign continued until the end of June 1959. It was learned subsequently that during the control period 41 violations per month were issued by the air police; during the countermeasure period 74 violations per month were issued.

The stepped-up on-base air police activity is clearly a contaminant in the evaluation of the countermeasure. But this contaminant has positive as well as negative aspects. On the positive side, if it had any effect, it was to reinforce the earnestness with which the base commander viewed airmen's behavior behind the wheel. On the negative side, it adds another dimension or attribute to a countermeasure which already has other elements that require more detailed investigation. It can be said that the program worked and worked very well in reducing lost-time injury accidents. But it is not known which attribute or combination of attributes contributed to its effectiveness. Further research on these components is clearly indicated.

Changes in Accident Reporting Criteria: It is conceivable that the apparent effect of the countermeasure merely reflects concealment by airmen of their involvement in off-base injury-producing accidents. Whether this could have played a part is difficult to determine satisfactorily. Careful inquiry of the air police, the ground safety office, hospital staff and squadron commanders revealed no evidence that such concealment did in fact occur. However, in future research it is believed desirable to include both lost time and property damage types of accidents as dependent variables and to monitor the classification procedure.

Other On-Base Activities: The base ground safety officer cooperated completely by keeping the nature and intensity of the activities of his staff during 1959 on a level comparable to that of 1958. Copies of the base newspaper were made available so that any major changes in types of personnel, pass or leave policies, or policies affecting private motor vehicles could be noted. No significant changes occurred which could account for the reduction in accidents.

Psychiatric Experience

The decline in accidents resulted in relatively few cases available for psychiatric processing. The fear that an influx of a large number of accident drivers might impose an undue burden on the limited psychiatric facility proved groundless.

None of the drivers was recommended for discharge on medical grounds. In only two cases was psychotherapy felt to be desirable.

One of the two psychiatrists to whom these cases were referred found that the circumstances of the referral contributed to somewhat defensive attitudes on the part of the drivers. The other psychiatrist did not have this experience. But even when defensiveness was experienced the psychiatrist expressed the belief that it could be dissipated during the initial few hours of psychotherapy. Both psychiatrists who took part in the study concluded that the doctor-patient relationship had not been interfered with significantly by the referral procedure. The administrative and psychiatric burden was relatively light.

Also in urgent need of objective evaluation are the safety education courses directed at school children, which rank among the most expensive of accident prevention measures. Millions of student and faculty hours that could be used for other purposes are thus employed on the grounds that the importance of childhood accidents justifies the expenditure. This argument, however, overlooks the fundamental question of the degree to which these courses actually prevent (or favor) the occurrence of accidents among the children subjected to them, a question that remains largely unexplored. This is not to suggest that these courses be eliminated but, rather, that the time is long past when they should have been scientifically evaluated in terms of method and curriculum instead of being justified on the basis of "obvious common sense." Although there is chronic controversy between the proponents of safety education and those who object to "frill" courses, neither group has presented evidence adequate to support its contentions.

In this area, as in those we have already discussed, it is essential to recognize the distinction between success in the communication of facts and success in the modification of pertinent behavior in the real world. Here again there can be little assurance, in the absence of appropriate research, that common sense, traditional wisdom, and folklore are an adequate or even desirable basis for attempts at prevention. Furthermore, experience with other forms of "health education" has shown that efficacy may be very low unless (1) the specific problems and target groups are carefully identified; (2) these groups are approached through the channels of communication that they customarily use for receiving information; and (3) the health practice being advocated is presented in such a way that the target groups come to regard it as valuable to themselves in their own frames of reference.

In addition, safety education often is not taught as a serious subject, as it might be if it were embodied in courses in physics and biology. Instead, it is commonly attached to courses in physical education or "hygiene." There are obvious implications as to the professional qualifications of the teachers of such courses and their preparation, attitudes, and involvement in so complex a field. Here too, as we emphasized in Chapter 1, the fundamental problem is not some difficulty inherent in the causation and prevention of accidents but, rather, the failure of society to choose to bring its resources to bear on the problem. Until these several issues are faced and thoroughly explored, it is unlikely that the most effective approaches to safety education will be identified or applied.

The driver education courses now given widely in secondary schools and by private groups suffer from many of the same shortcomings. At present millions of dollars and faculty-adolescent man-hours are annually invested in courses that vary widely in content and method—a huge investment in view of the dubious evidence as to its efficacy. Indeed, as McFarland and Moore (Chap. 8) have pointed out,

> Although it is clear that trained drivers initially have better records than the untrained, the studies made thus far do not conclusively indicate that the better record is solely the effect of training. . . . One study of teen-age drivers who were given psychological tests prior to their being old enough to learn to drive has indicated that those who later elected to take driver training were indeed different in several personality traits and adjustment tendencies from those who later rejected driver training and learned to drive in other ways.[22]

Such findings underscore the need for further research to determine not merely whether "driver education" is or is not effective but, more precisely, whether specific syllabuses, methods, and techniques are more effective than others. Positive research results would justify the standardization of courses and their elevation from their present marginal status in the high school curriculum. Negative findings, although they would imply the need for further experimentation, would support the claims of those who regard current courses as a waste of the student's instructional time. For these reasons it is hoped that the long-term studies of driver education being conducted by Dr. J. J. Conger and his colleagues at the University of Colorado will answer some of the major questions in this area.

Although many accidents are initiated by nonbehavioral factors, there have been few adequate evaluations of measures directed at such factors. In the vehicular accident area alone, for example, there apparently has been no scientifically adequate evaluation of the efficacy, in preventing accidents, of the vehicle-inspection programs enforced by various states.[23, 24] As another example, the effects on accidents among the elderly of improving their medical care and the environments in which they live also need to be studied so that preventive programs concerned with such factors can be based on appropriate evidence rather than guesswork, as is largely the case at present.† Although certain accident research methods have reached a considerable degree of sophistication, they are not often used for program design and evaluation, a deficiency which must be corrected if preventive measures are to be soundly based and available resources used with maximal effectiveness.

† For an introduction to some of the factors that should be considered in designing and evaluating programs directed at accidents among the elderly see Sheldon and Haddon *et al.* in Chap. 4, the discussion of injury thresholds in Chap. 9, and references 6, 7, and 25.

RELATION OF TRAFFIC SIGNALS TO INTERSECTION ACCIDENTS; CASE HISTORIES FROM MICHIGAN SIGNALIZATION EXPERIENCE

—J. Carl McMonagle

Environmental modifications that alter the distribution of hazards may influence accident incidence for better *or worse*. In the case of traffic control, as this selection shows clearly, environmental modification appropriate for one site may be inappropriate for another. This emphasizes once more the necessity of studying the real world in which accidents occur, of avoiding unsupported generalizations, and of evaluating the measures that are used.

THE UNPRECEDENTED EXPANSION of vehicular volumes since the war is putting the existing highway structure to a tremendous test and is revealing glaring deficiencies created by enforced neglect during the war years. The public is clamoring for relief and its demands can be met only by the construction of new facilities and by improving the operation of the old.

The traffic engineer has the responsibility for operating traffic on the plant as it exists. Regardless of the inadequacies, he must keep traffic moving as efficiently and safely as possible. In view of the importance and difficulty of his job, he not only must analyze his problems thoroughly but must examine the tools of his trade continuously and critically. This paper presents early results of some investigations in Michigan of one class of these tools—traffic control signals.

The conditions demanding rural traffic regulation and protection are, for the most part, concentrated in and about intersections. A recent study* of accident experience on a heavily traveled suburban trunkline in Michigan revealed that 70 percent of all the accidents occurred on the 30 percent of the route in intersection areas. This study had particular reference to the influence of roadside features in accident causation, but since roadside establishments cluster about practically every important intersection, the results are entirely characteristic.

It appears, then, that the requirements for the operation of traffic between intersections are understood and are not too hard to provide. But where traffic streams intersect, the problems of efficient, safe movement are multiplied. The difficulties inherent in this situation have long been recognized, and certain standard methods of intersection traffic regulations and protection have been developed and used.

Where traffic volumes are low, stop signs halt entering traffic for a convenient opportunity to cross or turn onto the main highway. Where both traffic streams are extremely heavy, grade separations permit uninterrupted movement on and interchange between both routes. But the real problems arise at intersections with volumes too large for stop signs to be effective, and yet not large enough to warrant a costly separation structure.

These intermediate locations constitute a twilight zone in which opinions as to proper traffic-engineering procedures jostle as violently as the vehicles themselves and some-

* See "The Effect of Roadside Features on Traffic Accidents," Chap. 4. Eds.

Reprinted, with permission, from the *Highway Research Board Bulletin*, *74*:46-53, 1953, published by the National Research Council, Washington, D. C. Photographs have been omitted.

times quite as unreasonably. Stop-and-go signals and flashers are the accepted means of traffic control at these intersections. The basic cause of the conflicts of opinion is a widespread confusion, and even ignorance, regarding the function and proper use of the first of these signals.

The signal salesman of the past offered the stop-go signal as a panacea for all traffic ills, and since safety was a condition sought by his customer, he labeled it a safety device. The public generally still holds to this belief.

As a matter of fact, the stop-go signal is nothing more than a regulator valve. Properly applied and operated, it can produce orderly flow in two intersecting traffic streams, and traffic safety is an important by-product of traffic order. But order, and not safety, is the functional purpose of signalization; neither orderly movement nor its by-product, safe movement, will be obtained unless the signal is applied to the right conditions in the right way.

Some years ago the Michigan State Highway Department began to suspect that signal installations do not necessarily end accidents. It appeared that what they really do is to alter traffic behavior and, for that reason, produce a different accident pattern. It was noted, moreover, that often accidents actually increased after signalization. In view of these experiences, it was deemed necessary to conduct a probing study of the whys and wherefores of accidents as pertaining to traffic signals.

With this thought in mind, a Traffic Analysis Section has recently been established in the Planning and Traffic Division. In addition to its function of determining needs for various types of traffic control and design, the new section also evaluates the efficiency and safety of such controls after placement. Although its work is only started, certain facts have already revealed themselves.

In the first place, it has become apparent that composite or mass grouping of accident data from many locations means little or nothing when applied to traffic-signal problems, because each location has individual conditions and characteristics which are basically important to an understanding and solution of its particular traffic problem. It was finally decided to isolate each case and diagnose each new signal installation by a before-and-after study of its accident experience. Michigan hopes, by this process, to establish a trend which may be used in the future to predict accident potential and to gain information helpful in determining the type of signal installation most conducive to safety.

The examples below are drawn exclusively from rural or suburban areas, because the conditions for which signals are used and under which they operate are radically different on congested city arteries from those that exist on isolated trunkline intersections in relatively open country. For one thing, signal control is a part of the process of movement through a crowded city district and drivers are conditioned to it. But usually the signal at a rural intersection is an exceptional feature of the rural trunkline and, as such, is often dangerously inconsistent with the driver's expectation of an unobstructed roadway for high-speed travel.

The remainder of this discourse will be centered on pictures and charts of carefully selected intersections. They were selected to prove that stop-go signals are not cure-alls. They are not presented as damning evidence against all such signals but as contributions to a record which, it is hoped, will grow in usefulness as it becomes more complete. The figures include a view and a before-and-after collision diagram of each intersection.

The first case is the intersection of M 112 with M 56, commonly known as Belleville Road. M 112 is the Detroit Industrial Expressway, but in this particular area it has lost its expressway characteristics and has intersections at grade, even though it remains a divided highway with controlled access.

The figure shows M 112 to be a four-lane divided highway with the medial divider having a width of 32 ft. at this point. Belleville Road is a well-developed asphalt-surfaced road. There are dual signal heads for both directions of M 112. Belleville Road has one signal head for both the near slab and far slab of M 112 in both directions. Signal head visibility is, therefore, better than average and not a contributing factor for the accident pattern to be discussed.

The volume of traffic on Belleville Road, plus the large number of angle collisions, indicated the need of a stop-and-go traffic signal installation. Consequently this project was completed on February 4, 1949. The accident study conducted over a 2-yr. period before and after the installation shows 30 accidents before and 40 accidents after the installation (see Fig. 2). Angle collisions were reduced from 16 to 8, while rear-end collisions were increased from 2 to 17. There are over twice as many rear-end collisions between westbound vehicles on M 112 as eastbound. An explanation is that motorists are coming out of the expressway section from this direction and are acclimated to high vehicle speeds and no cross traffic. Their time-speed sense apparently fails them when faced with the necessity of stopping for a red signal.

Another interesting fact to be noted is that most of the angle collisions in both the before and after periods occurred when vehicles from Belleville Road collided with vehicles on the far slab of M 112. This same condition has been observed at other locations, and we are running some observation and accident studies at certain selected intersections to determine a method of correcting this condition with signalization. We are providing a delayed far-side green, which means that a motorist can enter a divided highway and have a better than average chance of crossing both slabs, since the near signal will go red first, followed a short time thereafter by the far signal. We are doing this under the assumption that some drivers, when crossing a divided highway, will attempt to negotiate the entire crossing rather than stop in the medial area in case the green interval expires.

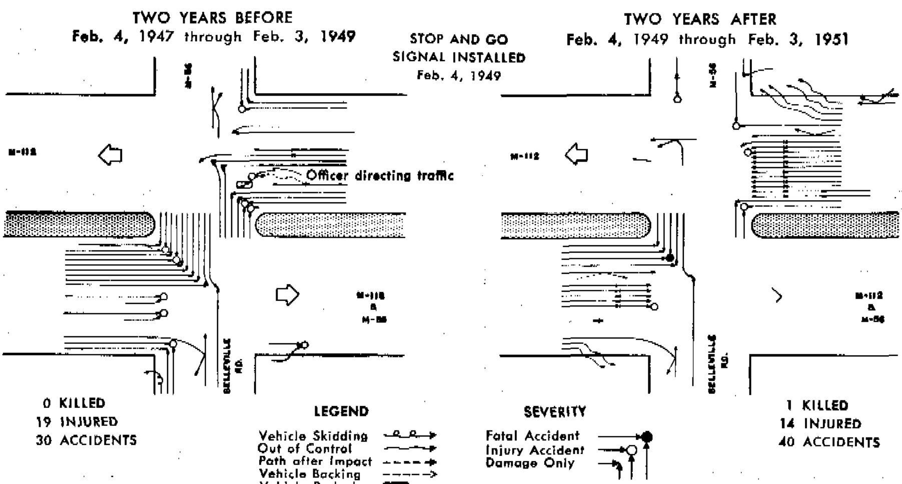


FIG. 2.

Since we have only conducted tests on this particular operation for a short period, before-and-after accident experience is not available, but operational-wise the plans seem to be obtaining good results.

Figure 3 [omitted] shows a view of US 10 and US 23 at the intersection of Clio Road, which is located in a rural area north of Flint. The view was taken looking north from Clio Road. The highway has dual signal heads, while Clio Road has single heads. The trunkline is a four-lane undivided

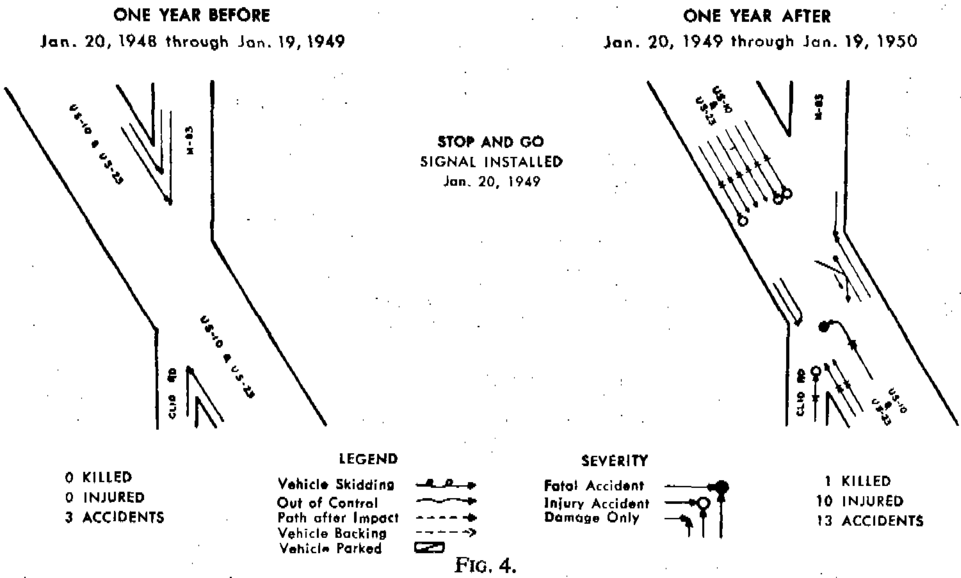


FIG. 4.

highway, while Clio Road is a two-lane concrete pavement. Clio Road north of this location is also state trunkline highway M 83 which serves a prosperous agricultural area.

Vehicle volumes on US 10, US 23 were very high, while the Clio Road volumes were above signal requirements. The collision diagram shows 3 accidents before and 13 accidents after installation (see Fig. 4). The vehicle speeds are high throughout this area, which accounts for the increase in rear-end collisions.

Maple Road runs west from Birmingham and intersects US 24 in a rural area although there is intersection development. Maple Road carries considerable traffic. The view shows US 24 as a four-lane highway; during the time covered in the accident study, US 24

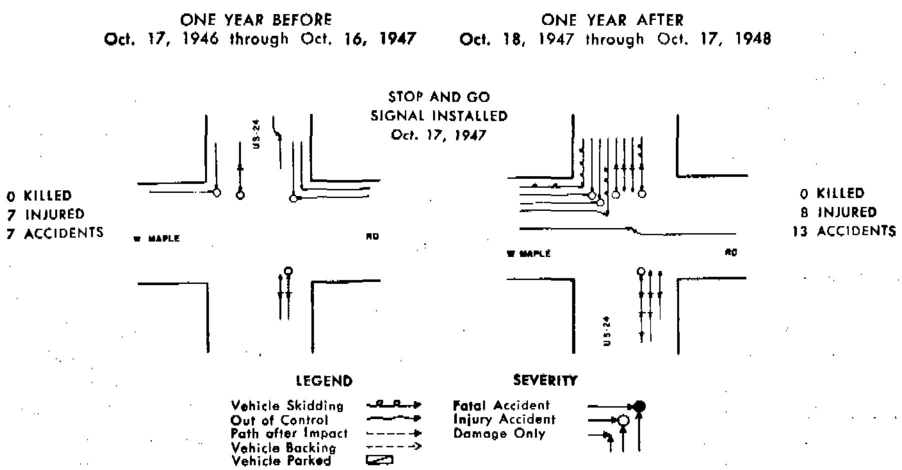


FIG. 6.

was a two-lane highway carrying near capacity vehicle volumes for such a roadway.

The collision diagram (Fig. 6) shows an increase in accidents from 7 to 13 after the installation of the traffic signal. Vehicle speeds are high on US 24, which again accounts for the increase in rear-end collisions. . . .

a business route going north into Adrian, while the bypass route continues west from the intersection.

The collision diagram (Fig. 10) shows an identical pattern in both the before and the after patterns. The number of accidents was constant at eight, while the injuries were reduced from seven to one.

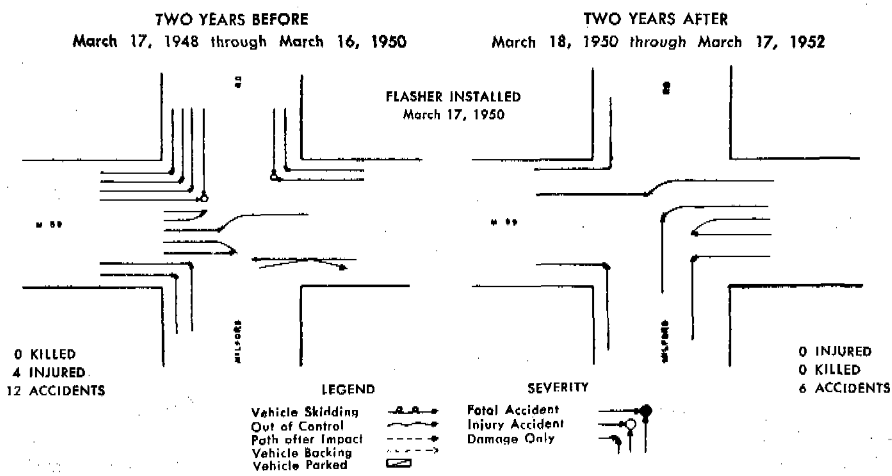


FIG. 8.

Figure 7 [omitted] shows the intersection of M 59 with Milford Road. A small village named Highland lies to the south of this intersection on Milford Road. A number of angle collisions were occurring due to the suddenness with which motorists found themselves upon M 59. It is a two-lane concrete pavement with moderate vehicle volumes traveling at a high average speed.

The collision diagram (Fig. 8) shows that positive results were gained by the installation of a flasher, since accidents were reduced from 12 to 6, while injuries were cut from four to zero in the 2-yr. periods before and after installation.

The intersection of US 223 with US 223 (Business Route) southeast of Adrian makes an interesting intersection to study, since it operates from a traffic standpoint like a T intersection, even though it has four approaching legs physically. US 223 splits into

SUMMARY

From the example shown of accident experience before and after installations of stop-go and flasher signals, it might be concluded that we should either abandon the stop-go installations or else improve our installation methods. It might be concluded that flasher signals should be substituted for present equipment at the existing stop-go locations. But the problem is not that simple, and is not to be solved by any easy answers.

To begin with, abandoning use of the stop-go signal under existing conditions would leave a wide and dangerous gap in the traffic engineer's array of control devices. Also, we believe that Michigan standards of signal installation are fully in line with the best accepted modern practice. And finally, the flasher signal itself has its own limitations. But there are still other factors to be considered:

All of these selected intersections have traffic-volume characteristics which place them in that intermediate range referred to earlier in this paper; traffic is too heavy for stop signs but does not quite warrant a separation of grades. However, there is a wide variation of volumes represented by these 13 selected intersections, ranging all the way from 1,100 to more than 14,000 vehicles per day on the main highway. The four flasher-equipped locations had average traffic of 3,800 vehicles, or only half the average of 7,600 vehicles per day for the nine stop-go intersections. Performance of the two types of controls is not directly comparable under these widely differing conditions.

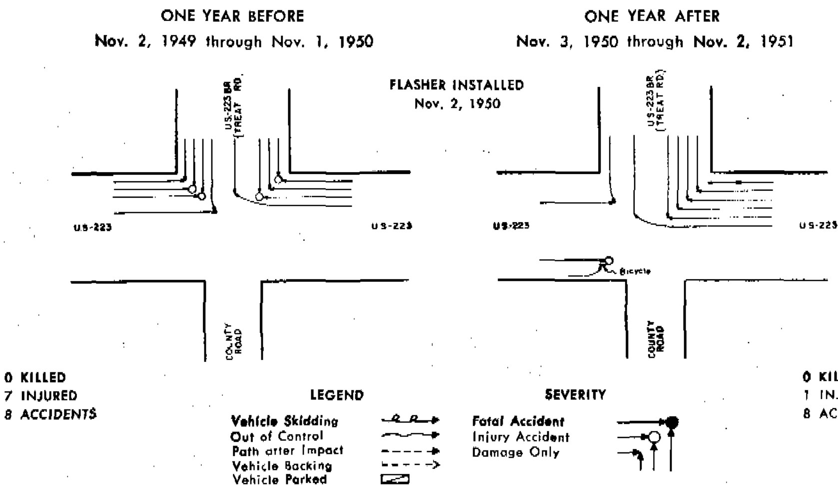


FIG. 10.

The stop-go intersections were selected to demonstrate that the stop-go signal is not a cure-all. They are only a few out of the 150 or 160 rural and suburban intersections in Michigan where similar control equipment is installed. While our analyses have not proceeded far enough to reveal the full performance record of these other locations, it is safe to say that a considerable number of them are operating with a fair degree of safety.

However, while the intersections shown may not be completely representative of Michigan's total experience with rural stop-go signals, they are thoroughly representative of the weakness of these devices in handling certain difficult conditions which are inherent in rural trunkline traffic operation. They clearly cannot be installed whenever and wherever public pressure dictates. They plainly show that we are playing with life and death when we place signals at locations to which they do not apply.

The figures reveal some of the conditions which render the operation of stop-go signals ineffective and dangerous. They suggest the probability that in some cases volumes are crowding the limit for this kind of control, and they show that most of the intersections are exposed to the hazards created by roadside mercantile development. But the most important unfavorable condition indicated by the collision diagrams is that these are isolated controls and that they intrude unexpectedly into the high-speed characteristics of rural trunkline traffic. This latter fact is the message spelled out by the huge increases in rear-end and turning collisions at several of these intersections.

This latter effect can be expected in some degree whenever a stop-go signal is installed in an isolated location on a high-speed rural trunkline. It is apparent, at least, that the present signal, installed according to the best currently approved methods, cannot be depended upon to command the attention of approaching drivers to a degree that assures consistent safety.

But even if the shortcomings of the stop-go signal were more glaring than is indicated by available experience, it does not mean that its use can be abandoned. Traffic must be regulated and protected at the many important rural intersections in this intermediate range. Flashers are unequal to the task of assigning use when volumes on the intersecting routes are in the upper brackets. It is totally unrealistic to dream that grades will be separated at any but the heaviest traveled of these locations—and at these not quickly.

It seems that the most practically constructive course is to focus some rather critical attention on this device whose operations we are analyzing. It is a highly standardized mechanism which has not been changed or improved in any basic way for at least 25 yr. Methods and procedures for using the stop-go signal have been developed and improved, and these also have become highly standardized.

Is it not possible that this standardization process has brought us to a dead end in the field of intersection control? It seems likely that what we are finding is that the same form of this device is not equally applicable to traffic conditions in both urban and rural areas. Do not all of the special conditions of rural trunkline traffic operation—higher speeds, isolated location, and intersections—point plainly to the need for signals specially designed for this service? It is even conceivable that further investigation, study and experiment might yield improvements in installation and operation methods.

These are some of the directions in which we believe our analyses of rural traffic signal operation are leading. With the alarming concentration of accidents at rural intersections, it is vital that highway and traffic engineers learn all they can about the conditions affecting intersection traffic design, operation and control.

Certain points have been soundly established. The stop-go signal, in spite of public confidence in its powers, is not primarily a safety device, it is not foolproof, and it is not a cure-all. These findings indicate that to install one of these devices just because the public demands it is like giving a child a loaded pistol just because he is crying for it.

Our investigations in this field will continue. The author strongly urges that other agencies undertake studies paralleling those reported in this paper. In the future we can unite our information and increase our understanding of these important phases of the traffic and safety problem.

Measures to Prevent or Reduce Damage

We have stated previously that one general means of preventing structural damage is either to prevent the various physical and chemical agents in the environment from reaching the structures to be protected or so to attenuate their impact that injury thresholds are not exceeded. The former is illustrated by electrical insulation, the latter by helmets, shields, fire nets, crash padding, and other devices which disperse the impinging forces over space or time. In view of the wide, though still far from adequate, use of such preventive measures, it is not surprising that an increasing number of evaluations of their adequacy have been reported in recent years. However, with but rare exceptions‡ this work has been concerned exclusively

‡ The work of Professor G. N. Quam, of Villanova University, concerned with the purpose and efficacy of student chemistry laboratory shielding is one such exception. In this context, it is important that the serious problem of student laboratory injuries produced by explosions should also be approached by study of (1) the factors that lead to such explosions; (2) the ways in which the amounts of energy involved can be reduced—for example, by decreasing the quantities of the reagents used; and (3) the substitution of less hazardous reactions.

with the prevention of mechanical injuries—chiefly those produced in vehicular crashes—by attenuating the forces involved. This is illustrated by the examples that follow.*

STATEMENT BEFORE A SUBCOMMITTEE ON INVESTIGATION OF HIGHWAY TRAFFIC ACCIDENTS†

—*A. L. Haynes*

This testimony by the executive engineer of the Ford Motor Co., though not a formal research report, serves as an excellent introduction to several aspects of the development and laboratory evaluation of preventive measures designed to attenuate mechanical forces reaching the body. It also shows clearly, as did Barmack and Payne's Lackland study, how the results of accident causation research can be used as the basis for specific countermeasures. The problems discussed by Haynes, however, still remain substantially unsolved, as is well demonstrated by the continuing toll of occupant deaths and injuries in crashes occurring throughout the entire range of operating speeds for which the automobile is designed.

It cannot be argued that injurious motor vehicle crashes are such rare events that it is unreasonable to anticipate them by safely packaging the passenger, since "*between one-fourth and two-thirds of all vehicles manufactured are at some time during their subsequent use involved in the tragedy of human injury and death.*"[33] Because of this high probability, vehicle designers should seek as their logical goal the production of vehicles that are "safe to have accidents in, if those accidents occur under the types of use for which the vehicles are designed."[33] The work described by Haynes was an early but logical step in this direction.‡

* An example of nonvehicular research concerned with the blocking of the delivery of harmful mechanical forces is the work of Haddon *et al.* (Chap. 9), which included an evaluation of the efficacy of skiers' release bindings in preventing mechanical injuries to the lower extremities. Another is McCorkle's investigation of the use and efficacy of safety glasses.[26] Also of interest is Dr. V. M. Coppleson's review of the use of electrical, chemical, and mechanical barriers in the prevention of shark attacks on man.[27]

It should be remembered that small but persistent fractions of motor vehicle injuries involve such nonmechanical insults as carbon monoxide, thermal and electrical energy, and submersion, and that measures for reducing their incidence should be developed and evaluated.

† Readers interested in the forces that influence the public safety should be thoroughly familiar with the hearings of the Roberts Committee in addition to the source cited here.[28-31] For related research on crash injury, see Chap. 9, Michelson *et al.* and Campbell later in this chapter, and reference 32.

‡ The problem of reducing injuries through adopting the best possible crash design applies also to vehicle exteriors, since more than a quarter of a million injure or kill pedestrians annually in the United States alone.[33]

CHAIRMAN ROBERTS, and members of the subcommittee, and ladies and gentlemen, I think that I have a much easier part in this program now after this horsepower picture has been cleaned up. I would like to spend the rest of the afternoon's discussion on this new dimension in safety. That is, the reduction of injury once a collision starts to take place. And in order to place the subject matter in the proper frame of reference, and remove all emotional feelings we may have about the problem, we are going to just look at the cold facts.

As Mr. MacPherson pointed out, it is relatively recent and, in fact, I would say in our experience with the problem it is only in the last year and a half that we really started to get information which the engineer could use and in which he could place confidence. For most of these facts I want to give credit to the Cornell Medical College, the research group headed up by Mr. John Moore.

These are the kind of facts, and if I may have the first slide, that really said to us, "Well, here is something that you can hang your hat on and you can do something about it."

The statistics furnished by Mr. Moore and his associates told us that 50 percent of the injuries, or 36 percent of the injuries caused in impact accidents, were because of the opening of doors, because of windshield structure, instrument panels, and steering wheel assemblies. All right, this is something that we can start working on.

Here is another type of statistic that was important to us. They told us that during a collision, the chances were 50 percent in a nonrollover accident that one or more doors would come open, and they also told us that in case of a rollover type of accident, 75 percent of the impacts would cause one or more doors to open.

Here is another bit of information. This was obtained from the Indiana State Police, a little bit before we got the information from Mr. Moore, and in effect it told us more or less the same thing. As a matter of fact, I understand this group now is working with Mr. Moore on a nationwide basis. This merely states that your chances are 2 to 1 of being a fatality if you are thrown from the automobile during the collision.

To reverse that, you are better off to stay with the car. Well, doors are opening according to statistics and let us see if we can do something about it. So we started to investigate first of all how do doors open, and the next picture will indicate to you how we approached the problem.

To slow this up now, you will actually see the door open. This is a shot of many collisions of this type that we held on our test ramp because we really had to find out how doors do open. I would like to start from the beginning on this one, and I would like to have you get the feeling for this technique that has been used here. . . .

Now, this is a slow motion part of it, and notice the door has opened. Now that we are able to open the doors out on the test range, we went to the laboratory and conducted many tests and we repeated the technique to make sure we were not making any mistakes. Here you will see a 3,500-pound steel ingot ramming into the car to convince us that it will open the door.

After we were sure that we could do this, then we had to devise mechanisms to see if we could now keep the door closed under the same type of conditions.

The next frame will indicate that we were successful in keeping the door closed. Now that we were able to do this in the laboratory, our job was now half complete and we had to convince ourselves and prove this out on our test track again.

In the next frame, it will show you that

Reprinted from Hearings before a Subcommittee of the Committee on Interstate and Foreign Commerce, House of Representatives, Eighty-fourth Congress, Second Session, on Investigation of Highway Traffic Accidents (pursuant to H. Res. 357), July 16, 23, August 8-10, 27-31, September 25-28, 1956, pp. 428-446. U.S. Government Printing Office, Washington, D. C., 1957. All but one illustration have been omitted.

we were successful in keeping the door closed. We will slow this up now and you will see that the door stayed shut. . . .

The next slide will show the fruits of our work. This is our safety door lock. It is a locking plate mounted in the striker element of the lock which is on the left part of this picture, colored in red. This engages behind the rotor element which is given there on the lock proper, and when the doors close that keeps the door from springing apart or springing away from the pillar and keeps the door shut during the collision.

The Cornell statistics told us that 40 percent of all injured drivers were hurt by the steering wheel. Well, how do we approach this problem? I know in your previous meetings quite a bit of discussion took place as to what happens when a driver impacted the steering wheel. This is how we started to approach it.

Here is a statistic about this. This is a fatality. The steering wheel, the wheel broke away and the hub of the wheel became the lethal instrument. This statistic, so-called, came from the Indiana State Police. How do we try to improve this problem? We went to the laboratory again, and we started just observing what happens when the mass of a human being, simulated as you will see in the next motion-picture frame, falls against the wheel.

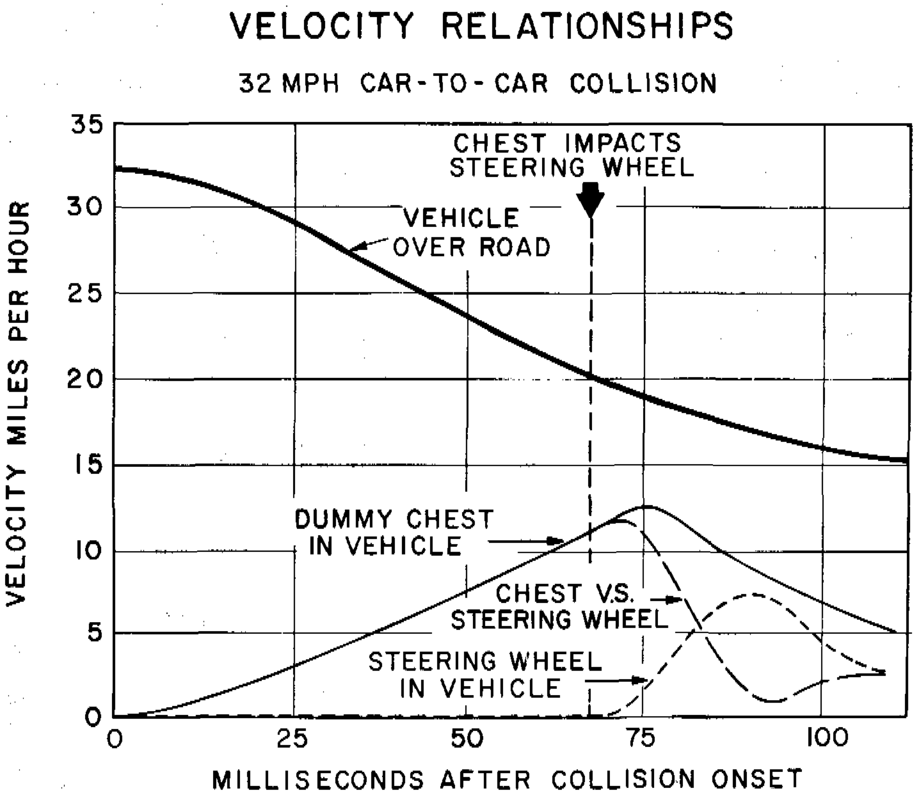
Notice when the weight is dropped, the rim breaks away and immediately the hub picks up the brunt of the load and that becomes the lethal instrument. What is involved in this type of problem? What is the phenomenon? Here is where the engineers are a little bit at the mercy of more information. I would like to mention this to the committee. This is the type of information that we lack in trying to solve this problem. We have been doing some work with Colonel Stapp at Holloman Airbase, and this is some of the preliminary work that he has accomplished which deals with the human tolerance to dynamic forces. We must know that if we are going to solve any part of this problem. He says, in effect, that there are three basic parameters, or basic characteristics, that are involved in this impact when a human body impacts an object. One is the rate with which the impact is applied, or the rate of deceleration, . . . and he also says that the time of the load application and how long the load is applied has an effect on the human body toleration.

Then he also says that the level of the force or the distribution of the force on the human body has something to do with its toleration to dynamic forces. Then he worked up three zones. [First there is the] effect zone, where the load would be applied to the body . . . when it is removed there is no effect, and the body goes back to a normal function; then there is the reversible disability zone where when the load is applied to it there is some disabled period that the body goes through, maybe a period of months perhaps, but then it returns to normal functioning. Then he says that there is the irreversible disability region where once the load is applied there, you will have permanent injury from minor perhaps to most serious and fatal injury. This is the information we must have, and we do not have enough of that.

. . . We thought if we could work out our devices (that we would include in our equipment, in our automobiles, and in our safety package) to function at a level of deceleration of about 20 G's, we would probably do a great deal in moderating the problem.

Once we have some of the information that the medical people could furnish us, then we can start applying engineering information that we need. This is the kind of information that we get from our impacts or our crash tests, one of which you will see tomorrow. Here we have actually what takes place in the event of a car collision. The [top] line indicates a car going into a crash, about 33 or 34 miles an hour, and then notice that at the moment of impact the . . . line starts to decrease, and it starts to slow up.

At the same time, notice the [steering wheel speed]. For the major portion of this event, it remains at zero velocity, and this is exactly what would happen if you were sitting in the car, and if you were an object in the car, you would be observing no relative speed between the wheel and the rest of the car. That takes place as long as the



dummy's chest, or in this case—we used dummies, of course, but we really mean the driver's chest—until the driver's chest strikes the wheel, the wheel itself has no velocity.

Notice the [chest speed]. The moment the car starts to hit somthing, or impacts another car or another object, the dummy's chest gains in speed. What we are interested in is the relative speed between the dummy's chest and the steering wheel, . . . the difference between the [chest speed] and the [steering wheel speed]. This is really the engineering information that we want.

You will notice that . . . the maximum during this particular event was about 12 miles an hour. Well, this is something, a number that we could use for extending the problem. We said, "Well, if 12 miles an hour is the maximum speed that is the maximum relative speed for a 33- or 34-mile-per-hour impact, 15 miles an hour would be a reasonable speed to take care of cases up to 40 miles an hour."

Then, as Mr. Raviolo pointed out in a previous discussion, most accidents—in fact about 67 to 70 percent—take place under 50 miles an hour, and I think the chart showed that almost 50 percent of the accidents take place under 40 miles an hour, and we said if we could start solving the problem below 40 miles an hour we would certainly be contributing to the safety of the automobile.

This is another bit of information that we have here. This helped us analyze the problem. When you saw that last bit of movie film, you saw that when the weight hit the wheel, it broke away and we actually measured from 1 to 2 inches where the rim broke away before the load was picked up by the hub. If the wheel broke away 1 inch, the collapsible depth was 1 inch, when estimated that there were 12,000 pounds of force developed on the chest of the dummy or the driver if the driver were being thrown against the wheel.

Now, that calculates out to 82 G's, and of course Colonel Stapp says it is far beyond the tolerance of the human body. So, we said that we would increase the collapsible distance and thereby reduce the 12,000 pounds to 3,000 pounds if we increase the collapsible distance to 4 inches, and reduce the 82 G's to 20 G's, and if we could increase the distance to 6 inches we would reduce to 14 G's.

We thought that was a ball park area to start working from, and this is exactly what we set out to do. You have heard a lot of discussion about collapsible wheels, and retracting wheels, and here is, I think, a good example of what a telescoping retracting wheel might be. We actually built these wheels and we decided that they are no good for several reasons. The first and important reason, as you will see in the next bit of film that I will show, is that it just did not do the job.

After the wheel started to collapse, the rim still broke away, and the hub would carry the load. The initial impact in it would become the damaging part of the system. Then, for secondary reasons we did not want to use this wheel, and they were fairly important. This is a complicated piece of machinery, and you will realize that a steering control is one of the most important controls in the automobile, and it certainly must be a safe mechanism.

To construct the tolerances required for such a mechanism, to build in the spring, the sliding lines, the telescoping tube, and the column in there, we had to build the shock absorber into the system to take the recoil and control the recoil after the impact takes place. This was just too complicated and too complex to do a good job of manufacturing and too troublesome, once you had installed this on a car.

We actually built these things. You will see now exactly what I mean when I say this did not do the job.

Notice the wheel telescopes and contracts and the rim breaks away but the load is still carried by the hub of the wheel. How else can you get energy absorption in the steering wheel?

The next slide, of course, indicates our approach to the problem. We reasoned that we could reset the hub away from the rim and then design the spokes in such a way that they would yield and bend out of the way of the body as it strikes it and also it would absorb energy in the yielding and bending. This is exactly what we set out to do.

The next strip of film will show you, and I say literally hundreds of tests that we went through, and we carried through in developing our design that we were after.

Now, this is a strip of film. You can almost see without measuring that this was still too stiff to do the job. We had slight improvement over the first few, but then we had welding problems and we had the attachment problem and it was still no good after our measurements.

Here we were getting a little bit closer, but watch this and notice how we still did not quite develop the design. It reversed and that caused serious injury to the driver's chest. Here we are getting still closer and still some problems to work out. Still the attachment to the hub was not quite refined, and it did not work out the way we wanted it.

This indicates that we played around with two spoke wheels and we found in our investigations that they could not do the job, and then also we did not have the symmetry that we wanted when the wheel was displaced, of 90 degrees from its normal position.

Here you can almost see—and of course we measured that—that this just received the mass almost like a catcher's mitt and it did a fine job. From this type of investigation, we went into a closer simulation, and in fact we will show you this later this afternoon, and we played around with this and investigated many configurations and many types of spokes, until we came up with what we thought was the answer at the moment.

Of course, this again is the fruit of our work, this was introduced with our 1956 package, and I am delighted or we are delighted to find out as was announced yesterday that another major producer is going to go to this kind of wheel, as was shown in some movies that you saw yesterday.

May I have the next slide, please. This is a study or analysis of what happens to an unrestrained occupant in a car when collision takes place. This is actually taken from an analysis of film . . . of collisions of these cars that we used for crashing purposes. You will see how we do this analysis again later on in the laboratory, and we would like to show you this as it is very interesting.

Notice that in the second frame, once the impact takes place, the occupant is lifted from his seat; in the third frame he hits the area above the windshield and his head strikes the area while his knees are striking the lower surface of the instrument panel, and then he bounces on down and hits the top surface of the instrument panel.

While we were making these studies, Mr. Moore and his associates were telling us that 38 percent of all injured right and center front-seat passengers were injured by instrument panels. He also said that the greatest percentage of these injuries were facial lacerations and the tearing of tissues and the breaking up of the bone in the head area. Well, this is an area we wanted to make some progress in.

These were some of the early types of investigations we carried on in our laboratories. The important element here . . . is what we call a calibrated beam, and inside this beam is a mounted specimen of an instrument panel. We may have had a steel panel, an aluminum panel, or it may have been cord with different types of padding materials.

The ball that is above this panel is dropped from various heights, and this ball must weigh the weight of a human head and maybe the upper part of the human torso. We then attained a feel for the problem by carrying on hundreds of such tests. . . .

While we were carrying on this kind of work in our own laboratories, again we had to call on the help of the medical profession, and we started a program with Wayne University Medical School to find out *actually what* does happen to human heads. We carried on a program where we used cadavers, and I would like to show you a film that is not very attractive to look at but this is the kind of work we must do to carry out this program.

This takes a lot of time and a lot of analysis to find out just what is happening. I will slow this up. Actually we were dropping this head from six stories to get phenomenon. We wanted to study the skull fracture characteristics and the concussion characteristics from experiments by the doctors. Then we were not satisfied with just using heads, and we wanted to use complete cadavers, because we thought the whole body would have an effect on this problem.

Again, this is a work that must be done before we finally solve this problem. This is without padding, and then we also ran some with pads. We were dropping these bodies at fairly low speeds, and we have to work up the technique for finding out exactly what happens.

I think the photographer could cut this a little bit. This one is now with a pad. We were able to make measurements on this, and the head was instrumented, and from various analyses we shall see what happens to the skull and the face and so on.

May I have the next slide, please. Then, after we made some progress and we gained more understanding—but still not enough, I must admit—we developed the technique—and you will see it later this afternoon—for measuring complete instrument panel phenomenon through different types of impacts. Here we used instrumented anthropomorphic heads and we impacted the heads at various speeds, then we made readings on the electric instrumentation in the back ground.

From this type of work we developed our 1956 safety padding material; it is an inch and a quarter thick at the leading edge and an inch and a quarter thick at the leading edge of the panel. It is made from this polyphenalcloric type of material with an energy-absorbing characteristic which is about 5 to 10 times better than sponge rubber. It contributes, I think, greatly to reducing and moderating the injury to the head once an impact takes place.

This is what happens again in an analysis of a restrained occupant in an automobile during an impact. You will notice if the occupant is childsize, he will jackknife during the impact and miss the instrument panel completely. We concluded thereby that this is the best thing that can happen. The lower part of the analysis shows what happens to an adult-sized person when he jackknifes, when he is restrained and he is jackknifed to the instrument panel.

Of course, I want to emphasize here that we were faced with the problem of solving . . . and moderating the conditions immediately, and we were able to place padding in the area. We were able to localize the problem and thereby place padding in the area where the head strikes.

I would like you to notice now a photograph of an actual crash and recall the picture I showed you earlier, first of all, the steering wheel. Compare this steering wheel with the wheel that we obtained from the Indiana State Police. Certainly, for this impact, this wheel did a much better job. But I would like you to notice the right-hand side of this photograph, and notice the great damage that was done to the instrument panel.

This was a crash, controlled or held for the purpose of studying what happens to the instrument panel struck by an unrestrained passenger. Now, the next photograph is the same type of impact, 30 miles per hour. Notice again the wheel similar to the one in the previous photograph, but notice the great moderation of damage to the instrument panel when the occupant was restrained.

We think definitely that restraining the occupant during an impact moderates the condition. I hope Mr. John Moore will be able to give you some testimony that will bear this out.

This is a typical photograph I am showing here just for the record. It shows how we evaluate and we actually check the specifications of the seat-belt assembly. Our specification is 4,000 pounds that this belt must meet under this loading condition, a body-block test as we call it. This compares to 3,000 pounds in the Civil Aeronautics Authority which they use as a specification for aircraft belts.

Also, our specification calls for 3,500 pounds release load on the buckle, once the belt has been loaded to its maximum and then released, as compared with 4,500 pounds that was used in the CAA specifications.

Again, this is some more of the information that we needed before we finalized our design information and our specifications in installing the seat belts. We had to test the floor structure. This is a static test but we actually loaded up the floor to the equivalent of 3 times 4,000 pounds; a total of 12,000 pounds was applied to this floor structure, and of course this floor was reinforced underneath by cross members to give us the needed extra strength.

Notice what happens after the floor is loaded. The structure is deformed and it actually yielded but notice that the fittings stayed unharmed and did not break. They carried the load. As I say, this is a static test, and of course we proved this out and finalized this in . . . dynamic testing . . . under actual conditions.

May I have the next one? I would like to show you now a strip of film indicating what happens. This gives you a general picture of a car-to-car type of collision. As a matter of fact, when we watched this, there were many people who claimed the dummy did not go through the roof. It was only when we saw the movies that we believed what we did not see. We will describe to you tomorrow in detail this technique of instrumentation used in carrying on this kind of test. We cut this roof area out for photographic purposes.

This is a car-to-car type of collision, the type that you would meet in a crossing type of accident, and after this we will show you exactly what takes place in a barrier-type crash, the most severe test.

This does bring out the importance of restraining the occupant in a collision. This is a severe type of accident, actually more severe than car to car, because in a barrier crash the barrier does not give one bit. When you have a car-to-car crash, both cars will give and take some energy, and thus absorb energy.

We were testing belt equipment and belt installations to make sure that they were meeting the specifications. We will show you in a moment the actual engineering information that we are seeking from this type of testing. This is the important thing with us.

This is 35 miles per hour into a barrier. We are working in the range now of—actually we are running 40 regularly—almost 50 miles per hour. This is a still shot of a car-to-car crash. I would like to show you the type of information that we are after, and notice the deceleration force that takes place at the front end of the car. In line with the hood ornament, when this is measured, actually on the frame of a car we are measuring in the order of 100 G's. But by the time the load reaches the passenger compartment, notice it is dissipated on down to where under this type of impact, it is less than 20 G's. It is in the realm of 15 to 20 G's.

Of course, this indicates that our present structure, or the structure that we have in our automobiles today, does act as a shock-absorbing, energy-absorbing, instrument. Of course, in order to take advantage of this low "G" level in the passenger compartment, we must restrain the passenger, because if he is let loose or permitted to fly loose and free on the inside, he will strike an object and develop high G's.

Work of the nature described by Haynes and others testifying at the same hearings has continued along similar lines, although much known then and learned since has not yet been universally applied.

DYNAMIC TESTS OF AUTOMOBILE PASSENGER RESTRAINING DEVICES*

—Irving Michelson, Bertil Aldman, Boris Tourin, Jeremy Mitchell

A considerable literature concerned with the efficacy, design, and evaluation of the various forms of automobile safety harness has developed in recent years. Since all of these harnesses have as their purpose the seemingly simple objective of restraining the occupant so that he will be decelerated with the vehicle rather than more abruptly by striking some interior or exterior structure, it might be assumed that the optimum characteristics of these devices would long since have been identified and evaluated. As this paper demonstrates, however, although knowledge and research in this area are well advanced, consensus has yet to be achieved—in part because of the many special interests involved and the many practical and research questions that remain unanswered.

* See also references 34, 35, and 36.

THE APPLICATION of dynamic test methods to automobile safety belts is not a new idea. Indeed, the type of seat belt in use in American cars today was developed through information from dynamic tests of restraining devices for aviation by Stapp and for automobiles by Mathewson and Severy. The latter performed their studies with controlled collisions of actual automobiles, with the vehicles, the passengers (usually dummies), and the seat belts instrumented to determine the magnitudes and durations of the forces produced in actual accidents. Similar controlled collision studies using automobiles have been done by research engineers of automobile companies. In addition to these research activities, laboratory dynamic tests have been used for several years for certifying seat belts for public sale in California by the California Highway Patrol and in Sweden by the National Institute for Materials Testing. Another laboratory dynamic test device has been used for some years in auto safety research at the University of Minnesota.

In spite of these examples of the use of dynamic methods in laboratory tests of automobile seat belts during recent years, there has been a general reluctance in the seat belt field to accept dynamic testing. During 1962, however, an increased interest developed, both here and abroad. The British Standards Institution, the Inland Transport Committee of the U. N. Economic Commission for Europe, and automobile and seat belt manufacturers in the United States and Great Britain have been considering adopting such a method. Indeed, several of these organizations have built dynamic test equipment during the past year. Very recently, the Society of Automotive Engineers and the Federal Supply Service (G. S. A.) have become interested. Nevertheless, dynamic testing is still not an accepted method in official or semi-official standards in the United States, except in California.

The objections raised against dynamic testing are mainly three: (a) that dynamic testing offers no advantage over static testing; (b) that even were it to offer advantages, it cannot be controlled well enough to constitute a standard test; and (c) that dynamic tests are much more expensive than static ones. Although the economic problem is not a subject for discussion at this forum, dynamic testing is sufficiently important to warrant an intensive effort to develop an economically feasible method, and there are indications that the problem can be solved satisfactorily. In any case, only the first two objections are discussed here.

The basic reason for even considering dynamic testing is that, in actual use, seat belts are subjected to dynamic loading conditions; that is, very large loads are applied in very short time intervals to elastic structures that respond to short-interval loading in a different way than they respond to slower loading. The mathematical physics of phenomena of this general type is described elsewhere. These treatises leave no doubt that qualitative and quantitative differences in effects exist between a transient and a relatively slow application of force to an elastic system; consequently, a dynamic test method should provide a closer simulation of actual use conditions than a static test. The question is really whether a particular laboratory dynamic test simulates the dynamic force and time conditions of severe car collisions more closely than the static tests now in general use.

Fortunately, controlled automobile collision studies have provided some data on the magnitudes and durations of forces generated in automobile collisions. These data can be used to judge whether a laboratory dynamic test, such as the

Reprinted, with permission, from the *Highway Research Record*, No. 4, pp. 62-75, 1963, published by the National Research Council, Washington, D.C. The appendix has been omitted.

Swedish one, gives a close simulation of actual collision conditions. Table 1 gives some of the dummy deceleration and lap-belt loading data from controlled car crashes at the Ford Motor Company and at the Institute for Transportation and Traffic Engineering. The data represent severe collisions into fixed barriers and head-on collisions.

TABLE 1.—FORCES GENERATED IN CONTROLLED CAR CRASHES USING LAP BELTS

TYPE OF COLLISION	REF. NO.	IMPACT SPEED (*mph*)	DUMMY DECELERATIONS Peak (*g*)	DUMMY DECELERATIONS *Duration** (*millisec*)	BELT LOADINGS *Peak* (*lb*)	BELT LOADINGS *Duration** (*millisec*)
Two-car, head-on	11	21	44	55	7,000	60
		21	40	60	5,000	60
		21	30	90	—	—
		21	34	60	4,500	65
		27	48	90	7,500	45
		27	38	80	5,000	80
		47	55	95	9,000	130
		52	72	90	9,000	135
		52	73	95	15,000	150
Car, fixed barrier	3	27	—	—	5,700	75
		29	—	—	5,800	65

* Durations of deceleration are for 5g and over; durations of loading are for 1,000 lb and over, because of indefiniteness of endpoints in many cases.

The Swedish dynamic test attempts to simulate severe crash conditions. It has been criticized as possibly being too severe because of the short stopping distance of its cart—a lead cone at the front of the cart causes it to stop in about 3 in. from an impact speed of 25 mph when it strikes the fixed barrier. The impact speed and the short stopping distance of the cart have been known, but because no other data have been available, it has been possible to speculate that dummy decelerations and belt loadings were very much higher, and durations very much shorter, than those typical of actual severe car collisions. For these reasons the Swedish test method was considered likely to be yielding a poor simulation of car collision conditions. Instrumented tests with the Swedish method have now been made and are reported herein, so that comparisons with controlled car collisions are now possible.

Before proceeding to make these comparisons, it must be noted that the two sets of controlled car collision data in Table 1 are themselves not strictly comparable. The UCLA data included both belt loads and dummy decelerations, but the seat belts used in their tests were 3 in. wide instead of the conventional 2 in. On the other hand, the Ford tests used belts of conventional width but furnished only belt load data; no dummy decelerations were reported. Comparison of the UCLA and Ford belt-load data indicate that the UCLA tests at 21-mph impact speed produced roughly the same forces as the 27- to 29-mph Ford tests.

Table 2 gives the results of instrumented runs made with the Swedish test rig during the summer of 1962; these runs were made with conventional lap belts. Comparison of Tables 1 and 2 shows that the dummy decelerations and the belt loadings generated in the 21-mph Swedish test, and the durations of these, were of the same order of magnitude as those observed in the UCLA and Ford controlled car collisions at 21- to 29-mph impact speeds; in fact, the ranges of the Swedish test data overlap the ranges found in the controlled car collisions in all factors involved.

TABLE 2.—FORCES GENERATED IN SWEDISH DYNAMIC TESTS USING LAP BELTS AND RIGID DUMMY

IMPACT SPEED (*mph*)	DUMMY DECELERATION *Peak* (*g*)	DUMMY DECELERATION *Duration** (*millisec*)	BELT LOADING *Peak* (*lb*)	BELT LOADING *Duration** (*millisec*)
21	32	60	5,400	52
21	25	47	5,400	50
21	35	50	6,600	50
21	27	63	5,700	63
15	22	45	3,800	42
15	23	43	4,000	49
15	30	40	5,200	42
15	—	—	5,300	50
15	20	45	4,400	58
15	29	43	5,400	44

* Durations of deceleration are for 5g and over; durations of loading are for 1,000 lb and over, because of indefiniteness of endpoints in many cases.

These comparisons and the graphs of the instrument data shown in the Appendix [omitted] demonstrate that, contrary to the earlier speculations, the Swedish cart's short stopping distance (and its consequently very high deceleration) do not control the deceleration of the dummy or the load on the belt. The major part of the stopping of the dummy occurs after the cart is completely stopped. The major factors that control the dummy deceleration and the belt load are the impact velocity, weight of the dummy, and the elongation characteristics of the seat belt itself. It is therefore not surprising that the Swedish test method generates forces of the same magnitudes and durations as those of actual collisions.

On the other hand, the static test methods prescribed in the official Federal Government specification and in the SAE standard, which is official in some states, impose loadings on the belt relatively slowly, so that the peak loading is reached in a time period of the order of 2 min, several thousand times as long as the loading durations observed in car collisions. Thus, it is clear that the Swedish test method produces a very much better simulation of actual belt loadings than the static method.

The reproducibility of the impact speed has been studied by the Swedish laboratory, using microswitches along the track. The variations have been found to be within ± 0.3 mph at 25 mph, so that the variation in kinetic energy does not exceed ± 2.5 percent. This demonstrates good control of the cart speed. At the three speeds studied (15, 21, and 25 mph), the stopping time remained fairly constant at about 18 millisec, but the peak decelerations increased considerably at the higher impact speeds (Table 3). (A detailed study of the deceleration characteristics of the cart will be published shortly by Aldman; the cart

TABLE 3.—SWEDISH TEST CART

IMPACT SPEED (*mph*)	STOPPING TIME (*ms*)	DECELERATION* (*peak g*)
15	19	54
21	17	86
25	17	150

* Durations of deceleration are for 5g and over; durations of loading are for 1,000 lb and over, because of indefiniteness of endpoints in many cases.

decelerations shown in Table 3 and in the graphs in the Appendix are average values and typical patterns included only to illustrate the order of magnitude and the time relationship between the decelerations of the cart and the dummy.)

In the testing of seat belts, it is the reproducibility of the dummy decelerations and the belt loadings which is of primary concern. Table 4 gives a summary of the dummy decelerations and belt loadings for various test conditions: impact speeds of 15 and 21 mph using lap belts and a rigid dummy, and impact speeds of 21 and 25 mph using harnesses and a jointed dummy. In addition to the average peak values, the ranges of the dummy decelerations and the belt loadings in each case are also presented. There is considerable overlapping of ranges among the various test conditions. However, when differences among the belts themselves are taken into account, a good deal of this overlapping is eliminated, and, moreover, some of the effects of differences in design of the belts are learned.

The term "harness" is used in this report to designate any seat belt that restrains the upper torso, whether it has a lap strap as well. It therefore includes diagonal chest straps, combinations of lap straps and diagonal chest strap, and combinations of lap strap and double shoulder straps.

The jointed dummy used by the Swedish laboratory for testing harnesses was found unsuitable for testing lap straps; the rigid dummy was made for the purpose of testing lap straps alone. The accelerometer was located on the lap of the rigid dummy and in the chest of the jointed dummy.

WEBBING ELONGATIONS

Table 5 gives the averages and ranges of dummy decelerations and belt loadings for various test conditions, but this time broken down into separate groups on the basis of known differences in webbing elongation. The "high elongation" webbings were those ranging from 21 to 29 percent elongation under a 2,500-lb load (using the

TABLE 4.—AVERAGE FORCES AND DURATIONS IN SWEDISH TEST

Device	Type of dummy	Position of accelerometer	Impact speed (mph)	No. of runs	DUMMY DECELERATIONS (peak g) Avg.	Range	BELT LOADS (peak lb) Avg.	Range	AVG. DURATION* (millisec) Decel.	Loads
Lap belts	Rigid	Lower body	15	6	25	20-30	4,700	3,900-5,400	43	48
			21	9	30	25-35	6,100	4,700-7,500	55	54
Harness	Jointed	Chest	21	8	62	35-82	6,000	4,000-8,100	50	60
			25	7	88	40-112	7,800	6,300-11,000	51	58

* Durations of deceleration are for 5g and over; durations of loading are for 1,000 lb and over, because of indefiniteness of endpoints in many cases.

SAE test method) and the "low elongation" webbings were in the 14 to 15 percent range. Most of the overlapping has been eliminated for any given set of test conditions by this grouping on the basis of webbing elongation, and the apparent reproducibility of the results is improved considerably.

Ranges within groups have not been entirely eliminated, however, because as was stated earlier, the characteristics of the belts themselves influence the deceleration of the dummy and the loads on the belt, and the belts within each group were not completely alike: variations of webbing elongation existed within each group (small variations are known to exist even among different pieces of webbing from a single roll), and, among the harnesses, the geometric configurations of the harnesses themselves varied; e.g., some had the chest strap anchored to the doorpost, others to the floor. The 25-mph runs are not separated into elongation groups because only high elongation webbings were used in these runs. The comparisons are not complete because of still another factor—the three low-elongation lap belts tested at 21 mph broke when they reached the peaks noted. Had they not broken, the 21-mph peaks may have been higher, and overlapping may have been completely absent as it was in the 15-mph tests in which no breakage occurred.

In spite of the qualifications just described, the separations between the high and low elongation groups were sufficiently consistent to demonstrate that differences in webbing elongation produce substantial differences in effects. In accidents of equal severity (that is, equal impact velocity and equal vehicle stopping distance), the low elongation webbing produced deceleration peaks and belt load peaks about one-third higher than the high elongation webbing. Because of this difference, it was observed that lap belts whose low elongation web-

TABLE 5.—EFFECTS OF WEBBING ELONGATION

Impact speed (mph)	DUMMY DECELERATIONS (peak g's) High elongation Average	Range	Low elongation Average	Range	BELT LOADS (peak pounds) High elongation Average	Range	Low elongation Average	Range	AVERAGE DURATIONS* (milliseconds) Decelerations High	Decelerations Low	Loads High	Loads Low
Lap belts†												
15	22	20-23	29	29-30	4,100	3,900-4,400	5,300	5,200-5,400	45	42	50	45
21	30	25-35		‡	5,800	5,400-6,600		§	55	‖	54	‖
Harnesses**												
21	55	40-70	78	72-82	5,500	4,000-7,300	6,800	5,600-8,100	50	53	65	58

* Durations of deceleration are for 5g and over; durations of loading are for 1,000 lb and over, because of indefiniteness of endpoints in many cases.
† Accelerometer on lap of rigid dummy.
‡ Broke at 30-35.
§ Broke at 6,200-7,500.
‖ Broke.
** Accelerometer in chest of jointed dummy.

bings were stronger in static tests (7,200 lb as against 6,800 lb) did not survive the dynamic tests at 21 mph, whereas the lower strength high elongation belts did survive.

The implication of these observations is very significant in weighing the respective merits of the static and dynamic tests. In static testing there is an assumption that belts withstanding equal static loadings are equally meritorious and can therefore survive equally severe crash conditions. But the tests show that in collisions in which all other conditions are equal, a difference in webbing elongation produces different loads on the belts. With respect to this phenomenon, the dynamic test is certainly superior in that it permits the elongation characteristics of the belt to influence the load on the belt, whereas the static tests ignore this factor completely.

Incidentally (inasmuch as it does not bear on the matter of test methods), these tests show that low elongation webbings place more load on the body being restrained than do high elongation webbings; a 33-percent decrease in elongation (e.g., from 21 to 14 percent elongation) produced an increase of roughly 33-percent in load, with virtually no change in the duration of the loading. This is a point that should be considered in designing seat belts from the medical point of view.

Locations of Third Anchor

Test results on harnesses, grouped in this case by location of the chest strap anchor (either shoulder-high on doorpost or on the car floor behind the seat) are given in Table 6. These data indicate that the location of the third anchor does affect the dummy deceleration and its rate of onset, the belt load, and the durations of the forces. Perhaps the most significant of these differences is that the floor-anchored chest straps produce the highest peak decelerations of the upper torso and yet require more time to stop the dummy than do those anchored to the doorpost shoulder high. The significance of these apparently contradictory effects of anchor location is revealed by a different kind of study (high-speed photography) described later. Only one point should be emphasized here—that a study of this type is possible with a dynamic test using a jointed full-size dummy, but a static test would reveal nothing along these lines.

TABLE 6.—EFFECTS OF LOCATION OF CHEST STRAP ANCHOR OF HARNESSES

	DUMMY DECELERATIONS (peak g)		BELT LOAD (peak lb)		AVERAGE DURATION* (milliseconds)			
					Deceleration		*Loads*	
Impact speed (mph)	*Doorpost anchor*	*Floor anchor*	*Doorpost anchor*	*Floor anchor*	*Door*	*Floor*	*Door*	*Floor*
21	59	64	7,000	5,400	32	61	48	67
25	72	100	8,800	7,000	37	61	48	66

* Durations of deceleration are for 5g and over; durations of loading are for 1,000 lb and over, because of indefiniteness of endpoints in many cases.

All of the observed effects of different webbing elongations, and some of those of different anchor locations, could have been derived from theoretical considerations, or, at least, they are reasonable and self-consistent in hindsight. For example, lower elongation webbing could be expected to decrease the stopping distance of the dummy, which would increase its deceleration rate, resulting in turn in an increased peak load on the belt. Nevertheless, the ability of the Swedish test method to demonstrate these reasonable effects with a high degree of consistency is another factor that increases confidence in the method.

Table 7 gives the pertinent test data obtained in 30 runs. Several typical graphs of instrument data are in the Appendix, which also contains descriptions of the Swedish dynamic test equipment and the instruments used in this study.

One final point deserves mention. With the Swedish test method, peak belt loads up to 11,000 lb and peak dummy de-

TABLE 7.—BASIC TEST DATA

Run no.	Impact speed (mph)	Dummy*	Belt type†	Elong. (%)	Third anchor	DUMMY DECEL. Peak (g)	DUMMY DECEL. Duration‡ (millisec)	BELT LOADING Peak (lb at anchors) Left	Right	Third	Total	BELT LOADING Duration‡ (millisec) Total	Notes§
3	25	J	LD	21	Door	85	25	2,500	4,400	2,000	8,900	37	
4			LD	21	Door	92	27	4,000	4,500	2,500	14,000	53	
5			LD	21	Floor	105	63	1,800	3,300	2,400	7,500	75	
6			LD	21	Floor	108	70	1,700	3,500	2,400	7,600	72	
7			L2S	21	Floor	112	60	2,400	2,400	1,500	6,300	58	
8			LD	—	Door	40	60	1,800	2,600	2,200	6,600	53	1, 3
9			LD	—	Floor	75	51	2,700	1,400	2,300	6,400	58	1, 3
14	21	J	LD	21	Door	70	26	1,900	3,100	2,300	7,300	53	
15			LD	14	Door	72	24	2,100	3,800	2,200	8,100	40	
16			LD	21	Floor	42	57	1,800	2,200	1,200	5,200	75	
17			LD	14	Floor	82	55	2,200	2,900	1,500	6,600	72	
18			L2S	21	Floor	55	66	1,500	1,600	900	4,000	68	
19			L2S	14	Floor	78	80	2,000	2,100	1,500	5,600	61	
20			LD	—	Door	35	46	1,500	2,200	1,900	5,600	50	1, 3
21			LD	—	Floor	65	48	2,000	2,100	1,200	5,300	58	1
51	21	R	L	21		32	60	2,600	2,800	—	5,400	52	3
52			L	21		25	47	2,800	2,600	—	5,400	50	3
53			L	21		35	50	3,300	3,300	—	6,600	50	
54			L	14		30	—	3,500	3,700	—	7,200	—	2
55			L	14		35	—	3,750	3,750	—	7,500	—	2
56			L	29		27	63	2,850	2,850	—	5,700	63	1, 3
57			L	23		35	—	3,100	3,100	—	6,200	—	1, 2
58			L	23		25	—	2,700	2,000	—	4,700	—	1, 2
59			L	15		30	—	2,700	3,500	—	6,200	—	1, 2
60	15	R	L	21		22	45	1,900	2,000	—	3,900	42	
61			L	21		23	43	2,000	2,000	—	4,000	49	
62			L	14		30	40	2,500	2,700	—	5,200	42	
63			L	14		—	—	2,400	2,900	—	5,300	50	
64			L	29		20	45	2,200	2,200	—	4,400	58	1
65			L	15		29	43	2,700	2,700	—	5,400	44	1

* J = jointed full-size dummy, without arms, with soft abdomen, 154 lb, accelerometer in chest; R = rigid dummy, torso and thighs only, seated position, 150 lb, accelerometer in lower back.

† L = lap strap only; D = diagonal chest strap only; LD = combination belt, lap and diagonal chest straps; L2S = combination belt, lap and two shoulder straps.

‡ For decelerations of 5g and over, for total belt loads of 1,000 lb and over, and for individual anchor loads of 500 lb and over.

§ 1 = belt included buckle; all others continuous webbing from anchor to anchor to minimize chance of slippage; 2 = webbing broke at or near peak load; 3 = slippage in one anchor.

celerations up to 112g were observed with the use of floor-anchored combination lap and chest strap belts. That these values are not atypical of severe automobile crashes is indicated by the controlled car crash data given in Table 1, in which peaks as high as 15,000 lb and 73g were observed with lap belts alone. The question of whether such forces are tolerable by the human body is beyond the scope of this report, but the data indicate the ways in which the designs of harnesses may be changed to avoid excessively high peaks.

Valuable as the test data presented herein may be, they also indicate that considerably more exploration of the performance of seat belts by dynamic methods would be profitable, particularly in the evaluation of current models and in the development of better designs and materials.

RESTRAINT OF UPPER TORSO

The data that have been presented up to this point have had a bearing on problems involving decelerations of the belt wearer and loads on the belts themselves. But such instrumented runs furnish no direct information on how well a particular type of belt limits the body's forward motion; that is, on how well it restrains the wearer. The problem of restraining the upper torso to minimize head and chest injuries is considered particularly important in automobile accidents.

It is still another advantage of the Swedish dynamic test method that it can be used to observe by means of high-speed photography the degree of restraint of parts of the body by belts of various geometric configurations. A study of this kind was also performed in the summer of 1962 on nine models of harnesses then available in the United States. This study revealed, among other things, that diagonal chest straps alone may permit the wearer to slide out from under the belt or suffer severe internal injuries, and that combination belts with shoulder-high anchors for the chest strap limit the forward motion of the upper torso to about one-half that permitted by floor-anchored chest straps. These tests are described in detail elsewhere; however, (a) the larger forward motion permitted by floor-anchored chest straps correlates well with the longer stopping time observed in the instrumented tests, and (b) the floor-anchored chest straps first permit the shoulders to lean forward into the chest strap and then produce a sudden very high peak deceleration when the dummy has leaned forward as much as it will go.

Not all types of dynamic test apparatus are capable of producing information of this type. For example, the California Highway Patrol equipment is unable to observe restraint of the upper torso because it uses as its dummy a body block that is equivalent to the hips alone; it is therefore suitable only for testing lap belts. Modification of this equipment to make it capable of testing harnesses is likely because several harnesses are now on the American market. It is also likely that laboratory apparatus for dynamic testing of seat belts for quality control in manufacturing (or other routine testing) will not need to be capable of demonstrating restraining characteristics if these characteristics have been shown to be satisfactory in the design development stage. But the restraining characteristics of any new design, even if it is only a "slight" modification of an old design, should be tested first for its restraining characteristics in equipment of the kind used in Sweden.

Summary and Conclusions

Results of instrumented tests of seat belts by the Swedish dynamic test method have shown that (a) the peaks and durations of the decelerations of the belt wearer, as well as the loads on the belts, are of the same order of magnitude as those observed in controlled car crashes of a severe nature, (b) the characteristics of the belt itself exert a major influence on the deceleration rate of the belt wearer and on the magnitude and duration of the load on the belt, and (c) belts made of webbings found to be equally strong by the standard static test are not necessarily equally resistant to the forces developed in collision conditions of equal severity. These facts taken together indicate a clear superiority of dynamic testing over static testing.

The results have also demonstrated that the laboratory dynamic test is capable of producing information on the performance characteristics of belts of different geometric configurations and of various materials, to aid in evaluation and development of better safety belts. The specific effects of different webbing elongations and of third-anchor locations (for combination lap and chest strap belts) have also been demonstrated. The standard static tests are incapable of furnishing research and development information of this type.

In view of the need to develop seat belts that are effective in restraining the upper torso and are convenient to install and to don, more extensive use of dynamic testing is clearly called for.

The final proof of the efficacy of an accident countermeasure is its ability to reduce the incidence or severity of accidents in the real world. This is something that cannot be settled in either armchair or laboratory but requires sophisticated field study, whether the measure attempts behavioral modification (as in the Lackland work), the attenuation of hazardous forces, or some other approach.

To date most of the nonexperimental, field evaluations of the efficacy of measures designed to block or attenuate environmental hazards have come from the Cornell University Aviation Crash Injury Research (AvCIR) and Automotive Crash Injury Research (ACIR) groups which derived from the work of De Haven (see Chap. 9). The latter group, for example, has documented the partial efficacy of lap belts[37] and of many other aspects of vehicle crash design.

A STUDY OF INJURIES RELATED TO PADDING ON INSTRUMENT PANELS

—B. J. Campbell, Ph.D.

The work of the Cornell Automotive Crash Injury Research has supplied much of our present knowledge of the ways in which automobile occupants' injuries are determined by the design of the vehicle structures with which they are surrounded.† No comparable source of information exists, although the group's work bears many relationships to the complementary work of Stapp, Severy, and others (see Chap. 9). At a time when much accident research has been largely divorced from reality, or at best based on secondary sources of dubious value, the Cornell workers have demonstrated a remarkable and continuing ability to study actual accidents on a nationwide basis. Despite serious practical problems, they have consistently produced work of a quality rare in accident research. It is also to their credit that their findings have increasingly influenced automobile design, as noted by Haynes (above).

Automotive Crash Injury Research (ACIR) has produced reports in which various automobile components were ranked in terms of the frequency and severity of injuries they produce in accidents. In each instance, the instrument panel was among the four causing most injuries (along with ejection, the steering wheel, and the windshield). It was in fact shown that in some situations the panel caused more frequent injury than any other part of the car.

In an accident, the instrument panel may cause injury when a car occupant is thrown onto it during the impact. One way of attempting to reduce such injuries has been to cover the panel with padding material designed to absorb energy and cushion the blow. Beginning approximately in 1956, American cars appeared with various kinds

[Automotive Crash Injury Research, CAL Report No. VJ-1823-R2, August 1, 1963. The foreword, the summary, and 3 appendixes have been omitted.]

† No comparable studies have been undertaken on injuries to occupants of buses, trucks, and other surface vehicles, even though inspection of such vehicles and of the crashes in which they are involved reveals the need for considerable improvement in the crash-injury protection they provide their users and the vehicles and pedestrians with which they collide. For some trucks this includes the need to protect their users and the public from fire and other effects of the spillage of dangerous materials.

of padding on the instrument panel and today most, if not all, U.S. cars can be purchased with some amount of padding in the forward areas of the car. Often the padding is an option available at extra cost, but in some (usually higher-priced models) it is standard equipment.

The question in this study is whether relevant injuries sustained by occupants in cars with padded instrument panels are any different from those sustained by occupants in similar cars without padding (and in similar accidents). The foregoing sentence, calling for "similar" situations, suggests sources of variability that must be taken into account in seeking a fair assessment of padding effectiveness.

Cars with and without padding should be selected on the basis of involvement in similar accidents. It would be undesirable to compare an occupant of a padded car in a mild accident to an occupant of an unpadded car in a severe accident. Next, since shape and construction of the instrument panel varies widely from model to model, the padded and unpadded cars compared should be of identical design. It would be undesirable to compare occupants in a car with a more smoothly rounded and padded panel against occupants of an unpadded car whose panel had more hard ridges of unyielding construction. Finally, the occupants compared should have similar physical characteristics, such as height and weight, since these partially determine where and how the panel will be struck. It would be undesirable to compare a short light female in a padded car to a tall heavy male in an unpadded car.

In selecting data for this study, comparability in the above three classes was a prime requisite.

Selection of Relevant Data

The total pool of ACIR data includes records of more than 60,000 occupants who were drivers and passengers in more than 30,000 cars in injury-producing accidents. From this pool of data, only the most relevant cases were selected for the final sub-sample. The process of selecting this sub-sample is described in three steps below.

Step 1: At the outset, all seated positions except Center and Right Front were eliminated on the assumption that occupants of these two seats are most likely to strike the panel. Even within the CF and RF categories, only normally seated occupants were retained. Thus, further exclusions consisted of occupants being carried on someone's lap, occupants carrying someone on their laps, occupants standing on the seat, and occupants lying in the seat.

Since the remaining CF and RF occupants were exposed to all varieties of accidents, further exclusions were necessary because it was desirable to retain only those accidents in which striking the panel was most likely; therefore, CF and RF occupants of any car involved in other than a simple, forward-force accident were also eliminated. Thus, CF and RF occupants were retained only if the accident did not involve rollover and was confined to a front impact from the 11, 12, or 1 o'clock directions, as illustrated below.

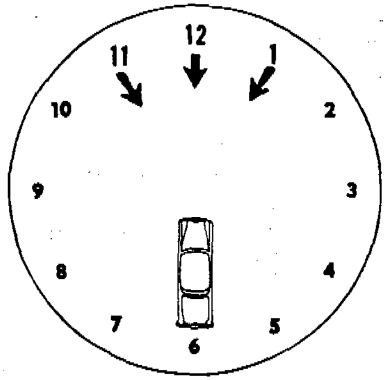


FIG. 1.

Finally, any remaining CF or RF occupant who was partially or completely ejected was also eliminated. Naturally, the foregoing exclusions resulted in elimination of tens of thousands of cases, but the remaining few thousands consisted of those in situations most likely to involve striking the panel in a direct manner. The study is, therefore, confined to CF and RF occupants, normally seated and in forward-force accidents.

Step 2: Occupants meeting the requirements of Step 1 (about 6,000) were divided into two groups—occupants of padded cars and occupants of unpadded cars. Next, a matching procedure was established to give assurance that occupants of padded cars would be compared only to occupants of unpadded cars in similar circumstances. The rule was established that a padded car occupant would be a "possible match" for an unpadded car occupant only if *all* of the following conditions were met:

1. they occupied the same seat (center front or right front) in their respective padded and unpadded cars; and
2. their cars were struck from the same direction (11 or 12 or 1 o'clock); and
3. the impacts were of closely similar severity (*identical* ratings on a subjective five-point accident severity scale); and
4. the cars were the same make (Lincoln, Plymouth, Oldsmobile F-85, etc.); and
5. the cars were the same model year (1956, 1960, etc.); and
6. the occupants of the padded and the unpadded cars had the same experience as to whether or not they also struck the windshield.*

Exact agreement on all six of these factors is a necessary but not sufficient condition for being considered as a matched pair of occupants. The final selection of matched occupant pairs was based on still further selection. The matching process outlined above reduced the 6,000 available occupants to approximately 2,600.

* In other words, if the occupant of the padded car was reported to have struck the windshield, he would be matched only with an occupant of a comparable unpadded car who had also struck the windshield (same procedure if reported as not having struck the windshield). A person may strike and be injured by the windshield, and this could influence the apparent effects of also striking the panel. Therefore, windshield involvement was included as a control factor.

Step 3: From among the "candidates," final selection was made on the basis of occupant height, weight, age, and sex. As a result of the matching described in Step 2 above, there were sometimes several padded-car occupants and several unpadded-car occupants who were identical with respect to accident-vehicle variables. These could be paired up randomly, but a better procedure was to attain further comparability on the basis of physical characteristics. Thus, from the available cases, those who were physically most similar were selected. It should be pointed out that while *exact* matching was required for the accident variables specified in Step 2 above, only the "best possible" match was feasible in Step 3. In many instances, age and height would be matched very closely, but in other cases, it was not possible to avoid a substantial mis-match. (It should be noted, however, that the direction of mis-match varied randomly; with respect to height, for example, it was about as likely that the padded-car occupant would be taller as it was that he would be shorter.)

After application of all the rules, the ACIR data pool yielded 792 matched pairs of occupants. One of each pair was an occupant of a padded car and the other was an occupant of an unpadded car. Each pair was matched exactly on six accident-vehicle variables and each pair was matched as closely as possible on four physical characteristics. The 792 pairs of occupants were allocated among more than 60 kinds of cars. There were, for example, 1956 Buicks, 1957 Buicks, 1958 Buicks, 1957 Dodges, 1958 Dodges, 1959 Fords, 1960 Fords, and so forth, through the range of cars on the highways. This means that there was considerable variety in construction and shape of the panels.

Sample Characteristics of the Matched Pairs

The matching process has both advantages and disadvantages. On the desirable side, it yields maximum control of other variables that might obscure or distort the apparent effects of padding. Accompanying this advantage is the inconvenient likelihood that the stringent matching requirements may result in the final small sample (792 pairs) being non-representative of ACIR data in general.

A comparison of the 792 pairs with the larger ACIR data pool shows that the specialized sub-sample tends to be somewhat concentrated among less severe accidents. This means that fewer cases are in the Severity 3 and higher range (17.6 percent as compared to 32.7 percent in the over-all data). This also leads to a somewhat lower rate of windshield involvement (25.1 percent in the sub-sample, compared to 38.2 percent in over-all ACIR data). These characteristics mean that on the *average*, the sub-sample consists of milder accidents. Nevertheless, there is still a variety of conditions, enough to permit examination of padding effects in most accident-severity ranges.

Treatment of the Data

Two kinds of analyses are shown regarding padding effectiveness. First, for each pair of occupants, it is possible to determine that the occupant of the padded car sustained injuries that were either less severe, of the same severity, or more severe than those of his counterpart. For example, if all 792 pairs were evaluated, the outcome might be as follows:

	FREQUENCY
Injury to padded-car occupant less severe	140
Tie	550
Injury to padded-car occupant more severe	102
Total	792 pairs

The non-tie outcomes can be compared for statistical significance by the Sign Test (Chi square, one degree of freedom). If padding had no effect at all, it would be expected that outcomes would be random and that a padded-car occupant would have less severe injuries about as often as he would have more severe injuries. In the above example this is not true and there are more outcomes favoring padding. This example was selected to show a statistically significant deviation. Thus, the idea of a random occurrence is rejected. In the example, the difference is significant at the 5-percent level (the level at which a chance occurrence is rejected in this study).

The second kind of analysis used in this study consists of comparing injury distributions for the 792 occupants of padded cars with the 792 occupants of unpadded cars. Injuries are classified on a five-point scale as follows: No Injury, Minor, Non-Dangerous, Dangerous, and Fatal. Data are analyzed for injuries to the head, neck, thorax, abdomen, arm, and leg. Obviously, padding is not expected to be beneficial for all of these body areas, but for purposes of testing the logic of data results all were included.

Findings—Part 1

The 792 pairs of occupants were examined as a total group, representing the many combinations of "vehicle-accident-occupant" circumstances within the limits defined earlier. With respect to over-all injury,* it was found that in 249 cases the occupants of padded cars had less severe injuries than their counterparts in unpadded cars. The padded-car occupants' injuries were more severe 220 times, and in 321 instances both sustained the same degree of injury (in two cases, the comparison could not be made because of lack of injury information). This is summarized below:

Padded-car occupant injury less severe	249
Tie	321
Padded-car occupant injury more severe	220
Not reported	2
Total:	792

$$\chi^2 = \frac{(249-220-1)^2}{249+220}$$

$$\chi^2 = 1.67$$

The split of 249 versus 200 (in the hypothesized direction) yields a Chi square of 1.67, not statistically significant at the confidence level selected for this study. Table 1 repeats this comparison and also shows the same layout for injury to each body area.

* A rating of the most severe injury sustained, regardless of its location.

Table 1 shows a significant association between presence of padding and lesser head injuries. There were 232 cases in which head injury to the padded-car occupant was less severe than that to his counterpart in the unpadded car. In 171 instances, the reverse was true and the unpadded-car occupant had a less severe head injury. The probability of occurrence of this departure from an even split is less than 1 percent. Analysis of data for other body areas showed more outcomes favoring padding than against padding, but none approached statistical significance. Of course, as was stated earlier, no effect was expected in some body areas, such as neck, abdomen, arms, and legs. It was unexpected, however, to find that in the case of thorax injuries, no significant effect emerged; and that, moreover, there was hardly a suggestion of a trend.

Further examination of the significant effect for head injuries is desirable, and Table 2 shows a distribution of head injury severity for the 792 occupants of padded cars in comparison to the 792 occupants of unpadded cars.

The head injury distributions of padded-car occupants differ significantly ($p < .05$), with head injuries less severe for occupants of padded cars. The difference does not show in the more severe range of injuries, but is definitely present among injuries at the low severity end of the scale. At the severe end of the scale, Dangerous and Fatal head injuries occur with virtually identical frequency for the padded and the unpadded groups. In the intermediate portion of the scale, however, more occupants of unpadded cars suffer Non-Dangerous injuries (greater than Minor but not dangerous to life). Of the unpadded-car occupants, 13.1 percent fell into this category against 9.7 percent of padded-car occupants. At the lower end of the scale, 42.1 percent of CF and RF occupants of unpadded cars escaped head injuries altogether, while 48.4 percent of occupants of padded cars sustained no head injuries.

Summarizing Part 1 of the analysis, there was no suggestion that padding was associated with lesser neck injuries, abdomen injuries, arm injuries, or leg injuries. This was not surprising, but it was rather unexpected to find that there was also no effect for thorax injuries. With respect to head injuries, however, there is a definite association between presence of padding and less severe trauma. Although the effect does not extend to the very serious head injuries, within the range of injuries that pose no threat to life, the addition of padding to a panel seems to indicate a reduction of about 22 percent in the chances of an injury greater than Minor in degree *and* this beneficial effect of padding is also manifested at the other end of the severity scale in a 15-percent increase in chance of escaping head injury.*

* There were 120 greater-than-Minor head injuries among occupants of unpadded cars, against 93 in padded cars. Reducing 120 to 93 constitutes a 22-percent reduction of 120.

TABLE 1.—COMPARISON OF OCCUPANTS OF PADDED AND UNPADDED CARS WITH RESPECT TO SEVERAL BODY AREA INJURIES

	Over-all	*Head*	*Neck*	*Thorax*	*Abdomen*	*Arm*	*Leg*
Padded-car occupant, injuries less severe	249 (.53)*	232 (.58)	35	118	53	140	206
Tie	321	386	722	558	693	513	383
Padded-car occupant, injuries more severe	220 (.47)	171 (.42)	32	114	42	137	201
Unable to compare (missing data)	2	3	3	2	4	2	2
Total	792	792	792	792	792	792	792
χ^2	1.67	8.93	—†	—	1.05	—	—
p	—	$<.01$	—	—	—	—	—

* This figure indicates that, of the classifiable outcomes, 53 percent favored padding. Thus, there were 249 "better," 220 "worse" for a total of 469. Of the 469, the 249 figure represents 53 percent. This technique is used throughout this and other tables to indicate direction of outcome when there was a trend.

† A dash is inserted instead of the χ^2 value when it is less than 1.0.

TABLE 2.—DISTRIBUTION OF HEAD INJURY SEVERITY FOR 792 OCCUPANTS OF PADDED AND UNPADDED CARS

	None	*Minor*	*Non-Dangerous*	*Dangerous-Fatal**	*Not reported*	*Total*
Padded car:						
Number	383	316	77	16	0	792
Percent	48.4	39.9	9.7	2.0	—	100.0
Unpadded car:						
Number	332	337	103	17	3	792
Percent	42.1	42.7	13.1	2.1	—	100.0

* Dangerous category and Fatal category combined because of low frequencies.

$\chi^2(3\text{df}) = 8.09 \qquad .05 > p > .01$

FINDINGS—PART 2

In Part 2 of the study, the over-all figures cited in Part 1 are subdivided according to several categories of relevant variables. The reasons for this further division of the data are twofold. First, with respect to head injuries, further sub-division may show whether the significant effect of padding is generalized and pervades the whole array of accident situations, or whether the significance derives from a very pronounced favorable effect in some specific situations and an absence of effects in others. Second, in the case of other body areas that failed to reflect any over-all effect of padding, it is desirable to subdivide and examine the data further in order to explore the possibility that a significant effect may yet be present in some particular segment, but is not strong enough to reflect in the total sample of 792 occupant pairs.

Accident Severity: The 792 occupant pairs were divided according to accident severity. There were 88 pairs exposed to accidents of Severity 1 (Minor), 565 pairs in Severity 2 (Moderate), 120 pairs in Severity 3 (Moderately Severe), and 19 pairs in Severity 4 or higher (Severe or greater).

Table 3 below is a breakdown of the data according to the accident severity categories mentioned above. Most of the body areas show the same results whether the data are examined totally or whether they are broken down by accident severity. For example, in the case of head injuries, the significant effect found over-all in Part 1 seems to be fairly generalized, and the trend shows in accidents of Severities 1, 2, and 3. Due to small numbers, however, statistical significance is seen only in Severity 2.

The findings for injuries to the neck, thorax, abdomen and arms are also consistent with the impression gained in Part 1: no padding effect is manifest.

The one exception is leg injuries. In the over-all data (Part 1) there is an almost exactly even split. That is, there are almost exactly as many instances in which the padded-car occupant had more severe leg injuries as instances in which he had less severe leg injuries. However, when the data are separated by accident severity, a pattern emerges. In the more severe accidents (Severity 3 and above) leg injuries seem to be lessened by padding. That is, in accidents of Severity 3 and higher, occupants of unpadded cars more often have greater leg injuries than their counterparts in padded cars. This occurs to a statistically significant degree.

Examining the leg injury distributions themselves (in accidents of Severity 3 and higher) indicates a trend not quite statistically significant but does suggest that leg injuries are fewer and less severe for occupants of padded cars. Table 4 shows this.

Direction of Force: As was stated, the data are confined to cases involving 11, 12, or 1 o'clock forces, and each of these was examined separately. The favorable trend for head injuries seemed consistent for each impact direction, though it was significant only for 12 o'clock (the "11" and "1" directions had too small frequencies to yield significance). When other body area injuries were examined by the direction-of-force breakdown, they showed no sign of an effect for any of the three impact conditions. Also, there was no appreciable trend.

Windshield Involvement: The 792 occupant

TABLE 3.—INJURIES TO SIX BODY AREAS BY ACCIDENT SEVERITY*

	HEAD					NECK				
	S1	*S2*	*S3*	*S4+*	*Total*	*S1*	*S2*	*S3*	*S4+*	*Total*
Padding better	17	167 (.58)	44 (.61)	4 (.31)	232 (.58)	6 (.75)	18	9	2	35
Tie	57	275	48	6	386	79	524	105	14	722
Padding worse	13	121 (.42)	28 (.39)	9 (.69)	171 (.42)	2 (.25)	21	6	3	32
Total	87	563	120	19	789	87	563	120	19	789
χ^2	—	7.03	3.12	1.23	8.93	1.12	—	—	—	—
p	—	<.01	—	—	<.01	—	—	—	—	—

	THORAX					ABDOMEN				
	S1	*S2*	*S3*	*S4+*	*Total*	*S1*	*S2*	*S3*	*S4+*	*Total*
Padding better	6	79	26	7	118	4	34	12	3	53 (.56)
Tie	73	409	67	9	558	81	499	98	15	693
Padding worse	8	76	27	3	114	2	29	10	1	42 (.44)
Total	87	564	120	19	790	87	562	120	19	788
χ^2	—	—	—	—	—	—	—	—	—	1.05
p	—	—	—	—	—	—	—	—	—	—

	ARM					LEG				
	S1	*S2*	*S3*	*S4+*	*Total*	*S1*	*S2*	*S3*	*S4+*	*Total*
Padding better	6	97	35 (.57)	2	140	16	138	43 (.65)	9	206
Tie	74	367	59	13	513	52	275	54	2	383
Padding worse	7	100	26 (.43)	4	137	19	151	23 (.35)	8	201
Total	87	564	120	19	790	87	564	120	19	790
χ^2	—	—	1.05	—	—	—	—	5.47	—	—
p	—	—	—	—	—	—	—	<.05	—	—

	S1 + S2	*S3 + S4+*
Padding better	154	52 (.63)
Tie	327	56
Padding worse	170	31 (.37)
Total	651	139
χ^2	—	4.82
p	—	<.05

* In each category 2-4 injured persons are unclassifiable as to degree of injury so that only 788-790 of the 792 cases are listed.

pairs were also divided according to whether or not they were reported to have struck the windshield. It was found that 593 pairs were reported as not having been involved with the windshield, while 199 pairs were reported as having struck the windshield. Dividing the data in this way does not particularly change the impression from earlier data analysis. Head injuries seem to be lessened by padding whether or not windshield contact was also reported. The other body area injuries continue to seem uninfluenced regardless of whether or not windshield contact was reported.

Seated Position: When the data were separated by seated position, the favorable padding effect for head injuries seemed more strongly concentrated among CF occupants, and although there was a similar trend for RF occupants, it was not strong enough to be statistically significant. This is the first suggestion that the desirable effects of padding on head injuries are other than generalized, showing effects in all situations. The other body areas showed no significant benefits attributable to padding either in the CF or the RF position.

A direct comparison of the distribution of head injury severity for CF occupants of padded and unpadded cars shows even more clearly the effects of padding. Table 5 indicates that head injuries for occupants of padded cars are significantly less severe than those of their counterparts in unpadded cars.

As can be seen in the . . . table [5], there was a substantially higher proportion of more severe head injuries among CF occupants of unpadded cars. Among these unpadded-car occupants, 15.9 percent suffered

TABLE 4.—LEG INJURIES IN ACCIDENTS OF SEVERITY 3 AND HIGHER

	None	*Minor*	*Non-Dangerous*	*Dangerous-Fatal*	*Total*
Padded cars:					
Number	76	45	18	0	139
Percent	54.7	32.4	12.9	0.0	100.0
Unpadded cars:					
Number	58	54	27	0	139
Percent	41.7	38.9	19.4	0.0	100.0

χ^2 (2df) = 5.04 $p > .05$ ($p < .10$)

TABLE 5.—HEAD INJURIES TO CF OCCUPANTS OF PADDED AND UNPADDED CARS

	None	*Minor*	*Non-Dangerous*	*Dangerous-Fatal*	*Total*
Padded cars					
Number	91	58	9	0	158
Percent	57.6	36.7	5.7	0.0	100.0
Unpadded cars:					
Number	70	62	18	7	157*
Percent	44.6	39.5	11.5	4.4	100.0

χ^2 = 12.0 (3df) $p < .01$

* One injury not reported with respect to degree.

head injuries greater than Minor in degree, whereas among padded-car occupants only 5.7 percent had such injuries. At the other end of the scale, 44.6 percent of unpadded-car occupants escaped head injury altogether, whereas among padded-car occupants 57.6 percent escaped injury. It appears that padding reduces more severe head injuries for CF occupants by about two-thirds, and also increases the chance of escaping head injury by about 30 percent.

It is particularly interesting to find that padding is even more effective for head injury to CF occupants, in view of the fact that occupants of this seat tend to be younger and smaller. Whereas the average height for all CF and RF occupants in this study is 63 inches, the average height of CF occupants is only 60.5 inches, indicating presence of a greater proportion of children. The average age of CF occupants is 23.3 years, contrasted to 32.2 years for the CF-RF combination.

The above indications led to further examination of the possibility that padding was particularly effective for children, whose height might position them so that they would be even more likely to benefit from padding as regards head injuries. There were 17 instances in which small children (48 inches or less) occupied padded and unpadded cars and were closely matched in height. In only one case did the padded-car occupant sustain more severe head injury than his counterpart in the unpadded car. In the remainder of the cases, the padded-car occupant was either injured less or injured the same as the unpadded-car occupant.

Year and Model of Manufacture: When the data were separated according to year of manufacture, the result was a severe dilution of data. Since cases were included for cars manufactured in 1956, 1957, 1958, 1959, 1960, 1961, and 1962, it is not surprising that frequencies were too small for statistically significant effects. It was noted, however, that in the case of head injuries, and where numbers were fairly large, the trends were favorable to padding throughout the range of model years.

When the data were separated by car make, the problem of data dilution was even more acute. The 792 pairs had to be allocated among 21 makes. The result was that there were trends in the expected direction (favoring padding) in most cases, though in only one was the trend strong enough for statistical significance. Numbers were generally too small. In only two makes was the direction contrary to that expected, and these involved small numbers of cases.

DISCUSSION

The instrument panel is one of the leading sources of injury in automobile accidents, ranking first in number of all classes of injuries caused and ranking third in number of fatal injuries caused. Padding has been added to instrument panels to cushion the blow if an occupant should strike it, and analysis shown in this paper indicates that under certain circumstances and within certain limitations, padding does indeed operate to prevent some head injuries and to reduce the severity of other head injuries. In order to obtain a clearer understanding of the role of instrument panel padding in injury reduction, a discussion of the cited "circumstances and limitations" is necessary.

A great deal of analytical exploration

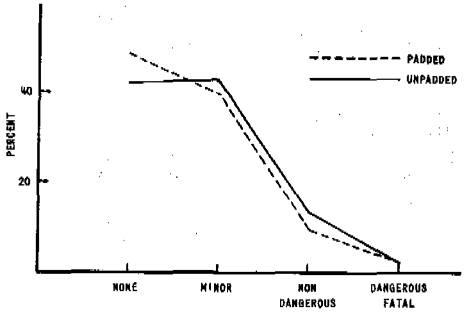


FIG. 2. Head injury severity.

preceded the current paper. This could be characterized as a lengthy process of systematic exclusion of less relevant cases until only the present select sub-sample remained. In the extensive earlier explorations of more general data, no clear-cut benefits of padding were observed. It was possible to detect the effects of padding only in the highly restricted sample finally included. This means forward force accidents in the not-too-severe range.

The strongest beneficial effect of padding was seen in the lower part of the injury scale—influencing injuries that are not so severe as to be dangerous to life. No effect of padding was detected in the range of injuries severe enough to threaten (or take) the occupant's life. At the low severity end of the scale, it was shown that under the "right" circumstances, padding appears to increase by 15 to 30 percent the chances of escaping a head injury, and it also appears to be responsible for a reduction of one-quarter to two-thirds in the frequency of head injuries of a severity greater than Minor (but still not in the Dangerous category).

Although these figures suggest a desirable state of affairs in the highly relevant situations studied, a word of caution should be injected to the effect that it is not reasonable to expect large and obvious changes in injury from the over-all injury spectrum. This is because when the "appreciable" effect of padding in the highly restricted situations studied herein is diluted by introduction of the many other accident situations in which it may be irrelevant, the "appreciable padding effects" fade to a point of being a rather low proportion of the total. Of course, even a perfect energy-absorbing device is of no benefit if something else is struck instead.

The principle of injury reduction by means of padding has been demonstrated in this study. Progress has been made but, nevertheless, the instrument panel still remains as one of the leading causes of injury. Severe and fatal injuries seem to have about the same relative frequency in cars with padding as in those without. Perhaps this indicates that the head can impact the padding, flatten it, and still sustain fatal injuries either from the deceleration experienced or from contact with the surface underneath.

Does this degree of progress represent the maximum that can be expected? It is easy to see that there are many variations in detail configuration of panels and that the design problems may be difficult. In many accidents, decelerative forces are quite high and only a small distance is available in which to bring the head to a safe stop. The instrument panel, of course, serves many functions other than that of "head stopping." It holds radios, clocks, glove compartments, lends structural support to the car and, of course, is one of the foremost objects of styling within the car. These considerations presumably make more difficult the problem of introducing a panel that is more effective in absorbing impact, but they should not discourage the search for better solutions.

This study is consistent with the engineering opinion that has prevailed that the structure under the padding—the stiffness of material and the small radius of curvature at the brow of the panel—can produce serious and fatal injuries when struck as the result of an accident. It would appear logical that a design approach further to reduce instrument panel injury must involve both improved padding and improved understructure from the standpoint of energy absorption consistent with human tolerance.

Another course of action is, naturally, to prevent people from striking the panel at all. It is becoming more and more obvious that the lap-type seat belt will not, in most instances of forward impact, prevent contact with the panel. Upper body restraint, in addition to pelvic restraint, would seem to offer a means of preventing panel contact, and recessing the panel so that a belted person would not strike it is another possibility. Nevertheless, the great majority of accident victims today do not have any kind of seat belt, and for them further improvement in the energy-absorbing capabilities of the panel seems called for.

It is important to note that research limited largely or exclusively to "injury-producing accidents" may produce substantially incorrect conclusions if its objective is the identification of the factors that determine the incidence of injury. Unless such research includes all crashes of given *physical* severity, *including those in which no injury is sustained*, it is often impossible to determine whether given shifts in injury patterns result from increases or decreases in the risks to individuals in such crashes.

A simple hypothetical example will make this clear. Let us suppose that Table I presents a comparison of the severity of injury of drivers *injured* in single-vehicle accidents of automobiles with Feature A and those with Feature B, that the vehicles had no other occupants, and that the purpose of the comparison is to determine which feature is less hazardous in crashes.

TABLE I

	FEATURE A		FEATURE B	
Injury severity	*Number injured*	%	*Number injured*	%
Serious	750	50	600	40
Less serious	750	50	900	60
Total	1,500	100	1,500	100

On this basis one might readily conclude that drivers of crashing vehicles with Feature B run substantially less risk of serious injury. This difference could have resulted, however, from (1) an *increase* in the incidence of Less Serious injuries among those *at risk* in vehicles with Feature B; (2) a *decrease* in the incidence of Serious injuries with Feature B; (3) *increases* in both injury categories; (4) *decreases* in both; or (5) various combinations of increase and decrease in connection with the vehicle characteristic under study.

To illustrate some of these alternatives, let us consider the following possibilities, each of which would produce the *same* summary data as in Table I. Table II presents the same comparison between Features A and B, but here, in addition to the summary data on number and types of injuries, we have included injury rates for the total number of occupants *at risk* in crashes of the two types of vehicles. To illustrate our point we have kept the injury rates and the population at risk in such crashes constant for Feature A and have shown for Feature B various combinations of populations at risk and injury rates which would produce the summary data in Table I. It is now seen that the truly significant figure in the table—the "injury rate per 100 occupants"—can fluctuate considerably and that the true situation can not be detected by considering only those injured.

This is a most serious problem when the issues are as important as those with

TABLE II

	FEATURE A	FEATURE B				
No. of occupants actually at risk in crashing vehicles	7,500	6,000	9,000	3,000	18,000	7,500
Injury rate per 100 occupants in colliding vehicles:						
Serious	10	10	6.7	20	3.3	8
Less serious	10	15	10	30	5	12
No. (and %) of injuries:						
Serious	750 (50%)	600 (40%)	600 (40%)	600 (40%)	600 (40%)	600 (40%)
Less serious	750 (50%)	900 (60%)	900 (60%)	900 (60%)	900 (60%)	900 (60%)

which such research is usually concerned. In the case of some measures (such as padding) where there is much evidence, both theoretical and observational, as to the direction in which the ratios should shift, interpretation may be less difficult. It is also less difficult where many of the individuals tabulated enter the series because of their involvement—whether injured or not—in accidents that are included in the series because someone else was injured. The same applies where the data show a high degree of internal consistency, as in the work reproduced above. The fundamental problem remains, however, to an unknown extent as long as the research design does not provide for the ascertainment of the population at risk in *all* pertinent crashes of given *physical* severity, including those in which no injuries are sustained.

In attempting to resolve this problem of interpretation it is not necessarily useful to demonstrate that data from different areas have similar characteristics, since this may merely show that the same biases are widespread. It is also difficult to interpret evidence showing that the makes and models of vehicles in injury-producing accidents are proportionally similar to those licensed in the same areas without information as to their feature-specific exposure and accident involvement rates. Finally, it is also difficult to interpret evidence that various characteristics of accidents of all types known to official agencies are similar to those in such special studies, since the official data are usually of unknown representativeness and reliability and seldom describe the specific features of interest. For these many reasons, studies of the influence on injury incidence of specific vehicle features should include a far wider range of scientifically collected information than merely data concerning the vehicles and individuals in injury-producing accidents.

The problem with research designs that fail to make such provision has also been discussed in Chapter 2 in connection with the work of King, where lack of knowledge of the population at risk permitted only "internal comparisons" in the data, with the result that conclusions about the sources of the patterns found among the data could not be reached with confidence. It is because of this fundamental problem that epidemiologists insist on information concerning the entire populations at risk and not merely those portions of such populations characterized by the pathology being studied. This is one of the chief differences between the epidemiologic and clinical research approaches.

ROADSIDE DESIGN FOR SAFETY‡

—Kenneth A. Stonex, M.A.

Although it is the collision of the occupant with the vehicle itself—or with its surroundings—that permits injury to occur, the severity of this collision is substantially influenced, especially in the absence of adequate vehicle crash design, by the abruptness with which the vehicle itself is decelerated. For this reason, measures such as separating opposite traffic streams by median strips and eliminating or modifying roadside obstacles have been widely adopted. However, only a few of these measures have been so thoroughly evaluated that the best ways to accomplish their objectives are known with certainty. In the absence of such evaluation, engineers are currently building into new highways such features as the standard guardrail end, which Stonex describes as "completely undesirable" and which proved unnecessarily dangerous when properly evaluated.[40]

The work described by Stonex demonstrates that roadside features can be objectively evaluated with a view to their modification. His paper points out some of the factors that such research must consider and the ways in which one exceptionally well-situated group has approached this problem. Similarly sophisticated and specific work is equally applicable to the *vehicle* features that caused or prevented injury to the occupants involved in the 170 off-the-road accidents that Stonex states occurred in the same facility during the years 1953 to 1958. In fact, injuries sustained in vehicles crashing on this and other proving grounds should serve as one index of the success of vehicle designers in providing the crash protection needed by the general population of vehicle users.

THE SIGNIFICANCE of the roadside in the highway safety problem is apparent from the National Safety Council's statistics. Year after year, between 30 per cent and 35 per cent of the highway fatalities—12,000 to 13,000—in the country occur in non-collision accidents, most of which involve vehicles leaving the roadway.

This factor is second only to the two-vehicle collision as the most important and deadly. A great deal of engineering design effort is being devoted constructively to reduction and elimination of the two-car collision. The divided highways of the Interstate System, the turnpikes and the expressways, grade separations, one-way streets, and signal installations and stop signs on intersections at grade are all attempts to eliminate this type of accident. Intensive driver training, public educational and enforcement programs are corresponding efforts in other fields.

In the policies on geometric design of the Interstate System, recognition is given to the importance of roadside hazards by the adoption of enlightened standards of roadside design with respect to slope, ditch cross-sections, and elimination of obstacles. In reconstruction and modernization of existing roads, some attention is given to this prob-

Reprinted, with permission, from *Traffic Safety Research Review*, 5:4:18-30, 1961, published by the National Safety Council. Portions of the text and most of the figures have been omitted.

‡ See also references 38, 39, and 40.

lem, but one cannot help but feel that the minimum standards are applied all too frequently in construction of portions of the Interstate System; one cannot help but feel that too little emphasis is given to the elimination of roadside hazards on primary and secondary and rural roads, even when they are being reconstructed.

It is the intention of this paper to discuss the problem in general and to give a number of specific examples of roadside hazards drawn in a large part from experience with the proving ground road system. It is the further intention to discuss some research results from which specific design criteria for roadside slopes and ditches are derived. The stability factors of passenger cars are related to observed roadside characteristics, especially with respect to the slope of roadsides and the coefficient of friction of the roadside surfaces. Some discussion is given to the hazards involved in conventional light poles and sign posts, with suggestions as to how this type of hazard may be reduced or eliminated. In addition, further observations on full-scale guardrail impact tests are given to supplement earlier publications on this subject. Particular emphasis is given to guardrail end installations.

Some of the examples of roadside hazards have been discussed previously; they are repeated here as background and to give appropriate emphasis to the problem.

Today the proving ground road system includes about 62 miles of all types of surfaces common to those of the public highway; portions of this system with the highest traffic volume were completed by the middle 1930's. During the past 35 years, over 200 million test miles have been operated, and the current rate is about 40,000 miles a day, or 10 million miles per year.

At the proving ground the normal standards of shop safety have always been employed. Thirty-five years ago, the attitude about industrial accidents, quite generally, was that accidents were something which would happen, and that we had to learn to live with them. The industrial safety engineer has shown that is not so. He has shown that accidents are preventable, and that accidents in the plant usually come about because of some human malfunction. Recognizing that, while accidents are preventable, some will not be prevented, he provides all safeguards he can imagine for all types of carelessness and inattentiveness which people may show so that the effect of an accident may be minimized. When, a few short years ago, we began to apply the concepts of the industrial safety engineer in our consideration of the proving ground safety program, it became evident that the most serious potential hazards lie in the operation of our vehicles on our road system, because both the masses and the speeds of cars exceed those of any of the other machines with which we work.

Generally speaking, the proving ground road system was quite well developed 20 years ago. In construction, the design standards which prevailed generally at the time of construction were used, and our practices were comparable with those of the state highway departments. The basic elements of highway safety were considered with admirable foresight in the design of the road system at the establishment of the proving ground, and one-way traffic, limited access, and avoidance of intersections of main test roads at grade has always been the policy. Thus, the basic design of the system avoided many of the types of accident which are now causing such great concern in the public transportation system. At the time, this approach was unique in its enlightenment and progressiveness.

As we reviewed our accident statistics, we found that, during the six calendar years 1953-1958, inclusive, covering approximately 65 million test miles, there was a total of 236 accidents, of which 72 per cent were off-the-road.

Our first and most obvious concern was to determine the reasons why drivers left the road. In many cases, the driver went to sleep. In other cases, he was obviously inattentive. Educational programs were undertaken, reprimands were given, and in the more flagrant cases, drivers were discharged.

In spite of all this, it became evident that drivers do leave the road, infrequently, it is true, but all too often, simply because they are people, and suffer normal human fallibility.

A further consideration led to the conclusion that it was not possible to keep all drivers on the paved surface all the time. *One of the fundamental principles of safety engineering is to anticipate every possible type of accident which may occur because of machine failure or human failure and then to establish safeguards to minimize the hazards or injury which may result when such a failure occurs.* When we looked at our road system from the same point of view, we were considerably disturbed to find that design standards provided little or no safeguard in the event that there should be a failure of some type. Our predecessors had pioneered in safety engineering by taking fundamental steps to avoid accidents, but they did not apply the second concept of the industrial safety engineer, to provide all safeguards in the event that an accident occurred because of human fallibility.

When we realized that our roadside design did not incorporate the safety features common to our machine shop, garage and maintenance tools, we sought to make amends at the earliest possible time by using the experience of other people. We made immediate comparisons with public highways in the adjacent area; we found that the public highway system repeated most of the shortcomings of our own road system. We looked at highway construction in other states, at the turnpikes, even the newest components of the Interstate System, and we found even there a lack of safeguards which would not be tolerated in any modern industrial operation. Almost every mile on any of these roads has one or more places where the occupants would suffer serious or fatal injury if the vehicle left the road at normal highway speeds.

An approach to the problem of safe roadside design from the attitude of the industrial safety engineer became our primary interest in providing the proper safeguards to our employes operating on our own road system. Every roadway accident includes as factors the driver, the vehicle, and the highway; and these factors were considered properly in our approach to our problem.

Our drivers are adult males in good health; they are selected on the basis of average or above competence and on the characteristics of desirability in an employe; they are all qualified drivers with several years of experience in normal highway driving; and it is fair to assume that they compare with the upper strata of drivers in the traffic stream on the basis of these characteristics. They drive on a closed road system with favorable geometry, relatively low traffic volumes, controlled access, one-way operation, and under relatively close supervision. Extensive training and educational programs have been conducted. Our minds are not closed to the possibility of more effective ways of improving driver performance; we have exhausted the obvious solutions and we have tried some which are not obvious. In spite of this, during the last six calendar years, we have had 170 off-the-road accidents.

The business at the proving ground is the development of automobiles. We are familiar with the scrupulous attention given to safety in design and manufacture, and we are confident that the current automobiles well maintained are practically free from hazardous failures in service.

Our confidence in the quality of the vehicles and our experience with the fallibility of the drivers left us with the firm conviction that the major deficiency in our effort to eliminate accidents was the failure to devote sufficient attention to the road itself. In our case, the concept of the original traffic pattern practically eliminated two-vehicle collisions and confined the problem almost entirely to the roadside.

Common Roadside Hazards

When one adopts the safety engineer's attitude and concepts, the most dangerous situation perceived immediately is that of obstacles adjacent to the road, that is, close enough to the traveled surface so that the driver who has lost control of his vehicle temporarily will strike the obstacle before he has an opportunity to regain control. On our road system, and on nearby public highways, the most obvious obstacles are trees. Trees are desirable and beautiful, and in the early days of the proving ground, we frequently modified the alignment so that a large, beautiful oak might be saved. However, a review of any newspaper shows that trees contribute almost every day to the statistics of injury and death.

Trees have played a part in the more serious proving ground accidents. The severity of an impact at normal speeds is indicated in Fig. 1 [omitted]. This car was driven by remote control into the tree at 35 mph; it was damaged seriously and it was evident to anyone who witnessed the accident that occupants of the car would have been seriously injured and probably killed. To the safety engineer, the immediate conclusion is that trees close to the roadside must be eliminated systematically. Fortunately, with modern equipment this is practical and relatively inexpensive. After the trees are removed and the roadside smoothed with a grader, no possibility of serious accident in this area remains.

This situation has its direct counterpart in the public highway system; on some major highways of relatively modern design, there are trees so close to the road as to be lethal obstacles, and in fact, on some of the newest highways, small trees are being planted as part of the landscaping program. As a result, in all too few years, the small trees will grow into big ones; significant hazards being cultivated immediately adjacent to the edge of the road.

Beauty is possible without large trees, or with trees well in the background; in exceptional cases, well designed guard-rail installations may be used if speed is well regulated.

On many of our city streets and even some of our rural highways, there are man-made obstacles in the form of utility poles and light poles immediately adjacent to the paved surface and signs to guide the traveler.

Conventional light poles are self-supporting structures normally erected on a concrete base and designed to withstand winds of hurricane velocity. Unfortunately, they become roadside obstacles of important dimensions. As a compromise, we have proposed use of a tripod structure constructed of light tubular material with shear mounts flush with the base.

The design was evaluated in a full-scale test. Fig. 2 [omitted] shows three frames from the motion picture record of this test. The car passed through beneath the pole with negligible impact and only superficial damage; one leg, whipping during the collision would have injured occupants of a convertible, but the design can be modified to control this.

Standard roadside signs, mounted at 42 in. above the pavement, are also a hazard. In a collision at 40 m.p.h. a sign of this type pierces the windshield partially and showers the front passenger compartment with glass. At higher speeds, the sign would not drop so far and occupants would find themselves running into a 25-lb. sign at whatever speed the car was traveling.

When the sign is mounted at 60 in., the car passes beneath harmlessly. Roadsigns at the proving ground are being relocated at 66 in.

Ditches

In all parts of the United States, it is necessary to provide some type of drainage system along the road to carry off surface water. These ditches are effective for carrying off the water but they may present a very serious hazard.

Fig. 3 [omitted] shows a typical situation on many miles of rural road; it doesn't take a safety engineer to recognize the seriousness of the inevitable accident when some driver becomes inattentive or falls asleep and leaves the road at this point.

Unfortunately, these practices are carried over to new roads; sharp V-ditches are still being graded almost immediately adjacent to the traveled path of roads being built according to the Interstate System's standards.

In some cases, careless inspection procedures or lack of detail in construction leaves a mound unnecessarily. On rural roads, additional right-of-way must be procured or agreements with the abutting landowners made to abolish such ditches and banks. However, on portions of a modern road system where adequate right-of-way is available, such construction is intolerable to the safety engineer.

The severity of ditch impacts is indicated by Fig. 4 [omitted], which shows a remotely controlled car driven off a road through a ditch with a 2:1 backslope. The car was severely damaged, and it is evident that the occupants would have suffered serious injury, at least.

Fig. 5 [omitted] shows a much milder degree of severity when the backslope is 4:1; in this case, the car climbed the bank with a rather minor impact, and the injury and damage would have been negligible.

The desirability of ditches with flat bottoms and smooth contours and flat slopes has been discussed before but no design criteria have been given. Some preliminary experimentation showed that a car could be driven through a flat ditch with a wide rounded bottom at 60 m.p.h. with ease and comfort.

Ditch Tests

In the test above, it was not clear whether the value of the slope and backslope, the width of the ditch bottom, or the depth of the ditch contributed most significantly to the severity of operation. We made a series of tests to evaluate the severity of crossing a ditch as a function of speed and ditch cross-section elements. In the initial series of tests, a ditch was dug in conformity with the Michigan State Highway Department standards for a median ditch on a divided road. Cars were driven through the ditch at moderate speeds; the driver noted the subjective severity as speed was increased and measurements of "vertical" accelerations were made so that the numerical values could be correlated with the driver's sensations up to the point where the operation became so severe it was unsafe. Remotely controlled tests were continued at increasing speeds through the point of minor to severe damage to the automobile; by extrapolation some estimate may be made of where serious or fatal injury may be produced.

Intuition suggests that a ditch cross-section should be of some curved form to minimize impact; as the suspension system deflects under impact, the unsprung mass of the car will follow a curved path. If the transition from the side slope to the bottom of the ditch is gentle enough that the bumper does not dig in, the unsprung weight and the sprung mass of the car should have a continuous, curvilinear motion. The simplest to consider would be a circular motion as indicated schematically in Figure 6 [omitted].

If the ditch cross-section is circular with radius r, the projection on the path at which the car may run through the ditch becomes elliptical in form; the path will make some angle with the axis of the road, possibly up to 20° or more. . . .

With a given ditch cross-section, the radius of curvature may be estimated graphically, and with the speed arbitrary, the value of the radical acceleration can be computed.

In order to verify the analysis and to develop values of radical acceleration or severity which could be tolerated, three test ditch sections were constructed as shown in Fig. 9 [omitted]. Two sections with a 4:1 slope are taken from the Michigan State Highway Department standards for a median ditch; these have ditch slopes of 4:1 with varying width of the bottom and varying depth to provide longitudinal drainage. Section 3 has the slope and backslope of 6:1 and the depth varying from $4\frac{1}{4}$ to $4\frac{3}{4}$ ft., with an 8-ft. wide bottom.

Typical values of the normal or vertical accelerations were computed from the test sections of 6:1 slope which are shown in Figure 10 [omitted].

The tests were conducted by laying out angles between the car path and axis of the ditch of 10°, 15°, and 20° and by driving the car through the ditch at increasing increments of speed. During each test, recordings were made of the normal acceleration, and the driver's opinion of the severity was noted. Tests were conducted up to the point of extreme discomfort, and an estimate was made of the tolerable value of normal acceleration. It should be noted here that considerable training was involved and the test engineer developed a considerable resistance to this type of operation. This is an experience of considerable severity at the higher speeds, and it may be anticipated that the unwary driver will suffer severe psychological damage before he suffers physical injury. Because of his natural alarm, he is apt to lose control of the car and precipitate an even more serious accident.

After the practical limits of driver tolerances have been reached, test cars equipped with remote control devices were operated in a limited series of tests to determine, if possible, the severity at which structural damage began to appear. It was intended to continue the tests up to the point where we assumed that serious injury would result to the passengers. A limited number of tests were made; these are not considered to be significant, because there should be no serious interest in a ditch section where the severity is beyond the driver's tolerance.

The test data consisted of the values of acceleration measured by a transducer carried on the car such that it measured the accelerations approximately normal to the longitudinal axis of the car and recorded them on an oscillograph; car speed was recorded simultaneously.

* * *

Thus it [was] apparent [from the results] that none of the three ditch sections tested would be acceptable for a primary road or a road where speeds above 50 m.p.h. might be anticipated.

Driving experience at moderate speeds indicates that the severity of impact begins to become uncomfortable at approximately the point where the suspension bottoms, and it approaches the intolerable level when the bumper strikes the ground. At higher speeds, it would be expected that even minor contact with the ground would be injury-producing; therefore, this is a condition which the ditch design should avoid. There is some evidence that suspension systems will bottom heavily under normal vertical accelerations in the order of 0.5 g's which appears to be in the range in which the calculated severity of operation is a first approximation to the average values observed.

Design criteria might therefore be based upon the development of ditch cross-sections which, when projected at reasonable angles of attack, would yield path profiles such that the curvature of the path of the center of gravity could be estimated reasonably accurately and first order approximation to vertical acceleration computed. Conservative design criteria should provide that calculated values of vertical accelerations should not exceed 0.5 g for a car passing through a ditch at an angle of 15° under the anticipated speeds of operation. This would assure reasonably comfortable operation at the design speed and provide a slight margin of safety for the driver who may have been unwise enough to exceed the design speed and unfortunate enough to leave the paved surface at 15° or even some higher angle.

The most important element of the ditch section design in controlling the severity is the length of the vertical curve between the side slopes and the ditch bottom. Obviously the radius of curvature is the controlling element; for design purposes, however, it is much simpler to use a circular vertical curve and employ criteria based on vertical curve length.

Figure 14 defines the elements which are considered in the development of criteria.

Figure 15 shows the relation between vertical curve length and the ditch slope for the arbitrary conditions of the speed of 65 m.p.h. and angle of attack of 15°, which will provide a severity or normal acceleration of 0.5 g. To be noted is the rapid increase in required length of vertical curve as the slope increases.

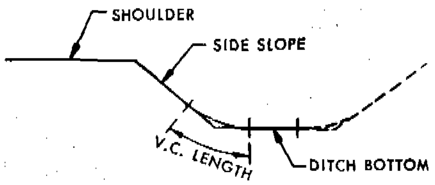


FIG. 14. Schematic elements considered in the development of criteria.

Figure 16 shows the influence of the angle of attack on vertical curve length for the arbitrary conditions of speed of 65 m.p.h. and a ditch with a slope of 6:1 which will produce a severity or minimum acceleration of 0.5 g. From this it will be noted that the severity increases much more rapidly than the angle of attack.

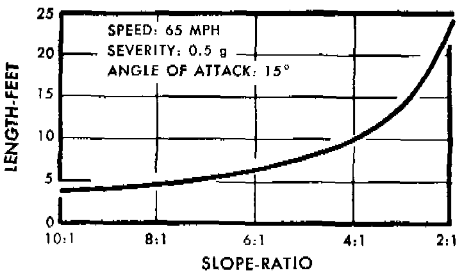


FIG. 15. Vertical curve length as a function of ditch slope for severity of 0.5 g's at an angle of attack of 15 degrees.

The importance of the vertical curve as a design element of the ditch section is emphasized by Figure 17 which shows variation in severity with vertical curve length; conditions assumed are 15° angle of impact, 65 m.p.h. and 6:1 side slope. The vertical curve cannot be left to chance in design, construction, or maintenance.

This concept of roadside safety for highways at current operating speeds provides that obstacles are cleared for a reasonable distance from the edge of the pavement to provide maneuver room for the driver who leaves the pavement, and that the roadside is traversable so he can maneuver satisfactorily; obviously the driver will be unable to control his car if the

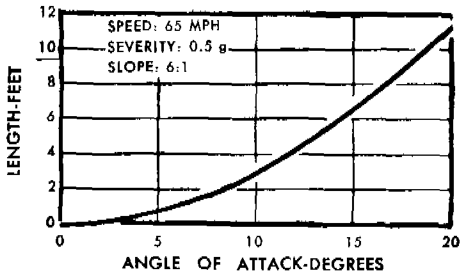


FIG. 16. Vertical curve length as a function of angle of attack for severity of 0.5 g's on a slope of 6:1.

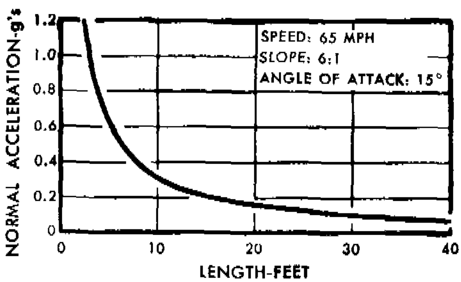


FIG. 17. Variation in severity with ditch vertical curve length.

roadside is so severe that the car is damaged or if he is injured or even severely shaken up or alarmed.

Roadside Slopes

There are little data upon which to base design criteria of the value of the side slopes on fill sections. In some cases, it is obvious that the slope is too steep, or that the transition from the side slope to natural grade was too abrupt. In many cases, we find on relatively flat, gentle slopes that the car slides rather than rolls over and, in certain cases, the car may roll over.

The force and moment systems on a car sliding down a side slope are indicated [in the section that follows. omitted here. Eds.]

* * *

In view of the relatively small reserve of stability provided by current automobiles with lower center of gravity height, careful design and construction of the roadside is a matter of great significance in the design for roadside safety. This leads to the suggestion that more sophisticated design and construction practices for roadside surfaces should provide for compact, smooth surfaces and that maintenance practices should give much more emphasis to preserving this smoothness. The importance of smooth, firm, low coefficient roadside surfaces can hardly be overemphasized in the consideration of roadside safety.

The effect of roadside slope in reducing the tripping deceleration level is of first order of significance; the 6-per cent to 18-per cent reductions noted on Figure 19 [omitted] when the car is sliding down the slopes of 6:1 and 2:1 at a 25° angle may indeed be of great significance.

The value of the slope also has secondary effects since the steeper the slope, the longer the velocity of the car will be maintained and thus the greater will be the possibility of striking some protuberance which will trip it. Furthermore, the steeper the slope, the greater is the weight transfer from the upper to the lower wheel and the greater the indentation into the ground will be and the larger the shear forces may be.

It must be concluded that, for safe roadside design, the slopes must be as flat as possible, not steeper than 6:1 and preferably flatter. They must be as smooth and firm as possible and provide the lowest possible reaction against a car sliding sideways down them.

Unfortunately, there is no manner of specifying roadside smoothness adequately; tentatively it may be said that slopes should be free from stumps, firmly embedded stones, and erosion channels, and smooth enough to be mowed comfortably. The apparent margin of stability factor of even the current automobiles with low center of gravity height is such that relatively small improvements in the flatness and the smoothness of the roadside slopes would make significant reduction in roadside hazards.

Guardrails

Under some circumstances, it will be impossible to eliminate the obstacles from beside the road. Bridge piers must be relatively close; in mountainous terrain, it will be impossible to have side slopes constructed according to the ideal which we have discussed. In other cases, on high fill, the slope of the natural ground will be such that it will be impossible to build a flat, gentle side slope, and a steep fill will be required. In these cases, some use of guardrail must be made to protect against more serious obstacles.

At the last meeting of the Highway Research Board, Messrs. Lundstrom and Skeels reported on a series of full-scale guardrail tests conducted at the proving ground. A major conclusion of their paper was that there was no such thing as a perfect guardrail; that a guardrail was a last resort, and that it should be used only when no other solution was possible. Mr. Beaton of California is reporting this year on a series of tests of median barrier installations which are comparable with guardrails, and thus we have a great deal more information than we had two years ago. However, since Messrs. Lundstrom and Skeels' report was incomplete, we have conducted some additional tests with particular reference to the design of the end installation. Fig. 27 [omitted] shows results of a full-scale impact of a car against the end of a guardrail installed according to the standard installation. This produces a shocking direct collision with an obstacle; it is a completely undesirable installation. We sought to improve on this by ramping the end sections down to the ground to allow the car to slide upward. The impacts were rather moderate, and this approach appeared to be rather promising.

A second test was made at 60 m.p.h. on an installation having a somewhat longer ramp. In this case, the ramp was too steep and the car was pitched violently up in the air. A *third installation* was made with a still longer ramp and with closely spaced posts extending six inches above the rail. The results were somewhat more favorable, but the impact was severe.

There may be other and better solutions to this problem; Figure 28 [omitted] shows probably a nearly ideal condition where the back slope of the ditch was approximately 30 in. above the pavement surface and the guardrail was taken across the ditch and started at approximately the top of the bank so that the end is protected completely. This solution can be applied to equal advantage where the back slope extends well above the level of the pavement, provided that there is a shallow ditch of good design.

In locations where there is no convenient ditch and back slope, a long, low earth mound ahead of the end of the guardrail would appear to have great advantage. Provision is made that one or the other of the wheels might run up on the bank, and when the car reaches the guardrail, it will simply slide along the top. If the car strikes the approach ramp dead center, it will simply slide up the long, gentle ramp with very low impact values.

A variation of this design might be to build a somewhat wider ramp, falling away more slowly as the end of the guardrail is reached, so that the whole car would drive up the ramp and the flat departure slope of the ramp would allow the car to settle down on top of the guardrail gently. We have not had an opportunity to evaluate the previous design. One is left with the intuitive impression that it should be a satisfactory solution to the problem for most installations. It seems clear that almost anything is better than no end treatment at all.

As noted also in the Lundstrom-Skeels paper, there is still uncertainty as to the type of installation which will produce the minimum hazard to the occupants of the car. There is some question about the compromise between minimum hazard to the occupants and to the other travelers on the highway. There is some question remaining about the optimum type of guardrail, whether it be a beam, cable-type, or some fixed impassable barrier. More information is needed on the best type of material, which may be steel, aluminum, fiberglass or wood. There is some question still about the optimum size of posts, their spacing, and the material of which they are made. There are still questions about the best way of mounting the guardrail, whether it should be mounted directly to the post, mounted with a spring connection, or with a solid block.

While it appears that there is a great deal not known about guardrail design, it is clear that hitting a guardrail is an accident, and that installation of guardrails should be avoided wherever possible.

* * *

The single most important point made by Stonex is his statement that "*One of the fundamental principles of safety engineering is to anticipate every possible type of accident which may occur because of machine failure or human failure and then to establish safeguards to minimize the hazards or injury which may result when such a failure occurs* [italics in the original]." This applies to accidents of *all* types, including those involving motor vehicles. Because of the high percentage of vehicles which are at some time involved in pedestrian or occupant injuries or deaths, Stonex's statement must be taken as an unequivocal justification for the need for appropriate modification not only of the environment around the vehicle but also of the vehicle itself so that the passenger and the pedestrian may be protected to the maximum possible extent, an objective which is at present far from realization.

INCREASE IN AUTOMOBILES WITH SEAT BELTS DURING A YEAR OF PROGRAM EFFORTS

—*Berthold Brenner, M.A., Alexander V. Monto, M.D.*

Behavioral factors are among the variables commonly involved not only in the causal sequences that lead to accidents but also in preventive measures. Thus, modification of behavior may not only prevent the initiation of accidents of some types but may also be necessary to ensure that the forces involved, once released, are anticipated and consequently attenuated as fully as possible. The latter objective is dealt with in the work that follows, which reports the evaluation of a behaviorally oriented experimental program designed to increase the installation of seat belts. This study was not concerned with the related question of the degree to which belts, once installed, are used. Exploratory investigation by the Cornell group shows, however, that in the areas of study a majority of belts were actually in use among those in accidents.[41] Despite these findings, it would obviously be preferable to design vehicles so that occupant restraint was either automatic or had to be used before the vehicle could be set in motion, since this would eliminate decision on the part of the user.*

IN VIEW of the evidence for the potential of automobile seat belts in preventing injuries, the Public Health Service, the American Medical Association, and the National Safety Council undertook, in 1958, a national program to encourage the use of seat belts. A community demonstration project was one of their first efforts. Its purpose was to gain experience with, and to evaluate the effectiveness of, a community seat belt promotion program. The place chosen as a community laboratory was Allen County, Ind., whose chief city is Fort Wayne.

Allen County was selected as the site on the basis of several considerations. The community chosen had to be small enough so that, with available resources, the program would have a sufficient impact to produce measurable changes. It had to be of a size which could economically be surveyed to evaluate program efforts. A strong local safety council and a cooperative health department and medical society were considered necessary. The community had to be self-sufficient, not a suburb of a metropolis, so that program efforts could be localized. Finally, it had to be reasonably accessible to the sponsoring national organizations.

The program was planned and organized during 1959. The Narrow Fabrics Institute, the trade association of seat belt webbing manufacturers, made a grant of $12,000 to the National Safety Council, which provided salaries for a full-time project director and secretary. The Accident Prevention Division of the Public Health Service was responsible for the evaluation surveys and related analyses. This included a survey of a control community where no major seat belt promotion program had been undertaken. The control community chosen was Vanderburgh County, Ind., whose chief city is

[Reprinted from *Public Health Reports*, *78*:1:22-26, 1963, published by the U. S. Public Health Service, Washington, D. C. One figure and the summary have been omitted.]

* See reference 33 for a brief discussion of public health measures requiring active as opposed to passive cooperation.

Evansville. Office space for the Allen County program was provided by the Veterans Administration Hospital in Fort Wayne.

A saturation program was planned to reach every citizen in Allen County repeatedly, not only through the mass media but also through his business or labor group, his church, his service and civic clubs, his children's school—through every possible channel. The program began in February 1960.

The Program

The Fort Wayne-Allen County seat belt program was launched at a kickoff luncheon with representatives of the press, radio, and television on hand. Throughout the campaign the newspapers kept news of the seat belt program before their readers. The radio stations scheduled spot announcements concerning seat belts. The television stations filmed every important seat belt event and produced a 30-minute show dealing with the use of seat belts. In addition, a state police officer who conducted a daily television program on road conditions made frequent references to seat belts.

The cooperation of other community groups was also obtained. One community service club offered to market seat belts within the club membership. A meeting with local law enforcement agencies of the Fort Wayne-Allen County area resulted in an agreement to supplement motor vehicle accident reports with judgments as to whether seat belts did or could have prevented deaths or injuries. Officers with stories on seat belt experience were among the speakers supplied by a specially formed speakers bureau.

An important contact was the city school system. Assemblies featuring well-known traffic officers were organized for the city's junior and senior high schools. A professional race car driver devoted part of his presentation at several assemblies to seat belts. A public speaking, poster, and essay contest held in the schools was also used to create interest in the campaign. Seat belts were installed in all driver-trainer cars, and the use of seat belts was discussed with the driving instructors.

The cooperation of industrial and commercial groups was obtained. A meeting with the gasoline retailers association resulted in key service stations installing seat belts bought elsewhere in the city, thus allowing the sale of seat belts to be carried on in such outlets as grocery stores and department stores. All casualty insurance groups were given an outline of the campaign and were offered information on seat belts for their organizational meetings. The companies reacted to these approaches by donating a sufficient number of seat belts to equip local police cars and official vehicles. Letters were sent to the seat belt manufacturers inviting their cooperation, and several sent representatives to aid the program staff. One company supplied a demonstration bus equipped with seat belts and display material. Parked at a busy control intersection, the bus aroused the public's attention and was publicized by the press, radio, and television. For 1 week the bus was used as a central headquarters for the distribution of seat belt information. In addition, the manufacturers contracted for radio and television spots featuring seat belts. Local advertising agencies assisted in developing radio spot announcements and video spot tapes.

Displays and demonstrations were also employed. A board mounted with photographs of automobile wrecks was standard display. At a Scout-O-Rama, a car seat, one side equipped with a belt, the other without, was mounted on rollers. With two Boy Scouts in place (one belted in, the other not) the device was released down an incline. When it jerked to a stop at the bottom, the belted boy remained sitting in his seat while the other occupant tumbled forward onto a straw mat.

Packets containing brochures, face sheets, and posters were distributed to numerous individuals and service, educational, and religious organizations. For groups desiring more information, the speakers bureau provided qualified speakers and leaders for discussions.

Program Evaluation Surveys

Three sample surveys were conducted to evaluate the Fort Wayne-Allen County seat belt program. The first survey was conducted in Allen County (including Fort Wayne) shortly after the seat belt program was launched. It was designed to estimate the percentage of passenger vehicles with seat belts as of the beginning of the program, in mid-February 1960. The second survey was conducted in Allen County shortly after a year of program efforts and was designed to provide a comparable estimate for mid-February 1961. In addition to these before-and-after surveys in Allen County, a third survey was conducted in Vanderburgh County, Ind. (including Evansville), in order to estimate the percentage of passenger vehicles with seat belts as of mid-February1961 in a comparable county but without a year of intensive local seat belt program efforts. In each of the surveys the owners of a 10-percent sample of the registered passenger vehicles were queried by mail as to (*a*) whether seat belts were installed in their cars, (*b*) the seat positions in which the belts were installed, and (*c*) when the belts had been installed.

Persons who failed to respond received a follow-up mailing about 2 weeks after the first mailing. Those not responding to the second mailing were contacted by phone or in person about 2 weeks after this second mailing. A reply as to whether seat belts were installed was received from about 90 percent of those included in the samples. In deriving estimates it was assumed that vehicles for which the required information was not obtained were distributed in the same proportion as similar vehicles for which information was obtained.

Since the estimates presented in this paper are based on sample surveys, they are subject to sampling errors. The chances are about 95 out of 100 that a complete enumeration would have yielded a percentage well less than one unit higher or lower than the estimated percentage of vehicles with one or more seat belts. For a specific category which includes relatively few vehicles, such as all 1960 model vehicles, chances are 95 out of 100 that a complete enumeration would have yielded a percentage less than about 3 units higher or lower than the estimated percentage.

Increase in Vehicles With Seat Belts

As of February 1960, an estimated 4.0 percent of the 76,000 passenger vehicles in Allen County were equipped with one or more seat belts. By February 1961, after a year of program efforts, this percentage had risen to 8.3, whereas only 4.5 percent of the 53,000 passenger vehicles in Vanderburgh County had one or more seat belts as of that time. Most of the difference between the two communities is undoubtedly due to the effects of the Fort Wayne-Allen County safety seat belt program. It must be noted, however, that there was apparently a somewhat lower percentage of vehicles with seat belts in Vanderburgh County at the beginning of the period in question. A necessarily rough estimate, based on replies to the question as to when seat belts were installed, indicates that the percentage of vehicles with seat belts in Vanderburgh County as of February 1960 was closer to 3 percent than to the 4 percent estimate for Allen County.

It is estimated that the rate at which current model cars in Allen County were being purchased with seat belts or equipped with seat belts shortly after purchase increased from 7.9 to 16.2 percent (fig. 1).

Among vehicles of various makes and model years, the higher the percentage with seat belts at the beginning of the program, the greater tended to be the increase in the percentage with seat belts during the program year. This is illustrated in figures 1, 2, 3, and 4, which show both higher initial percentages and greater increases for late model and foreign cars. Apparently social, economic, and automotive factors associated with the installation of seat belts in the absence of an

THE CASE OF FLIGHT 320

—*Morton M. Hunt*

The crashes of modern aircraft invariably involve complexities which make their investigation exceptionally difficult. It is to illustrate the approaches employed in research of this type that we have chosen as our final selection Hunt's documentary account of the investigation of the crash of a commercial aircraft.

This work also illustrates many of the points we have previously emphasized: the importance of considering accidents in intimate relation to the environments in which they occur; the need to consider factors of many types and to use a variety of data sources; and the hazards in basing conclusions on common-sense presumptions. It also illustrates dramatically the difficulty in uncontrolled work of reaching firm conclusions about causation in the absence of unequivocal evidence of a mechanical failure or some other change in the system incompatible with its continued functioning.

LIKE MANY OTHER facets of present-day technology, the operation of passenger airplanes has developed so rapidly as to leave public understanding far behind. The average person still thinks about flying in rather primitive terms, and on a stormy winter night, hearing the drone of a plane somewhere in the turbid maelstrom above, he is apt to burrow deeper under the bedcovers and think pityingly of the poor souls aloft, as though they were battling for their lives in a wallowing schooner with its canvas in shreds, its mainmast splintered, and its seams leaking freely. The night of Tuesday, February 3, 1959, was just such a night over much of the northern part of the East Coast, and in particular over the New York City area. Muddy clouds hung low all evening, trailing still lower tatters of scud that wiped clammily across the ground; hour after hour, drizzle alternated with chill rain showers or stinging sleet. Patches of fog and mist filled the hollows of the suburbs and slithered across the city's rivers and bays. At LaGuardia Airport, a raw, fitful wind whipped freezing rain into the faces of those entering the terminal to catch a plane or to meet incoming passengers. In the chrome-and-plastic waiting room of American Airlines, a score of fidgety people were awaiting the delayed arrival of Flight 320 from Chicago. According to schedule, it should have arrived at 11:05 P.M., but shortly before that time an airline agent had erased "11:05" from the board and chalked "11:40" in its place, and a little later he amended this to "11:55." Every now and then, someone would walk over to the windows and peer into the opaque ocean of air that somewhere contained the oncoming plane.

That night, the wind being generally from the southwest, planes coming in to LaGuardia approached from the northeast, letting down from their normal flight altitude over Westchester County and descending in a straight path over Larchmont, New Rochelle, Pelham Bay Park, and Clason Point, in the Bronx. At 11:50 P.M., just north of the Larchmont-Mamaroneck line, Frank Swenson, a thirty-seven-year-old assistant export manager of a pump and generator company, was about to enter his apartment house, near the Boston Post Road, when he heard a plane passing overhead. Because it seemed to be flying rather low, he paused to listen; he had noticed ice forming on the ground, and he wondered whether the unknown plane was in any trouble from ice forming on its wings. A minute or so later, in a house about three miles to the west, at 31 Inverness Road, New Rochelle, Dr. Gerhart Schwarz, a forty-six-year-old radiologist, was awakened by a peculiar-sounding and apparently very low airplane; he rushed out onto his terrace and peered into the black skies, but could see nothing. Shortly afterward, Mrs. Julian Allen, a middle-aged housewife, heard a plane passing over her house, near Pelham Shore Road and Pelhamdale Avenue, just below the southern boundary of New Rochelle; it seemed to her that the engines were laboring and that the plane sounded as if it were "lacking in buoyancy." She wondered what was wrong, but the sound died away, and so did her moment of apprehension. Some four miles farther to the south, at about 11:54 P.M., Percy Tumber, a forty-nine-year-old greenkeeper at the Oakland Golf and Country Club, in Bayside, was riding in the back seat of a car headed down the Hutchinson River Parkway toward Whitestone Bridge, when suddenly he was deafened and shaken by a torrent of strange high-pitched sound. Looking up through the car's back window, he saw a large plane directly overhead, skimming through the fringes of the cloud bottom, with the lights on the ground reflecting from its gleaming belly. It was making an unearthly whining sound, and it seemed to him to be no more than a hundred feet up; he had a disturbing certainty that something was very wrong.

THERE is little doubt that what at least some of these people were hearing was Flight 320, but the troubles they imputed to it existed principally in their own uninformed minds. Flight 320—a stubby-winged giant aluminum cylinder bearing the registry number N 6101A and named Flagship New York—was in perfect condition in the final minutes of a routine trip from Chicago. The plane, a type known to the trade as an L-188 and to the public as a Lockheed Electra, was brand-new, from its rounded crimson nose to its red rudder assembly, a hundred and four feet aft. Its air speed of a hundred and seventy-five knots was perfectly safe and proper for this part of the approach to LaGuardia (in full flight it cruised at close to four hundred miles an hour); its four Allison turboprop engines, far from laboring, were loafing along at fourteen hundred horsepower each (they could develop nearly four hundred horsepower each when necessary); and the queer, high-pitched sound heard by people on the ground was the normal but unfamiliar one caused by the turbines of the four engines spinning at 13,820 revolutions per minute and expelling great blasts of hot gases. As protection against ice—which was not then forming on the plane, even though the ground was freezing—there was a hot-air system for the wings, a defrosting spray for the propeller blades, electric heaters for the Pitot tubes (through which air reaches the air-speed indicator), and a warning light on the pilot's panel to inform him of any build-up of ice on the front of the fuselage. The Electra at that time was the very latest type of plane to go into commercial service. It had been on the drawing boards and in the design laboratories at Lockheed for several years; in its subsequent testing phase, with maddening caution, the Civil Aeronautics Administration (later rechristened the Federal Aviation Agency), had spent nearly twenty thousand hours inspecting its parts and test-flying it, to make sure that every particle of it was up to federal requirements. On August 22, 1958, the C.A.A. had finally certified the Electra for passenger operations, and on December 5, 1958, American Airlines had received from Lockheed the first of thirty-five Electras it had ordered. A strike by American's pilots, lasting from December 19, 1958, through January 9, 1959, delayed the introduction of the Electras into standard operations, but on January 23rd, after the strike was settled, the New York-Chicago run of the Electras was inaugurated, and by the night of February 3rd the new American Airlines planes had made nearly two hundred uneventful trips between the two cities.

The three men in the cockpit of Flagship New York would have smiled tolerantly at the apprehensions of Mr. Swenson and the others. Captain Albert DeWitt, a strong-chinned man of fifty-nine whose dark hair and trim mustache made him look far younger than his age, was qualified to fly nine kinds of commercial airliners, and in the past three decades had spent 28,135 hours, or more than three years, off the ground. Twenty-five hundred hours of this had been "instrument time," in which he was sealed in by weather or darkness and unable to fly by direct observation. The co-pilot, Frank Hlavacek, and the flight engineer, Warren Cook—men in their thirties—had accumulated far less flying time than DeWitt, but by all current standards both were seasoned veterans. As the plane headed southwest over Westchester, Captain DeWitt was at the wheel, Cook was monitoring a broad panel of dials and indicators that gave him a clinical picture not only of the health of the four engines but of all the mechanical and electrical systems of the plane, and Hlavacek, in the co-pilot's seat, was working the radio. On a control pedestal at his left hand were the tuning handles of three different radio receivers, and en route he had talked with federal air-traffic controllers and with American Airlines employees at Chicago's Midway Airport, at Idlewild, and at LaGuardia, reporting the plane's position and receiving weather news and verbal clearance to continue the flight.

To laymen, the night of February 3rd was a formless mixture of wind, sleet, and clouds; to the three crew members it was a neat pattern of narrow, interlacing radio beams. Each beam had a fixed direction and frequency, and was identified by a Morse-code signal given at regular intervals. Tuning his radios to the frequencies of these beams, Hlavacek could tell, by the sound he picked up and by the markings on an aeronautical chart in his lap, where the plane was at any moment. The LaGuardia radio beam, which Hlavacek had tuned in while the plane was still over the middle of New Jersey, emanated from an antenna located in an open lot in the Clason Point section of the lower Bronx, just 2.8 nautical miles across the East River from the airport. From that antenna, four spokes of steady tone on 209 kilocycles stretched out reassuringly into the storm. One, reaching almost forty miles up toward Ridgefield, Connecticut, was in a direct line with LaGuardia's Runway 22, on which Flight 320 was to land. But far more precise than this beam was another one, on 109.9 megacycles, which originated in a tiny hut at the near end of Runway 22. This was called the "back-course" beam, to distinguish it from the standard, or front, beam, which guided planes coming in from the opposite direction. As soon as the Electra flew into the range of this precision beam, the electronic ears of the plane's I.L.S. (Instrument Landing System) radio equipment would automatically activate a dial on the instrument panel, its vertical needle moving to one side or the other, according to whether the plane was veering off in either direction from the straight path down the beam. And that was not all. As the plane moved down the LaGuardia beam over New Rochelle, it crossed a "fan marker," or highly localized transmitter, which directed a fan-shaped radio signal straight up into the air. As soon as the plane passed through this beam, a white light would show on the instrument panel, indicating that the runway lay 7.6 nautical miles ahead; when it passed over the LaGuardia radio-beam transmitter at Clason Point, the white light would show again, indicating that the runway was now 2.8 miles away and the plane could safely start its final descent. And, in addition to these electronic safeguards, Hlavacek was getting instructions by voice from LaGuardia. As a plane heads in for LaGuardia from Chicago, it enters the jurisdiction of the field's Approach Control officer somewhere over New Jersey; when it is on the final approach to the runway, it is transferred to Local Control. Hlavacek had been in the knowledgeable hands of Approach Control ever since passing over Jersey City. Altogether, a man feeling his way across his own bedroom in the dark could have been no more at home than Hlavacek and DeWitt were in the blackness of that winter night.

Unaware of all that was being done to get them in safe and sound, the sixty-eight passengers in the cabin buckled their seat belts and waited for the landing. By and large, they were successful professional people, travelling in pursuit of their careers. The group included, among others, a television producer; a public-relations executive; an authority on photosynthesis from the University of Illinois; a magazine editor; a lawyer from the Wall Street district; a Bell Telephone research engineer from New Jersey; four Protestant clergymen; a steel man, a valve man, a paper man, and a fat-and-oils man; several old men; several young women; a college senior; a boy of eight; and an eleven-month-old infant. The two stewardesses finished clearing away the remains of the snack served during the flight and then took seats in the lounge, belted themselves in, and waited for the touchdown.

Since 11:34 P.M., when Flight 320 passed over Jersey City, it had been in the charge of an Approach Control operator named John Grula, who was sitting before a radar screen in a dimly lit room in the tower above the LaGuardia Administration Building. (Like all other air-traffic-control personnel, Grula works for the Federal Aviation Agency.) In response to his orders, the spot of light representing Flight 320 on his screen had obligingly crawled up toward the Tappan Zee Bridge of the New York State Thruway and had then begun a large, slow, counterclockwise circle over Rockland County, descending according to his instructions.

"Turn left now, heading three-six-zero for the LaGuardia back course," Grula said at 11:39.

"Roger," replied the voice of Hlavacek. "Turning left, three-six-zero." And the little spot of light began to inch around in a circle.

"Now, American Three-twenty," called Grula two minutes later, after shepherding a couple of other flights along, "turn left, heading two-seven-zero, two-seven-zero. . . . Barometric pressure now two-nine-seven-seven, two-nine-seven-seven."

Again Hlavacek acknowledged reception. Step by step, Grula swung the spot of light around in a complete circle while ordering it to lower altitudes; then he brought it eastward across upper Westchester, and finally swung it around to head southwest, until he could see, by its position on his screen, that it was coming into the narrow zone where it would pick up both the LaGuardia radio beam and the I.L.S. back-course beam. "American Three-twenty," he called at 11:48. "Cleared for a back-course approach to the airport. Turn right, heading two-one-zero, to intercept the LaGuardia back-course. When you do, take over and complete your approach."

A moment later, Hlavacek called back to say they were now on the I.L.S. back-course beam and were taking over, and another moment later he announced that they had just passed over the New Rochelle fan marker.

At this point, Flight 320 was ordered to switch its radio to 118.7 megacycles and report to Local Control. Fred Prawdzik, the F.A.A. employee in charge of Local Control, was sitting upstairs in the glassed-in tower room, looking out through the slanting windows toward Runway 22. At 11:53, he heard on his radio, "LaGuardia Tower, American Three-twenty. We're by New Rochelle." Prawdzik had only one other plane in his sector at the moment—a Northeast Airlines DC-3 several miles closer to the runway and in a position to land safely before the advent of Flight 320—so he left the Electra's course unchanged. "American Three-twenty, LaGuardia Tower," he radioed. "Report the range station [the second marker, over Clason Point]. Straight in, Runway Two-two, wind south-southwest, seven." In another minute or two, Prawdzik saw two lights coming down out of the clouds over the East River as the DC-3 broke through and drifted toward the runway. Just then, Flight 320 announced that it was over the

Clason Point transmitter, 2.8 miles from the end of the runway, and as soon as Prawdzik had seen the DC-3 roll down the runway and turn off, he called back. "American Three-twenty," he said. "Cleared to land straight in, Runway Two-two, wind south-southwest, eight."

"Three-tw—twenty," replied Hlavacek's voice at 11:55:27.

Prawdzik, knowing the position of the plane and the normal approach speed of an Electra, figured that it would break through the cloud ceiling in about half a minute and touch down about half a minute after that. He waited, watching attentively. Nothing appeared. He glanced at the sweep second hand of his clock, moving imperturbably around the dial; the half minute passed, then three-quarters, and then the minute, and still there was nothing to be seen but the runway lights and the haze-dimmed red lights of a gas tank across the river. Prawdzik pushed his microphone button. "American Three-twenty, LaGuardia Tower," he said. "Do you read?" He paused, but there was no answer. He called again, peremptorily; there was no answer. He called a third time, a trifle stridently. There was no answer. On the assumption that the plane had missed the runway and was going around for a new approach, he told Flight 320 what altitude to climb to and then began calling the other controllers on his intercom and telling them, with some urgency, that he no longer knew 320's position. Within two or three minutes, the controllers had stopped all inbound and departing traffic at the field; every approaching plane would have to hold its position until Flight 320 was found. Prawdzik and the other controllers kept calling the plane, on several different frequencies, but without success. A radar scanner spotted an unidentified bright dot moving west from New Rochelle and thought it might be 320, but an Idlewild controller reported that it was a plane under his control. "Three-twenty," Prawdzik called again and again. "Come in. Do you read? Three-twenty, do you read?"

In the plane itself, when Hlavacek called LaGuardia to announce that they were over New Rochelle and, later on, over Clason Point, everything was in perfect order for the final approach. The landing gear and flaps were lowered, and after Clason Point was passed, DeWitt eased the power back to twelve hundred horsepower per engine, reduced the air speed to a hundred and forty knots, and established a gentle rate of descent. Flight Engineer Cook had his hands on the alternate set of throttles, ready to adjust them if there was any deviation from the power setting that the Captain had chosen. DeWitt was guiding the aircraft. Hlavacek was monitoring the flight instruments and outside conditions, and standing by to report them to DeWitt. Glancing at his altimeter, he called out, "Six hundred feet." From here on, he would call out the altitude every hundred feet; if DeWitt did not break out of the cloud ceiling at four hundred feet, he was required by federal air regulations to apply power, start climbing, and call the tower for new orders. All of a sudden, Hlavacek saw lights appear outside DeWitt's window. The next instant, there was a terrific concussion, a vast wash of water, and a blackness and coldness as of the end of time. Hlavacek was first aware of a tearing pain in his midriff, where his seat belt held him fast; then he noticed that he was underwater. He fumbled at his seat-belt buckle, somehow got it open, and was astonished to find himself floating straight upward and breaking the surface of the East River, off Riker's Island. Despite the darkness, he could see a segment of wing floating nearby, and he climbed onto it; there he sat, stupefied and infinitely bewildered, hearing all around him a babble of cries and screams.

Flight Engineer Cook heard the mighty sound of the plane hitting; then everything was pitch-black and he was underwater, his legs trapped against the control pedestal. After wriggling wildly and without hope for an endless span of frozen time, he came loose, floated up with a searing pain in his lungs, and burst into the night air near Hlavacek, who feebly pulled him onto the wing.

"Where is Al?" mumbled Hlavacek, through a smashed jaw.

"I don't think he made it," Cook gasped.

In the cabin, everything had seemed to rip open all at once, and through huge rents in the fuselage the paralyzingly cold salt water rushed in. Herbert Forman, a thirty-six-year-old engineer, had been dozing when he felt the impact; immediately, an incredible weight of water poured over him. He clawed free of his seat belt and trousers, flailed around a bit, and to his amazement found himself floating in a dark river, with faint lights visible in the distance. Nearby, he made out a dark shape and swam for it; it was the wing segment, and Hlavacek and Cook soon dragged him onto it. Seymour Kemach, a display salesman from Brooklyn, had been chatting in the lounge with one of the stewardesses; with the crash he blacked out for a second, and then came to his senses to find the broken cabin rapidly filling with water. In the darkness, he was aware of a tangle of seats, floating floor sections, and screaming people. Stumbling to the door, he yanked at it several times, finally got it open, and fought his way out, pulling with him a young boy and a couple of women. They all climbed onto what seemed to be part of the tail assembly, which was floating nearby. Altogether, nine people got out of the wreck alive. One died the next day. Sixty-four were killed in the crash.

While Flight 320 was in the final minute of its journey, the tugboat H. Thomas Teti, Jr., was slowly hauling two empty barges down Riker's Island Channel in the East River. The lights of Runway 22 were about a mile ahead, but because of the rain and low-lying fog they could not be seen from the tug; the Teti was being steered by radar, with visual checks of channel markers and buoys when they were close enough at hand. In the pilothouse, Captain Samuel Nickerson, a spare, sombre man of fifty-seven, was talking with his co-captain, Everett Phelps, who had just come in to relieve him, when they were startled by a tremendous concussion off to starboard, toward Riker's Island. On deck, someone yelled "Plane wreck!" and Nickerson, who had long fretted about the lowness of the planes coming over this section of water in bad weather, responded with an immediate order. "Let go the scows!" he shouted, and deckhands scrambled out to cast off the hawsers. Nickerson called for full power and a hard turn to starboard, and threw on the searchlight. The tug, her light probing spongily through the fog, made for the scene of the crash, and after several minutes Nickerson and the deckhands began to see floating wreckage and the pale circles of desperate faces. For the next ten minutes, all was the ghastly confusion of a sea rescue at night. In the midst of it, the tug radioed the news to the Coast Guard, which soon notified the controllers at LaGuardia where their missing plane was. At ten minutes past midnight, a police launch appeared, but the rescue work had been done; nine survivors were aboard the Teti, and nothing else was visible but small scraps of wreckage. The Teti headed for a police dock at College Point, Queens, a mile from the airport. There the survivors—six passengers, plus Hlavacek, Cook, and a stewardess named Joan Zeller—were put ashore and taken by ambulance to Flushing Hospital or Queens General Hospital. Newspaper reporters kept asking what had happened. No one knew.

THE task of finding out what happened was one of the duties assigned by Congress twenty-two years ago to the Civil Aeronautics Board, a small, independent federal agency whose province is the regulation of air-carrier routes, airmail subsidies, and the cost of air travel, in addition to the investigation of all accidents involving civilian aircraft and the determination of their causes. On the night Flight 320 crashed, the C.A.B.'s chief investigator for the New York area was Joseph O. Fluet, a stocky, wavy-haired man of fifty-one who had been, successively, an airplane mechanic, a private pilot, the operator of a small-town airport, and a federal check pilot before coming to the C.A.B. in 1942. Fluet, whose broad shoulders and dented nose suggest an ex-boxer but whose manner is gentle and a bit professorial, had gone to bed in his house, in Great Neck, shortly before midnight. A few minutes later, the phone rang. It was a flight controller at LaGuardia with the bad news. "Oh, good Lord!" Fluet said. "Any indication of difficulty during the flight? . . . None? . . . No word yet about survivors, I suppose? . . . All right, now tell me how the weather was at the time. . . . Did any flight land just before him? . . . Aha! Please send someone off to catch the crew of that DC-3 and tell them I want written statements about conditions during their approach and landing. I'll make the tower my headquarters until morning. O.K.? . . . I'm on my way." It was the beginning of his hundred-and-twenty-sixth air-crash investigation.

At the field, Fluet hurried up to the large glass-enclosed control room at the top of the tower, where a group of F.A.A. men and airport officials were waiting for him. His first action was to go to the west windows, facing Runway 22, and take notes on the ceiling and the visibility. Then he asked if anyone had ordered a check of the back-course transmitter, the main LaGuardia radio beam, the fan marker, and the controllers' radar. He had no particular reason to think that any of them had misled the plane, but some possibilities had to be investigated at once, the evidence being perishable; if any radio device had been malfunctioning, routine maintenance might fix it and leave no clue. An F.A.A. man said that technicians were already working on the job.

"Good," Fluet said. "Next, will one of you please phone the police and see if they can give us a list of all survivors and their whereabouts?" Another man volunteered to see to it. "Now," said Fluet, "who were on Approach Control and Local Control at the time?" Grula, Prawdzik, and several men who had been on alternate positions spoke up. Fluet led them over to a vacant desk and began to question them about whether Flight 320 had mentioned any mechanical troubles, obstreperous passengers, icing, or turbulence; then he queried them closely about the plane's conduct as seen on the radar screen. Grula said he thought that Flight 320 had been somewhat slow in responding to orders, but Prawdzik, who used no radar screen at the Local Control position, said that he had noticed nothing unusual on the radio, except that the co-pilot's voice had faltered a bit on the last transmission.

It was soon clear that the controllers could provide Fluet with no obvious clue to the cause of the accident. "I'd like each of you to make a full written statement," he said, "and I'd like to hear the tapes myself right now." (Controller-aircraft talk is continuously recorded at all major airports.) A few minutes later, he was seated in the F.A.A. supervisor's office, on the third floor of the Administration Building, listening to the playback and timing the intervals with a stopwatch. He heard nothing revealing or incriminating, but still he was making progress by exclusion; judging by the words and the sound of Hlavacek's voice, he could tentatively assume that the crash was not due to engine failure, a runaway propeller, fire, or structural failure—though the final elimination of these would have to wait upon examination of the wreckage. Similarly, he could temporarily rule out the idea of illness or asphyxiation of the crew. If it was not exactly a satisfying result, it was at least a beginning.

One obvious and indisputable fact was that the weather had been poor at the time, so Fluet's next stop was the Weather Bureau office, on the same floor of the building, where he catechized a meteorologist named James Dillon about the variability of the ceiling that evening, the extent of icing, the degree of turbulence near the surface, and the times of observed fog patches. The weather, he learned, had been terrible only from the layman's point of view, for the rain had never been heavy, the ceiling had hardly ever dropped below four hundred feet, and neither the icing nor the turbulence had been enough to disable even a small private plane, let alone a powerful, modern Electra. The only suspicious fact that emerged was that the barometer had been falling rapidly at the time, a condition that would have caused the plane's altimeters to read too high unless the crew had reset them before making the approach. A few minutes before landing time, Grula had told Hlavacek that the barometer read 29.77, and this was correct according to Dillon's records, but whether Hlavacek and DeWitt had properly reset their altimeters—a perfectly routine chore—would not be absolutely certain until the instrument panel was found.

Fluet next headed downstairs and along the corridors to the American Airlines flight-dispatch room. The atmosphere in the cluttered, brightly lit office was like that in an army command post where word had just arrived of a collapsed right flank. Losing no time, Fluet asked William H. Miller, American's regional operations officer, to get from Chicago the load manifest, flight plan, and passenger list of Flight 320, so he could look into the possibilities of overloading or dangerous cargo, unusual routing or flying conditions, persons on board who might be the object of sabotage, and so on. He also asked Miller to arrange for salvage operations as soon as possible, and instructed that word be passed to all interested parties that the C.A.B. would be conducting a full-scale investigation and that he would hold an organization meeting at 9 A.M. in the F.A.A. controller-training classroom on the third floor of the Administration Building.

At around 4 A.M., Fluet went back to the tower and got on the phone—to the police, who told him where the survivors were, to the hospitals (neither Cook nor Hlavacek was in any condition to talk), to an F.A.A. communications man (the radio and radar equipment had checked out satisfactorily), and, finally, to his superior, James Peyton, the chief of the C.A.B. accident-investigation division, in Washington, whom he asked for half a dozen expert assistants. Peyton promised to have the necessary investigators in New York by 9 A.M.

By 9 A.M., the controllers' training room at LaGuardia was overflowing. Fifty or sixty men had crowded in and were sitting on the chairs, the training desks, and the window sills. Fluet took a desk at the front of the room and called the meeting to order. There being no clear-cut indication of the cause of the crash, he said, it would be necessary to conduct a complete investigation, with a full complement of specialized teams.

The C.A.B., which has only about seven hundred employees, is one of the smallest of the federal agencies. (The F.A.A., which handles air-traffic control and pilot licensing and maintains the national network of air-navigation facilities, has some thirty thousand.) Moreover, of the seven hundred, only about sixty-five are actual investigators. But the C.A.B. has a powerful lever that multiplies its effectiveness—the self-defensiveness of everyone who might be implicated in a crash. The Air Line Pilots Association, eager to clear its members of blame in an accident, assiduously points out every imperfection in airline maintenance, the equipment of the plane, and so on. The airline, for its part, vigorously draws attention to the faults in such things as F.A.A. navigational facilities and the limits of aircraft design. And the F.A.A., the airport operator, the manufacturer of the airframe (the plane itself, not including the engines), and the makers of the propellers, the engines, and the instruments all have a position to defend. In investigating a crash, the C.A.B. sets up various technical teams on which each of the hostile groups is represented, and harnesses their mutual faultfinding into an effective investigative effort.

Fluet began by organizing the operations group, which would study the crew's flying records, their training on the Electra, the plane's flight plan, the evidence concerning the actual flight path, the evidence of witnesses and survivors, and the like. He named Joseph Zamuda, a C.A.B. investigator from Washington, chairman of the group, and then called for representatives from the F.A.A., American Airlines, the Lockheed Aircraft Corporation, the Air Line Pilots Association, and the Flight Engineers International Association. In a similar manner, he pieced together a structures group, to examine the wreckage for evidence of mechanical failure; a power-plant group, to conduct the post-mortem on engines and propellers; an instrument group, to study the instruments and radio devices; a weather group; and a group to collect reports of maintenance trouble with Electras in general, and with American's Electras in particular. At the end of the session, Fluet again told them all that there was no ground for any one theory about the crash yet, and that every area deserved a thorough study.

EVEN as the organizational meeting was getting under way, the huge derrick barge *Constitution*, of the Merritt-Chapman & Scott Corporation, was being towed out to a police buoy marking the scene of the crash; from the position of the buoy, the C.A.B. men on board the Constitution could see that DeWitt had been considerably off course to the right. When the barge was finally anchored, an elderly, bespectacled diver named Gus Markelson was bolted into his helmet and dropped into the murky river. Steel cables were lowered to him, and after a long wait he surfaced and gave a signal to haul away. The crane took up the slack, and very slowly the long, lovely silver fuselage began to appear. To the men on the barge it looked almost intact, but a policeman in a launch on the far side of it shouted, "Christ! She's split down the middle like a broiled lobster!" Then the cables began to slide, and an instant later the huge shape slowly sank back into the water. By now the tide was flowing too swiftly for further work, and operations were halted until the next slack water, at about 7 P.M. Then Markelson went down again, under the glare of searchlights. This time, as the fuselage was being lifted clear, it broke in two, and the larger section fell back and sank at once. The rest—a piece about twenty feet long, which had been the body of the plane forward of the wings, but from which the whole cockpit had broken off—was slowly hoisted, swung over the deck, and lowered. There was nobody in it. The sleek, polished skin was ripped open and all the seats had been torn loose.

So it went at every slack tide. In the opaque waters, Markelson groped his way about, not always sure what he had found and was sending up. Once it was an immense section of fuselage. Another time it was a piece of wing with two engines. Sometimes it was a tangle of electric wiring, or merely a twisted sheet of aluminum skin. And once in a while it was a seat with a passenger still buckled into it.

By Wednesday night, when Fluet had canvassed his field commanders for all they knew, and had gone home to get his first sleep in thirty-six hours, he had still developed no theory about the accident, but he had begun to reduce the area of uncertainty. He had acquired a considerable amount of negative information, not the least important of which was that the pilot and co-pilot of the Northeast Airlines DC-3 preceding Flight 320 had told Zamuda's operations group that there had been no icing and little turbulence; they had broken out of the cloud ceiling at four hundred feet and had made a routine landing in light rain, without difficulty. Even so, no more than five or ten per cent of the information Fluet wanted had so far been gathered, let alone evaluated. He had long since schooled himself not to think in terms of a tentative answer at an early stage, for fear of overlooking some inconspicuous but crucial item. "What I have to keep always in mind," he told a friend a few nights later, "is that sixty-five human beings died in that crash. Now, what is the objective? Not just concocting a plausible explanation but doing a real job of prevention. It mustn't happen again the same way. If I overlook anything—*Jesus*! The next one will be my fault. I *have* to keep my mind completely open." This attitude, widely cultivated among C.A.B. investigators, causes cases to drag on for weeks, and often for months, to the great annoyance of the public, certain congressmen, and lawyers ravenous for damages. In the end, however, the C.A.B. locates what it cautiously calls the "probable cause" in ninety-one per cent of its airline-accident cases. About half of these findings have resulted in specific "fixes"—perhaps a change in the control wiring of a runaway propeller, perhaps the strengthening of a tail member, perhaps a revision of some bit of approach procedure that had caused a near collision. Collectively, these fixes have been in large part responsible for the decrease in the airline death rate from 2.28 per hundred million passenger miles in 1939, to 1.6 in 1946, and to less than one from 1952 on.

THURSDAY morning, Zamuda phoned the Flushing Hospital and learned that Hlavacek was still in critical condition and was under heavy sedation. But at the Queens General Hospital, doctors said that Flight Engineer Cook, despite a fractured breastbone and internal injuries, would be able to talk. Gathering his operations group and taking along a tape recorder, Zamuda led the way to Cook's bedside. Cook was flat on his back, feeble and pale, with his right cheek and chin bandaged and a drainage tube up his nose. Zamuda, a tall, casual, square-faced man who used to be a pilot, greeted Cook with solicitude, and the flight engineer smiled and replied in a weak, halting voice. Placing a microphone near Cook's head, Zamuda started by asking him about the plane's operating condition. Cook said that everything had been mechanically fine during the flight. He also said that until the very instant of the crash DeWitt had been controlling the plane not by means of the wheel and the rudder pedals but by operating the autopilot manually; that is, by constantly resetting its controls, which then adjusted the plane's flight to his prescription. The investigators were mildly surprised; such a technique, though not forbidden, was rather unorthodox. Zamuda then asked Cook to describe the events of the final descent. Struggling against sedation and weakness, he answered, "Uh . . . we were cleared for the back-course approach. . . . We were right on it, and I was holding the power for Al. He said to keep it at one hundred and forty miles per hour—uh, *knots*, I mean one hundred and forty *knots*. I remember it was taking about twelve hundred horsepower per engine to maintain that, and I was concentrating on that, and I just happened to look up and then I saw Al's altimeter, and it was between zero and the hundred-foot mark on the drum—you know, the . . . and I was just going to yell, and we hit. If I had just seen . . ." The members of the group were electrified, and started hotly whispered side-discussions. The Electra was equipped with new precision altimeters, which indicated hundreds of feet by means of a hand that swung around a dial, and thousands of feet on a rotating drum that appeared through a little square window at the right side of the dial; the older type had three hands swinging around the dial—one for hundreds, one for thousands, and one for tens of thousands of feet. Cook said that he had seen an altitude of less than a hundred feet—yet he said that he'd read it off the drum, which had only thousand-foot marks on it.

Noticing the effect of his statement, Cook tried to clarify the matter. "The little drum," he said. "It was between zero and one hundred, and the big hand was on about five hundred, I think it was. . . . Let me think. . . . I realized, just too late, if I'd hollered . . . We hit just then." He lapsed into silence. Zamuda brought him back to the details of the approach. The plane had been in cloud until the last second, Cook said, but everything had been in perfect order—flaps, gear, Pitot heat, engine power, and a score of other things.

"And you're sure the clock hand was at five?" asked one of the F.A.A. representatives.

Cook struggled to speak. "That's . . . and the clock read . . . I was looking right at it when we hit." A doctor bent over him for a moment, then stepped back. "The box was . . ." Cook mumbled. "The box was between zero and the . . . Was it? . . .

Well, it was between zero and one on the drum and then the big hand was on five. We were right at five hundred feet when we hit, I'd say." He sighed and fell silent. After a few more questions, Zamuda thanked him and the group left.

It had been a peculiarly frustrating interview. Many a crash leaves no survivors to tell what they saw and heard. This time, a highly trained crewman had been in the cockpit and had survived the crash, yet he could tell the investigators of nothing that was wrong—except, of course, the altimeter reading, but what he had actually seen on the altimeter, and what it meant, was irritatingly uncertain.

As soon as Fluet learned of Cook's testimony about the altimeter, he decided to redouble his efforts to collect ground-witness testimony that might either support or rebut it. He set up an interview group under the leadership of a tall, hawk-nosed young C.A.B. man named Claude Schonberger. Edward E. Slattery, Jr., a veteran press officer of the C.A.B., ran an appeal for witnesses in the press and on the air, and within a day calls were pouring in. Schonberger and his team—one man from American Airlines, two from the Allison Division of General Motors, and one from the F.A.A.—travelled around the city and the suburbs from morning to night, taking down testimony both in longhand and on tape. In the end, when Schonberger's group plotted out the positions of the witnesses who had offered the most likely-sounding testimony, eight of them lay in a straight line approximating Flight 320's final back-course approach. All eight had thought the plane abnormally low, but only one had actually seen it clearly and could be considered a key witness. He was Percy Tumber, the greenkeeper. He estimated the altitude of the plane he had seen as about one hundred feet. What made his testimony important was that at that geographic point, in a normal approach, the plane should have been at eight hundred feet. Yet if what Tumber had seen was indeed the Electra, and if he had correctly estimated its altitude—and if the ceiling really was four hundred feet, as the Weather Bureau and the crew of the DC-3 had both said—why had DeWitt and Hlavacek failed to see the ground and realize their situation in time? Schonberger's group had other testimony that might explain this puzzle, at least tentatively. Captain Nickerson, of the H. Thomas Teti, had said that because of rain and fog he could see markers and buoys only when they were hard by, and had not been able to spot survivors even with his searchlight until he was within two hundred feet of them. In other words, although the general weather profile showed a ceiling of four hundred feet, there may have been a small, localized patch of fog, low scud, and rain along the line of approach over the Bronx and the East River. It was therefore possible—though by no means proved—that Flight 320 had been abnormally low during its entire approach, and that the crew had not known this from outside evidence.

By now, all phases of Fluet's investigation had been organized and were proceeding simultaneously. Fluet brought the entire fifty-one-man investigative force together every other morning, so that each group chairman could brief everyone else on what his group had learned so far. Sometimes the different groups verified each other's findings; sometimes evidence by one group stimulated another group to try to prove the opposite. Schonberger's witness-interview group tended to feel that they had fairly well shown the plane to have been low all along, but Zamuda's operations group decided to concentrate on the opposite possibility; namely, that the plane had passed over Clason Point at the proper altitude and had descended too rapidly thereafter, because of DeWitt's relative unfamiliarity with the Electra's rate of sink and because of the weather complications.

What all the groups were waiting for was the recovery of the plane's cockpit, which Markelson had not yet been able to find, and while waiting, Zamuda and his men spent many hours studying DeWitt's long flying record. The Captain had never had an accident before, and his whole record was excellent. He had gone to ground school on the Electra for eighty-four and one-half hours, and had flown the plane itself a total of eleven hours and fifty-nine minutes in practice for his official F.A.A. flight check, which he passed on December 16, 1958. But in going through the entire file the operations group also found a couple of provocative items. By F.A.A. regulation, Electra pilots, until they had a hundred hours of time on the plane or a special exemption from their airline, were allowed to land only when the ceiling was higher and the visibility greater than required for other planes. Exemption from this restriction had been given to DeWitt by his airline after he had had only twelve and a half hours of scheduled operations in the Electra. Even more provocative was a flight-check report on him dated December 15, 1958, which noted that he had failed his I.L.S. procedure on the Electra that day. Lawrence E. McShane, an F.A.A. check pilot, had jotted down the comment that DeWitt had twice come in too high over the end of the runway. McShane advised him to practice for a couple of hours more; DeWitt did so, and passed his check the next day. But three days later the American Airlines pilots had gone on strike for almost a month. Conceivably, DeWitt could have been out of practice on the new plane when operations began again. Conceivably, he had made too few instrument approaches in the Electra to override habit patterns acquired during his thousands of hours in the slower-descending DC-7s. Fluet and Zamuda, like all experienced fliers, knew that a pilot who has dropped down to the four-hundred-foot legal minimum without breaking through a reported cloud ceiling is powerfully impelled to try going down fifty feet more to see if he isn't just about to pop out. If DeWitt had done that, he might have flown himself right into the water.

On Saturday, the fourth day after the accident, the police phoned Fluet in the afternoon to say that the diver had found the cockpit and that it would be raised late that night. Fluet immediately alerted all his group chairmen, and by 10 P.M. they were crowded onto the barge Constitution. At last, a little after midnight, the crane began to reel in its cables, and slowly the cockpit came to the surface, upside down, and was lifted out, with water cascading from it—a beautiful egglike shape, shorn horizontally across as if by some mighty scissors. With infinite caution, it was lowered to the deck of the barge. As it was settling on the deck, a barge hand played a flashlight inside. "He's in there!" he cried out. No one said a word.

The instant the cables had been cast off, Zamuda crawled under the cockpit in order to get a look at the instruments and controls before either handling or exposure to the air could change anything. He found himself in the grisly position of having to peer over the shoulder of Captain DeWitt, but he carried on, shouting out readings in the arcane jargon of his trade: "Captain's altimeter minus fifteen hundred feet. Setting twenty-nine point eighty-three. Radio-compass heading two hundred and ten, slave indicator centered. Air-speed indicator showing twenty knots. Landing-gear handle in down position. Pitot heat switch on. Elevator trim minus nine degrees, nose up. Rudder trim minus one degree, nose right. Aileron trim zero." For five minutes, he went on calling off instrument and control positions. Then he relinquished his place to a C.A.B. electronics specialist named Rudolph Duncan, who read the roster all over again; Duncan was the chairman of the instruments group, and since the area of his investigation was so highly significant, it was important for a second complete set of readings to be on record. As Zamuda listened to Duncan's confirmation of his findings, he was painfully aware that the case was still proving an investigator's nightmare—devoid of clues, or even hints of trouble. All indications pointed only to a safe, uneventful landing. Both altimeters, to be sure, showed bizarre readings, but that was undoubtedly because of the water pressure they had been subjected to. The barometric settings on them, however, were still as Hlavacek and DeWitt had left them. Hlavacek's was almost exactly right, while DeWitt's was off a bit; still, his error would have amounted to only eighty feet, not five hundred. Once again the investigators had made progress only by reducing the zone of uncertainty a trifle more. At last, DeWitt's body was taken out and the instrument panel was detached from the cockpit. The body was sent to the New York Medical Examiner for autopsy, at Fluet's request, but the findings were negative; the possibility that DeWitt had become suddenly ill in the closing moments of the flight was ruled out. The instrument panel was carefully carried to American's flight-dispatch office, and there the instruments were removed from it and locked in a steel cabinet.

By the second week of the investigation, the various groups were pursuing their studies more or less independently, meeting only occasionally to compare progress. A C.A.B. aeronautical engineer named Bernard Doyle scrutinized every moment of the plane's recorded operational history (it had flown a little less than three hundred and two hours in all) and retrospectively stood by at every inspection and repair job it had ever undergone. There was nothing unusual to be found.

The structures group, charged with the examination of all the recovered material except the instruments, radio devices, and power plants, was going over every scrap of salvage for signs of failure or imperfection. Every trim tab, hydraulic piston, spar, and cable was examined clinically, but nothing incriminating turned up. In a vacant LaGuardia hangar, the group reconstructed, as best it could, some forty tons of wreckage—ranging from a forty-foot segment of fuselage to scraps of tubing no larger than a man's little finger—so as to see more clearly the precise pattern of impact and destruction. Bit by bit, over a three-week period, a master diagram was filled in, which showed beyond any doubt how the plane had hit the water—in a nearly level position, and travelling moderately fast. There was no longer any need to keep open the possibility of a sudden dive, spiral, or stall. DeWitt had flown the plane right into the river, and not even in the last terrible instant had he known the truth, for the underside of the tail showed no excessive destruction indicating that he had tried to pull up.

The work of the power-plant group was delayed for several weeks while Markelson tried to find the Electra's remaining two engines. The engineer's control panel had been examined piecemeal and the instrument readings had been found normal; perhaps the engines themselves would furnish a clue. When the other two engines were finally located and raised, all four were carefully crated and shipped off to the Allison plant in Indianapolis, followed by the C.A.B. power-plant chairman and his seven assistants—two men from Allison, two from Aeroproducts (the propeller manufacturer), one from the airline, one from the F.A.A., and one from the engineers' union. In Indianapolis, the engines were hoisted up in slings, photographed for the record, and then disassembled. Everything was bent and corroded, but with much tapping and straining the Allison mechanics got nearly all of it to come apart, while the members of the power-plant group stood around like doctors watching an autopsy. After hours of study, the group concurred in a nineteen-page report crammed with figures, tables, and descriptions. Its net message could have been put in two words: Findings negative.

The only group that remained to be heard from was Rudolph Duncan's instruments group, and an added responsibility had been imposed upon them: A few days after the crash, General Elwood Quesada, Administrator of the F.A.A., had ordered all drum altimeters removed from commercial planes until further notice. He was merely being cautious, but to the public his order sounded rather like an indictment for murder. Duncan started work on the hypothesis that the altimeters had failed. "I began with the assumption that they suffered some mechanical defect of a sort that would mislead the crew by five hundred feet," he explained later. "But you understand, I neither believed nor disbelieved this. I merely *assumed* it, in order to test the possibilities." All altimeters operate on the principle of an aneroid, or flexible sealed chamber containing air, which expands as it is carried to higher altitudes and in so doing turns a delicate watchlike series of gears that move a hand around the dial. In good condition and properly set for barometric pressure, the altimeter of a commercial transport plane flying at low altitudes and at landing-approach speeds is accurate to within fifty feet—this being the "static error" caused by the changing air flow around the inlet holes in the surface of the plane. All altimeters, of course, can be affected by dust, moisture, foreign particles, wax, or grease, any of which might cause them to jam completely or—what is worse—to stick temporarily at a crucial time. The old type seldom encountered such troubles, but Duncan thought that perhaps the new type, since they were more complicated and precise, might be more subject to sticking and jumping. While Duncan was waiting for the cockpit to be found, he had begun to search for reports of mechanical difficulties with drum altimeters. The instruments were made by the Kollsman Instrument Corporation, of Elmhurst, Long Island, which had spent years on their design and development and had only recently begun to sell them. Three or four hundred had been delivered to civilian customers, and about seven hundred to the military. At the Kollsman plant, Duncan and his group called on Walter Angst, the designer of the drum altimeter. Angst, a middle-aged Swiss-born engineer, was voluble about the merits of his altimeter. A search of the Kollsman records concerning drum altimeters that had been returned to the company for repair revealed nothing significant—a few complaints of an altimeter's jumping fifty or a hundred feet, but nothing like five hundred. The airlines using drum altimeters were asked to extract from their records all pilot complaints about them. Again, nothing of significance. If anything incriminating were to be found, it seemed, it would have to be in the cockpit.

On the Monday morning following the recovery of the cockpit, the entire instruments group, polite but guardedly mistrustful of one another, convened at Angst's workbench to watch him dissect Flight 320's controversial altimeters. Angst cautiously opened the cases and from each drained off nearly a tumblerful of liquid containing a cloudy white suspension, which proved, on subsequent analysis, to be sea water and the corrosion product of aluminum. But of wax (which might just conceivably have got in through the air inlet when the surface of the plane was being polished) there was no trace, nor of detergents (which might have got in when it was being washed). Angst then rinsed the mechanisms in distilled water, dried them in an oven, and started cautiously cleaning off the dried aluminum oxides with a tiny pointed wooden tool and a fine wire brush. Two hours later, when he had finished the first altimeter, he gently turned one of the wheels by hand, and the entire intimate little cluster of meshed gears moved lightly and silently, while the hand spun around the dial without hesitation. To the disappointment of some of the watchers, but the great relief of Angst, there was not the least flaw—certainly no indication of faulty manufacture. The second altimeter gave a like result.

Meanwhile, the charges against the altimeters had been renewed, this time by the person from whom everyone had been waiting to hear. Frank Hlavacek, the copilot, was finally removed from the critical list and allowed by his doctors to talk—as best he could through a fractured jaw that had been wired together—to Zamuda's operations group on February 20th, seventeen days after the accident. Nine men, plus a doctor and Hlavacek's wife, crowded into his small hospital room that afternoon to hear his version of the crash. Everything Hlavacek told the group fitted in with the story told by Flight Engineer Cook. He confirmed the details of the speed and power used on the approach, the radio contacts, and DeWitt's use of the autopilot to control the plane. But his most interesting statements concerned the final part of the descent. Hlavacek said that he remembered reading exactly nine hundred feet on his altimeter as they passed over the marker of the range station on Clason Point and headed out across the East River. "As we went by the range, everything was normal," he said. "We were descending—the first part of the descent I recall being a little fast, but that was just for an instant. Al stopped it right there and put it back to a normal rate of descent for that approach."

"What rate is that?" prompted Zamuda.

"About two or three hundred feet a minute," Hlavacek said. "A slow—very slow—rate of descent." The plane had been in the clouds until the last second or so, he said, and he had not had the slightest suspicion that anything was wrong until they hit. Unlike Cook, Hlavacek got tangled in no contradictions or confusions concerning the last reading he had seen on the altimeter. The drum had shown between zero and one, and the hand had been on six, for a reading of six hundred; before the hand reached five hundred, he saw lights outside, and the plane hit.

Far from clarifying the case, Hlavacek's testimony threw the operations group into a new turmoil. The argument started as soon as the men left the hospital room, and it went on for hours. No man was better qualified than Hlavacek to report on the crucial readings over Clason Point and just before the crash; his recollections, furthermore, coincided with Cook's. But something was distressingly wrong all the same. If the altimeters had shown nine hundred feet at Clason Point, and the plane had descended at only two hundred or three hundred feet per minute from there on, it would never have got down to the runway in time but would have overshot it seriously. Either DeWitt had been guilty of flying much too high or Hlavacek's memory had played him false, possibly in his anxiety to exonerate DeWitt of any charge of flying too low.

FLUET had by now made much progress since his arrival at LaGuardia on the night of the disaster. Scores—perhaps hundreds—of possibilities had been slowly excluded as a good deal of valuable information was collected. Unfortunately, the information still left room for two different answers. A considerable amount of evidence tended to show that the plane had been flying too low from New Rochelle onward. If so, the altimeters must have read wrong- yet no precedent and no sufficient mechanical explanation for this had been found, nor was Hlavacek's testimony about the altimeter readings internally sound. On the other hand, there was a fair amount of evidence tending to show that DeWitt was new to the Electra and not very familiar (relatively speaking) with poor-weather I.L.S. approaches in the plane, and that the weather that evening had denied him the chance to recognize any errors he was making—yet for a man of his experience to have lost five hundred feet in any contemporary airliner without noticing it was almost unthinkable; moreover, no one in the plane had felt any stall or dive, and the structures group had proved that the plane had hit almost flat.

Fluet's major concern now was to exclude one of these two possibilities, if he could. Because he had no way of adding to his stock of information concerning the chance of pilot error, he figured he would start by trying to incriminate or exonerate the altimeters. Since nothing had yet been proved faulty in the design or assembly of the plane's drum altimeters, Fluet reflected, and since it was established that the two had read pretty nearly alike, perhaps the answer lay in some outside factor common to both—possibly in the static-air system, the independent tubes that lead from air-inlet holes to each altimeter. Any icing, for instance, that might have affected one set of tubes could easily have affected the other.

Calling Jack Real, the chief flight-test engineer at Lockheed, in California, Fluet asked him to arrange tests and demonstrations of the static-air system under severe icing and sleet conditions. On March 9th, Zamuda and the operations group were all at the Lockheed test center, in Burbank. Real first delivered a lecture, illustrated with slides and films, to review for his select little audience the testing that had been done on the Electra static-air system before the plane was certified by the F.A.A. He showed them films of a refrigerated nose section with water being sprayed on it, and told how a prototype plane had for months chased freezing weather all over the Northwest and Alaska. But the designers had done their work well; nothing seemed to make the inlet ports or the tubing retain water or ice. Lockheed engineers, he said, had partly obstructed the inlets with metal plugs, tape, and various balsa-wood carvings simulating ice, and even then the greatest error they had been able to produce was around a hundred and fifty feet.

The following day, the committee went out to the Lockheed Air Terminal, where an Electra was waiting for them. A member of the ground crew squirted a powerful stream of water on an inlet port of the plane's static-air system, as though he were washing the plane carelessly, but all the water drained out quickly without getting near the altimeters. Next, he heaved buckets of frozen slush against the ports; the results were the same. Finally, the whole group climbed into the plane and took off for a rendezvous with a converted B-29 borrowed from the Air Force. At about fifteen thousand feet, the B-29, which carried a great water tank in its belly, put forth a spray of water that froze in the outside air; the Electra tagged along inside the stream of spray, sometimes flying close to the tanker and sometimes far behind it, in order to duplicate a wide range of natural icing conditions. For four hours the test continued, while ice built up on the exterior, now in the form of fine frost, now in the form of frighteningly heavy solid masses. But none of it caused the altimeters to vary from their proper readings. The group had come to another dead end. Still, if Fluet had not found what he was looking for, one of his alternative solutions to the accident had become a little less likely.

WHATEVER the tenor of Fluet's own thinking, he was only the chief investigator, not the judge and jury. In major accident cases, after the investigators have done most of their fact-gathering work, the C.A.B. holds a public hearing, at which the evidence is put on record, under oath, and the various interested parties have a chance to question witnesses and introduce new evidence. Afterward, analytical specialists at the Washington headquarters of the C.A.B. make a complete review of the evidence and prepare a written opinion, which is then subject to the critical revisions of as many as a score of other C.A.B. specialists. This labored-over analysis is finally presented to the five Presidentially appointed board members of the C.A.B., who, sitting as a tribunal, consider whether the evidence has been fairly dealt with by the investigators and review both the logic and the law of the written report. When they are satisfied with it, they sign it, and it becomes the C.A.B.'s official finding in the case.

At the end of February, most of Fluet's investigative groups had finished their studies and disbanded, and he and his several group chairmen busied themselves with the preparations for the hearing. All the reports, letters, maintenance records, maps, and witness statements they had collected were duplicated, collated, and made up into indigestible bundles, each half a foot thick and weighing several pounds. Each party to the case was given a bundle, so that each would know in advance of the hearing everything that would be put into evidence.

The public hearing on Flight 320 took place on Wednesday, March 18th, in the ballroom of the Governor Clinton Hotel, in New York. It lasted for six long, talk-filled days, sometimes intolerably dull, sometimes crackling with argument and punctuated by bangings of the C.A.B. Hearing Officer's gavel. Each of the group chairmen narrated his group's experiences on the case, and placed in evidence the documents he had gathered and the final report he had written. The F.A.A. traffic controllers, the Weather Bureau observers, the pilots of the DC-3, the F.A.A. check pilot who had O.K.'d DeWitt's test on the Electra all told their stories again, under oath. So did everyone else who had had a part in the drama, either as a leading player or as an extra—Hlavacek, Captain Nickerson, Jack Real, and a score of others. Flight Engineer Cook, healthy once again, and no longer confused, testified that he now distinctly remembered seeing a reading of just over five hundred feet on DeWitt's altimeter the last time he glanced at it. His evidence was not taken at face value by everyone at the hearing, however; a representative of the Kollsman company read back what Cook had said from his hospital bed—"I saw Al's altimeter, and it was between zero and the hundred-foot mark on the drum . . . and I was just going to yell, and we hit"—and asked why Cook had been impelled to yell if he truly thought the plane had been at five hundred feet. "That was forty-eight hours after the accident," Cook said. "I was in shock and considerable pain." But an implication remained that Cook's last glimpse of DeWitt's altimeter—whatever its reading—had powerfully alarmed him.

One of the last witnesses to appear was one of the most important—Percy Tumber, the greenkeeper. A leathery, rawboned fellow dressed in sports clothes, he groped for words that would convey his impression of what he had heard and seen. "The noise was terrible," he said. "'It was an eerie sound." When he looked up through the car's rear window, he said, he "saw this plane and it was terribly low and it looked eerie-like." Someone snickered, and Tumber bristled noticeably. Fluet asked him to describe the sound of the engines more fully. "It was like whistling in fir trees," he said. "It was like something you may hear in a ghost story." No one snickered.

On the afternoon of March 25th, the hearing was recessed and the accumulated heap of transcripts, reports, photographs, depositions, and charts was turned over to the C.A.B. analysts in Washington. The writing of an aircraft-accident report belongs to a species of literary art that is indigenous to our era. It involves the endless quibblings of multiple authorship, the calcification of style by official jargon, and the adulteration of clear thought by maddening digressions and qualifications. Yet, for all that, it represents a deliberate process jointly entered into by men with many special viewpoints, and is probably the most sensible way of handling such highly technical studies. The C.A.B. analysts, however, labor as they would behind their mountains of evidence, could get no closer to a clear-cut answer than Fluet had got before the hearing. Everyone was dissatisfied with the first draft of the report, and conferences and expert criticism did not help. "The whole thing would make such good sense and be so simple if it were only the altimeters," an instrument man said to Fluet at one point. Fluet sighed. "It certainly would," he replied, "but that isn't reason enough to say they were to blame. We just have to have more evidence, one way or the other."

So Kollsman's altimeters were put on the rack once again. A young C.A.B. instrument specialist named Thomas Collins was assigned to subject them to the most exhaustive tests his ingenuity could devise, and he went to work with a will. He spread mashed-up insects over the air inlets and measured the deviation from accuracy. He mounted altimeters in banks of ten and sent them through thousands upon thousands of simulated flights. He undertook immensely complicated studies of the effects of vibration. Week after week, he put altimeters by the score through every hoop he could conceive of.

Meanwhile, Cook's and Hlavacek's statements about the final moments of the flight were being separately rechecked. Fluet—who by late spring had been promoted to chief of the Operations Division of the C.A.B. but continued as chief investigator on the case—decided to fly out the conditions described by Hlavacek, rather than rely upon computations from them. In June, he and Zamuda took off in an Electra flown by an American Airlines pilot and made repeated approaches to Runway 22 under the conditions sworn to by Hlavacek and Cook. A movie camera in the plane and another in a police launch moored at the site of the crash recorded the flight path. In order to cross Clason Point at nine hundred feet and approach the point in the water where the launch waited, the plane had to be steeply dived—at a rate of two thousand to twenty-five hundred vertical feet per minute. The resulting angle of the plane and the rapid buildup of its speed—from a hundred and forty knots to a hundred and ninety-five—could not have been missed by anyone but an absolute novice. On the other hand, at the slow rate of descent attested to by Hlavacek, the plane, in order to hit the water at the site of the crash, would have had to cross Clason Point at about five hundred feet. This disclosure advanced the reasoning one step further: If the plane *had* crossed Clason Point at about five hundred feet, then Hlavacek's statement that he had read nine hundred on the dial there became insupportable—always assuming that the altimeters were exonerated.

By autumn, Collins' tests were complete: The altimeters must be considered unequivocally exonerated. Not one altimeter had registered an error of five hundred feet at Flight 320's final altitudes—much less two altimeters at the same time. As these findings were summed up in the fifth, and final, draft of the C.A.B. report:

> An identical and simultaneous malfunction . . . of the magnitude suggested by the crew testimony would involve such an extreme mathematical improbability that the Board is compelled to reject it. In rejecting the possibility of dual and simultaneous altimeter error, the Board must, as a consequence, reject portions of the testimony of one or both flight crew members.

Then, in mitigation of this harsh pronouncement, the report adds:

> Considering that the flight crew members received physical injuries and that they were also under great emotional stress, such questioning of their testimony has a rational basis. Under such circumstances, the Board has frequently found that the recollection, particularly of events immediately preceding such an accident, is very difficult and often erroneous. Furthermore, we are mindful of the natural human tendency to assume conformance with standard operating procedures to fill in the voids or hazy areas of one's memory.

So all that was left was operational error. But explaining the accident in these terms might prove the hardest job of all. As Oscar Bakke, at that time director of the Bureau of Safety of the C.A.B., has said, "Where an investigation leans to operational error, there is almost never any physical evidence of what the pilot or co-pilot did, and we're practically limited to the use of deductive reasoning. And *that's* difficult work."

The investigators began with the now established altitude of five hundred feet over Clason Point, and deductively extended the flight profile in each direction. Evidently, the plane had dropped down too rapidly between New Rochelle and Clason Point; given this error, a routine descent from Clason Point on would produce a crash unless the pilot or co-pilot became aware of the situation. But there, of course, lay the real question: Why had neither one noticed it? Even at night and in foul weather, there are all sorts of safeguards and warnings at work. If a pilot fails to notice that his vertical-speed indicator, for instance, is showing too much of a dive, he can recognize the drop from his increasing forward speed, shown on the air-speed indicator, and his loss of altitude, shown on the altimeter, and similar cross-checks exist for the other major factors involved in flight. But once in a while, owing to an unfortunate combination of conditions, minor errors may go unnoticed and grow larger, or augment each other, until the situation is beyond the point of recovery. "In most cases of pilot error," Bakke has noted, "we've found that it has taken six, seven, or even more unfavorable circumstances, all working together, to cause the accident. If any single one of them had not been working against the pilot, he would have recognized and corrected his situation in time—which, I suspect, is just what happens all the time in normal flights, and in our daily lives."

DURING the late fall, Fluet and half a dozen other investigators combed through the voluminous evidence from this point of view, trying to identify all the conditions that might have been unfavorable to both DeWitt and Hlavacek. From the papers dealing with DeWitt's training on the Electra, they culled the pertinent fact that although he had received five hours of ground practice in a Link trainer that incorporated the Electra's new autopilot system, he had not been exposed to the new drum altimeter; the trainer had been equipped with the old three-needle indicator. Moreover—and while this had not seemed serious until now, it was one more unfortunate circumstance—the vertical-speed indicator on the Link trainer was of a type widely used on planes slower than the Electra; for any given rate of climb or dive, the vertical-speed needle on the Electra moved less than half as far across the face of the dial. During an instrument approach, a pilot's eyes must flick incessantly across his instruments, and he tends to rely on needle angle rather than on a careful reading of numbers. DeWitt's ground training, therefore, had been of no help in preparing him for the two new instruments. And DeWitt's actual in-flight experience with the Electra had been not only sparse but interrupted by flights in DC-6s, with the old-style instruments. In the stiff phrases of the final report:

> We regard it as significant that the ground trainer in which the captain received initial training . . . had installed the conventional vertical-speed indicator and not the instrument which was actually installed on the Electra. . . . The Board [also] finds it difficult to understand why American did not at least incorporate [the drum altimeter] in the Electra cockpit trainer used by the crews during their Electra training.

Thus there emerged the strong probability that DeWitt—with only forty-odd hours on the new instruments and more than twenty-eight thousand hours on the old ones—had dropped down too low between New Rochelle and Clason Point simply by misreading the vertical-speed indicator and passing over the evidence of the unfamiliar altimeter. (If so, it was deduced, his plane must have picked up a little extra forward speed—and a recomputation of the winds aloft that night, and the times at which Hlavacek called in both check points, verified this hypothesis.) Being preoccupied with the effort to head the plane down the back-course beam (he was off course at Clason Point), and using the autopilot to steer by, which put him in an awkward posture and gave him no "feel" of the plane, DeWitt was probably unaware of the excess loss of altitude. Hlavacek, meanwhile, was busy with navigational problems—making contact with Approach Control, retuning his set for Local Control, and studying his navigation diagrams for local landing procedure.

The reasonable conclusion was that Flight 320 had flown over Clason Point in good order but at an altitude of some three hundred feet below the prescribed eight-hundred-foot minimum for that point. And here, again, two small but unfavorable conditions worked the wrong way for DeWitt. First, perhaps because of his preoccupation with making an instrument approach in a relatively unfamiliar plane, he had neglected to correct his altimeter to the latest barometric pressure, and it read eighty feet too high. Second, the "static error" inherent in every altimeter at this altitude and speed was in DeWitt's case an error in the wrong direction, amounting to perhaps forty-five feet. Over Clason Point, his altimeter therefore could well have read something like six hundred and twenty-five feet—a hundred and twenty-five feet too high. As soon as Flight 320 passed over the Clason Point marker, DeWitt started down, undoubtedly expecting at any moment to break out of the four-hundred-foot cloud ceiling over the

East River and see the runway lying a mile ahead. But the unhappy fact, established by Captain Nickerson of the Teti, was that the ceiling over this section of the flight path was less than four hundred feet.

And now the minor disadvantages of the new cockpit environment, the dark night, and the poor weather, plus DeWitt's own small faults of judgment and practice, began to fuse into disaster. Probably because he again misread his vertical-speed indicator, DeWitt at first started down too rapidly (Hlavacek had himself testified to that). He was still having some difficulty staying on the back-course beam. Having been off to the left at Clason Point, he had swung a good bit too far to the right. (From the final report: "Since the captain was utilizing the autopilot, his corrections of altitude and direction were somewhat slower than would normally be expected in a manual approach.") Had he been under orders to go no lower than five hundred feet while still in cloud, he might have been a little less concerned about his back-course approach and a little more concerned about his altitude, but he had been granted a waiver from that extra margin of safety—and of inconvenience—by his own airline. (From the final report: "The Board questions the wisdom of the Company in exempting Captain DeWitt when he had but 12:32 hours of flying the Electra in scheduled operations.") The plane slid down to an altitude of about three hundred feet—more than four hundred on DeWitt's altimeter, if he was looking—within twenty seconds after passing the Clason Point marker. Now DeWitt had to correct that drift to the right, or the landing would be impossible; meanwhile, the plane continued down. Still intent upon the autopilot and the back-course needle, he may have been relying on Hlavacek to keep him informed of altitude. But Hlavacek called out no altitudes below six hundred feet. Although he testified that he had been monitoring the altimeter and air speed, the investigators now refused to accept this:

> It is also not at all unlikely that the co-pilot was giving careful attention to the captain's efforts to maintain the localizer path, especially in view of the apparent difficulties being experienced. . . . Although preoccupation with this or any of the several elements of a new cockpit environment could reasonably explain the failure of Mr. Hlavacek to follow the procedure required in the Operations Manual with respect to monitoring and calling out altitude and air-speed below six hundred feet, the Board believes it more likely that he was anticipating breaking out beneath the overcast and, thereafter, having seen lights on the ground or water, was focussing particularly on visual identification of the airport and was no longer monitoring the flight instruments.

Even so, why did he not see the runway's threshold lights? One more unfortunate circumstance: A dike at the end of Runway 22, which keeps Flushing Bay from flooding it, cuts off the threshold lights below a certain point. DeWitt, by descending too low too soon, had made it impossible for Hlavacek to see them. And anything else that Hlavacek or Cook or DeWitt may have seen probably seemed reassuring because of a common sensory illusion; in the C.A.B.'s experience, flying over water in poor visibility often impairs pilots' sense of perspective. In any case, time passes quickly in such a situation, and there remained only a few seconds. Then, as the relentless logic of circumstance reached its conclusion, Flight 320 reached *its* conclusion, in the waters of the East River.

ON January 6, 1960, more than eleven months after the accident, the final report was signed by the five board members of the C.A.B., and on Sunday, January 10th, it was given to the press for release. A number of people had strong reactions to it. Walter Angst was delighted with it. (Sweetening his weekend even more was the news that the F.A.A. had just rescinded its ban on the use of the drum altimeter in Electras.) American Airlines was glum but silent. The Air Line Pilots Association was less inhibited; the report, the Association president said, "maligned a dead pilot" and thus "conveniently wrote the accident off the books." As for Hlavacek, who received the gist of the report by telephone at his home the night before it appeared in the newspapers, he responded with a burst of temper; he was "furious," he said, and the C.A.B. was "trying to take the easy way out by blaming a dead pilot." Several days later, having read the report and found that he, too, was criticized, he described it to the press as "highly inaccurate and totally useless," adding, "This is one of the few bad accidents they've gotten a couple of crew members out of alive—and they simply disregard our testimony." A reporter asked what effect the report would have on him. "I don't expect it to have any," he said confidently. "In fact, I've just been O.K.'d by the F.A.A. as captain on DC-6s. That ought to prove something."

Fluet, for his part, experienced a sense of weary relief rather than one of triumph. "Well, it's done, but I don't exactly feel like cheering," he said. "It's a lot less agonizing for us when there's been, say, a fire, and we try to find out how it started, or an explosion, and we put the wreckage together to find out what caused it. There's always something more *satisfying* about a neat, simple conclusion like that, and generally we can work out a good quick fix to remedy the situation. The case of Flight 320 was something else. Within the first few days, I suspected it would narrow down to two major possibilities—instrument failure and operational error, both difficult to prove—and either way somebody had to be hurt. On the one hand, you had a poor, decent dead guy who couldn't defend his reputation, along with a lot of living pilots who would feel damaged by a finding of pilot error. On the other hand, you had a fine company—the Western world's leading maker of altimeters—with its good name and its latest and finest product at stake. It was a painful choice, and bound to result in a lot of bitterness either way. But every time I get discouraged, I try to remind myself of the statistics. Slowly but surely, it's getting safer and safer for the individual passenger to fly, and the C.A.B. can take some of the credit for that. No matter who criticizes our work, the statistics make it possible for us to be proud of what we're doing."

The care and intensity with which aircraft accidents are investigated from all seemingly pertinent standpoints stands in sharp contrast to the customary superficial or nonexistent investigation of accidents of most other types, many of which are not only far more important as causes of morbidity and mortality than the crashes of commercial aircraft but are also easier to approach from the research standpoint. Motor vehicle accidents of most types, for example, occur in sufficiently large numbers and in circumstances that make the obtaining of adequate case series and controls a relatively simple matter. Nonetheless, there have been only a few attempts to study the full range of seemingly pertinent variables even in motor vehicle accidents. Unfortunately these attempts have all involved uncontrolled studies, usually based on samples of unknown representativeness.[43-47] As a result of these and other methodological shortcomings, the results have often been difficult or impossible to interpret with confidence.

A further shortcoming of studies of this type has been the tendency, also illustrated by portions of the work reported by Hunt, of some research workers a priori to favor conclusions of certain types. Accident research will not make its full contribution to the alleviation of injury and death until these deficiencies and those we have documented earlier are replaced with work of at least the quality customary in the pertinent collateral fields. It is to help attain this objective that we have prepared this volume.

REFERENCES

1. Moore, J. O., and Tourin, B., *A Study of Automobile Doors Opening Under Crash Conditions,* T. R. 2. Automotive Crash Injury Research, Cornell University, August 1954.
2. Garrett, J. W., *An Evaluation of Door Lock Effectiveness: Pre-1956 vs. Post-1955 Automobiles.* Automotive Crash Injury Research, Cornell University, July 1961.
3. Hirsch, J., "Public Health and Social Aspects of Alcoholism," in *Alcoholism,* ed. G. N. Thompson. Springfield, Ill.: Charles C. Thomas, 1956.
4. Haddon, W., Jr., "Alcohol and Highway Accidents," in *Alcohol and Road Traffic,* Proceedings of the Third International Conference on Alcohol and Road Traffic. British Medical Association, London, 1963.
5. Borkenstein, R. F., Crowther, R. F., Shumate, R. P., Ziel, W. B. and Zylman, R., *The Role of the Drinking Driver in Traffic Accidents,* ed. A. Dale. Department of Police Administration, Indiana University, Bloomington, 1964.
6. Haddon, W., Jr., "The Prevention of Accidents," in *Textbook of Preventive Medicine,* eds. D. Clark and B. MacMahon. Boston: Little, Brown & Co. In press.
7. Haddon, W., Jr., "A Note Concerning Accident Theory and Research with Special Reference to Motor Vehicle Accidents," *Annals of the New York Academy of Sciences, 107*:635-646, 1963.
8. Dalziel, C. F., "The Effects of Electric Shock on Man," Institute of Radio Engineers *Transactions—Medical Electronics,* PGME-5, pp. 44-62, May 1956.
9. Safety Devices on Household Refrigerators. Hearing before Subcommittee 6, Business and Consumer Interests, of the Committee on Interstate and Foreign Commerce, 83rd Congress, April 27, 1954, on S. 2876 and S. 2891, a bill to require inside latches on the doors of household refrigerators shipped in interstate commerce.
10. Moynihan, D. P., "The Legal Regulation of Automobile Design," in *Passenger Car Design and Highway Safety.* Association for the Aid of Crippled Children, New York, and Consumers Union, Mount Vernon, 1962, pp. 265-285.
11. *Behavior of Young Children under Simulated Refrigerator Entrapment.* U. S. Department of Commerce, National Bureau of Standards, March 26, 1957.
12. Standard Announced for Refrigerator Safety Devices. Press release DC—13245, Office of the Secretary, U. S. Department of Commerce, April 16, 1957.
13. Standard for Devices to Permit the Opening of Household Refrigerator Doors from the Inside, *Federal Register,* p. 6058, August 1, 1957, Title 15, Chap. II, Part 260.
14. Refrigerator Safety Devices Standard Published. Press release D.C.—6563, Office of the Secretary, U. S. Department of Commerce, August 2, 1957.
15. Lederer, J., *Perspectives in Air Safety.* Flight Safety Foundation, New York, 1963.
16. Haddon, W., Jr., "Research with Respect to Fatal Accident Causes: Implications for Vehicle Design," a paper presented at the 1961 Summer Meeting (June 6), Society of Automotive Engineers. SAE Preprint 366A.
17. Anonymous, "Let's Ban the Traffic Tramps," *Journal of American Insurance, 39*:14-15, 1963.
18. Goldberg, L., "Drunken Drivers in Sweden," in *Alcohol and Road Traffic,* Proceedings of the Second International Conference on Alcohol and Road Traffic. Toronto: Garden City Press Co-Operative, 1955.
19. Fisher, R. A., *The Design of Experiments,* 7th Ed. New York: Hafner, 1960.
20. Irby, T. S., and Jacobs, H. H., "An Epidemiological Approach to the Control of Automobile Accidents: Experimental Patrol Intensification at a Military Base," *Traffic Safety Research Review, 4*:1:4-7, 1960.
21. Domey, R. G., and Duckworth, J. E., Work in progress on non-battle accidental

death and injury in the military services and the Department of Defense. Harvard School of Public Health.

22. Rainey, R. V., Conger, J. J., and Walsmith, C. R., "Personality Characteristics as a Selective Factor in Driver Education," *Highway Research Board Bulletin, 285*:23-28, 1961.
23. Lowrey, F. P., "Vehicle Condition and Periodic Safety Inspections," in *Passenger Car Design and Highway Safety, op. cit.*, pp. 114-122.
24. Chayet, Neil L., Work in progress on the evaluation of motor vehicle inspection. Law-Medicine Institute, Boston University.
25. Thomson, L. G., and Goldstein, L. G., "Home-Accident Research: A No Man's Land," National Safety Congress *Transactions, 6*:56-60, 1962.
26. McCorkle, T., *Eye Injuries, Safety Gear, and Values in Iowa Farm Life*, Bulletin 6, Institute of Agricultural Medicine, College of Medicine, State University of Iowa, July 1960.
27. Coppleson, V. M., *Shark Attack*, 2nd Ed. Sydney: Angus and Robertson, 1962.
28. Research Needs in Traffic Safety. Hearings before a Special Subcommittee on Traffic Safety of the Committee on Interstate and Foreign Commerce, House of Representatives, 85th Congress, April 23, 1958.
29. Automobile Seat Belts. Hearings before a Subcommittee of the Committee on Interstate and Foreign Commerce, House of Representatives, 85th Congress, April 30, August 5-8, 1957.
30. Motor Vehicle Safety. Hearings before a Subcommittee of the Committee on Interstate and Foreign Commerce, House of Representatives, 86th Congress, July 7-9, 1959.
31. To Establish a National Accident Prevention Center. Hearings before a Subcommittee of the Committee on Interstate and Foreign Commerce, House of Representatives, 87th Congress, February 6-8, 20-21, 1962.
32. Proceedings of the Stapp Car Crash and Field Demonstration Conferences. Published annually by The Center for Continuation Study, General Extension Division, University of Minnesota and Aerospace Medical Division, U. S. Air Force.
33. Goddard, J. L., and Haddon, W., Jr., "An Introduction to the Discussion of the Vehicle in Relation to Highway Safety," in *Passenger Car Design and Highway Safety, op. cit.*, pp. 1-6.
34. See Michelson *et al.* in this chapter.
35. Grime, G., "Effect of Seat Harness on Movement of Car Occupant in a Head-on Collision," *Highway Research Record*, No.4, Publication 1068, pp. 76-90, 1963.
36. *Seat Belts and the Safety Engineer*, Special Report, reprinted from the American Society of Safety Engineers Journal, July-September 1963.
37. Tourin, B., and Garrett, J. W., *Safety Belt Effectiveness in Rural California Automobile Accidents.* Automotive Crash Injury Research, Cornell University, February 1960.
38. Stonex, K. A., "Vehicle Aspects of the Highway Safety Problem," *Traffic Safety Research Review, 6*:2:15-24, 1962.
39. Stonex, K. A., and Skeels, P. C., "Development of Crash Research Techniques at the General Motors Proving Ground," *Highway Research Record*, No. 4, Publication 1068, pp. 32-49, 1963.
40. Stonex, K. A., "Relation of Cross-Section Design and Highway Safety," *Traffic Safety Research Review*, 7:4:2-14, 1963.
41. Garrett, J. W., *A Study of Seat Belts in Wisconsin Automobile Accidents.* Automotive Crash Injury Research, Cornell University, Cal Report No. VJ-1823-R3, 1963.
42. Doull, J., Plzak, V., and Brois, S. J., *A Survey of Compounds for Radiation Protection.* Publication 62-29, School of Aerospace Medicine, Brooks Air Force Base, Tex., 1962.
43. Baker, J. S., *Experimental Case Studies of Traffic Accidents.* Traffic Institute, Northwestern University, Evanston, Ill., 1960.
44. Moseley, A. L., *et al.*, "Research on Fatal Highway Collisions," *Papers 1961-62.* Harvard Medical School, Boston.
45. Moseley, A. L., *et al.*, "Research on Fatal Highway Collisions," *Papers 1962-63.* Harvard Medical School, Boston.
46. Huelke, D. F., and Gikas, P. W., Work in progress at University of Michigan Medical School, Ann Arbor, Mich.