1111

TRANSPORTATION RESEARCH RECORD

Traffic Accident Analysis, Visibility Factors, and Motorist Information Needs

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1987

Transportation Research Record 1111
Price $20.00
Editor: Naomi Kassabian
Compositor: Harlow A. Bickford
Layout: Betty L. Hawkins

mode
1 highway transportation

subject areas
51 transportation safety
52 human factors
54 operations and traffic control

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Traffic accident analysis, visibility factors, and motorist information needs.

(Transportation research record, ISSN0361-1981 ; 1111)
1. Traffic accidents—Congresses. 2. Traffic safety—Congresses. I. National Research Council (U.S.) Transportation Research Board. II. Series.
TE7.H5 no. 1111 [HE5614] 380.5 s 87-28204
ISBN 0-309-04463-4 [363.1'2565'072]

Sponsorship of Transportation Research Record 1111

GROUP 3—OPERATION, SAFETY, AND MAINTENANCE OF TRANSPORTATION FACILITIES
James I. Taylor, University of Notre Dame, chairman

Facilities and Operations Section

Committee on Visibility
Richard E. Stark, Illinois Department of Transportation, chairman
Richard L. Vincent, Lighting Research Institute, Inc., secretary
John B. Arens, Albert Burg, Charles W. Craig, Eugene Farber, Mark Freedman, Paul H. Greenlee, Daniel G. Henderson, Ronald J. Hensen, Marvin H. Hilton, Antanas Ketvirtis, Gerhart F. King, L. Ellis King, Ken F. Kobetsky, Roger W. McNees, Herbert A. Odle, R. H. Priddy, Earl C. Williams, Jr., Henry L. Woltman, Helmut T. Zwahlen

Committee on Methodology for Evaluating Highway Improvements
William D. Glauz, Midwest Research Institute, chairman
William T. Baker, Federal Highway Administration, U.S. Department of Transportation, secretary
William D. Berg, Brian L. Bowman, Eugene D. Brenning, Charles Philip Brinkman, Myung-Soon Chang, Forrest M. Council, Alfred-Shalom Hakkert, Fred R. Hanscom, Ezra Hauer, David L. Helman, Ruediger Lamm, King K. Mak, Ernst Meyer, David D. Perkins, Jerry G. Pigman, Ramey O. Rogness, Roy W. Taylor, Harold T. Thompson, James M. Witkowski, Paul H. Wright

Users and Facilities Section

Committee on Vehicle User Characteristics
Helmut T. Zwahlen, Ohio University, chairman
Martin G. Helander, State University of New York at Buffalo, secretary
Vivek D. Bhise, Albert Burg, Robert E. Dewar, Peter H. Hehlen, Slade Hulbert, Margaret Hubbard Jones, Rodger J. Koppa, Jan Moraal, James F. O'Hanlon, Richard Arnold Olsen, Michael Perel, Thomas A. Ranney, Thomas H. Rockwell, Kare Rumar, Michael Sivak, Anthony C. Stein, Burton W. Stephens, Charles B. Stoke, E. Donald Sussman, Robert C. Vanstrum, Paul F. Wasielewski

Committee on User Information Systems
Conrad L. Dudek, Texas Transportation Institute, Texas A&M University System, chairman
Michele Ann McMurtry, National Transportation Safety Board, secretary
Kenneth A. Brewer, Richard D. Desrosiers, Robert E. Dewar, Ronald W. Eck, Fred R. Hanscom, David L. Helman, Warren E. Hughes, Gerhart F. King, David B. Knies, Harold Lunenfeld, Truman Mast, Roger W. McNees, Peter S. Parsonson, Thomas A. Ranney, Arthur W. Roberts III, H. Douglas Robertson, Allen E. Shinder, Richard A. Staley, Leslie Ann Whitaker, Eugene M. Wilson, David M. Zaidel

Committee on Traffic Records and Accident Analysis
Benjamin V. Chatfield, Chatfield Associates, Inc., chairman
C. Arthur Geurts, Utah Department of Transportation, secretary
J. M. Barnett, Thomas E. Bryer, Myung-Soon Chang, John T. Dempster, Jr., Nicholas J. Garber, James H. Hedlund, Donald R. Herd, Alan F. Hoskin, Jack A. Hutter, A. Dewey Jordan, Larry G. Karsten, John C. Laughland, Martin E. Lipinski, Philip P. Madonia, Joseph C. Marsh IV, Clarence W. Mosher, James O'Day, Olga Pendleton, Darrell E. Roach, Robert R. Roberts, Alex Sorton, John G. Viner, Charles V. Zegeer, Ralph W. Zimmer

James K. Williams and David K. Witheford, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1986.

Transportation Research Record 1111

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

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

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

Contents

v **Foreword**

1 **Video Collision Reconstruction Using Physical Evidence**
Zygmunt M. Gorski, Alan German, and Edwin S. Nowak

7 **Analysis of High-Hazard Locations: Is an Expert Systems Approach Feasible?**
Gary S. Spring, John Collura, and Paul W. Shuldiner

17 **Safety Implications of Truck Configuration**
Oliver Carsten

27 **Relationships Between Vertical and Horizontal Roadway Alignments and the Incidence of Fatal Rollover Crashes in New Mexico and Georgia**
Paul Zador, Howard Stein, Jerome Hall, and Paul Wright

42 **Evaluation of Opportunity-Based Accident Rate Expressions**
Mark Plass and William D. Berg

48 **Demonstration of Regression Analysis with Error in the Independent Variable**
Richard M. Weed and Ricardo T. Barros

54 **Validation of a Nonautomated Speed Data Collection Methodology**
Fred R. Hanscom

62 **Standard Target Contrast: A Visibility Parameter Beyond Luminance To Evaluate the Quality of Roadway Lighting**
F. W. Jung and A. Titishov

71 **A Method of Calculating the Effective Intensity of Multiple-Flick Flashtube Signals**
Marc B. Mandler and John R. Thacker

79 **Evaluation of Alternative Sign-Lighting Systems To Reduce Operating and Maintenance Costs**
Jonathan E. Upchurch and Jeffrey T. Bordin
DISCUSSION, *Michael S. Janoff*, 91
AUTHORS' CLOSURE, 92

93 **Evaluation of the Effectiveness of Crash Cushion Delineation**
F. Thomas Creasey, Conrad L. Dudek, and R. Dale Huchingson

104 **Special Traffic Control To Meet Motorist Information Needs on Long, Steep Grades**
Joseph L. Henderson, Eugene M. Wilson, George A. Dale, and Claudia L. Scrivanich
DISCUSSION, *Andrew G. Macbeth*, 109
AUTHORS' CLOSURE, 109

110 **Advisory Speed Signs and Curve Signs and Their Effect on Driver Eye Scanning and Driving Performance**
Helmut T. Zwahlen

121 **Alternatives to Enforcement in Modifying the Speeding Behavior of Drivers**
Stephen Maroney and Robert Dewar

126 **Excess Travel: Causes, Extent, and Consequences**
Gerhart F. King and Truman M. Mast

ABRIDGMENT
134 **A Study of Route Selection from Highway Maps**
Gerhart F. King and Ajay K. Rathi

138 **CRASH Revisited: Additions to Its Clarity, Generality, and Utility**
Albert G. Fonda

Foreword

The papers in this Record all pertain to traffic safety, but each addresses the problem from a different perspective. The first eight papers deal with traffic accident analysis with emphasis on accident reconstruction. Gorski et al. illustrate one technique of accident reconstruction in which scale models of the involved vehicles are overlaid on a scale diagram of the collision scene. The reconstructed motion of the vehicles is then videotaped by using a cutout animation technique. Spring et al. report on efforts to develop computerized analysis of traffic accident data by using expert systems concepts. Fonda expands on previously developed equations for the popular CRASH digital computer program for accident reconstruction. Carsten and Campbell examine the effect of truck configuration on accident characteristics, and Zador and Stein analyze data on curvature and grade collected at the sites of fatal single-vehicle rollover crashes and at random comparison sites in New Mexico and Georgia. Weed and Barros use computer simulation to demonstrate regression analysis with error in the independent variable. Plass and Berg discuss the recent development of opportunity-based accident rate expressions as sensitive indicators for use in safety studies. Hanscom reports on developing a validated spot speed study procedure that does not rely on automated equipment.

Four papers explore visibility factors for safe driving. Jung and Titishov report on the use of a visibility parameter beyond luminance to evaluate the quality of roadway lighting. Mandler and Thacker explain work to devise a method of determining the effective intensity of light flashes composed of multiple pulses of light. Upchurch reports on a study to identify a sign lighting system that has a lower electric power cost and reduced maintenance requirements and that provides adequately for the motorists' needs in terms of legibility and illumination level. Creasey et al. summarize the results of field studies involving the effectiveness of crash cushion delineation techniques at freeway gore areas.

The last five papers discuss motorist information needs with some attention to modification of driver behavior. Henderson et al. report the results of a study on a special message information sign for vehicles such as tractor-semitrailers, single-unit trucks, motor homes, and vehicles pulling trailers. Zwahlen reports on a study to determine the effectiveness of advisory speed signs used in conjunction with curve warning signs in Ohio. Maroney and Dewar describe two experiments conducted to examine alternatives to enforcement as a means of reducing speeding behavior. King and Mast summarize a recent FHWA research project designed to develop estimates of the proportion of all highway travel that is excessive or "wasted" and an estimate of the economic costs generated by such excess travel. King and Rathi report on an experiment designed to assess the ability of subjects to plan long trips in unfamiliar areas using maps only.

Video Collision Reconstruction Using Physical Evidence

ZYGMUNT M. GORSKI, ALAN GERMAN, AND EDWIN S. NOWAK

The reconstruction of a motor vehicle collision is often achieved through investigation and interpretation of the physical evidence left in the aftermath of the collision. Such evidence can be in the form of tire marks and gouges at the collision scene, the structural damage to the vehicles, and occupant contacts with the vehicle interior. Collisions can be extremely complex and yet it is often necessary for investigators to explain the precise nature of the collision events to individuals who have little expertise in interpreting physical evidence. Historically reconstructions have been illustrated by means of schematic diagrams to provide a static representation of collision events. With the availability of reasonably priced videocassette recording equipment, it is now possible for investigators to readily adapt animation techniques normally used by film makers to the field of collision reconstruction. Scale diagrams of the involved vehicles are overlaid on a scale diagram of the collision scene and the reconstructed motion of the vehicles is videotaped by using a cutout animation technique. The resulting visual presentation provides an excellent means of communicating complex events to nonspecialists. The method is demonstrated by the reconstruction of a multiple-vehicle, multiple-impact collision.

The basis for producing a videotape of animated vehicle movements is a detailed documentation of all the collision-related physical evidence. The investigator must inspect the vehicles involved and the collision scene as soon as possible, because some types of physical evidence disappear quite quickly. The precise locations of various portions of evidence should be documented extensively. The vehicles should be measured to determine their damaged profiles. Sample vehicles may be used to obtain the original dimensions. These points will be expanded later in the discussion.

COLLISION RECONSTRUCTION

Documenting the location of physical evidence found at the collision scene is relatively straightforward. Various techniques may be utilized to make the required measurements (*1*, pp. 451–462). The method favored by the authors is the use of a rectangular coordinate system, because simple measuring tools can be utilized very effectively (2, pp. 335–336). The process involves making a series of careful and detailed measurements at the collision scene. Often, some measure of skill is involved in determining the scene evidence that is relevant to the collision being investigated.

Multi-Disciplinary Accident Research Team, Faculty of Engineering Science, University of Western Ontario, London, Ontario N6A 5B9, Canada.

Various factors make accurate measurements of the damaged vehicle difficult. The vehicle will normally be measured in the field, typically in a towing compound, where the terrain may not be level. The vehicle may be so deformed that none of the original contours remain intact. Also, damage profiles at different vertical levels may be of critical importance.

A technique for obtaining measurements of the profile of a damaged area on a vehicle by using a contour gauge has been described by Tumbas (*3*). A variant of this method (*4*, pp. 352–368) is to place a reference rectangle around the entire vehicle. If practicable, the original length and width of the vehicle concerned should be used as the dimensions of the sides of the rectangle in order to provide a particularly useful reference frame. Measurements are taken from the sides of the rectangle to damaged areas of the vehicle at right angles in both the horizontal and vertical planes. Some adjustments to the vehicle or rectangle must be made where the ground or vehicle is not level or some compensation in the measurements must be made to avoid errors.

The measurements required to produce a replica of a damaged vehicle for the purposes of video recording will depend somewhat on the type of collision involved. In any reconstruction, the investigator should take a number of measurements to establish a good representation of the areas of direct damage.

An area of direct damage is a region that was in direct contact with the vehicle or object struck in the collision. Within such a region, measurements should be taken to establish the precise location of points of mutual contact, which are specific locations on one vehicle for which contact evidence can be identified on specific locations of another vehicle or object. Typically, a crease, transfer, or imprint on a vehicle is identified as originating from contact by a bumper bolt, license plate, grille, roof pillar, and so on, of another vehicle. These points must be identified in their original positions on an undamaged vehicle as well as on the damaged vehicle being measured.

Points of mutual contact that have been identified and so documented may be utilized in the animation of the collision events. The reason for moving the colliding vehicles from one position to the next during recording will be based on evidence from points of mutual contact in addition to scene evidence.

Additional measurements should be taken to document the overall shape of the damaged vehicle. As a guide it is suggested that the following be measured as providing valuable information relating to vehicle position, orientation, and motion:

1. The amount of end-shifting of the vehicle's structure;
2. The location of the corners of the hood, provided that the hood is not hanging loose;

3. The locations of the outboard surfaces of the wheel hubs;
4. The positions of the roof pillars, measured at the top (at the junction with the roof) and at the level of the base of the window glass; and
5. The locations of the ends of both front and rear bumpers.

Once detailed measurements of the vehicles and collision scene have been obtained, the next step is to produce scale diagrams of each of these items.

No ideal scale exists for all situations; it is based on the investigator's needs, which differ from case to case. When an animated videotape of collision events based on physical evidence is produced, it is desirable to choose the largest scale possible without making the subjects so big that the investigator cannot move around them.

A suggested scale is 1:10, which will produce passenger car diagrams that are 50 to 55 cm (19 to 22 in.) long. Such a scale will produce large but manageable scene diagrams for the majority of collisions. The suggested scale is also useful in that most passenger car diagrams produced to this scale will fit on two sheets of 8.5 × 11-in. paper.

With rudimentary drafting instruments (an engineer's scale, protractor, and straightedge) the investigator can construct a rectangle on the sheet of drawing paper in which he will reconstruct the dimensions of the damaged vehicle that had been measured previously. Initially, a pencil drawing of the vehicle in plan is produced.

In a similar manner the investigator draws the original shape of the same vehicle. Original measurements of the case vehicle are readily obtained by visiting a large parking lot and taking the measurements directly from a parked vehicle of the same year and model. By matching the characteristics of the Vehicle Identification Number (VIN) of the damaged and sample vehicles, the investigator can be assured that both have the same relevant body dimensions.

Once the pencil drawings of the damaged and the original vehicles have been completed, the investigator will trace the drawings onto clear plastic sheets with a set of colored marking pens. The collision scene should be drawn on a large roll, or a large sheet, of thin white paper. If large sheets are not available, smaller sheets can be taped together with transparent tape, which is not visible on the videotape. The relevant roadway markings should be drawn in pencil on this large sheet of paper. Once the amount of detail in the drawing is satisfactory, the investigator can trace the pencil markings with a thick, black marker.

Thus the scene outline and evidence will be black on a large white area of paper, and each of the vehicles will be drawn in a specific color on clear plastic sheets. The scene and vehicle diagrams are now ready for recording.

VIDEOTAPE RECORDING EQUIPMENT

In videotape recording, the audio and visual information is encoded onto magnetic tape. Broadcast-quality recording normally utilizes tape 1 or 2 in. wide on which separate tracks are provided for video, audio, and frame identification information. The equipment used in conjunction with this tape includes high-resolution cameras, broadcast-standard videotape recorders, and computer-aided editing machines.

Recent advances in technology have provided relatively inexpensive videotape recording equipment designed for amateur use. Generally cameras and recorders in such systems use ½-in. videotape. The narrower tape format provides for considsiderably less sophistication than is found in professional equipment; however, reasonably good quality can be achieved with these less expensive systems.

Typically, animation effects are produced by using the video camera in a stop-start mode. A subject is recorded for a short period of time. Then it is moved slightly and an additional recording is made. The simple process of stringing together individual sequences of material is referred to as "assemble editing." This method of recording usually produces interference in the picture during playback because there is no synchronization of the individual frames from the completion of one sequence to the beginning of the next. Such interference can be quite distracting to the viewer.

A technique referred to as "insert editing" minimizes these effects. To accomplish this, it is necessary to produce a control track on the videotape to encode the video information during the actual recording session. The control track consists of a string of electronic pulses, usually located on one of the tape's audio tracks. These pulses identify each frame of the tape so that the video and audio material can be played back properly.

The control track is readily produced by videotaping a black, nonshiny surface throughout the portion of tape that will be used for the reconstruction segment. Individual segments are now recorded onto the preprocessed tape by using the video-dub feature of the videotape recorder. The dubbing process uses the line of control pulses on the tape to synchronize the individual frames in the recording, which results in a more stable picture.

The investigator may be able to use the raw videotape produced by insert editing as described earlier. If still better quality is desired for the finished product, additional editing of the tape will be required. Such editing may involve the use of a controller unit governing the operation of two videotape recorders. At a minimum, editing may be performed between two separate videotape recorders. In the latter case it will not be possible to perform the editing at an exact frame and so some interference will still be present in the final picture. For a more complete discussion of editing techniques, the reader is referred to texts on this subject (*5,6*).

The quality of the final recording will depend to a great extent on the sophistication of the equipment used and on the skill of the operator. The results that can be obtained using inexpensive recording equipment are quite reasonable. Furthermore, the recordings produced are vastly superior to static diagrams in their ability to display complex collision situations.

ANIMATION AND RECONSTRUCTION TECHNIQUE

Historically animation has primarily been associated with film making. Numerous techniques are used in film animation (*7*). The basis for all of these is that a motion picture camera is used to make single-frame exposures of a subject, which is moved slightly between frames. The camera and the background remain static, and when the film is projected at normal speed, persistence of vision provides the illusion of fluid motion of the subject. This basic technique has been adapted for use with a video camera and videocassette recorder to provide a visual display of vehicle movements in a collision situation.

The basic methodology is to overlay scale diagrams of the vehicles on a scene diagram drawn to the same scale. The vehicle diagrams are essentially transparent so that physical evidence on the scene, and even on the vehicles themselves, can be observed. This enables the viewer to identify locations at which portions of physical evidence match up. These can be the location of a vehicle's wheel on a tire mark in the roadway or the conjunction of points of mutual contact between two colliding vehicles. The scene generally is left stationary and the vehicle diagrams are moved in stages to produce the animation.

Before the recording process begins, the movements of the vehicles must be established for each recorded segment. It may be necessary to calculate detailed position-time histories of the vehicles over the entire collision sequence in order to be able to produce animated vehicle motion in real time or in precise slow motion. Such data may be required if a knowledge of the relative locations of vehicles or objects, or both, is of primary interest in the reconstruction. This information may well be desired if the main question to be considered is one of driver's perception time, the location of a pedestrian with respect to the striking vehicle at some point before the impact, or some similar issue.

In cases where there are no time-specific issues, trial-and-error positioning of the vehicle diagrams on the collision scene may be used to reconstruct the sequence of vehicle motion. In such instances it may only be necessary to illustrate how the physical evidence with respect to vehicle damage and marks at the collision scene is to be interpreted.

Specific positions for the vehicle diagrams should be marked on the collision scene for each segment being recorded. It is convenient to draw small pencil lines on the collision scene at the desired intervals. The investigator will advance his vehicle diagram from one interval to the next along this line of movement intervals. The front edge of the plastic sheet containing the vehicle diagram might be used as the reference line that is placed at each pencil mark.

For a two-vehicle collision, the investigator might begin videotaping the collision sequence at some point before the first impact. The vehicle diagrams will be moved toward each other at a rate based on an estimate of their velocities. If one vehicle is traveling faster than the other, it will have to be positioned farther away from the point of impact (POI) than the other if they are both to meet at the designated POI.

It should be noted that as a vehicle makes contact at the POI, its speed will be reduced rapidly and the distance between the penciled movement intervals will become shorter as the vehicle decelerates. Also, this vehicle will likely begin to rotate and establish a new line of travel based on the forces acting on it. Thus a rotation rate will have to be marked through some curvilinear set of penciled intervals. If the case vehicle is to be involved in a secondary impact with a third vehicle, the rotation rate will be governed by the interval rate that will bring the vehicle to the POI at the required time.

The desirability of large-scale diagrams now becomes apparent. On such a large scale, small reference marks penciled in to identify specific vehicle positions are not visible on the videotape. In addition, when the vehicle diagrams are moved, it is difficult to keep them oriented in a straight line from one interval to the next. Slight deviations from the path of travel can be very noticeable when small-scale vehicle diagrams are used. With large-scale diagrams these slight orientation errors become indistinguishable and the vehicle movement appears more fluid.

When the vehicles strike each other at the POI, some crushing should be shown in the vehicle diagrams. This crushing can be imitated by substituting vehicle shapes with progressively greater crushing until the final one shown is the crushed shape that resulted from that particular impact. (This substitution technique is shown in Figure 2.) Usually the intermediate crush shapes are estimates of how the vehicle would be expected to crush in the real collision; however, the amount of crush should be accurate in positions where physical evidence is to be matched.

Before recording, the investigator should choose a large enough work site. He will need to move around the scene diagram and will need room for the camera and lighting. If possible, the camera should be mounted on top of the collision scene, pointing straight down. In a film studio this is not a problem because there is usually a ceiling grid for the attachment of lighting sources. In a regular office environment it may be possible to mount the camera on a large tripod and still shoot straight down at a collision scene that passes between the outstretched legs of the tripod. If the collision scene is too wide to be accommodated by this arrangement, the tripod may be placed on a large work table.

Direct overhead lighting of the scene should be avoided because the transparent sheets containing the vehicle diagrams reflect light into the camera mounted above the scene. The intent is to produce an environment with little shadow and an even exposure around the entire surface being recorded. At a minimum, two floodlights placed on opposite sides of the video camera can provide a reasonable amount of balanced lighting.

One difficulty with the use of a large scale is that for collision events taking place over a considerable length of roadway it will probably not be possible to place the entire collision scene in the view of the camera. Moving the camera to various locations along the scene diagram is awkward at best. Instead, the collision scene should be moved underneath the stationary camera, giving the illusion that the camera is being moved. The scene diagram should be moved at regular intervals. This can be accomplished by establishing a reference point on the floor next to the edge of the scene diagram. The scene diagram can be scrolled for an established distance interval for each recorded segment relative to the reference line.

CASE STUDY

The animation technique will be illustrated by means of a case study of a real-world collision that was investigated and reconstructed by the authors. It should be noted that this particular reconstruction is based solely on matching physical evidence observed from the damage to the striking vehicles and that identified at the collision scene. The intent in producing a videotape of this particular collision was to show how the pieces of physical evidence fit together to support the reconstructed vehicle movements in a complex series of events. Furthermore, the reconstruction is not meant to reproduce the collision events absolutely with respect to time (real or slow motion), but rather to show the contact between physical evidence of damage and that at the scene in the correct sequence. The accuracy of the vehicle movements between the contacts is not important in this case.

Figures 1 through 8 show vehicle movements during an example recording session. The example involves an offset head-on collision between two passenger cars, a 1978 Dodge Aspen and a 1978 Buick LeSabre. The Aspen entered a median-divided, limited-access highway, traveling the wrong way in thick fog. The two lanes shown are designated for travel in the same direction; the lanes for travel in the opposite direction are not shown in this study. The LeSabre rotated counterclockwise from this initial impact and was struck by a tractor-trailer.

A detailed inspection of the collision scene and vehicles allowed for the documentation of a large number of individual pieces of physical evidence that enabled specific vehicle positions and contacts between the involved vehicles to be identified.

Figure 1 shows the two passenger cars traveling toward each other just before the impact. The front portion of a tractor-trailer may be seen in the lower right-hand corner. The general area of the first impact between the passenger cars contained a

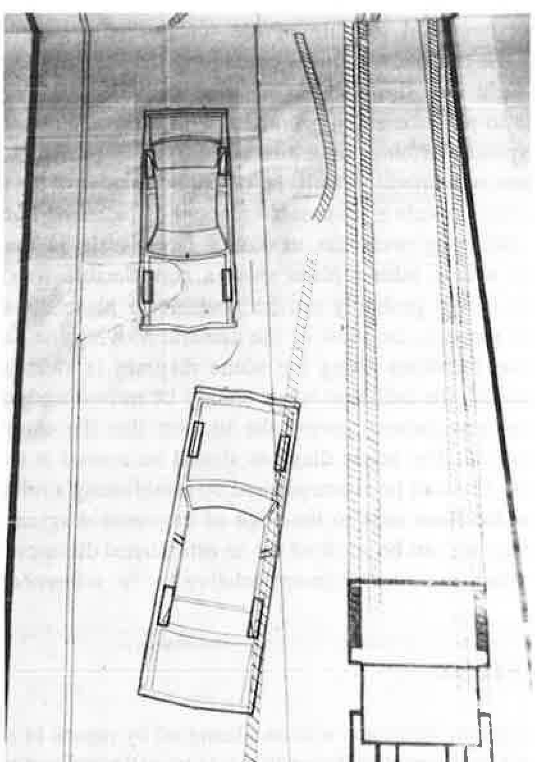


FIGURE 1 Preimpact vehicle locations.

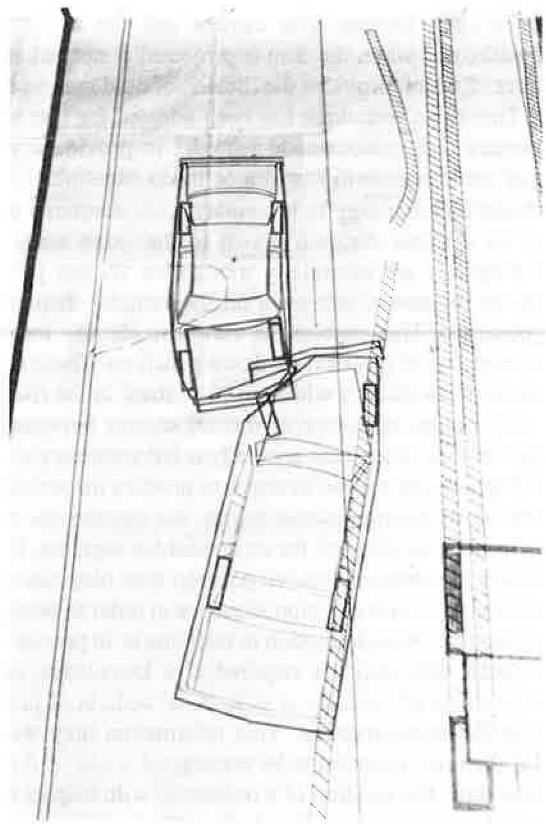


FIGURE 2 Frontal impact between Aspen and LeSabre.

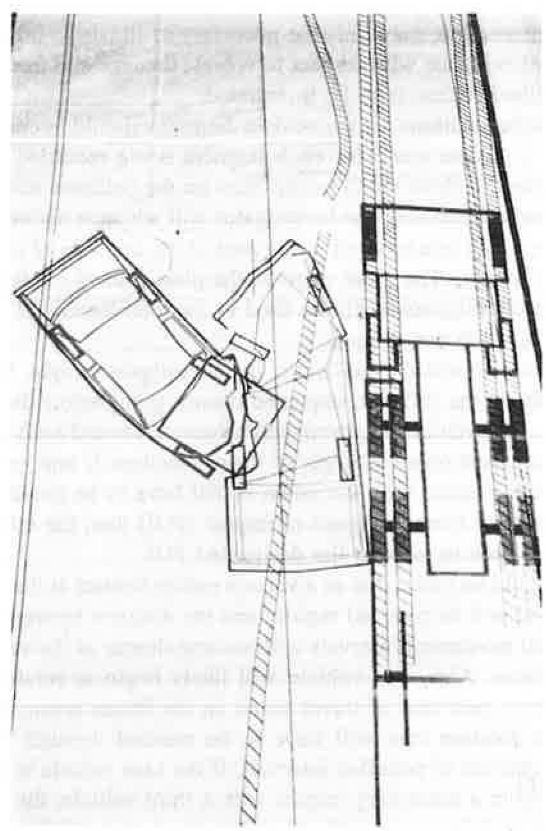


FIGURE 3 Conversion to sideswipe-type impact.

tire skid mark (long hatched lines) and gouges. A deep gouge in the center of the passing lane of the roadway could be traced to the driveshaft of the Aspen, which was found fractured during the vehicle inspection. In this way the physical evidence at the scene pointed to a specific location for the initial point of impact.

The damage to the front ends of both vehicles was also closely scrutinized to establish the overlap between the vehicles as they first made contact. The right edge of the direct damage could be easily identified on the front end of the LeSabre. Also, the LeSabre's front bumper was gouged deeply at its left corner. The location of the gouges along the bumper and their relative positions were measured accurately. Scale diagrams, which were drawn of both vehicles, allowed the gouges on the LeSabre's front bumper to be matched exactly with the end of a suspension bar on the front undercarriage of the Aspen. This suspension bar contained ridges in a pattern that exactly matched the set of gouges observed on the LeSabre's front bumper.

Not only was it possible to match the front ends of the two passenger cars at the initial contact, but also when the vehicles were placed on the collision scene, the fractured portion of the Aspen's driveshaft matched the location of the deep gouge mentioned previously. In addition, the LeSabre's right front tire was located exactly on the skid mark that was also identified previously. Thus the vehicle damage evidence and the scene evidence complemented each other (Figure 2).

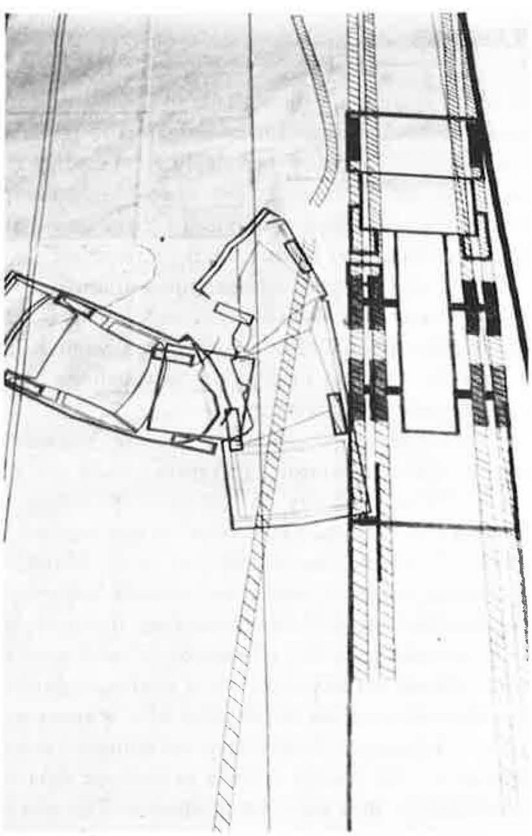


FIGURE 4 Brushing contact of right side of LeSabre against truck-tractor's wheels and an impact of the trailer's landing gear with right rear end of car.

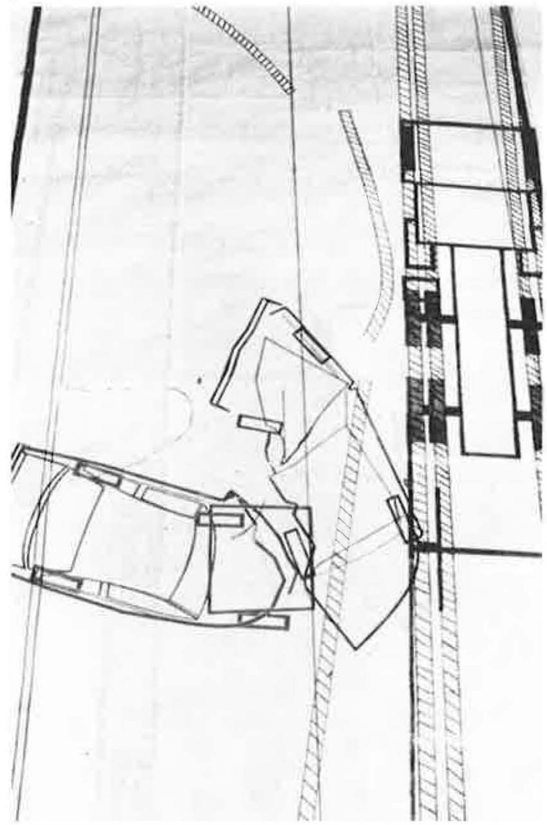


FIGURE 5 Crush to the right rear end of LeSabre.

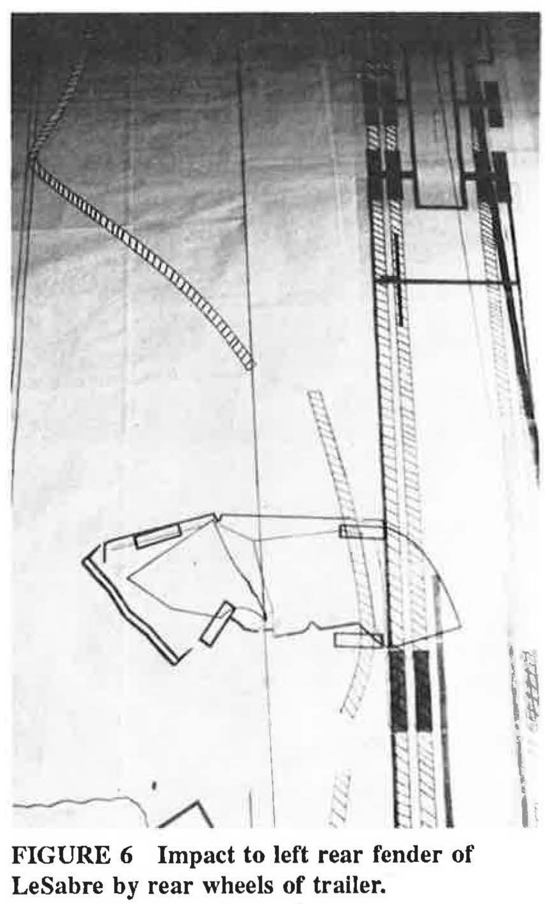


FIGURE 6 Impact to left rear fender of LeSabre by rear wheels of trailer.

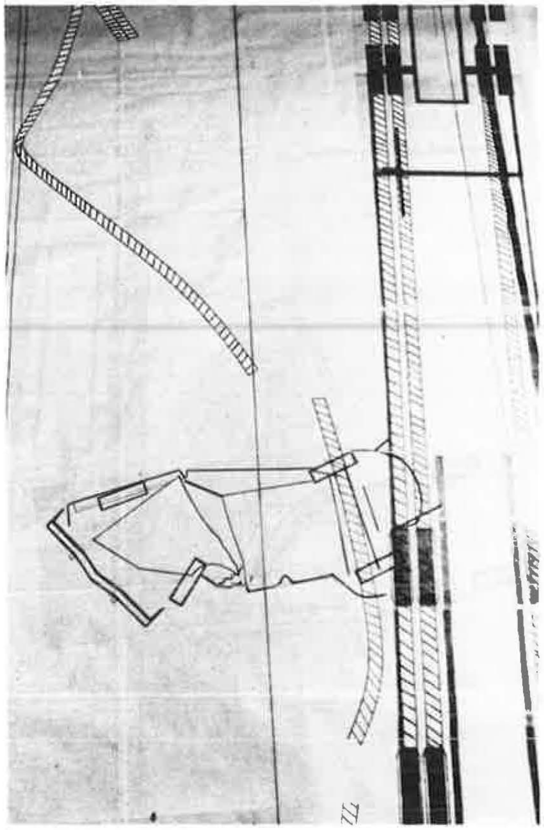


FIGURE 7 Crush to left rear fender of LeSabre.

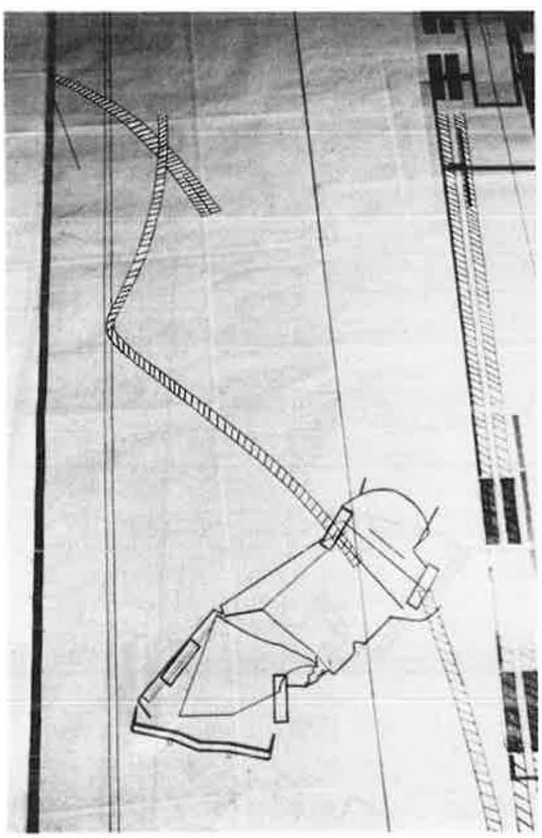


FIGURE 8 LeSabre rotates strongly counterclockwise toward final position.

The rest of the physical evidence that led to the detailed reconstruction of this collision will not be described at length. All other vehicle movements were based on interpretations of physical evidence similar to that just described for the initial impact.

After the initial frontal impact, the Aspen rotated counterclockwise and contacted the left side of the LeSabre (Figure 3) and the LeSabre rotated counterclockwise toward an adjacent lane containing the approaching tractor-trailer. The LeSabre sustained a brushing contact to its right side from the wheels of the truck-tractor, followed by a substantial impact to the right rear end by the trailer's left-side landing gear (Figures 4 and 5).

The LeSabre continued to rotate so that its rear end slid underneath the trailer, and the left rear wheels of the trailer struck the rear left side of the car (Figures 6 and 7). (Note that at this impact the collision scene had to be scrolled to keep the vehicles in view of the camera.)

After this last impact, both vehicles followed paths along scene evidence to their final positions. The tractor-trailer's wheels were locked from braking and produced a set of skid marks at the collision scene, which allowed the determination of its final position. The LeSabre rotated strongly after the third impact, and this produced a set of curved tire scuff marks and fluid sprays on the scene, which allowed the car's precise location to be determined on its way to its final position (Figure 8).

CONCLUSIONS

With the technological advances made in videotape recording equipment it is now possible for investigators to produce animations of collision events of reasonably good quality at moderate cost. However, although the capital cost of the equipment is not great, there is a hidden cost in terms of the time that must be spent in recording and editing the reconstruction.

The benefits of this type of videotaped animation can be substantial. Such graphical displays provide a useful medium for describing the importance of evidence from vehicle damage and the collision scene as these relate to situations involving quite complex vehicle dynamics.

The use of computer programs for collision reconstruction has become quite common. Programs such as SMAC (*8*, pp. 155–173) rely heavily on matching the output of the collision algorithm to the physical evidence documented by the investigator. Reconstructing a collision using SMAC is an iterative process; the input conditions are modified in the light of the accumulated results from various runs. Because physical evidence is an exact indicator of position, a videotaped reconstruction produced entirely from such evidence should give valuable information to the investigator who is about to input data into such a program. Conversely, the output of computer-based reconstructions can be utilized as the base data for the vehicle movements in a collision animation. The videotaped reconstruction brings the computer output to life in a much more meaningful way than a static plot of the vehicle dynamics. Computer-generated graphics have been used directly for videotaped reconstructions of collision situations (*9*, *10*). An extension of the two-dimensional method reported in this paper has been the recording of reconstructed collision events with the use of three-dimensional models (*11*).

ACKNOWLEDGMENTS

The authors wish to express their appreciation to John Tryphonopoulos and Maridon Duncanson, both members of the research team, for their investigative contributions to this paper; to Jim Barrie, Ontario Police College, for his assistance in producing an example of this video reconstruction technique; and to the Road Safety and Motor Vehicle Regulations Directorate of Transport Canada for their financial support of the team's collision investigation research programs.

REFERENCES

1. S. M. Leung and D. Borden. An Alternative Method to Measure Traffic Accident Scenes. In *Proceedings of the Canadian Multidisciplinary Road Safety Conference IV*, 1985.
2. A. German, Z. M. Gorski, and E. S. Nowak. Collision Reconstruction Techniques: Collision Scene Analysis. In *Proceedings of the Canadian Multidisciplinary Road Safety Conference IV*, 1985.
3. N. S. Tumbas. *Instruments for Determining the Trajectory of a Vehicle in a Collision and Measuring the Damage It Sustains.* SAE 850246. Society of Automotive Engineers, Warrendale, Pa., 1985.
4. Z. M. Gorski, A. German, and E. S. Nowak. Collision Reconstruction Techniques: Vehicle Damage Analysis. In *Proceedings of the Canadian Multidisciplinary Road Safety Conference IV*, 1985.
5. L. McQuillan. *The Video Production Guide*. Howard W. Sams & Company, Indianapolis, Ind., 1983.
6. M. D. Shetter. *Videotape Editing: Communicating with Pictures and Sound*. Swidevski Electronics Inc., Ill., 1982.
7. K. Laybourne. *The Animation Book*. Crown Publishers, Inc., New York, 1979.
8. I. S. Jones and D. J. Segal. The Application of the SMAC Accident Reconstruction Program to Actual Highway Accidents. In *Proceedings, 18th Conference, American Association of Automotive Medicine*, AAAM, Arlington Heights, Ill., 1974.
9. J. E. Martinez. *Computer Aided Automobile Accident Reconstruction: Theory, Graphics, Animation*. SAE 850250. Society of Automotive Engineers, Warrendale, Pa., 1985.
10. W. S. Reed, A. T. Keskin, and C. J. Sing. Simulating Vehicular Collisions. *Mechanical Engineering*, Oct. 1985, pp. 63–68.
11. R. A. Christopher and K. R. Feder. Modelling of Vehicular Accidents. *Journal of Forensic Sciences*, Vol. 26, No. 2, April 1981, pp. 424–430.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

Analysis of High-Hazard Locations: Is an Expert Systems Approach Feasible?

GARY S. SPRING, JOHN COLLURA, AND PAUL W. SHULDINER

The focus of this paper is the detailed analysis of specific highway locations that have been identified as hazardous in the framework of a state Highway Safety Improvement Program (HSIP). A methodology is proposed for implementing a microcomputer-based prototype expert system that would perform the location analyses described here. A prototype system would be used to assess the feasibility of building usable microcomputer-based expert systems for this application and to make recommendations for the design and implementation of such systems. By automating these activities with low-cost, easy-to-use-computer technology, it is hoped that the effectiveness of state highway transportation agency operations will be enhanced and that the provision of consistent and comprehensive analysis procedures will improve the overall safety and efficiency of the highway network. For automation to be feasible, certain minimum requirements must be met. With those in mind, a review of current state HSIPs was conducted. It was concluded that computerization of these analyses by using expert systems concepts on a microcomputer is technically feasible. A methodology to develop such a system for a state highway agency is proposed.

Department of Civil Engineering, University of Massachusetts, Amherst, Mass. 01003.

Last year 43,607 people died in traffic accidents on the nation's highways (*1*). Traffic accidents are one of the major causes of death in the United States today and have been since the beginning of this century. However, it was not until the late 1950s and beyond that the numbers began to grow to alarming proportions. The combined effects of the growing highway system and the nation's growing affluence caused a tremendous increase in the number of people and cars on the road, increasing chances for accidents to happen.

The provision of safe highways presents a challenge to today's highway professional. Essential to meeting that challenge is an organized approach to identifying and correcting highway safety problems. A chronology of the nation's safety efforts is given in Table 1 (*2*). Although there has always been a universal concern for highway safety, before the mid-1960s there was no central coordination of efforts. Safety considerations rested largely with the individual states (*3*). States developed programs such as motor vehicle inspection, highway design standards, and high-frequency accident location identification and correction. However, not all states were active in all areas. Federal efforts were equally fragmented. Although the Bureau of Public Roads had the main responsibility for highway safety, several other governmental agencies had programs that were concerned directly or indirectly with the subject.

Finally, in the mid-1960s coordination efforts began in earnest. The 1966 Highway Safety Act authorized the federal government to provide financial assistance to states that

TABLE 1 NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION HISTORICAL AND LEGISLATIVE BACKGROUND

Year	No. of Traffic Fatalities	Significant Event
1924	18,400	National Conference on Street and Highway Safety (convened by Secretary of Commerce Herbert Hoover)
1937	37,819	Second National Conference on Street and Highway Safety; report: *Guides to Traffic Safety*
1946	31,874	Third National Conference produced Action Program for Highway Safety
1954	33,890	President's Committee for Traffic Safety established and adopted Action Program
1956	37,965	First Congressional interest: Subcommittee on Health and Safety, House Committee on Interstate Commerce
1958	35,331	Secretary of Commerce authorized to assist in carrying out the President's Action Program and to cooperate with the states in furthering highway safety
		Interstate Compacts for Traffic Safety (Beamer Resolution)
1959	36,223	Report by the Secretary of Commerce to Congress on magnitude of traffic safety problems and the role the federal government should play in attacking them
		Requirements for passenger-carrying motor vehicles purchased for use by the federal government to meet certain safety standards
		Prohibition of use in commerce of any motor vehicle that discharges substances in amounts found by the Surgeon General to be injurious to human health
1960	36,399	Registration of automobile license revocations (National Driver Register)
1961	36,285	Requirements for passenger-carrying motor vehicles for use by the federal government to meet certain safety standards
		Hydraulic brake fluid specifications
1962	38,980	Standards for seat belts in automobiles sold or shipped in interstate commerce
1965	47,089	Amendment to the Federal-Aid Highway Act providing for voluntary state highway safety standards (Baldwin Amendment)
1966	50,894	National Traffic and Motor Vehicle Safety Act of 1966; established the National Traffic Safety Agency in the Department of Commerce
		Highway Safety Act of 1966; established the National Highway Safety Agency in the Department of Commerce
		Department of Transportation Act of 1966
1967	50,724	Executive Order 11357 combined the two foregoing agencies in the Department of Transportation as the National Highway Safety Bureau
1968	52,725	Report of the Secretary's Advisory Committee on Traffic Safety
		National Traffic and Motor Vehicle Safety Act of 1966, amendments
1970	52,627	National Traffic and Motor Vehicle Safety Act of 1966, amendments
		Federal-Aid Highway Act of 1970
"		Report of the President's Task Force on Highway Safety: *Mobility Without Mayhem*
1972	54,589	Motor Vehicle Information and Cost Savings Act
		National Traffic and Motor Vehicle Safety Act, amendments
1973	54,052	Federal-Aid Highway Act of 1973
1974	45,196	Motor Vehicle and Schoolbus Safety Amendments of 1974
1975	44,525	Federal-Aid Highway Act of 1973, amendments
		Energy Policy and Conservation Act established the Automotive Fuel Economy Program by adding a new Title V to the Motor Vehicle Information and Cost Savings Act
1976	45,523	Federal-Aid Highway Act of 1976
		National Traffic and Motor Vehicle Safety Act, amendment and authorization
		Motor Vehicle Information and Cost Savings Act, amendments
1977	47,878	Automobile Fuel Economy Program amendment, contained in the Department of Energy Organization Act
1978	50,331	Highway Safety Act of 1978 (included as Title II of the Surface Transportation Assistance Act of 1978); includes an amendment to section 158(b) of the National Traffic and Motor Vehicle Safety Act of 1966
		Automobile Fuel Economy Program amendments, contained in the National Energy Conservation Policy Act

SOURCE: For 1924–1974, National Center for Health Statistics, Department of Health, Education, and Welfare, and state annual summaries (adjusted to 30-day deaths). For 1975–1976, Fatal Accident Reporting System (FARS), NHTSA.

NOTE: Federal government entities concerned with highway safety include the Department of Health and Human Services, Department of Commerce, Department of Defense, U.S. Postal Service, General Services Administration, Interstate Commerce Commission, Interdepartmental Highway Safety Board, and President's Committee for Highway Safety.

develop and maintain a safety program. The basic features of the act cast the mold for national highway safety policy (*3*). Imposing uniformity on state and local programs provided much-needed coordination of the nation's highway safety efforts. Subsequent highway safety acts have expanded the guidelines set forth in 1966. The U.S. Department of Transportation (DOT) developed the Highway Safety Program Manual (HSPM) in 1974 to guide state and local agencies in conforming with the acts. The HSPM contains 18 safety standards. Standard 9 (*4*)

> requires the development of a program for identifying and maintaining surveillance of locations having high accident experience. After identifying hazardous locations, the state must take appropriate measures to reduce accidents and to evaluate the effectiveness of safety improvements at these locations. Also, a program must be developed to maintain surveillance of the roadway network for potentially high accident locations and for correcting problems at these locations. Each state is required to periodically evaluate their program and provide the Federal Highway Administration (FHWA) with an evaluation summary.

Efforts have paid off. The fatality rate in persons per million vehicle miles peaked in 1966 at around 5.5 and has since decreased to its present value of about 2.6 (*1*,*2*).

To aid state and local agencies in the design and implementation of highway safety programs within the framework of HSPM Standard 9, FHWA formally defined a Highway Safety Improvement Program (HSIP) [Federal-Aid Highway Program Manual, Volume 8, Chapter 2, Section 3 (FHPM 8-2-3), March 1979]. The HSIP consists of components for the planning, implementation, and evaluation of safety programs and projects. The three components consist of processes related to

- Collecting and maintaining data,
- Identifying hazardous locations,
- Analyzing those locations (diagnosing their problems and developing countermeasures to the problems),
- Developing improvement projects,
- Establishing project priorities,
- Scheduling and implementing projects, and
- Determining the effects of safety improvements.

The arrows in Figure 1 indicate the flow of information in the HSIP at the process level (*4*).

RATIONALE AND SIGNIFICANCE OF AUTOMATION

In these days of fiscal austerity, it is essential that public agencies increase the efficiency and effectiveness of their operations. The location analyses (see Figure 2) (*5*) that are part of the HSIP are tedious and time consuming, making their computerization very desirable to the analyst. However, these analyses are based primarily on experience and good engineering judgment, and so conventional computer programs do not quite fill the bill. A microcomputer-based expert system would combine the best of both worlds. It would provide all the advantages intrinsic to microcomputers in the workplace as well as allow much of the work to be done by technicians.

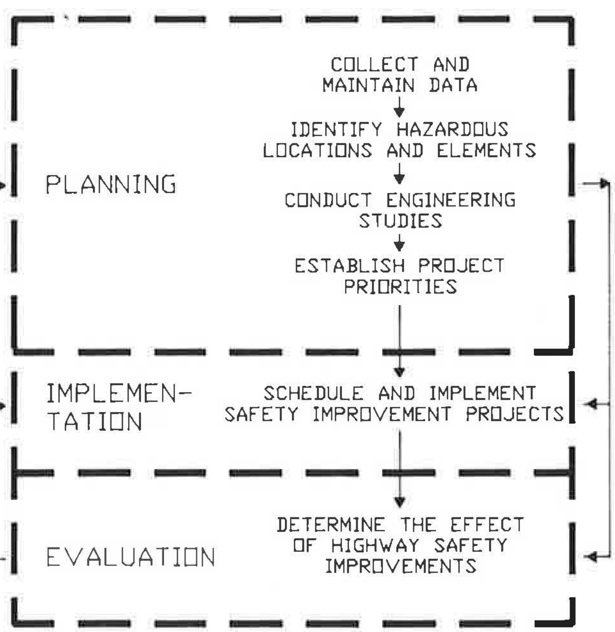


FIGURE 1 Highway safety improvement program at the process level.

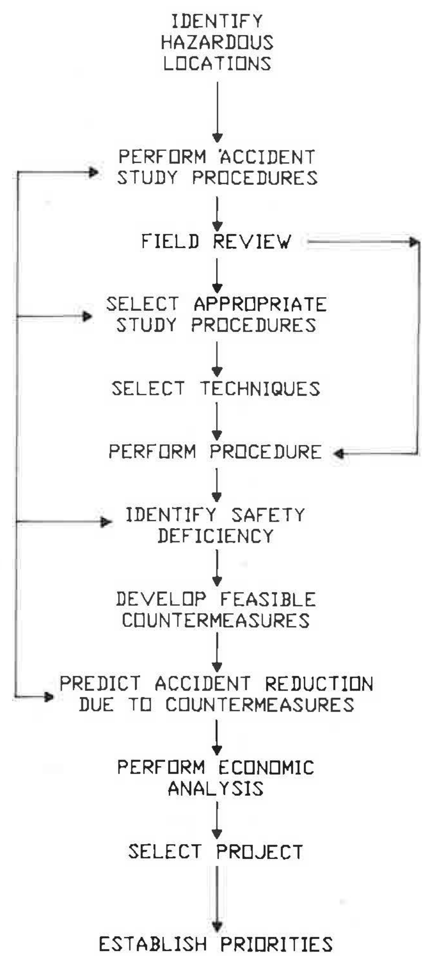


FIGURE 2 Engineering investigation model.

Expert systems are computer programs that solve problems too complex for conventional software, in other words, problems that cannot be represented by a model and that are based on judgment. Perhaps the most useful characteristic of expert systems is that they allow much of the expert's work to be done by technicians, thus freeing professional staff for other duties. They expose the nonexpert to the reasoning processes of the expert and offer advice to improve the nonexpert's understanding of the problem situation by proposing strategies for dealing with it. They never forget a rule or commit a simple oversight. They never go on vacation or get sick. They are easily reproduced, just by copying a few floppy disks, so that their expertise is readily available to anyone who needs it.

The primary objective of this paper is to propose a methodology for implementing a microcomputer-based prototype expert system that would perform the location analyses described here. Such a system would be used to assess the feasibility of building usable microcomputer-based expert systems for this application and to make recommendations for the design and implementation of such systems. By automating the location analyses on low-cost, easy-to-use computer technology, it is hoped that the effectiveness of state highway transportation agency operations will be enhanced. Further, by providing consistent and comprehensive analysis procedures, it is hoped that the overall safety and efficiency of the highways will improve.

The type of system considered here is for state-level highway safety programs. States are required to maintain the large accident data bases necessary to monitor their highway systems. Their safety programs are very large and therefore much more difficult to maintain than smaller-scale systems. However, they are plagued with the same money and staff shortages as are other public agencies. Consequently, it appears to these researchers that their need for an automated analysis procedure is even greater than it is at other levels of government.

IS AUTOMATION FEASIBLE?

Before the feasibility of a system such as the one proposed here can be examined, it is necessary to define what is meant by

Dear sir:

We are presently reviewing and assessing the status of our nation's Accident Records Systems. Our area of interest pertains to the procedures currently being used to identify, analyze and improve high hazard locations on our highway network. We would appreciate any information about your system that you would be able to share with us. Items of specific interest are:

- What data files are used in your computerized ARS (e.g. Accident reports, traffic volumes, location file, geometric information)?
- What software is used to manipulate these files (e.g. canned statistical packages, data base managers, or in-house programs written in Fortran, Cobol or some other programming language)?
- How are highway locations specified in the location file (e.g. mile markers, or coordinates) and what increment is used (e.g. 0.01 miles)?
- How does your system identify high hazard locations? Is an established method such as Rate Quality Control used or do you specify criteria of your own?
- When an hazardous locations list has been generated, what analyses are performed for problem diagnsis?
- How are appropriate improvements, which result from the diagnosis, identified and implemented?
- How are Before and After studies conducted?

This and any other information you send us about your system will be a great help to us in conducting our research. The objective of this research is to develop microcomputer based software (perhaps with the use of expert systems concepts) which will interface with mainframe computer accident data files and which will interactively perform identification and analysis procedures. By automating these procedures on low cost, easy to use computer technology, it is hoped that the effectiveness of State highway transportation agency operation would be enhanced. Further, by providing consistent and comprehensive analysis procedures, it is our intention to improve the overall safety and efficiency of our highway network.

Thank you for your interest. We look forward to hearing from you.

FIGURE 3 Letter of inquiry.

"feasible." A system that could and would be used by a majority of state highway agencies is a feasible one. The general process of automating location analysis on a microcomputer will be examined first. The applicability of an expert systems approach will then be examined.

For automation of these analyses on a microcomputer to be feasible, a majority of states must meet the following requirements:

- Data maintenance and identification of hazardous locations should be automated;
- It should be possible to process a location's accident data automatically;
- Manual analysis procedures currently in use should, at least in part, use downloaded accident data and should have the same general form; and
- Current automation efforts should not be duplicated.

General Approach

To determine whether these requirements are met, the safety improvement program of each state was reviewed. The letter shown in Figure 3 was sent to the 50 state departments of transportation (DOTs) across the country. The results of the survey are presented in Tables 2–5. The implementation component of the HSIP is not included in the results because it is largely administrative rather than analytical in nature. Thirty-three responses were received and supplemented with a review

TABLE 2 PROCESS ONE—PLANNING COMPONENT

STATE	REFERENCE METHOD					COMPUTER FILES					
	MILEPT	REF PT	LINKNODE	COORD	LORAN C	ACCIDENT	TRAFFIC	RD CLASS	HAZ LOC	GEOMETRY	OTHER
ALA	X*		X			X	X	X		X	
ALSK				X		X	X	X			X
ARIZ	X*	X		X		X	X	X		X	X
CAL	P					X	X	X		X	X
CONN	P*					X	X	X		P	
DEL	P*	P				X	X	X		X	X
FLA	X		X			X	X	X		X	X
GA	X					X	X	X			
IND		X				X					
KAN	X					X	X	X		X	
KEN	X					X	X	X			
LA	P					X	X	X		X	X
MASS		X		X*		X	X	X			
ME			X			X	X	X			X
MICH	P					X	X	X		X	
MINN	X					X	X	X		X	X
MISS	P	P				X	X	X	X		X
MO	P					X	X	X	F	F	F
MONT	X					X	X	X			
NC	P					X	X	X		X	X
ND	X		X			X	X	X		X	X
NEB		X				X	X	X			
NEV	X*	X				X	F	M		F	
NH			X			X	X	X			X
NJ	X*		X			X	X	X	X	X	X
NY		X*	X			X	X	X		X	X
OKLA		X*		X		X	X	X			
OREG	P					X	X	X			
PENN		X				X	X	X		X	X
TEX	X					X	X	X			X
UTAH	X*	X				X	X	X		X	X
VA	P		F			X	X	X			X
VT	X					X	X	X		X	X
WASH+	X*					X	X	X			X

* State Highway only
P indicates document method
F indicates future implementation
M indicates manual procedure
\+ Data files are not linked automatically

TABLE 3 PROCESS TWO—PLANNING COMPONENT

STATE	IDENTIFICATION METHOD							
	FREQ	RATE	FRQ/RATE	RQC	SEVERITY	HAZ IND	HAZ FEAT	OTHER
ALA				X				
ALSK								X
ARIZ		X			X			
CAL	X							
CONN				X				
DEL			X					
FLA				X			X	X
GA				X	X			
IND	X							
KAN				X	X			
KEN				X				
LA				X				
MASS			X+			X		
ME				X				
MICH	X							
MINN	X	X			X			X
MISS				X		X		
MO			X					
MONT			X		X			X
NC								X
ND			X		X			
NEB				X				
NEV						X		
NH					X			X
NJ						X		X
NY				X				
OKLA				X	X			
OREG						X		
PENN						X		X
TEX			X					
UTAH		X					X	X
VA				X				
VT				X	X			
WASH+					X			

of safety literature. The results of this effort were used to answer the following four questions:

1. What activities are currently on line and what types of data sorts are available from mainframe data bases?

Clearly, data must be available for downloading to the microcomputer in order for automation to be advantageous.

2. Do these location analyses lend themselves well to automation?

Not only must the data be available for downloading, but the analysis procedure must use those computerized data; otherwise, automating it would not be worth the effort.

3. What types of analyses are currently performed?

An automated analysis procedure would probably not be acceptable to a majority of users if manual procedures currently used are not primarily based on downloaded data and are not fairly uniform.

4. Are there any procedures that are automated now?

It is also necessary to know what efforts, if any, have been made to do what is proposed here, so the work will not be duplicated, and to ensure that all efforts complement each other whenever possible.

TABLE 4 PROCESS THREE—PLANNING COMPONENT

	COLLECT & ANALYZE DATA				DEVELOP COUNTERMEASURES				CHOOSE IMPROVEMENT ALTERNATIVE					
STATE	ACC BASE	TRAFF OP	ENVIRONS	SPECIAL	ACC PATT	FAULT TR	FIELD	OTHER	COST EFF	BEN/COST	RATE-RET	TIME-RET	NET BEN	OTHER
ALA	X				X		X							X
ALSK								X						X
ARIZ	X				X		X							X
CAL	X				X		X			X				X
CONN	X	X	X		X		X			X				X
DEL	X	X			X		X							X
FLA	X	X	X		X		X			X				
GA	X				X		X			X				
IND	X	X	X		X		X			X				
KAN	X				X					X				
KEN	X				X		X							X
LA	X				X		X							X
MASS	X				X									
ME	X	X	X		X		X		X					
MICH	X	X	X		X		X		X					
MINN	X				X		X							X
MISS	X	X	X	X	X		X			X				
MO	X	X	X		X		X							
MONT	X				X		X			X				
NC	X	X			X					X				
ND	X		X		X									
NEB	X	X	X		X		X			X				
NEV	X	X	X		X		X		X					
NH	X	X	X		X		X							X
NJ	X				X		X							X
NY	X				X		X			X				X
OKLA+														
OREG+	X													
PENN	X				X					X				
TEX	X						X	X						
UTAH	X				X		X			X				
VA	X				X		X			X				
VT	X		X		X									X
WASH	X				X									

+ Automated analysis

TABLE 5 PROCESS FOUR—PLANNING COMPONENT; PROCESS ONE—EVALUATION COMPONENT

	PRIORITIZATION					EVALUATION				
STATE	PROJ DEV	INC BC	DYN PROG	LIN PROG	OTHER	ACC BASE	NON ACC	PGM EVAL	ADM EVAL	OTHER
ALA			X							X
ALSK					X	X				
ARIZ	X					X				
CAL		X				X				
CONN	X					X		X		
DEL					X	X				
FLA					X	X		X		
GA	X					X				
IND	X					X				
KAN	X					X				
KEN						X				
LA	X									
MASS										
ME	X									
MICH	X					X				
MINN	X					X				
MISS	X					X	X	X	X	
MO	X					X		X		
MONT	X					X				
NC										
ND					X					X
NEB	X					X				
NEV	X					X		X		
NH					X	X				
NJ						X				
NY	X					X				
OKLA	X					X				
OREG										
PENN										
TEX										
UTAH	X									
VA					X	X				
VT						X				
WASH						X				

Discussion

Activities On Line and Data Sorts Available (Question 1)

In order to perform any kind of location analysis, it is necessary that, at the very least, states have accident and highway classification data on line. Further, an essential element of a traffic records system is an accurate highway location reference system. If an agency cannot pinpoint the location of accidents or other roadway data, problem locations cannot be accurately identified. Consistency of the reference system between the files was stated to be a real problem by many states. However, all but one have these data on line. Of those, only one does not provide automatic linkage of the files.

Some of the respondents have complex, state-of-the-art data base management systems (DBMS) and many others plan to implement such systems in the future. A more complete discussion of the establishment of comprehensive data systems may be found elsewhere (*6,7*). Those without a DBMS in place have programs written in a programming language (FORTRAN or COBOL) that were developed in house to manipulate their data files. In any event, all can supply a wide range of data sorts to the user for analysis. Some also perform statistical analysis of data, for example, to identify accident types that are overrepresented at a particular location. In addition, all the respondents except the two mentioned previously have automated the identification process and regularly generate listings of suspect locations. Most listings are made according to some criterion such as the ratio of accident rate to critical rate or by frequency. Those few that do not generate listings use interactive computer systems to perform cluster analysis or some other user-specified type of analysis.

Automation Possibility, Studies Now Performed, and Work Now Under Way (Questions 2–4)

All respondents use accident-based data as input to the analysis process and accident pattern analysis as a major component of the process, and all analyses have basically the same form. Most states perform location analysis manually. Few have automated analyses. Alaska has an interactive system in place on their IBM XT/370 and Tektronix 4125 that allows the user to perform cluster analysis to identify specific problems at each location. Texas has contracted with the Texas Transportation Institute to develop microcomputer software that analyzes accident experience at high-hazard locations. The software identifies factors overrepresented in accident occurrence at these locations relative to the average for similar highways in the area (*8*).

Similarly, although most states rely on collision diagrams as part of their analysis, only three or four can generate them by computer. Most also perform one of the standard economic analyses to choose the desired improvement alternatives as well as to rank projects by priority.

Current automation efforts use conventional software that either processes and selectively sorts data or performs some type of statistical analysis on the data. Although useful to the expert, the outputs are of limited use to the unskilled technician.

Summary

The feasibility of automating location analysis, as defined in the FHWA's HSIP, on a microcomputer was examined. Four questions were asked to determine that feasibility and a review of current state HSIP efforts was conducted to answer them. As a result of the review, the following observations may be drawn:

- All but two respondents have essential accident and highway data on line, can provide a wide range of data sorts for any specified location, and have a computerized high-accident location identification system.

- The location analysis performed by all respondents is, at least in part, accident based. This means that the most common procedure uses accident data that can be downloaded as input, which in turn means that data input can be automated.
- Most respondents use accident patterns as input to a field review to diagnose problems and to develop countermeasures at suspect locations. This means that only one location analysis procedure need be automated for a majority of states to use it.
- Very few states use automated location analysis procedures. Those few require an expert to run them and to interpret their output. Further, they are designed to be used with a specific system.

In short, automation is feasible. The minimum requirements for automating the location analysis portion of the HSIP are met. Because all states perform the analyses but few have automated them, automation is desirable as well.

IS AN EXPERT SYSTEMS APPROACH APPLICABLE?

To deal with the question of applicability, it is necessary first to know about expert systems, namely, what they are, what they do, how they work, where they have been applied successfully, and, perhaps most important, for what types of problems they apply. Second, it is necessary to examine the location analysis problem and, finally, to compare problems handled by expert systems with the location analysis problem to determine whether expert systems are amenable to this type of problem.

Expert Systems

Artificial intelligence (AI) is that part of computer science concerned with designing "intelligent" computer systems. That is, AI systems exhibit the characteristics usually associated with intelligence in human behavior—understanding language, learning, reasoning, solving problems, and so on. Expert systems are computer programs that apply AI problem-solving techniques to complex real-world problems normally done by experts. They attempt to use the knowledge of human experts to solve problems (*9*). Their use of domain-specific knowledge, in contrast to other AI applications that use more general reasoning methods, gives them an enormous amount of problem-solving power by greatly reducing the solution space that must be considered.

Knowledge in any specialty is usually of two sorts: public and private. Public knowledge (also referred to as "deep" knowledge) includes published definitions, facts, and theories typically found in texts and references in the domain of study. Private knowledge ("surface" knowledge) is heuristic, experiential knowledge that comes from successfully solving many problems in a specific domain, that is, doing things again and again, getting a feel for the problem, learning when to go by the book and when to break the rules. Heuristics enable the human expert to make educated guesses, to recognize promising approaches to problems, and to deal effectively with faulty or incomplete data.

Expertise consists of knowledge about a particular domain, an understanding of domain problems, and skill at solving them. An expert is distinguished not only by how much he knows about his domain, but also by how quickly he recognizes patterns and brings rules to bear (*2*). So an expert system also requires a knowledge-processing component in order to perform expertly. This component is called the system's inference engine (sometimes called the system's interpreter). It is a computer program used for deriving conclusions about problem characteristics by using knowledge in the knowledge base. Finally, the system requires a user interface to enable the user to communicate with it. Hence, expert systems have three essential components: a user interface, a knowledge base, and an inference engine (*3*).

The knowledge base consists of facts and rules representing the heuristic knowledge about the problem domain. Rules often have the form IF (premise) THEN (conclusion), whereas facts are represented as assertions of the form (variable name) = (value). An example of this structure is Rule 31 taken from PUFF, a pulmonary function disorder diagnosis expert system (*10*):

> IF:
> 1) The severity of obstructive airways disease of the patient is greater than or equal to mild, and
> 2) The degree of diffusion defect of the patient is greater than or equal to mild, and
> 3) The tlc observed/predicted of the patient is greater than or equal to 110, and
> 4) The observed-predicted difference in rv/tlc of the patient is greater than or equal to 10
>
> THEN:
> 1) There is strongly suggestive evidence (.9) that the subtype of obstructive airways disease is emphysema, and
> 2) It is definite (1.0) that "OAD, Diffusion Defect, elevated TLC, and elevated RV together indicate emphysema." is one of the findings.

The inference engine employs search procedures to manipulate and use these rules. Two common strategies are backward chaining and forward chaining. Backward-chaining (or goal-directed) strategies require selecting a goal and then scanning the rules to find those whose consequent actions will achieve that goal, trying to satisfy those rules from facts or from the conclusions of other rules, and so on until the goal is met or not met. In the forward-chaining (or data-driven) approach, the rules are searched to determine what conclusions can be drawn from information provided by the user, facts in the knowledge base, and previous conclusions. As conclusions are reached, the premises of other rules are satisfied, and the search process continues until no more conclusions can be made or until a goal is met, whichever comes first. As a simple example, if the PUFF system were provided a patient's signs and symptoms that matched the premises of the sample rule shown above, it would deduce that the patient is likely to have emphysema (i.e., it would use its knowledge about pulmonary disorders to interpret the given problem attributes).

A fundamental difference between this type of system and other types of computer systems lies in the nature of the problems that they solve. Conventional computer programs solve well-defined, well-understood problems. They use a small amount of knowledge (e.g., a mathematical model) and apply it over and over again in their solution of a problem. The expert system is applied to ill-defined, poorly understood problems. It uses an heuristic knowledge base to narrow the number of alternative solutions to a set of the most likely ones.

The Location Analysis Domain

Location analysis problems are like Sherlock Holmes mysteries: all the pieces are there, but the expert—Sherlock—is required to put them together in a meaningful way so that he can figure out what is going on (i.e., solve the problem). The safety engineer solves location analysis problems in much the same way. He uses his knowledge about why accidents happen to figure out (or to deduce) what is wrong at a particular highway location. This suggests an expert systems approach.

Another feature of the location analysis process that recommends it to an expert systems approach is the availability of expert knowledge in the domain; it should be remembered that expert systems use large amounts of it. Several efforts have been made to write down common "rules of thumb" used in the location analysis process. One such effort was made by FHWA in its *Highway Safety Engineering Studies Procedural Guide* (*5*). It includes a review of general countermeasures for accident patterns and their probable causes. The items tabulated are typical bits of knowledge long used by safety engineers to solve the mystery of what is happening. Box, in his article in *Traffic Engineering* (*11*), also presents some insights into what to look for when analyzing a problem location. Much has been written on this subject (*5,12*), all very similar to the FHWA study and the Box article. A great deal of this work is based on good common sense and years of experience.

Applications of Expert Systems

Expert systems have been successfully applied in many different areas; the general types of systems to which they can be applied are listed in Table 6. The domains in which expert systems have been applied successfully include medical diagnosis, mineral exploration, natural language understanding, and many more. All deal with problems that are poorly defined, not well understood, and data poor. The fact that a great many transportation problems are of that sort suggests some very exciting possibilities for the future of expert systems applications in that domain. Takallou (*13*) points out in his review of expert systems applications in civil engineering that there is a lack of ongoing research in the area of transportation engineering. In fact, no expert systems have been developed dealing with highway safety.

TABLE 6 EXPERT SYSTEMS APPLICATIONS

Category	Problem Addressed
Interpretation	Inferring situation descriptions from sensor data
Prediction	Inferring likely consequences of given situations
Diagnosis	Inferring system malfunctions from observables
Design	Configuring objects under constraints
Planning	Designing actions
Monitoring	Comparing observations with plan vulnerabilities
Debugging	Prescribing remedies for malfunctions
Repair	Executing a plan to administer a prescribed remedy
Instruction	Diagnosing and repairing student behavior
Control	Interpreting, predicting, repairing, and monitoring system behaviors

Summary

From the preceding discussion, an expert systems approach to performing location analyses makes sense. Not only have these location analysis problems been neglected but it appears that they are tailor made for an expert systems approach. In addition, there is a well-documented knowledge base from which to draw.

METHODOLOGY

In the previous two sections the feasibility and desirability of automating and the applicability of an expert systems approach to performing location analyses on a microcomputer have been examined. It was concluded that automation is feasible and desirable and that the expert systems approach does apply for this domain.

The basic expert systems structure that could be used for a prototype system is shown in Figure 4. A detailed description of the major components follows.

Knowledge Base

Perhaps the most critical issues that must be addressed when building an expert system are knowledge acquisition and representation. Knowledge acquisition [i.e., the process by which expert knowledge is captured for use in a knowledge-based expert system (KBES)] is not a simple linear process. One is not trying to capture fixed algorithmic approaches to problem solving. Rather, the knowledge to be acquired is heuristic, judgmental, subjective, and not necessarily organized. Further, the organization of the knowledge for application is not always consciously understood by the expert himself. To facilitate the

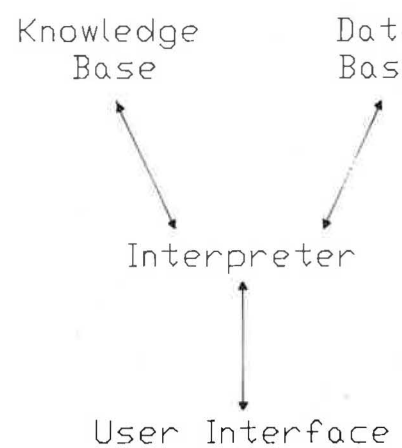


FIGURE 4 Basic structure of an expert system.

efficient and accurate acquisition of knowledge for this application, the process is best defined in the following stages (*14*):

- Identification, in which the important stages of the problem are characterized and goals for the entire project set;
- Conceptualization, during which the key attributes of location analysis are made explicit (some initial thought could be given to knowledge representation issues at this point); and
- Formulation, in which a formal model of the location analysis procedure and its key properties and relationships are mapped into a representation scheme.

The production system representation (an example of which was presented earlier) has been used with great success in many of the expert systems that have been built to date. The basic idea of this type of representation is that the data base consists of rules, called productions, in the form of condition-action pairs. The utility of the formalism comes from several facts: first, the conditions under which each rule applies are made explicit; second, the system's chain of reasoning can easily be traced, which makes it fairly simple to include explanation facilities in the system (see the discussion of transparency in the following section); and, finally, the knowledge is represented in a modular form, which facilitates system learning. For these reasons it is recommended that this system be designed, at least initially, by using the production formalism.

For a prototype system, the expert knowledge used could simply be the rules of thumb developed by FHWA in their engineering guide mentioned previously (*5*). As the system developed, it would of course be necessary to augment those rules with interviews conducted with safety experts.

The Inference Engine

The inference engine manipulates the knowledge base for presentation to a nonexpert user. The choice of inference engine is strongly coupled to the nature of the task that the system is designed to perform. The system, especially for transportation applications in which problems are generally unstructured and therefore cannot be completely captured by a model, must exhibit the quality of transparency. That is, the user should be able to see the chain of reasoning that led to a given outcome or recommendation. Transparency will be considered essential in developing this system.

The fact that the inference engine (the executive that runs the expert system) is separate from the knowledge base allows the use of "shells," which are general inference engines that can operate on different knowledge bases. A great number of shells have been developed just in 1986 alone (*15*). They range in price from around $50 for McGraw-Hill's Microexpert to $5,000 for Teknowledge's M.1 (*16*). A recent issue of *Computerworld* (*16*) includes a review of several of them and presents many of their essential characteristics. Whether to use any of these shells (and if so, which one) or to create the inference engine by using an appropriate programming language such as LISP depends largely on the goals and resources of the system's developer.

System Design and Implementation

To provide a system that can be used by any state that is interested, the system should be designed so that it can run in a wide variety of environments with its component hardware and software kept as inexpensive as possible. It should be designed and implemented for a pilot project in a specific state. The subject state in which to implement the system should have the following characteristics:

- Has requisite data available on mainframe,
- Has problems with HEP submissions because of staff shortages,
- Has microcomputer technology available to use the system,
- Is easily accessible to the builder of the system, and
- Has an interest in automating these analyses and in using its microcomputers.

The problem of interfacing with the chosen state's mainframe computer can be addressed at this point. First, a mainframe computer program would need to be written that prepares the accident information for downloading to a microcomputer. The communications format chosen depends on what type of link is available (i.e., hard-wired versus telephone line). The appropriate communications software can be chosen at this time. Once the microcomputer is linked to the mainframe, the file format must be ascertained so that data can be downloaded to the microcomputer. The uploading question, although potentially useful, does not have to be addressed at this time (information uploaded could be used to rank safety projects by priority statewide).

Testing provides feedback for problem reformulation, redesign of knowledge representation, and other refinements to the system. Therefore, the prototype system should be exercised by using a library of already-solved problem locations. The conclusions of the expert system could then be evaluated by comparing them with the human expert's solutions.

CONCLUSIONS

An expert systems approach to performing the location analysis portion of state HSIPs appears feasible. The true test, though, will come after a prototype system has been developed and implemented. Clearly, the benefits offered by expert systems are many. Nevertheless, there is more to developing them than simply buying a shell, hiring an expert, and writing some rules.

The tasks to which they are applied must be reasonably well defined and fairly narrow in scope. The chief reason for the success of expert systems lies in their specificity. If this is missing, this approach is not appropriate.

In addition, expert systems are expensive in terms of both time and money. So even if it turns out that the tasks are appropriate, they must be performed fairly often to make the system cost-effective. Also, the rules used to represent the knowledge base must be generally accepted by experts in the field. For such a large investment to be worthwhile, the results must be usable by more than just a handful of people. Most important, however, is that it must be possible to represent the knowledge by a set of rules.

To conclude, then, the system described here is theoretically feasible. Whether it can be reasonably designed and implemented is a question that must still be answered. It is hoped that the methodology presented here will be helpful in answering that question.

REFERENCES

1. *U.S. News and World Report*, Vol. 100, Feb. 17, 1986, p. 10.
2. *Highway Safety 1979*. NHTSA, U.S. Department of Transportation, May 1980.
3. *TRB Special Report 178: Future of the National Highway Safety Program*. TRB, National Research Council, Washington, D.C., 1977.
4. *Highway Safety Improvement Program*. Report FHWA-TS-81-218. FHWA, U.S. Department of Transportation, Dec. 1981.
5. *Highway Safety Engineering Studies Procedural Guide*. FHWA, U.S. Department of Transportation, June 1981.
6. *TRB Circular 293: Introduction to Comprehensive Computerized Safety Recordkeeping Systems*. TRB, National Research Council, Washington, D.C., July 1985.
7. *State and Community Program Area Report: Traffic Records*. NHTSA, U.S. Department of Transportation, Sept. 1985.
8. K. K. Mak, T. Chira-Chavala, and B. A. Hilger. *Automated Analysis of High Accident Locations*. Texas Transportation Institute, Austin, Tex., Jan. 1986.
9. F. Hayes-Roth, D. A. Waterman, and D. B. Lenat. *Building Expert Systems*. Addison-Wesley Publishing Company, Inc., New York, 1983.
10. E. A. Feigenbaum. The Art of Artificial Intelligence: Themes and Case Studies of Knowledge Engineering. *Proc., 5th International Joint Conference on Artificial Intelligence*, W. Kaufmann, Inc., Los Altos, Calif., 1977.
11. P. C. Box. Accident Pattern Evaluation and Countermeasures. *Traffic Engineering*, Aug. 1976.
12. T. K. Datta. *A Procedure for the Analysis of High Accident Locations for Traffic Improvements*. Wayne State University, Detroit, Mich., 1976.
13. H. Takallou. Knowledge-Based Expert Systems: What's Happening in Transportation Engineering. *ITE Journal*, Oct. 1985.
14. C. Dym. *Expert Systems: New Approaches to Computer-Aided Engineering*. Xerox Corporation, Calif., Palo Alto Research Center, 1984.
15. *PC Magazine*, May 27, 1986, p. 151.
16. Expert System Development Shells. *Computerworld*, July 14, 1986.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

Safety Implications of Truck Configuration

OLIVER CARSTEN

The relative safety of single and double tractor-trailer combinations is examined in the light of recent findings on the performance characteristics of the two classes of vehicle. In particular, the accident data are searched for evidence of a safety deficit for the doubles resulting from the phenomenon of rearward amplification. Although there is no conclusive evidence of an overall difference in fatal and injury accident involvement rates between singles and doubles, this is tempered by the finding of a generally safer operating environment for the doubles. There are strong indications that the doubles have a rollover problem in property-damage accidents. The overall conclusion is that the handling characteristics of large trucks are reflected in their accident experience.

Transportation Research Institute, University of Michigan, Ann Arbor, Mich. 48109–2150.

In the last 15 years a considerable body of literature has appeared on the dynamic performance of truck combinations. One major focus of this literature has been the phenomenon of rearward amplification for combinations with one or more trailers (*1–3*). Rearward amplification is defined as the tendency in multitrailer combinations traveling at highway speeds for motions of the tractor to be exaggerated further in each successive trailer. The phenomenon is particularly severe in emergency maneuvers, when the motion of the tractor may be both abrupt and of large amplitude. But it may also occur in negotiating tight curves, such as those encountered on exit ramps, or even in regular highway driving if travel speed is sufficient. The major effect of the rearward amplification is to cause the second (or third) trailer to have a lower rollover threshold than the first trailer or, in turn, the tractor.

The advances in knowledge of the handling characteristics of combination vehicles have not been fully reflected in studies using accident data. In particular, there has been little success in exploring the issue of whether the theoretical dynamic handling problems for twin trailer trucks are reflected in the national safety experience of combination trucks. A number of recent studies comparing the accident experience of singles and doubles exist, and most, if not all, of these have been evaluated in the recent Double Trailer Truck Monitoring Study by the Transportation Research Board (*4*). However, some of these prior studies had serious deficiencies (*5*), whereas others depended on data that did not have complete coverage or did not clearly distinguish singles from doubles (*6*). With the availability of new data from the University of Michigan Transportation Research Institute (UMTRI), it is possible to examine issues of vehicle configuration using an accurate, national accident database. UMTRI has for several years been conducting a large-truck research program, focused primarily on vehicle issues. This program is using survey research to enhance the data on large-truck involvements in fatal accidents and to collect exposure data on the use of large trucks. The aim of this program is to address the area of vehicle safety, while controlling for environment (e.g., road class) and for use (e.g., carrier type). Although the exposure data collection is not yet complete, several years of accident data have been compiled and will be used here, along with other sources, to assess the safety experience of singles and doubles.

DATA SOURCES AND DATA VALIDATION

The first step in the analysis here was an attempt to corroborate and reconcile the accident databases. The recently developed UMTRI file of Trucks Involved in Fatal Accidents (TIFA) has been used as a yardstick here. The TIFA database provides detailed descriptions of all medium and heavy trucks (greater than 10,000 lb gross vehicle weight rating) that were involved in a fatal accident in the continental United States, excluding Alaska. The file combines the coverage of the Fatal Accident Reporting System (FARS) with the descriptive detail of the Bureau of Motor Carrier Safety (BMCS) accident reports. The detailed vehicle and carrier descriptions are obtained either by matching a FARS case with the corresponding BMCS report or by conducting one or more telephone interviews. Extensive editing and consistency checking is performed on all information obtained by interview. For example, vehicle identification numbers are decoded to confirm that the make and model information and the power unit description conform to published model specifications. Overall, the TIFA files have a very low missing-data rate for the variables that document the truck configuration. In the 1980–1982 file, the vehicle combination type is unknown for only 1 percent of the cases. Given this low rate of missing data combined with the complete coverage of fatal involvements and the extensive checking performed for accuracy, there is every reason to believe that the TIFA data provide an accurate description of the relevant vehicles and accidents.

In performing the analysis on singles or doubles, only those sources or reporting levels that could be reconciled to match TIFA have been regarded as appropriate for use in calculating numbers of accidents. In addition, the various data sources have been examined for internal consistency and reasonableness. Sources or reporting levels that did not meet these requirements have been used for descriptive information where this was unlikely to be affected by bias from underreporting. The sources used in this assessment were the BMCS accident reports and data on large-truck involvements from the National Accident Sampling System (NASS).

Tables 1 and 2 show the comparison between NASS and

TABLE 1 TRACTOR-TRAILER ACCIDENT INVOLVEMENTS BY NUMBER OF TRAILERS: COMPARISON OF NASS AND TIFA

Data Source	Number of Trailers	
	Single	Double
NASS 1981-84[a]		
Property damage only	465,521	5,996
Injury (excl. fatal) . . .	210,486	10,898
Fatal	12,806	673
TIFA[b]	13,103	627

[a]The cases in NASS where the vehicle had a trailer but the number of trailers was unknown were distributed proportionately to the cases with a known number of trailers within each accident severity level.

[b]The numbers in the TIFA file for 1981 through 1983 were inflated to four-year estimates.

TABLE 2 ICC-AUTHORIZED TRACTOR-TRAILER ACCIDENT INVOLVEMENTS BY NUMBER OF TRAILERS: COMPARISON OF BMCS AND TIFA

Data Source	Number of Trailers	
	Single	Double
BMCS 1980–83		
Property damage only	39,673	1,968
Injury (excl. fatal) . . .	41,071	1,786
Fatal	5,106	241
TIFA 1980–83	6,475	296

BMCS on the one hand and TIFA on the other. They also show the number of involvements reported at different accident severities. Because of the small number of cases of large-truck involvement in any single year of NASS, a 4-year file of all the tractor-trailer involvements was created. The counts obtained are shown in Table 1 and the good correspondence on the fatal accidents between NASS and TIFA should be noted. This shows that, in spite of small sample size, the NASS estimates for tractor-trailer involvements at the fatal level, and by inference at the injury level, are reasonable.

If the TIFA numbers for fatal involvements are combined with the NASS estimates of injury and property-damage involvement, one can calculate a ratio of property-damage to injury to fatal involvements for each class of vehicle. This works out to 36:16:1 for the singles and 10:17:1 for the doubles. If these numbers are to be believed, then for every fatal involvement of a single-trailer truck there are 16 injury involvements and 36 property-damage involvements. For each double-trailer truck fatal involvement, there are 17 injury involvements and 10 property-damage involvements. The very large difference between the two classes of vehicle in the ratio of property-damage to fatal involvements does not appear credible. This difference is apparently an artifact of the data and can be attributed to doubles units not being identified in NASS property-damage accidents. There is difficulty in identifying any kind of large truck in an accident for the NASS database because the vehicle has frequently left the area before the investigation begins. It appears reasonable that this problem would be more acute in the less severe accidents and that doubles, which are more likely to be on a long haul, would have a greater tendency than singles to have left the area. Therefore, the NASS estimates of property-damage accident involvement will be excluded as unreliable.

In Table 2 BMCS counts of involvements for tractors with trailers are shown by accident severity and number of trailers. They can be compared with numbers obtained from the TIFA database. Accidents in Alaska and Hawaii were excluded from the BMCS data because they are not covered by TIFA. In addition, a recode was performed on the BMCS combination type field in the UMTRI file. Examination of the BMCS cases incorporated into the TIFA file indicated that all but a handful of the vehicles reported as tractors with full trailers and tractors with other trailers are in fact tractors with semitrailers. Similarly, almost all of the tractors reported as pulling a semitrailer and some trailer other than a full trailer were in fact pulling a semitrailer and a full trailer. Therefore the appropriate recode was performed and Table 2 reflects the result. Because of known underreporting of accidents to BMCS by nonauthorized carriers, the counts have been restricted to the carriers authorized by the Interstate Commerce Commission (ICC). The TIFA numbers have been similarly restricted.

Comparing the BMCS counts of fatal accident involvements with the numbers from TIFA, it is clear that even for fatal accidents there is a certain amount of underreporting. However, it is almost identical for singles and doubles: for the former it is 21.1 percent and for the latter 18.6 percent. Thus any estimates of injury accident involvement rates derived from BMCS are not likely to suffer from differential reporting. There does, however, appear to be very substantial underreporting of property-damage accidents to BMCS. Even given the reporting threshold of $2,000 of damage, the roughly equal numbers of injury and property-damage accidents do not appear credible. To validate the rejection of the BMCS property-damage counts, a comparison was made with Texas and Michigan state accident files. BMCS defines an injury accident as one that requires medical treatment away from the scene, but both state files use the police KABCO coding for severity in which a C-injury is defined as "possible injury." These injuries are unlikely to require treatment and, for the purpose of making the comparison, the C-level injury involvements in the state data were grouped with the property-damage-only involvements. In 1984 Texas reported 4.5 times as many noninjury (property and C-injury) tractor combination involvements as injury involvements (A- and B-injury); Michigan reported 6.6 times as many noninjury involvements as injury involvements. The BMCS property-damage threshold of $2,000 clearly has an impact in the ratio of noninjury to injury accidents, but hardly appears capable of reducing the ratio to 1:1.

If the counts of BMCS-reported property-damage accidents are to be disregarded, this does not mean that all the information on them provided by the file has no value. The descriptive information would only be questionable if one could hypothesize a bias effect from missing data, that is, a situation in which the unreported cases might change one's conclusions about, for example, the proportion of rollover accidents by number of trailers or the amount of property damage from rollover accidents as compared with nonrollover accidents. In many situations the effect of such bias is unlikely to be great and the data

TABLE 3 TRACTOR-TRAILER FATAL ACCIDENT INVOLVEMENT RATES BY DATA SOURCE AND NUMBER OF TRAILERS

Data Source for Involvement Counts	Number of Involvements		Total VMT in Millions[a]		Involvement Rate per 100 million VMT	
	Single	Double	Single	Double	Single	Double
TIFA 1982						
All	3,139	131	45,817	1,968	6.9	6.7
ICC only ...	1,611	66	16,490	854	9.8	7.7
TIFA 1980-82						
All	9,914	448	137,451	5,903	7.2	7.6
ICC only ...	4,808	221	49,470	2,563	9.7	8.6

[a]From 1982 Truck Inventory and Use Survey. The mileages for 1980-82 are three times the 1982 mileages.

from the BMCS property-damage accidents can be used for the description of accidents and their consequences.

THE OVERALL SAFETY EXPERIENCE OF SINGLES AND DOUBLES

With data from TIFA, NASS, and BMCS an overall comparison can be made between the safety experience of tractor and semitrailer combinations and of tractor and twin trailer combinations. Such a comparison will not, given existing use data, be able to take into account the operating environment in which the two classes of vehicles are used, but it will enable the observation of any differences in safety that are of sufficient magnitude to affect the overall picture.

In Table 3 counts of tractor-trailer fatal accident involvements from TIFA are combined with exposure estimates from the 1982 Truck Inventory and Use Survey (TIUS) to provide fatal accident involvement rates. With 1982 TIFA alone, the doubles units appear to have a slightly lower rate of fatal accident involvements, both overall and for the vehicles operated by the ICC-authorized carriers. However, if instead accident data from 3 years are used because of the relatively small number of doubles units involved in fatal accidents in a single year (130 in 1982), the doubles have a slightly higher rate overall, but a somewhat lower rate for the ICC-authorized carriers. A reasonable conclusion would be one of no difference in fatal accident involvement rate between the singles and doubles.

Table 4 provides rates of involvement in accidents that resulted in at least one injury. The two sources of the involvement counts here are the 1981–1984 combined NASS file and the 1982 BMCS file limited to ICC-authorized carriers only. Because the definition of an injury accident varies between the two data sources (BMCS requires reporting of an injury accident involvement only if there is medical treatment away from the scene, whereas NASS uses the observation of the police officer), no comparisons should be made between the BMCS and NASS counts. According to both sources, the doubles have a slightly lower rate, but the difference is small enough and the data quality is uncertain enough to lead to a conclusion of no difference in injury accident involvement rates. Thus the overall assessment is one of no difference in either fatal or injury accident involvement rates between singles and doubles. However, these numbers do not take into account the operating environment in which the vehicles are used. If one class of vehicle was used more often in a safer operating environment, the overall accident involvement rates, which are roughly similar, would conceal a real difference in safety.

Table 5 shows fatal accident involvement rates by operating environment for all combination trucks. The involvement counts are from TIFA, and the exposure figures are the esti-

TABLE 4 TRACTOR-TRAILER INJURY (including Fatal) ACCIDENT INVOLVEMENT RATES BY DATA SOURCE AND NUMBER OF TRAILERS

Data Source for Involvement Counts	Number of Involvements		Total VMT in Millions[a]		Involvement Rate per 100 million VMT	
	Single	Double	Single	Double	Single	Double
NASS 1981-84 (All) ..	225,769	9,094	183,268	7,870	123.2	115.5
BMCS 1982 (ICC only)	11,881	527	16,490	854	72.0	61.7

[a]From 1982 Truck Inventory and Use Survey. The mileages for 1980-84 are five times the 1982 mileages.

TABLE 5 COMBINATION-TRUCK FATAL ACCIDENT INVOLVEMENT RATE BY ROAD TYPE: TIFA, 1980–1982

Road Type	Number of Involvements[a]	Total VMT in Millions[b]	Involvement Rate per 100 million VMT
Urban interstate . . .	917	25,551	3.6
Urban non-interstate	1,979	27,164	7.3
Rural interstate	1,750	60,554	2.9
Rural non-interstate	5,678	66,078	8.6
Unknown	276	–	–
All	10,600	179,347	5.9

[a]From TIFA.

[b]From Federal Highway Administration *Highway Statistics*, 1980 and 1982.

mates calculated by FHWA. Substantial differences in safety are revealed between operating environments, with the rural Interstates having the lowest fatal accident involvement rate. The involvement rate on rural Interstates is one-third that on rural non-Interstates. The rate on all Interstates is less than half that on all non-Interstates. If doubles log a greater share of their mileage on the relatively safe Interstates than did singles, one might conclude that the finding of no difference in overall accident involvement rates was unfavorable to the doubles. In the absence of true exposure data comparing the use of singles and doubles by operating environment, this cannot be tested directly. It is possible, however, to infer differences in use from the accident data. This procedure is by no means perfect, because it ignores the interactions of other factors beyond those being directly observed. But, if it does not permit an estimate of the *size* of differences in exposure, it does at least permit an estimate of the *direction* of differences.

Table 6 shows the proportions of fatal accident involvements by road class for singles and doubles. (All the data presented in Tables 6 through 15 are from two censuses of accident involvements and are therefore not subject to sampling variance. However, in order to establish that the dimensions shown in each of these tables were not independent of each other, the chi-square test was run for each table. For every table the chi-square was significant at the .05 level or less.) Here a classification into divided and undivided, which is not available in the FHWA exposure estimates, is used. This classification is more appealing from a safety viewpoint and is the only one common to both the TIFA fatal data and the BMCS accident data. According to Table 6, 48 percent of doubles fatal involvements occur on divided roads as opposed to 41 percent for singles. Table 7 shows the same comparison using all BMCS-reported involvements by ICC-authorized carriers. Here a remarkable 70 percent of the doubles involvements are on divided roads as compared with 52 percent of the singles involvements.

The distributions of involvements by road class point out the need for more detailed exposure data. But pending better data, it is still possible to test some hypotheses by using current data. One possible explanation for the very large concentration in Table 7 of doubles involvements on divided highways might be that rearward amplification is more of a problem on high-speed roads. However, if the ICC doubles involvements reported to BMCS are broken out by accident severity, the divided roads account for 60.5 percent of the fatal involvements, 72.5 percent of the injury involvements, and 68.8 percent of the property damage involvements. Rearward amplification, which may be a major causal factor in a few fatal accidents and perhaps in some injury accidents, cannot be expected to account for all of the observed distribution of accidents by road class. This distribution appears to be rather a reflection of use.

TABLE 6 TRACTOR-TRAILER FATAL ACCIDENT INVOLVEMENTS BY ROAD CLASS AND NUMBER OF TRAILERS: TIFA, 1980–1982

Road Class	Number of Trailers			
	Single		Double	
	N	%	N	%
Divided	4,057	40.9	215	48.0
Undivided	5,783	58.3	231	51.6
Unknown	74	0.7	2	0.4
Total . . .	9,914	100.0	448	100.0

TABLE 7 ALL ICC-AUTHORIZED TRACTOR-TRAILER ACCIDENT INVOLVEMENTS BY ROAD CLASS AND NUMBER OF TRAILERS: BMCS, 1984

Road Class	Number of Trailers			
	Single		Double	
	N	%	N	%
Divided	13,029	51.6	959	70.0
Undivided	10,383	41.2	364	26.6
Unknown	1,819	7.2	47	3.4
Total	25,231	100.0	1,370	100.0

From Table 5, it has been shown that travel on Interstates and presumably on divided highways is safer than on other kinds of roads. Because there is no evidence that the divided highways are any less safe for doubles than for singles, the differing split of accidents by road class between singles and doubles can only be explained by exposure. The data clearly imply that doubles log a greater share of their travel on divided roads than do singles. Hence one would expect doubles to have a lower overall accident involvement rate than singles. The fact that the rate is roughly equal to that of the singles is cause for concern. In mitigation, it should be noted that, according to Table 3, the doubles units operated by ICC carriers do have lower accident involvement rates than their singles counterparts. Hence, the expectation from the road class information of lower rates for doubles does appear to be met.

ACCIDENT TYPE

Although the analysis of accident involvement rates of single-trailer and double-trailer vehicles cannot be carried any further pending the availability of more detailed exposure data, the accident data alone can be examined for indications of areas in which the safety performance of current doubles is deficient when compared with that of singles. The focus here, as indicated in the introduction, will be on handling-related factors.

Tables 8 and 9 show the proportions of single- and multi-vehicle accident involvements for the two classes of vehicle. Here the hypothesis is that if the current doubles fleet has greater handling problems than the singles fleet, the doubles should be overrepresented in the single-vehicle accidents. This indeed appears to be the case. For both fatal accidents (Table 8) and overall accidents in ICC-authorized vehicles (Table 9), the data show an excess of doubles involvement in single-vehicle accidents.

In the next two tables the distribution of the first harmful event and the most harmful event for fatal involvements is examined. In Table 10 (first harmful event) doubles are underrepresented in collisions with motor vehicles in transport, which follows from their overrepresentation in single-vehicle accidents. There is an excess of collisions with pedestrians and bicyclists, which may hint at some urban-related problems for doubles. As regards handling issues, the doubles are overrepresented in collisions with fixed objects, which might result from loss of control, but there are proportionately fewer first-event rollovers for doubles than for singles. (A first-event rollover is the primary event in the accident, whereas a subsequent-event rollover occurs after some other primary event.) The picture is not very different in Table 11, which gives the distribution of the most harmful event. Once again doubles demonstrate an excess of fatal collisions with pedestrians and bicyclists and an excess of collisions with fixed objects. Now, however, doubles slightly exceed singles in the proportion of overturns. This suggests that doubles have a tendency to roll over once an

TABLE 8 TRACTOR-TRAILER INVOLVEMENTS BY NUMBER OF VEHICLES INVOLVED AND NUMBER OF TRAILERS: TIFA, 1980–1982

Number of Vehicles Involved	Number of Trailers			
	Single		Double	
	N	%	N	%
One vehicle	2,159	21.8	119	26.6
More than one vehicle	7,753	78.2	329	73.4
Unknown	2	0.0	0	0.0
Total	9,914	100.0	448	100.0

TABLE 9 ALL ICC-AUTHORIZED TRACTOR-TRAILER ACCIDENT INVOLVEMENTS BY NUMBER OF VEHICLES INVOLVED AND NUMBER OF TRAILERS: BMCS, 1984

Number of Vehicles Involved	Number of Trailers			
	Single		Double	
	N	%	N	%
One vehicle	12,203	48.4	786	57.4
More than one vehicle	13,028	51.6	584	42.6
Total	25,231	100.0	1,370	100.0

TABLE 10 TRACTOR-TRAILER INVOLVEMENTS BY FIRST HARMFUL EVENT AND NUMBER OF TRAILERS: TIFA, 1980–1982

First Harmful Event	Number of Trailers			
	Single		Double	
	N	%	N	%
Collision with:				
motor veh. in transport	7,245	73.1	301	67.2
pedestrian	682	6.9	44	9.8
pedalcycle	92	0.9	9	2.0
parked motor veh.....	131	1.3	9	2.0
other non-fixed object .	254	2.6	14	3.1
fixed object	845	8.5	43	9.6
Overturn	603	6.1	25	5.6
Other non-collision	62	0.6	3	0.7
Total	9,914	100.0	448	100.0

TABLE 11 TRACTOR-TRAILER INVOLVEMENTS BY MOST HARMFUL EVENT AND NUMBER OF TRAILERS: TIFA, 1980–1982

Most Harmful Event	Number of Trailers			
	Single		Double	
	N	%	N	%
Collision with:				
motor veh. in transport	6,775	68.3	295	65.8
pedestrian	722	7.3	47	10.5
pedalcycle	90	0.9	9	2.0
parked motor veh.....	79	0.8	3	0.7
other non-fixed object .	180	1.8	10	2.2
fixed object	423	4.3	24	5.4
Overturn	964	9.7	46	10.3
Other non-collision	287	2.9	14	3.1
Unknown	394	4.0	0	0.0
Total	9,914	100.0	448	100.0

TABLE 12 TRACTOR-TRAILER INVOLVEMENTS BY ROLLOVER AND NUMBER OF TRAILERS: TIFA, 1980–1982

Rollover	Number of Trailers			
	Single		Double	
	N	%	N	%
None	8,251	83.2	357	79.7
First event	618	6.2	23	5.1
Subsequent event	1,045	10.5	68	15.2
Total	9,914	100.0	448	100.0

accident has begun and that these rollovers are associated with fatal injury.

This conclusion is reinforced by the distribution of rollovers for fatal involvements (Table 12). Doubles have a somewhat lower probability of a first-event rollover but a considerably higher probability of a subsequent-event rollover. In Table 13 another handling-related factor, jackknifing (which, as coded in FARS, includes trailer swings) is examined. Here doubles have an excessive number of first-event jackknifes but a slightly lower probability of a subsequent-event jackknife. Thus, from fatal accidents, at least, there is clear substantiation of handling-related problems for doubles.

In Tables 14 and 15 data are presented to examine whether the indication of handling problems for doubles in the fatal data is borne out by information on all involvements reported by ICC-authorized carriers. Table 14 shows the distribution of noncollision accidents for the involvements reported to BMCS in 1984. Doubles have a smaller proportion of involvements in collision accidents. They are overrepresented in every major type of noncollision accident, particularly overturns. The probability of a rollover for a double is two-and-a-half times greater than the probability for a single. Table 15 makes the same comparison for property-damage-only accidents reported to BMCS by the ICC-authorized carriers. Here less than half the doubles involvements are in collision accidents, compared with almost three-fourths of the singles involvements. For these accidents, doubles have a probability of rollover that is more than four times greater than that for singles.

In the BMCS injury-level (not including fatal) involvements reported by the ICC-authorized carriers, doubles have about a 25 percent higher probability of rollover. Because there is evidence (Table 11) that doubles rollovers are correlated with injury, one might expect doubles accidents to result in somewhat more serious injuries than singles accidents. An examination of the NASS data, shown in Table 16, tends to confirm this. Here the distribution of the maximum on the Abbreviated Injury Scale (AIS) for any injury incurred in the accident is shown. Injuries of unknown severity (AIS-7) have been added to the AIS-2 group. (Of the singles involvements 13.5 percent had MAIS-7; of the doubles involvements, none. Adding the 13.5 percent to the MAIS-1 proportion for the singles would have resulted in concluding, purely on the basis of reallocating the MAIS-7 involvements, that there was a difference in the distribution of MAIS-1 and MAIS-2 involvements between the singles and the doubles and that this difference was unfavorable to doubles. It was believed that it was more conservative here to allocate the MAIS-7 involvements to the MAIS-2 category, because it is improbable that any of the AIS-7 injuries are really AIS-3 or greater.) According to Table 16 doubles are involved in a lower proportion of MAIS-2 accidents but a higher proportion of MAIS-3 accidents. Thus the most severe injury incurred is likely to be more severe in an injury accident involving a double-trailer combination than in an injury accident involving a single-trailer combination. Whether this difference is entirely attributable to handling-related accidents or whether it is a by-product of road class cannot be concluded

TABLE 13 TRACTOR-TRAILER INVOLVEMENTS BY JACKKNIFE AND NUMBER OF TRAILERS: TIFA, 1980–1982

Jackknife	Number of Trailers			
	Single		Double	
	N	%	N	%
None	8,966	90.4	384	85.7
First event	719	7.3	56	12.5
Subsequent event	229	2.3	8	1.8
Total	9,914	100.0	448	100.0

TABLE 14 ALL ICC-AUTHORIZED TRACTOR-TRAILER ACCIDENT INVOLVEMENTS BY NONCOLLISION TYPE AND NUMBER OF TRAILERS: BMCS, 1984

Non-Collision Type	Number of Trailers			
	Single		Double	
	N	%	N	%
Ran off road	1,616	6.4	117	8.5
Jackknife	1,749	6.9	138	10.1
Overturn	1,942	7.7	262	19.1
Separation of units .	130	0.5	16	1.2
Fire	172	0.7	5	0.4
Cargo loss or spillage	132	0.5	2	0.1
Cargo shift	97	0.4	2	0.1
Other non-collision . .	47	0.2	1	0.1
Collision	19,346	76.7	827	60.4
Total	25,231	100.0	1,370	100.0

TABLE 15 ICC-AUTHORIZED TRACTOR-TRAILER PROPERTY-DAMAGE ACCIDENT INVOLVEMENTS BY NONCOLLISION TYPE AND NUMBER OF TRAILERS: BMCS, 1984

Non-Collision Type	Number of Trailers			
	Single		Double	
	N	%	N	%
Ran off road	724	6.1	45	6.7
Jackknife	1,177	9.9	79	11.8
Overturn	813	6.8	191	28.5
Separation of units .	111	0.9	14	2.1
Fire	159	1.3	3	0.4
Cargo loss or spillage	102	0.9	1	0.1
Cargo shift	62	0.5	2	0.3
Other non-collision . .	31	0.3	1	0.1
Collision	8,700	73.2	334	49.9
Total	11,879	100.0	670	100.0

TABLE 16 MAXIMUM AIS (MAIS) FOR INJURY-LEVEL TRACTOR-TRAILER INVOLVEMENTS BY NUMBER OF TRAILERS: NASS, 1981–1984

MAIS	Number of Trailers					
	Single			Double		
	N	Weighted N	%	N	Weighted N	%
MAIS-1	543	105,565	64.7	15	5,248	62.1
MAIS-2	226	45,921	28.2	5	1,579	18.7
MAIS-3	65	9,628	5.9	5	1,566	18.5
MAIS-4	10	1,401	0.9	1	55	0.6
MAIS-5	6	614	0.4	0	0	0.0
Total . .	850	163,129	100.0	26	8,446	100.0

from the NASS data. Unfortunately, there are insufficient cases to examine any accident factors.

Thus, although doubles have approximately the same overall accident involvement rate as singles, there are clear indications in the accident data that in certain areas of performance, conventional double-trailer vehicles do not perform as well as singles. Rollovers, in particular, are more common for these vehicles than for the tractor-semitrailer combinations. There is some evidence that these rollovers are related to injury and they tend to be costly. According to the 1984 BMCS data, the ICC-authorized carriers reported cargo spillage for 31 percent of their rollover involvements, but only for 4 percent of their nonrollover involvements. The same group of carriers reported a mean property damage of $12,846 for doubles involvements in which the primary event was other than a rollover and $15,540 for involvements in which the primary event was a rollover.

CONCLUSIONS

Research on the physical handling of doubles combinations has indicated the potential for safety problems in the normal use of these vehicles. The findings from actual highway experience do not show a higher fatal or injury accident involvement rate for doubles. Nevertheless, this must be tempered by evidence that doubles are used more in safer operating environments. It must also be tempered by indications in the accident data of handling-related problems for doubles, and particularly by a finding of large overinvolvement in rollovers at the property damage level as compared with singles.

The comparison of singles and doubles demonstrates the influence of vehicle characteristics on accident experience. One might presume that drivers are able to compensate for differences in vehicle handling. The fact that the objective performance measures for doubles are reflected in the accident data indicates that drivers are unable to compensate fully. It seems likely, therefore, that improving the handling of double-trailer combinations will provide significant safety benefits.

REFERENCES

1. T. Hazemoto. *Analysis of Lateral Stability for Doubles.* SAE Paper 730688. Society of Automotive Engineers, Warrendale, Pa., June 1973.
2. R. D. Ervin et al. *Ad Hoc Study of Certain Safety-Related Aspects of Double-Bottom Tankers.* Report UM-HSRI-78-18-1. Highway Safety Research Institute, University of Michigan, Ann Arbor, 1978.
3. C. Mallikarjunarao and P. Fancher. *Analysis of the Directional Response Characteristics of Double Tankers.* SAE Paper 781064. Society of Automotive Engineers, Warrendale, Pa., Dec. 1978.
4. *Special Report 211: Twin Trailer Trucks: Effects on Highways and Highway Safety.* TRB, National Research Council, Washington, D.C., 1986.
5. G. R. Vallette, H. McGee, J. H. Sanders, and D. J. Enger. *The Effect of Truck Size and Weight on Accident Experience and Traffic Operations,* Vol. 3: *Accident Experience of Large Trucks.* Report FHWA/RD-80/137. FHWA, U.S. Department of Transportation, 1981.
6. T. Chirachavala and J. O'Day. *A Comparison of Accident Characteristics and Rates for Combination Vehicles with One or Two Trailers.* Report UM-HSRI-81-41. Highway Safety Research Institute, University of Michigan, Ann Arbor, 1981.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

Relationships Between Vertical and Horizontal Roadway Alignments and the Incidence of Fatal Rollover Crashes in New Mexico and Georgia

PAUL ZADOR, HOWARD STEIN, JEROME HALL, AND PAUL WRIGHT

Survey data on curvature and grade collected at the sites of fatal single-vehicle rollover crashes and at random comparison sites in New Mexico and Georgia were analyzed to determine the relationship of horizontal and vertical alignment to such crashes. The results showed that road sections with extreme horizontal and vertical alignments were as much as 50 times more common at crash sites than at comparison sites. Although sharp left curves and steep downgrades were over-represented in both states, the relative importance of down-grades was greater in New Mexico than in Georgia. Because the relative importance of the two alignments can be expected to vary in other states as well, no overall set of priorities for hazard identification was developed. It is recommended that each state develop its own priorities for hazard identification based on comparisons between the bivariate curve-grade distributions of fatal single-vehicle crash sites and those of a representative roadway sample. A method for setting such priorities that can be used by individual states is presented.

In a recent review of the condition of highway systems in the United States (*1*) it was concluded that about two-thirds of all rural roadways were deficient in terms of pavement condition, geometric design, cross section, or operational features. Some geometric design deficiencies were present in about one-third of all rural roadways. Almost 90 percent of the deficient road sections were on rural collectors. In view of the substantial "substandard safety and geometric characteristics" found on segments of the rural collector system, it is not surprising that although rural collectors accounted for only less than 10 percent of all vehicle miles traveled in 1981, their share in fatal crashes was more than 15 percent (*2*). This situation is unlikely to improve substantially in the foreseeable future; it has been estimated by the U.S. Department of Transportation that less than half of the annual expenditure of $3.8 billion needed to eliminate all deficiencies will be available over the next 20 years (*1*). There is no evidence that safety-related projects will be funded above average levels; thus it is of paramount importance that available funds be allocated in the most effective manner. A critical step in the cost-effective allocation of the available funds for the improvement of the geometrical design of the highway system is the identification of hazardous sites.

Most methods for identifying hazardous sites rely on past crash experience or on inventories of roadway and roadside features. The methods based on crash rates or crash counts assign a high priority to upgrading sites that had more crashes than was typical of other roadways with similar characteristics (*3–5*). Such sites with high crash rates are often referred to as "black spots." However, most short roadway sections rarely have more than one or at most two crashes in a given period, regardless of how hazardous they may be. Moreover, some sections that are not particularly hazardous could also have one or more crashes due to driver error, weather conditions, or a combination of very unusual circumstances. Because of such random fluctuations, crash rates for short sections based on short time periods are not effective measures of hazardous operational features such as adverse road geometry. Aggregating crashes over longer sections and longer time periods, or both, would reduce the fluctuations but only at a cost. Because roadway geometry typically varies substantially along most roads, aggregation over long stretches of roadways would dilute the effect of severely adverse geometrical hazards and therefore make it impossible to set optimally cost-effective priorities for their reduction. Waiting for crash data to accumulate before a hazardous site is corrected is rarely cost-effective or even tolerable. In any case, as Hauer and Persaud have shown (*6*) in their study of the regression-to-the-mean phenomenon, the severe selection bias that arises in studies of black spots persists even if the crash histories extend over many years.

Inventory-based models typically involve developing indices to rank the severity of specific roadside hazards and their potential for involvement in a collision (*7–11*). In these models it is recognized that the probability that any specific roadway section will be the site of a crash is low. The crash event is thought of as a chain of minor events with low individual probabilities. These probabilities are estimated and multiplied, and their products are summed to obtain the overall probability of the crash. For example, in the Glennon model (*7, 8*) an injury-producing roadside fixed-object crash is defined as the sequence of four conditional events:

1. The vehicle is within the incremental part of the roadway where collisions with roadside objects are possible,
2. There is an encroachment onto the roadside,
3. The lateral displacement of the vehicle is sufficiently large to permit a collision with the object, and
4. The collision is of sufficient magnitude to produce an injury.

P. Zador and H. Stein, Insurance Institute for Highway Safety, Watergate 600, Washington, D.C. 20037. J. Hall, Bureau of Engineering Research, University of New Mexico, Albuquerque, N. Mex. 87106. P. Wright, School of Civil Engineering, Georgia Institute of Technology, Atlanta, Ga. 30332.

This type of model is difficult to validate, because the incidence and severity of injuries depend on many factors such as occupant age, restraint use, and vehicle design and because implementing such models requires collecting vast amounts of data. These include measures for estimating the potential for roadside encroachments, complete inventory (e.g., type and location) of all roadside features and fixed objects, the injury-reducing potential of improvements, and the costs associated with each candidate roadside improvement scenario.

In 1974, Wright and Robertson (*11*) compared the roadway alignments found at the sites of fatal fixed-object crashes with the alignments at comparison sites chosen 1.6 km (1 mi) away from these crash sites. In subsequent studies with similar designs, the alignments at fatal rollover crash sites were assessed by Wright and Zador (*12*) and Hall and Zador (*13*). All three studies reported that the two most important features that distinguish crash sites from comparison sites are horizontal and vertical roadway alignment. Specifically, sharp curves (5 to 6 degrees or greater) and steep grades (3 percent or greater) are significantly associated with crash sites.

In other research, the influence of geometric design on crash rates has been analyzed in terms of single factors without detailed analysis into the combined effects of grade and curvature. Most crashes occur on straight and level sections of the roadway; however, curved sections with steep grades generally have higher crash rates. Several studies have computed the crash rates of roadway sections by curvature and have reported that higher crash rates are associated with sharper curves, more frequent curves, and isolated sharp curves (*14*). It was also found that crashes on curves tend to be more severe than crashes on tangent sections (*3*). Studies of vertical alignment have found that roads with steep grades, particularly in combination with sharp curves, have higher crash rates (*14*). However, crash rates of tangent sections were reported to be not significantly influenced by grade. Despite the vast body of evidence that poor geometric features are associated with crash sites, only a few of the hazard location models explicitly consider them (*9, 11*). A survey of state highway departments found that although most employ formal guidelines for selection of sites for implementation of low-cost countermeasures, few consider roadway geometry as a factor (*15*). The most common factors were crash history and traffic volume.

In this paper detailed comparisons between fatal rollover crash sites and comparison sites are presented in terms of both vertical and horizontal alignments. A procedure for incorporating results of this type in an effective strategy for reducing the frequency of fatal single-vehicle crashes is also outlined.

METHODS

Data Collection

The data for these analyses are from four independent sources: studies of fatal single-vehicle rollover crashes in Georgia (*12*) and New Mexico (*13*) and surveys of randomly selected sites in these states. Engineering surveys were conducted, usually by three-person teams, at the locations of fatal rollover crashes and at the comparison locations. The surveys were confined to a 0.3-km (0.2-mi) section at each of the locations. The measurements at the crash location were referenced to the point along the roadway edge at which the rollover of the vehicle commenced, which is termed the crash reference point. As shown in Figure 1, a point 1.6 km before the crash reference point was designated as the comparison site. In the location of comparison sites, turn choices at T- or Y-intersections were made at random (by flipping a coin).

Measurements of curvature were made beginning 15 m (50 ft) from the crash and comparison sites and at 30-m (100-ft) intervals for 137 m (450 ft) both upstream (before the site in the direction from which the vehicle approached) and downstream (beyond the site in the direction in which the vehicle was traveling) from these sites. The gradient was measured every 30 m for 152 m (500 ft) both upstream and downstream from the sites. Thus, 10 curvature and 11 gradient measurements were obtained for each crash site and its comparison site.

A 30-m cloth tape was used for measuring distances. Horizontal curvature was measured by the middle ordinate method. The curve measurements were usually taken on the edge of the roadway. The middle ordinates were converted to degrees of curvature of the centerline of the roadway. Gradients were measured at the center of the lane used by the driver approaching the crash location. Measurements were made with a specially designed instrument consisting of a 1.2-m (4-ft) carpenter's level with an adjustable calibrated leg. On Interstate highways, curvature and gradient data were taken from plan and profile sheets.

Rollover crash and comparison data provide insight about the role of various geometrical and roadway features relevant to crashes, but they do not necessarily provide a representative sample of roadways in either state. Typically, the comparison was on the same road as the crash site and thus many features were similar. For example, roadway characteristics such as pavement width and shoulder width and delineation were typically the same in both the crash and comparison sites. Features such as roadside object density and characteristics were also similar. More important, the influence of the terrain on horizontal and vertical alignment was similar. Although previous analyses had found that the crash sites had more severe alignments than the comparison sites (*12, 13*), these differences may have been underestimated because of the proximity of the comparison sites to the crash sites.

To address these issues, random sample surveys of the rural road systems were performed at 300 sites in each state. A two-way classification of rural roadways by average daily traffic and roadway function was obtained for Georgia and New Mexico. One-half of the survey sites were assigned in proportion to roadway mileage alone; the other half were selected in proportion to estimated miles traveled. After the number of sections was computed, specific sections were identified by randomly selecting milepost locations from computerized roadway files maintained by the states. The geometric data collected at the random sites also included the middle ordinate (curvature) and the gradient 15 m (50 ft) before and after the random site. The distribution of all the sites included in this analysis (i.e., crash, comparison, and random survey) by roadway functional class and average daily traffic is shown in Tables 1 and 2, which also include the distribution of rural system miles in each state.

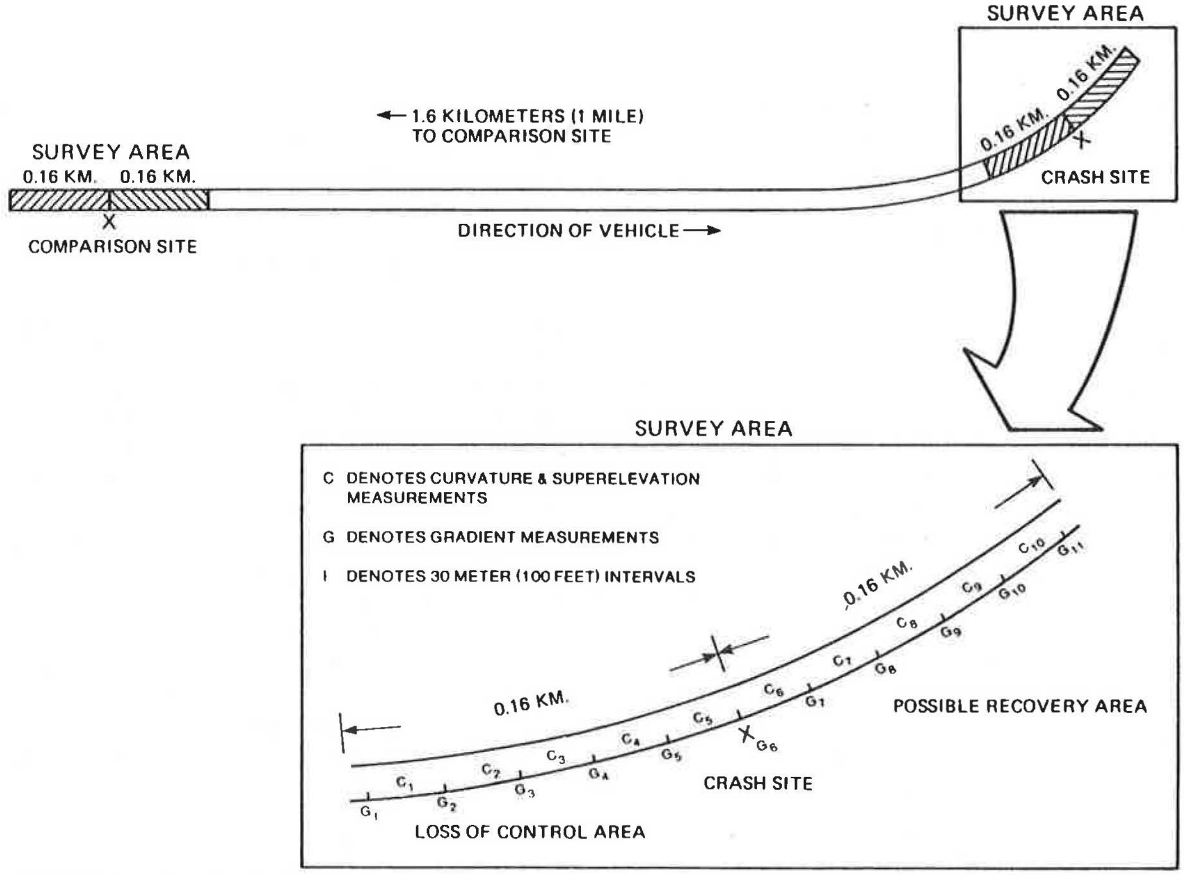


FIGURE 1 Hypothetical fatal crash and comparison sites.

Analysis

As Figure 1 shows, curvature and grade were measured at alternate intervals 15 m apart. For the purpose of the present analyses, the curvature for left-turning roads and the grade for downhill roads were assigned minus signs. The 10 curvature measurements were taken in consecutive pairs and the paired curvature measurements at the beginning and end of the nine 50-m-long sections were averaged. The corresponding (weighted) average grade was calculated as one-fourth times the grade of the preceding section plus one-half times the grade for the section and one-fourth times the grade for the following section. These averages were used to represent the sections' curvature and grade.

Within each state, the sections surveyed at crash sites were grouped in terms of their position with reference to the actual crash reference point, marked *X* on Figure 1, where the vehicle left the road. Three of the sections immediately upstream from position *X* on Figure 1 were classified as the crash sections, and the last four, which were among those occupying the potential recovery area in Figure 1, were classified as the downstream sections. Sections surveyed 1.6 km upstream from the crash sites are termed comparison sections and sections chosen for the random survey are termed random sections. Thus, four section types were defined in each of the two states.

The analysis consisted of three main steps: determination of curvature and grade percentile distributions, summation of weighted sections, and comparison of the joint distributions of crash and comparison sections. As the first step in the analysis, selected percentiles of the curvature and grade distributions were determined separately for each of the section types in both states. (There is no prior reason for the frequency of sections with right turns or upward slopes to exceed the frequency of sections with left turns or downward slopes at random survey sites. The symmetry of these alignment distributions was achieved by including each random site twice in the analyses, once with the sign as measured for the alignments and once with the opposite sign.) For the random sections, the percentiles were determined by using two different methods for assigning section weights. With one method the sections were weighted in proportion to the total road length they represented. With the other method the section weights were proportional to the aggregate miles traveled on the part of the road system that the section represented.

As the second step, the weights of all sections with curvatures and grades subject to selected constraints were summed by section type. However, regardless of the section type, the constraints were defined in terms of the grade and curvature distributions of the segments from crash sites so that the resulting sums could be compared among the section types. Because extreme curvature and grade values are of primary interest in setting priority rules for hazard location, combined weights were computed for sections with both grade and curvature below or above selected extreme percentiles.

Table 3 shows the method of presentation for the cumulative distributions of road sections given in Tables 4–8. Entries are arranged in four quadrants (shaded) corresponding to the four combinations of the lower and upper tails of the grade and curvature distributions; the unshaded areas represent less extreme combinations of curvature and grade. The percentages shown in the upper-left quadrant correspond to the lower tails of both curvature and grade distribution; those in the upper right correspond to the upper tail for curvature and lower tail for grade; the lower-left quadrant corresponds to the lower tail for curvature and upper tail for grade; and the lower-right quadrant corresponds to the upper tail for both curvature and grade.

Each quadrant representing the extreme combinations of grade and curvature in Table 3 has 25 cells. There are 11 cells on the border of each quadrant representing the corresponding less extreme combinations of grade and curvature. Each cell contains the percentage of the grade and curvature distribution that would fall within the range of values specified for that cell.

The variability of cell estimates depends on sample size. For counted data that can be expected to follow the Poisson distribution, the ratio for the estimates of the standard deviation and the mean is approximately equal to the reciprocal of the square root of the sample size (n), so that for $n \cong 10$ this ratio is about 1/3. For n's much below 10, cell estimates can be quite variable; however, the proposed standard table format with 144 cells need not be changed even for small data sets. This is because cells closer to the center of the table include all the data from the cells farther away from its center. For example, the data in the 100 cells not on the boundary of the table represent the data set as collapsed into its 10-by-10 subtable.

Examples of cell types are given in the following for grade and curvature values that are below the median; examples for the other sections are provided in the notes to Table 3. The 25 cells of the upper-left shaded section represent the extreme combinations of values for both curvature and grade. For example, Cell A is the (weighted) percentage of sections with curvature at or below the fifth percentile and grade at or below the first percentile. Cell D at the border of the lower-right corner is the percentage of sections with both curvature and grade between the 10th percentile and the median. The other five border cells in the same row as Cell D (e.g., Cell B) represent the percentage of sections with grade between the 10th percentile and the median and with curvature corresponding to the percentage given in the column heading. The five other border cells in the same column as Cell D (e.g., Cell C) represent the percentage of sections with curvature between the 10th percentile and the median and with grade corresponding to the row heading for that cell. The bottom row is the marginal distribution of curvatures and the right-hand column is the marginal distribution for grades.

As the third step in the analysis, the joint distribution of

TABLE 1 DISTRIBUTION OF NEW MEXICO RURAL ROAD SYSTEM AND SURVEY STUDY SITES

Rural Roadway Classification	Average Daily Traffic 0-999	1,000-1,999	2,000-3,999	4,000-7,999	8,000-16,000	Total
Interstate						
State Miles	0	3	299	365	220	887
Crash Sites	0	0	13	13	14	40
Comparison Sites	0	0	13	13	14	40
Random Sample Sites	0	0	14	28	30	72
Principal Arterials						
State Miles	519	717	684	154	76	2,150
Crash Sites	7	14	12	2	3	38
Comparison Sites	7	14	12	2	3	38
Random Sample Sites	12	23	32	12	11	90
Minor Arterial						
State Miles	1,258	382	90	31	3	1,764
Crash Sites	7	6	1	2	1	17
Comparison Sites	7	6	1	2	1	17
Random Sample Sites	28	12	4	3	2	49
Collector						
State Miles	5,263	348	166	54	1	5,832
Crash Sites	18	6	2	1	0	27
Comparison Sites	18	6	2	1	0	27
Random Sample Sites	67	11	7	4	0	89

TABLE 2 DISTRIBUTION OF GEORGIA RURAL ROAD SYSTEM AND SURVEY STUDY SITES

	Average Daily Traffic						
Interstate	0- 9,999	10,000- 19,999	20,000- 29,999	30,000- 39,999	40,000- 60,000		Total
State Miles	226	292	205	142	22		887
Crash Sites	1	7	7	3	7		25
Comparison Sites	1	7	7	3	7		25
Random Sample Sites	4	13	15	14	3		49
Principal Arterials	0- 4,999	5,000- 9,999	10,000- 14,999	15,000- 19,999	20,000- 29,999		Total
State Miles	2,176	416	78	13	9		2,692
Crash Sites	31	4	0	0	2		37
Comparison Sites	31	4	0	0	2		37
Random Sample Sites	28	11	3	1	1		44
Minor Arterials	0- 999	1,000- 1,999	2,000- 2,999	3,000- 4,999	5,000- 9,999	10,000- 39,999	Total
State Miles	1,199	2,355	1,224	893	473	115	6,259
Crash Sites	11	32	5	11	5	2	66
Comparison Sites	11	32	5	11	5	2	66
Random Sample Sites	8	24	15	15	12	6	80
Major Collectors	0- 999	1,000- 1,999	2,000- 2,999	3,000- 4,999	5,000- 9,999	10,000- 19,999	Total
State Miles	10,932	2,019	602	370	282	59	14,264
Crash Sites	22	4	5	0	1	0	32
Comparison Sites	22	4	5	0	1	0	32
Random Sample Sites	82	21	8	6	7	3	127

TABLE 3 BIVARIATE CUMULATIVE DISTRIBUTION OF SECTIONS BY CURVATURE AND GRADE PERCENTAGES

		Curvature Percentages												
		Lower Tail					Median			Upper Tail				
Grade Percentages	0.0	0.5	1.0	2.5	5.0	10.0	50.0	10.0	5.0	2.5	1.0	0.5	0.0	All
	0.5													
	1.0				A									
Lower Tail	2.5													
	5.0						C							
	10.0													
	Median (50.0)		B				D							
	10.0							F						
	5.0	E												
	2.5													
Upper Tail	1.0													
	0.5													
	0.0						G							
	All							H						

In this table, roadway sections are accumulated from low to high values below median (lower tail) and from high to low above median (upper tail). Entries are percentages of all sections with both curvature and grade bracketed by the corresponding lower tail (e.g., P0.5) or upper tail (e.g., UP0.5) percentiles, for example:

A: (Curvature $\leq$ P5.0, Grade $\leq$ P1.0),
B: (Curvature $\leq$ P1.0, P10.0 < Grade $\leq$ Median),
C: (P10.0 < Curvature $\leq$ Median, Grade $\leq$ P5.0),
D: (P10.0 < Curvature $\leq$ Median, P10.0 < Grade $\leq$ Median),
E: (Curvature $\leq$ P0.5, UP10.0 < Grade),
F: (Median < Curvature $\leq$ UP10.0, Median < Curvature $\leq$ UP10.0),
G: (P10.0 < Curvature $\leq$ Median, UP0.5 < Grade),
H: (UP10.0 < Curvature).

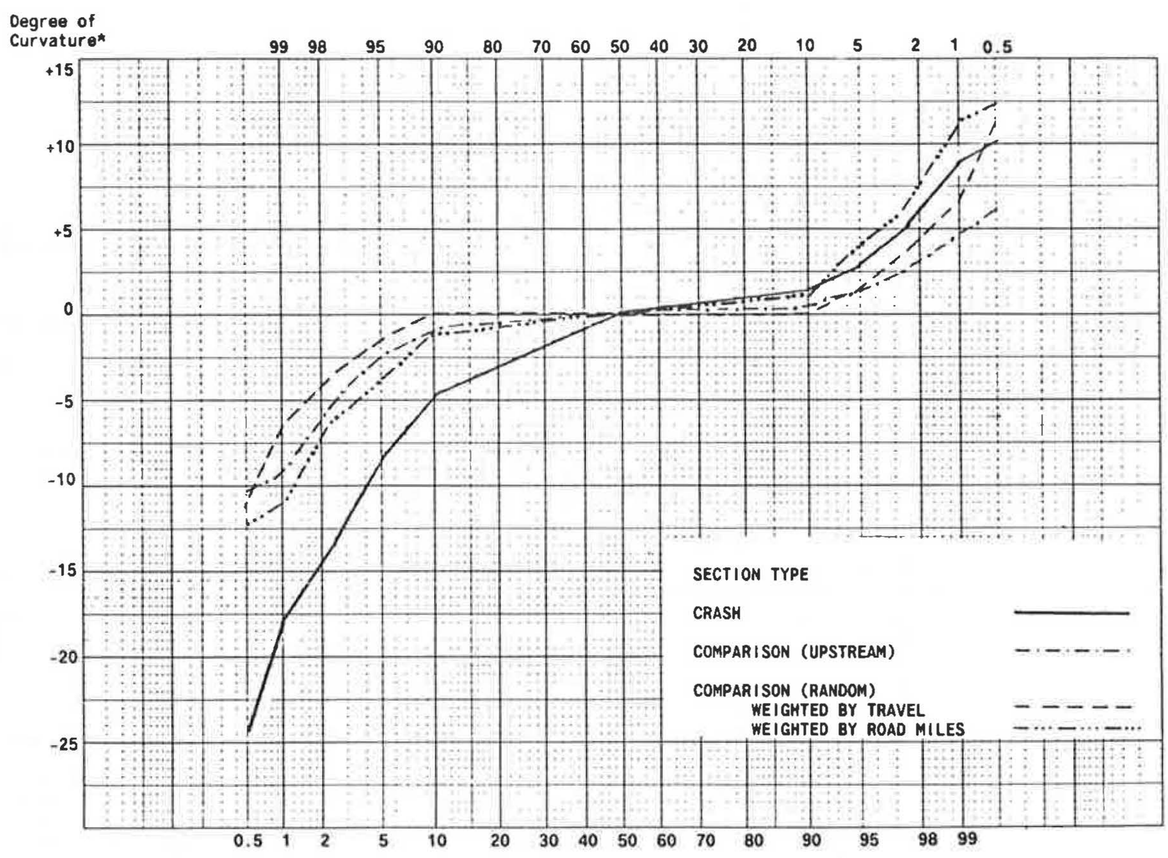
crash sections by curvature and grade was compared with the joint distribution of random comparison sections. (Comparisons with downstream sections and comparison sections were also made but are not discussed here.) This was done by taking the base 2 logarithms of the ratios of summed weights in corresponding cells. (Use of base 2 logarithms allows quick calculation of ratios to within a factor of 2 for order-of-magnitude comparisons.) Large (positive) values in the resulting ratio table (see Tables 7 and 8) are indicative of more crashes than would have been expected on the basis of proportionality to the weights in the "denominator" table that correspond to the random sections. Large negative values indicate fewer than the expected number of crashes. Zero indicates precisely the expected number of crashes. In the log ratio tables the cells that had zero weights both for the numerator and the denominator are marked by a period. The cells in which there were no crash sites are marked with a minus sign, and cells in which there were no random survey sites are marked with a plus sign. The minus sign in the tables is a reminder of an extreme deficiency in crash sites and the plus sign is a reminder of an extreme excess of crash sites compared with comparison sites.

Thus, entries in the ratio tables based on travel volume make it possible to compare crash rate estimates for travel over roads with differing geometries. For example, if α and φ, say, are two entries in the same ratio table, then 2^{b-a} is an estimate for the ratio of the crash rate per volume of travel corresponding to the cell containing φ divided by the crash rate corresponding to the cell containing α. Similar calculations using the ratio table based on road miles allow comparisons of crash rates per road mile. It should be noted that straight and flat sections tend to be underinvolved in crashes. Here both underinvolvement and overinvolvement refer to the average, and therefore the overinvolvement of sections with adverse geometry compared with flat and straight sections would be even higher than the numerical values in the table indicate.

RESULTS

The curvature and grade distributions are plotted in Figures 2–5 on normal probability paper. The estimated percentiles are shown on the vertical axes and the corresponding percentages are plotted on the horizontal axes. In these figures, the horizontal axes are scaled so that normally distributed data would give rise to straight lines. The figures point to marked departures from normal distributions, especially for curvature. In interpreting these figures, it should be kept in mind that the middle positions of the distributions between the 10th and 90th percentile are represented by the 50th percentile only. Because this investigation was concerned with the identification of geometrical hazards, the details of the distributions for the normal ranges of curvature and grade were not explored.

Figure 2 presents the curvature distributions for crash sections, comparison sections, and random sections weighted by



* Positive sign denotes right curves, negative sign denotes left curves

FIGURE 2 Probability distribution of curvature values in New Mexico by section type.

both road miles and travel volume in New Mexico. The comparable distributions for Georgia are given in Figure 3. The most remarkable feature of both figures is the evidence for very long left tails corresponding to left curves of the crash site curvature distributions. For example, for both states none of the comparison sites included left curves of 15 degrees or sharper but about 2 percent of the crash sites did. The differences between the right tails of the distributions corresponding to right curves are less pronounced, although this effect is still clear for the Georgia data.

Figures 4 and 5 present the grade distributions for New Mexico and Georgia, respectively. As these figures show, sharp downgrades were considerably more common at crash sites than at any of the other site types in both states, except in Georgia where the upstream sites and the crash sites had nearly identical grade distributions.

The joint curvature and grade distributions for crash sites are given in Table 4 for both states. Cell percentages based on at least 10 sections are marked with an asterisk and have expected standard errors less than or equal to about 30 percent of the estimate. The joint curvatures and grade distributions for the New Mexico random survey sites are given in Table 5 using both travel volume and road miles as the weights. These distributions are presented for Georgia in Table 6. Because the cumulative distributions presented in Tables 4–6 are accumulated from the most extreme to the least extreme cases, they can be used conveniently for setting priorities for roadside hazards.

Tables 7 and 8 present comparisons between the crash and the random survey data. The logarithms of the ratios of the corresponding percentages in Table 4a divided by those in Tables 5a and 5b are given for New Mexico in Tables 7a and 7b, respectively; the corresponding log ratios for Georgia are given in Tables 8a and 8b. As in Table 4, cells based on 10 or more sections are marked with an asterisk.

HOW TO SET IMPROVEMENT PRIORITIES

Tables 4–8 and Figures 2–5 can be used to assess the importance of improving road sections that have particular combinations of geometric hazards by using the following four-step procedure. (These tables, however, are not intended for direct use in states other than Georgia and New Mexico.)

1. The rate of overinvolvement of sections with a selected combination of curvature and grade is read from Tables 7 and 8.
2. The estimated percentage of travel and road miles corresponding to this level of hazard is determined by reference to Tables 5 and 6.
3. The curvature and grade percentiles are determined from Figures 2 through 5.
4. Table 4 is used to estimate the percentage of fatal rollover crashes that would be reduced by correcting the designated geometric hazards.

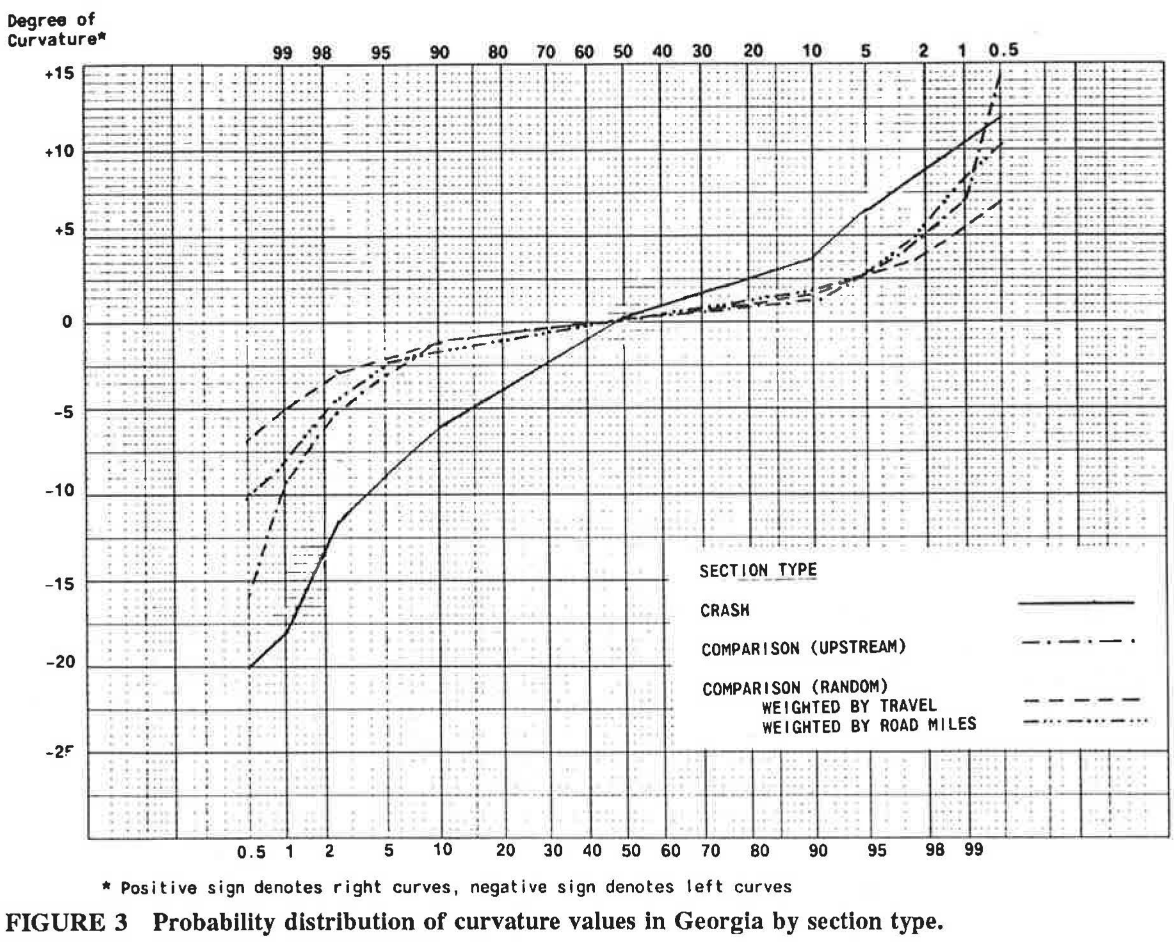


FIGURE 3 Probability distribution of curvature values in Georgia by section type.

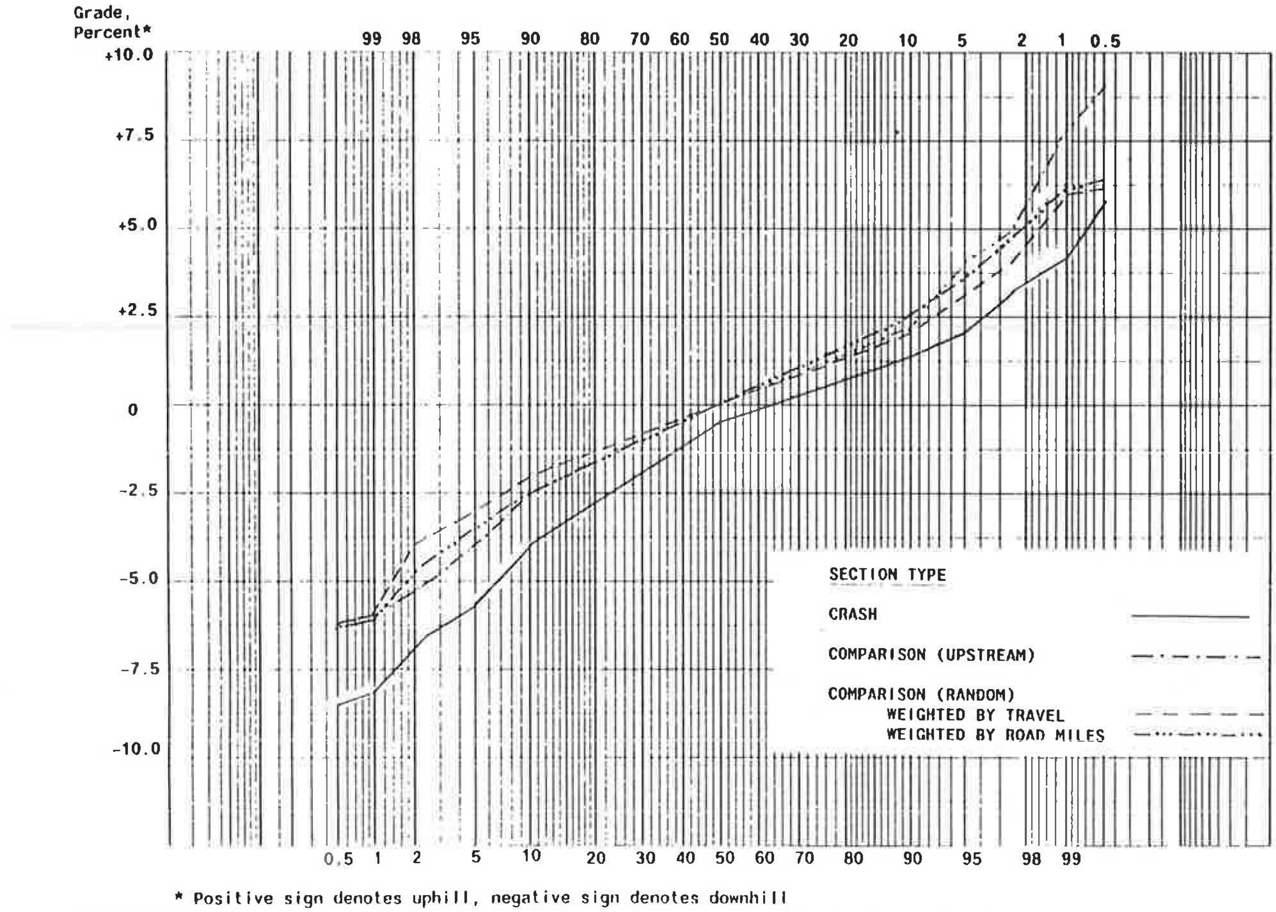


FIGURE 4 Probability distribution of gradient values in New Mexico by section type.

This four-step procedure was applied to New Mexico data for sections with curvature and grade at or below the lower tail 10th-percentile cutoff as follows:

1. The value of the logarithm (base 2) in Table 7 is 3.9; therefore, the overinvolvement is $2^{3.9} = 14.9$. This means that such sections have fatal rollover crashes about 15 times as frequently per volume of travel as do the average road sections. The corresponding overinvolvement per mile of roadway is by the factor of $4.6 = 2^{2.2}$.
2. Table 5 shows that 0.24 percent of the travel volume and 0.76 percent of the roadway miles are subject to this level of extreme geometry.
3. Figures 2 and 4 show that the 10th percentiles of curvature and grade are about –5 degrees and –4 percent.
4. Table 4 shows that overall about 3.5 percent of all fatal rollover crashes in New Mexico occurred at crash sites of similar extreme curvature and grade.

Applying the foregoing procedure for Georgia did not produce similarly dramatic results. The overinvolvement rates could not be explicitly estimated because there were in fact no such extreme sections found among the comparison sites in a random sample of 300 sections. However, about 0.4 percent of the crash sections did have curvatures sharper than the 10th percentile for the curvature distribution (6.4 degrees left) and grades steeper than the 10th percentile for the grade distribution (3.3 percent downgrade). Thus, eliminating this small number of geometrical hazards could be expected to reduce fatal rollover crashes by about one-half of 1 percent.

Roadway sections in Georgia with extreme left curves with slight downgrades are of more concern.

1. Sections with curvature below the 10th percentile and grade over the 10th percentile but below the median were overrepresented by a factor of $18.4 = 2^{4.2}$ in terms of travel volume (Table 8a) and by a factor of $55.7 = 2^{5.8}$ in terms of road miles (Table 8b).
2. Only about 0.25 percent of all travel on only about 0.08 percent of all roads occurred at these extremely hazardous sites (Table 6).
3. The 10th percentile of the curvature distribution was –6.4 in Georgia (Figure 3). The 10th and 50th percentiles of the grade distributions were –3.3 and –0.5 percent (Figure 5).
4. These sections accounted for about 4.6 percent of all fatal rollover crashes (Table 4).

As these comparisons between sets of data on fatal rollover crashes in New Mexico and in Georgia show, severe curvature and severe grade may have substantially different relative effects on these events in different states. Part of the New Mexico road system is at very high elevations and severe

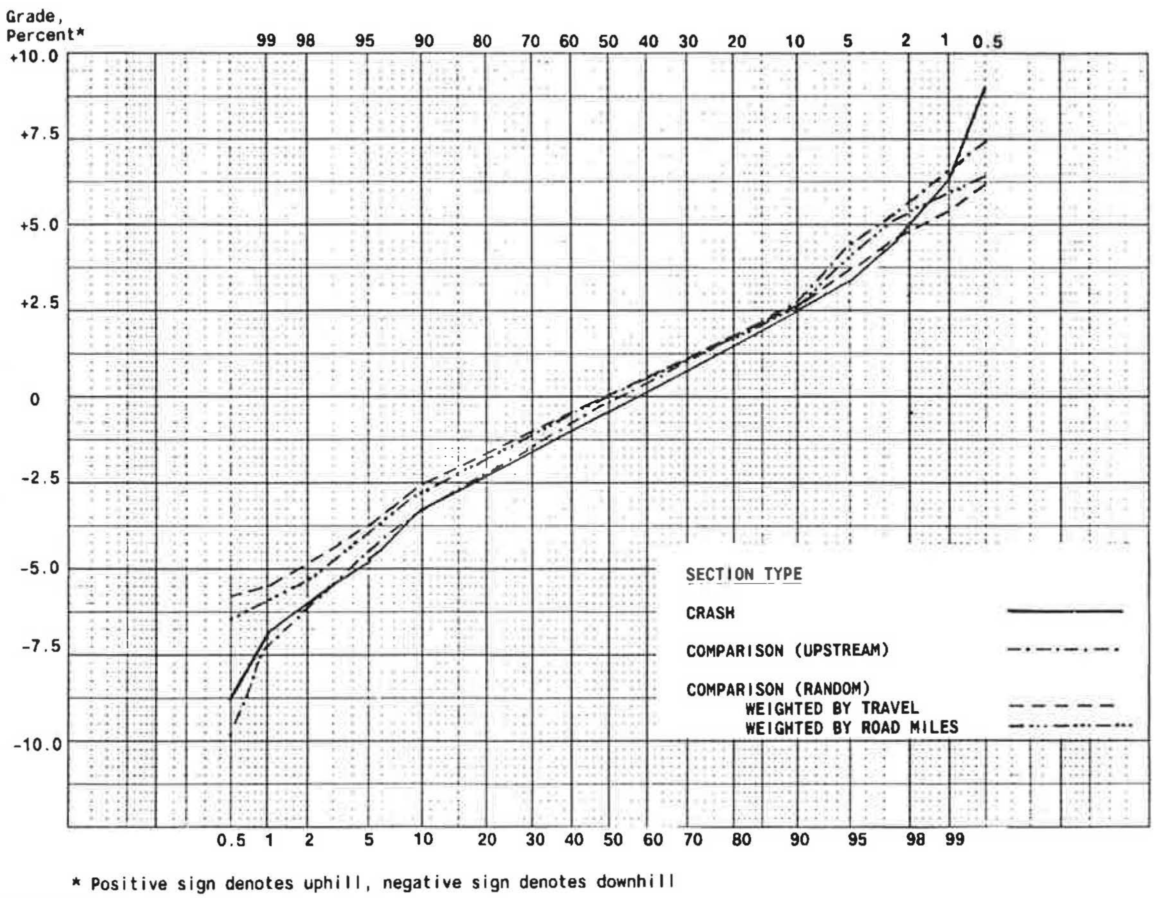


FIGURE 5 Probability distribution of gradient values in Georgia by section type.

grades appear to be a significant factor in many of the fatal rollover crashes there. Georgia is generally of lower elevation and severe grades tend to be less important than severe curves. It is probable that each state or geographical region in the United States has its own unique distribution of curvature and grade problems related to fatal rollover or, more generally, fatal single-vehicle crashes. The procedure for setting priorities outlined below is based on the assumption that data bases similar to those assembled in Georgia and New Mexico can be developed for analyses. In the absence of such a data base, some weighted combination of the data from New Mexico and Georgia could be selected to describe the situation in other regions.

This procedure was designed to be both practical and relatively cost-effective. To be practical a procedure must generate candidate sites for improvement in sufficient numbers to allow allocation of available funds. However, it is not necessary to assign priorities to all parts of the roadway all at once. Systematic surveys of all geometric features throughout the state may result in a wasteful allocation of resources because such surveys can be costly even when they identify the right kind of candidate sites.

For the procedure to be cost effective, only candidate sites with very high rates of overinvolvement should be included in the list of proposed improvements and only limited funds and effort should be spent on sites with less than extreme rates of overinvolvement. However, in most states the variation in crash involvement rates due to curvature and gradient is not known. Therefore, the definition of what constitutes overinvolvement in a given state and the selection of sites for proposed improvement need to be carried out at the same time, at least at the outset. Thus the basic steps toward a cost-effective allocation of roadway improvement, described below, will need to be performed repeatedly. Although the preparation of an operational plan is beyond the scope of this paper, a description of the basic steps needed for a cost-effective allocation of roadway improvement funds is provided.

In this paper, overinvolvement rates were compared in terms of miles traveled and in terms of road miles. Although differences between the two measures may exist, they tend to be more of degree than of kind. In any case, the final choice of improvement projects cannot be made without reference to their estimated reduction in risk. Consideration of these factors was, however, outside the scope of this study.

A cost-effective allocation of roadway improvement funds should involve the following steps.

- Collect a geometric inventory of short roadway sections that includes potential candidates for improvement. Only sections with very adverse geometry (e.g., curvature and grade above some locally chosen thresholds) need to be included in the inventory.

TABLE 4 BIVARIATE CUMULATIVE DISTRIBUTION OF CRASH SECTIONS BY CURVATURE AND GRADE PERCENTAGES: NEW MEXICO VERSUS GEORGIA

a. New Mexico

Curvature Percentages

Grade Percentages		Lower Tail					Median		Upper Tail					
		0.0–0.5	0.5–1.0	1.0–2.5	2.5–5.0	5.0–10.0	10.0–50.0	50.0–10.0	10.0–5.0	5.0–2.5	2.5–1.0	1.0–0.5	0.5–0.0	All
Lower Tail	0.0–0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.27	0.27	0.27	0.00	0.27
	0.5–1.0	0.00	0.27	0.27	0.27	0.27	0.00	0.00	0.55	0.55	0.55	0.27	0.00	0.82
	1.0–2.5	0.27	0.82	0.82	0.82	0.82	0.55	0.00	1.09	0.82	0.55	0.27	0.00	2.46
	2.5–5.0	0.27	0.82	0.82	0.82	1.37	1.91	0.27	1.37	1.09	0.82	0.27	0.00	4.92*
	5.0–10.0	0.27	0.82	0.82	1.37	3.55*	3.55*	0.55	2.19	1.64	0.82	0.27	0.00	9.84*
Median (50.0)	10.0–50.0	0.00	0.00	1.09	3.01	4.92*	15.57*	15.30*	4.37*	1.37	0.27	0.00	0.00	40.16*
	50.0–10.0	0.00	0.00	0.55	0.55	1.37	16.39*	20.49*	1.91	1.09	0.55	0.00	0.00	40.16*
Upper Tail	10.0–5.0	0.00	0.00	0.00	0.00	0.00	4.64*	3.83*	1.37	0.82	0.82	0.55	0.27	9.84*
	5.0–2.5	0.00	0.00	0.00	0.00	0.00	2.73	1.37	0.82	0.82	0.82	0.55	0.27	4.92*
	2.5–1.0	0.00	0.00	0.00	0.00	0.00	1.37	0.27	0.82	0.82	0.82	0.55	0.27	2.46
	1.0–0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.82	0.82	0.82	0.55	0.27	0.82
	0.5–0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.27	0.27	0.00	0.00	0.27
	All	0.27	0.82	2.46	4.92*	9.84*	40.16*	40.16*	9.84*	4.92*	2.46	0.82	0.27	100.00

See text and Table 3 for interpretation of this table.

*Percentage based on ten or more crash sections.

b. Georgia

Curvature Percentages

Grade Percentages		Lower Tail					Median		Upper Tail					
		0.0–0.5	0.5–1.0	1.0–2.5	2.5–5.0	5.0–10.0	10.0–50.0	50.0–10.0	10.0–5.0	5.0–2.5	2.5–1.0	1.0–0.5	0.5–0.0	All
Lower Tail	0.0–0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.21	0.21	0.00	0.00	0.00	0.42
	0.5–1.0	0.00	0.00	0.00	0.00	0.00	0.21	0.42	0.21	0.21	0.00	0.00	0.00	0.84
	1.0–2.5	0.00	0.00	0.00	0.00	0.00	0.63	1.47	0.21	0.21	0.00	0.00	0.00	2.31*
	2.5–5.0	0.00	0.00	0.00	0.00	0.00	1.68	2.52*	0.63	0.42	0.00	0.00	0.00	4.82*
	5.0–10.0	0.00	0.00	0.00	0.00	0.42	4.40*	3.56*	1.47	0.63	0.21	0.21	0.21	9.85*
Median (50.0)	10.0–50.0	0.21	0.63	1.05	2.10*	4.61*	6.77*	14.88*	3.98*	2.31	1.26	0.42	0.00	40.25*
	50.0–10.0	0.00	0.00	0.63	1.68	3.35*	4.88*	18.56*	3.14*	1.05	0.63	0.21	0.21	40.04*
Upper Tail	10.0–5.0	0.21	0.21	0.63	1.05	1.47	3.98*	3.14*	1.26	0.84	0.21	0.00	0.00	9.85*
	5.0–2.5	0.21	0.21	0.63	1.05	1.47	1.05	1.05	1.26	0.84	0.21	0.00	0.00	4.82*
	2.5–1.0	0.00	0.00	0.21	0.42	0.84	0.63	0.63	0.21	0.00	0.00	0.00	0.00	2.31*
	1.0–0.5	0.00	0.00	0.00	0.00	0.00	0.63	0.21	0.00	0.00	0.00	0.00	0.00	0.84
	0.5–0.0	0.00	0.00	0.00	0.00	0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.42
	ALL	0.42	0.84	2.31*	4.82*	9.85*	40.04*	40.25*	9.85*	4.82*	2.31*	0.84	0.42	100.00

[1]See text and Table 3 for interpretation of this table.

*Percentage based on ten or more crash sections.

TABLE 5 BIVARIATE CUMULATIVE DISTRIBUTION OF RANDOM SURVEY DATA IN NEW MEXICO BY CURVATURE AND GRADE PERCENTAGES: NEW MEXICO VERSUS GEORGIA

a. Travel Volume

Grade Percentages	Curvature Percentages: Lower Tail 0.0–0.5	0.5–1.0	1.0–2.5	2.5–5.0	5.0–10.0	Median 10.0–50.0	50.0–10.0	Upper Tail 10.0–5.0	5.0–2.5	2.5–1.0	1.0–0.5	0.5–0.0	All
Lower Tail 0.0–0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5–1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.0–2.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2.5–5.0	0.00	0.00	0.00	0.12	0.12	0.74	0.00	0.34	0.34	0.00	0.00	0.00	1.20
5.0–10.0	0.00	0.00	0.00	0.12	0.24	1.25	0.07	0.78	0.66	0.33	0.14	0.14	2.34
10.0–Median (50.0)	0.00	0.00	0.00	0.00	0.50	18.31	7.06	1.82	0.82	0.51	0.26	0.26	27.69
Median (50.0)–10.0	0.00	0.00	0.00	0.24	0.36	34.90	16.49	1.69	1.01	0.38	0.00	0.00	53.44
Upper Tail 10.0–5.0	0.00	0.00	0.00	0.47	0.59	10.06	4.31	1.58	0.85	0.47	0.12	0.12	16.54
5.0–2.5	0.00	0.00	0.00	0.35	0.47	5.76	2.40	1.26	0.78	0.47	0.12	0.12	9.89
2.5–1.0	0.00	0.00	0.00	0.21	0.33	2.28	1.26	0.78	0.54	0.24	0.12	0.12	4.64
1.0–0.5	0.00	0.00	0.00	0.21	0.33	1.12	0.65	0.12	0.12	0.12	0.12	0.12	2.22
0.5–0.0	0.00	0.00	0.00	0.00	0.00	0.82	0.26	0.12	0.12	0.12	0.12	0.12	1.20
ALL	0.00	0.00	0.00	0.83	1.68	64.52	27.93	5.87	3.33	1.68	0.52	0.52	100.00

See text and Table 3 for interpretation of this table.

b. Road Miles

Grade Percentages	Curvature Percentages: Lower Tail 0.0–0.5	0.5–1.0	1.0–2.5	2.5–5.0	5.0–10.0	Median 10.0–50.0	50.0–10.0	Upper Tail 10.0–5.0	5.0–2.5	2.5–1.0	1.0–0.5	0.5–0.0	All
Lower Tail 0.0–0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5–1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.0–2.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2.5–5.0	0.00	0.00	0.00	0.38	0.38	0.53	0.08	0.26	0.26	0.00	0.00	0.00	1.25
5.0–10.0	0.00	0.00	0.00	0.38	0.76	1.03	0.30	1.33	0.96	0.70	0.10	0.10	3.42
10.0–Median (50.0)	0.00	0.00	0.00	0.00	1.24	17.97	7.90	2.76	0.88	0.60	0.53	0.53	29.88
Median (50.0)–10.0	0.00	0.00	0.00	0.07	0.44	30.22	12.41	2.96	1.83	0.86	0.00	0.00	46.02
Upper Tail 10.0–5.0	0.00	0.00	0.00	0.85	1.23	11.59	5.60	2.26	2.16	1.51	0.38	0.38	20.68
5.0–2.5	0.00	0.00	0.00	0.47	0.85	6.76	2.39	2.02	1.94	1.51	0.38	0.38	12.02
2.5–1.0	0.00	0.00	0.00	0.32	0.70	3.74	1.27	1.25	1.18	0.76	0.38	0.38	6.96
1.0–0.5	0.00	0.00	0.00	0.32	0.70	1.48	0.48	0.38	0.38	0.38	0.38	0.38	3.04
0.5–0.0	0.00	0.00	0.00	0.00	0.00	0.87	0.00	0.38	0.38	0.38	0.38	0.38	1.25
ALL	0.00	0.00	0.00	1.30	3.67	60.81	26.20	9.31	5.82	3.67	1.01	1.01	100.00

See text and Table 3 for interpretation of this table.

TABLE 6 BIVARIATE CUMULATIVE DISTRIBUTION OF RANDOM SURVEY DATA IN GEORGIA BY CURVATURE AND GRADE PERCENTAGES: TRAVEL VOLUME VERSUS ROAD MILES

a. Travel Volume

Grade Percentages	0.0	Curvature Percentages: Lower Tail 0.5	1.0	2.5	5.0	10.0	Median 50.0	10.0	Upper Tail 5.0	2.5	1.0	0.5	0.0	All
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.06	0.00	0.00	0.06
Lower Tail	2.5	0.00	0.00	0.00	0.00	0.00	0.11	0.27	0.06	0.06	0.06	0.00	0.00	0.44
	5.0	0.00	0.00	0.00	0.00	0.00	1.22	0.58	0.06	0.06	0.06	0.00	0.00	1.86
	10.0	0.00	0.00	0.00	0.00	0.00	3.15	1.22	0.22	0.22	0.06	0.00	0.00	4.59
Median	(50.0)	0.00	0.00	0.00	0.00	0.25	21.94	10.33	0.67	0.30	0.17	0.06	0.00	33.20
	10.0	0.00	0.00	0.00	0.06	0.24	33.70	16.98	0.73	0.00	0.00	0.00	0.00	51.65
	5.0	0.00	0.00	0.00	0.11	0.28	6.71	3.04	0.53	0.25	0.00	0.00	0.00	10.57
	2.5	0.00	0.00	0.00	0.06	0.22	2.94	1.21	0.22	0.00	0.00	0.00	0.00	4.59
Upper Tail	1.0	0.00	0.00	0.00	0.06	0.06	1.98	0.34	0.17	0.00	0.00	0.00	0.00	2.54
	0.5	0.00	0.00	0.00	0.06	0.06	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.17
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ALL	0.00	0.00	0.00	0.17	0.77	65.51	31.57	2.15	0.77	0.23	0.06	0.00	100.00

See text and Table 3 for interpretation of this table.

b. Road Miles

Grade Percentages	0.0	Curvature Percentages: Lower Tail 0.5	1.0	2.5	5.0	10.0	Median 50.0	10.0	Upper Tail 5.0	2.5	1.0	0.5	0.0	All
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28	0.28	0.28	0.00	0.00	0.28
Lower Tail	2.5	0.00	0.00	0.00	0.00	0.00	0.55	0.40	0.28	0.28	0.28	0.00	0.00	1.23
	5.0	0.00	0.00	0.00	0.00	0.00	1.12	1.38	0.28	0.28	0.28	0.00	0.00	2.77
	10.0	0.00	0.00	0.00	0.00	0.00	3.21	2.66	0.43	0.43	0.28	0.00	0.00	6.30
Median	(50.0)	0.00	0.00	0.00	0.00	0.08	17.99	11.13	1.31	1.03	0.83	0.28	0.00	30.51
	10.0	0.00	0.00	0.00	0.28	0.75	32.30	16.26	1.23	0.00	0.00	0.00	0.00	50.55
	5.0	0.00	0.00	0.00	0.55	0.71	7.29	3.85	0.80	0.08	0.00	0.00	0.00	12.64
	2.5	0.00	0.00	0.00	0.28	0.43	3.54	1.89	0.44	0.00	0.00	0.00	0.00	6.30
Upper Tail	1.0	0.00	0.00	0.00	0.28	0.28	2.46	0.84	0.16	0.00	0.00	0.00	0.00	3.74
	0.5	0.00	0.00	0.00	0.28	0.28	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.83
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ALL	0.00	0.00	0.00	0.83	1.54	60.79	33.89	3.77	1.54	1.11	0.28	0.00	100.00

See text and Table 3 for interpretation of this table.

TABLE 7 LOGARITHM OF THE RATIO OF THE BIVARIATE DISTRIBUTIONS OF CRASH AND RANDOM SEGMENTS IN NEW MEXICO BY CURVATURE AND GRADE PERCENTAGES: TRAVEL VOLUME VERSUS ROAD MILES

a. Travel Volume

Curvature Percentages

Grade Percentages	Lower Tail					Median		Upper Tail					
0.0	0.5	1.0	2.5	5.0	10.0	50.0	10.0	5.0	2.5	1.0	0.5	0.0	All
0.5	.	.	.	.	.	.	.	+	+	+	+	.	+
1.0	.	+	+	+	+	.	.	.	+	+	+	.	+
Lower Tail 2.5	+	+	+	+	+	+	.	+	+	+	+	.	+
5.0	+	+	+	2.8	3.5	1.4	+	2.0	1.7	+	+	.	2.0 *
10.0	+	+	+	3.5	3.9*	1.5*	3.0	1.5	1.3	1.3	1.0	-	2.1 *
Median (50.0)	.	.	+	+	3.3 *	-0.2*	1.1*	1.3*	0.7	-0.9	-	-	0.5 *
10.0	.	.	+	1.2	1.9	-1.1*	0.3*	0.2	0.1	0.5	.	.	-0.4 *
5.0	.	.	.	-	-	-1.1*	-0.2*	-0.2	-0.0	0.8	2.2	1.2	-0.7 *
2.5	.	.	.	-	-	-1.1*	-0.8	-0.6	0.1	0.8	2.2	1.2	-1.0 *
Upper Tail 1.0	.	.	.	-	-	-0.7	-2.2	0.1	0.6	1.8	2.2	1.2	-0.9
0.5	.	.	.	-	-	-	-	2.8	2.8	2.8	2.2	1.2	-1.4
0.0	.	.	.	.	.	-	-	1.2	1.2	1.2	-	-	-2.1
ALL	+	+	+	2.6	2.5	-0.7	0.5	0.7	0.6	0.5	0.7	-0.9	0.0

Table a (b) is based on Table 4.a and Table 5.a (5.b).
See Text and Table 3 for the interpretation of this Table.
Entry in Tables is "." if both numerator and denominator were 0, entry is "+" if denominator was zero and it is "-" if numerator was zero.
*Percentage based on ten or more crash sections.

b. Road Miles

Curvature Percentages

Grade Percentages	Lower Tail					Median		Upper Tail					
0.0	0.5	1.0	2.5	5.0	10.0	50.0	10.0	5.0	2.5	1.0	0.5	0.0	All
0.5	.	.	.	.	.	.	.	+	+	+	+	.	+
1.0	.	+	+	+	+	.	.	+	+	+	+	.	+
Lower Tail 2.5	+	+	+	+	+	+	.	+	+	+	+	.	+
5.0	+	+	+	1.1	1.9	1.8	1.7	2.4	2.1	+	+	.	2.0*
10.0	+	+	+	1.9	2.2*	1.5 *	1.9	0.7	0.8	0.2	1.4	-	1.5*
Median (50.0)	.	.	+	+	2.0*	-0.3 *	1.0*	0.7*	0.6	-1.1	-	-	0.4*
10.0	.	.	+	3.1	1.6	-0.9 *	0.7*	-0.6	-0.7	-0.7	.	.	-0.2*
5.0	.	.	.	-	-	-1.0 *	-1.0 *	-0.7	-1.4	-0.9	0.5	-0.5	-1.1*
2.5	.	.	.	-	-	-1.2 *	-1.1	-1.3	-1.2	-0.9	0.5	-0.5	-1.3*
Upper Tail 1.0	.	.	.	-	-	-1.8	-1.2	-0.6	-0.5	0.1	0.5	-0.5	-1.5
0.5	.	.	.	-	-	-	-	1.1	1.1	1.1	0.5	-0.5	-1.9
0.0	.	.	.	.	.	-	.	-0.5	-0.5	-0.5	-	-	-2.2
ALL	+	+	+	1.9*	1.4*	-0.6 *	0.6 *	0.1 *	-0.2 *	-0.6	-0.3	-1.9	0.0

Table a (b) is based on Table 4.a and Table 5.a (5.b).
See Text and Table 3 for the interpretation of this Table.
Entry in Tables is "." if both numerator and denominator were 0, entry is "+" if denominator was zero and it is "-" if numerator was zero.
*Percentage based on ten or more crash sections.

TABLE 8 LOGARITHM OF THE RATIO OF THE BIVARIATE DISTRIBUTIONS OF CRASH AND RANDOM SEGMENTS IN GEORGIA BY CURVATURE AND GRADE PERCENTAGES: TRAVEL VOLUME VERSUS ROAD MILES

a. Travel Volume

Curvature Percentages

	Lower Tail					Median		Upper Tail					
0.0	0.5	1.0	2.5	5.0	10.0	50.0	10.0	5.0	2.5	1.0	0.5	0.0	All
0.5	.	.	.	.	.	.	+	+	+	.	.	.	+
1.0	.	.	.	.	.	+	+	1.9	1.9	-	.	.	3.9
Lower Tail 2.5	.	.	.	.	.	2.5	2.4	1.9	1.9	-	.	.	2.4*
5.0	.	.	.	.	.	0.5	2.1*	3.5	2.9	-	.	.	1.4*
Grade Percentages 10.0	.	.	.	.	+	0.5*	1.5*	2.7	1.5	1.9	+	+	1.1*
Median (50.0)	+	+	+	+ *	4.2*	-0.4*	0.5*	2.6*	3.0	2.9	2.9	.	0.3*
10.0	.	.	+	4.9	3.8*	-1.2*	0.1*	2.1*	+	+	+	+	-0.4*
5.0	+	+	+	3.2	2.4	-0.8*	0.0*	1.2	1.7	+	.	.	-0.1*
2.5	+	+	+	4.2	2.7	-1.5	-0.2	2.5	+	+	.	.	0.1*
Upper Tail 1.0	.	.	+	2.9	3.9	-1.7	0.9	0.3	.	.	.	.	-0.1*
2.5	.	.	.	-	-	2.5	+	.	.	.	.	.	2.3
0.0	.	.	.	.	.	+	.	.	.	.	.	.	99.9
ALL	+	+	+ *	4.8*	3.7*	-0.7*	0.4*	2.2*	2.7*	3.3*	3.9*	+	0.0

Table a (b) is based on Table 4.b and Table 6.a (6.b).
See Text and Table 3 for the interpretation of this Table.
Entry in Tables is "." if both numerator and denominator were 0, entry is "+" if denominator was zero and it is "-" if numerator was zero.
*Percentage based on ten or more crash sections.

b. Road Miles

Curvature Percentages

	Lower Tail					Median		Upper Tail					
0.0	0.5	1.0	2.5	5.0	10.0	50.0	10.0	5.0	2.5	1.0	0.5	0.0	All
0.5	.	.	.	.	.	.	+	+	+	.	.	.	+
1.0	.	.	.	.	.	.	+	-0.4	-0.4	-	.	.	1.6
Lower Tail 2.5	.	.	.	.	.	0.2	1.9	-0.4	-0.4	-	.	.	0.9*
5.0	.	.	.	.	.	0.9	0.6*	1.2	0.6	-	.	.	0.8*
Grade Percentages 10.0	.	.	.	.	+	0.7*	0.1*	1.8	0.5	-0.4	+	+	0.6*
Median (50.0)	+	+	+	+ *	5.8*	-0.1*	0.4*	1.6*	1.2	0.6	0.6	.	0.4*
10.0	.	.	+	2.6	2.2	-1.1*	0.2*	1.4*	+	+	+	+	-0.3*
5.0	+	+	+	0.9	1.0	-1.0*	-0.1*	0.7	3.4	+	.	.	-0.4*
2.5	+	+	+	1.9	1.8	-1.5	-1.2	1.5	+	+	.	.	-0.4*
Upper Tail 1.0	.	.	+	0.6	1.6	-1.6	-1.0	0.4	.	.	.	.	-0.7*
0.5	.	.	.	-	-	0.6	.	.	.	.	.	.	0.0
0.0	.	.	.	.	.	+	.	.	.	.	.	.	+
ALL	+	+	+ *	2.5*	2.7*	-0.6*	0.2*	1.4*	1.6*	1.1*	1.6*	+	0.0

Table a (b) is based on Table 4.b and Table 6.a (6.b).
See Text and Table 3 for the interpretation of this Table.
Entry in Tables is "." if both numerator and denominator were 0, entry is "+" if denominator was zero and it is "-" if numerator was zero.
*Percentage is based on ten or more crash sections.

• Collect geometric data on crashes that have occurred at sites with adverse geometry. Prepare the local version of the ratio tables (cf. Tables 7 and 8) for estimating overinvolvement.

• Define types of candidate sites in terms of the extent of overinvolvement. These should include all sites with extremely adverse geometries such as those marked with a positive sign (+) in the ratio tables. It is probable that most states will have large numbers of sites with estimated overinvolvement rates in excess of 10 or more or even 50 or more.

• Identify individual candidate sites for improvement. This master list could include sites of single-vehicle crashes, not necessarily fatal ones only, with sufficiently adverse geometry as well as sites with adverse geometry but no crashes.

SUMMARY AND RECOMMENDATIONS

Survey data on curvature and grade collected at crash and comparison sites in the states of New Mexico and Georgia were analyzed. The results showed that road sections with extreme geometry were far more common at the locations of fatal rollover crashes than at comparison sites. Numerical values for the extent of crash overinvolvement for sections with the most extremely adverse geometries could not be assigned because such sections were simply not found in the randomly chosen comparison samples, although 300 comparison sites had been surveyed in both states. However, such sites were overinvolved by a factor of 50 for some of the most extreme combinations of curvature and grade values in both states. The results also showed that the relative roles of extreme curvature and grade in causing fatal rollover crashes could vary between states, possibly because of differences in terrain or other factors. Specifically, although sharp left curves and steep downgrades were found to be more common at crash than at comparison sites in both states, the prevalence of steep downgrades at crash sites was greater in New Mexico than in Georgia. Because such differences in the relative roles of these factors are likely to be found in other states as well, no attempt was made to define an absolute priority scheme for hazard identification. Each state or geographic region should develop its own cost-effective set of priorities for hazard identification following the procedure outlined earlier.

Data for comparisons of bivariate curve-grade distributions at crash sites and at representative comparison sites may be available from construction plans or photologging surveys or could be routinely collected as part of existing state highway programs (e.g., maintenance or planning) that involve personnel already out in the field. The curvature and grade characteristics should be collected for known crash sites as well as for randomly selected sites representative of the state road system. Roadway sections included in the Highway Performance Monitoring System (HPMS) might be used to generate the bivariate distributions for representing the state's roadway system (*16*). However, because current data requirements for these sample sections do not allow for direct association of curve and grade on a specific roadway, the geometric data would have to be reanalyzed to construct the actual curve-grade bivariate distribution.

In addition to adverse vertical and horizontal alignments, inadequate superelevation was also shown to be associated with the incidence of fatal rollover crashes (*17*). This suggests that the bivariate curve-grade distribution recommended in the present paper for the identification of geometric hazards could be further improved by the incorporation of measures of superelevation deficiency. However, road sections with the most adverse vertical and horizontal alignments are extremely overinvolved in fatal rollover crashes and rate already high priorities for improvements regardless of their superelevation. In any case, the currently recommended design limits for superelevation rates preclude the adequate banking of curves 10 degrees or sharper for typical travel speeds (*17*). In states with primarily level terrain, very sharp curves are likely to be infrequent and, correspondingly, the role of superelevation is likely to be greater in causing fatal rollover crashes. Such states should appropriately modify the procedure recommended for hazard identification in this paper to include superelevation in their priority scheme from the outset. It is also recommended that all states collect data on superelevation deficiencies at the sites where curvature and grade are surveyed so that the current recommendations could be further refined in the future.

The present study has shown that extreme roadway geometry can raise the likelihood of fatal rollover crashes, and probably of all fatal single-vehicle crashes, by up to a factor of 50 or higher. Although the identification of specific measures for reducing the hazards at such sites was beyond the scope of this work, it is clear that improvements should be targeted to sites with such extreme risks.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the following assistance in the preparation of this work: Brian O'Neill and Sharon Rasmussen for editorial comments; JoAnn Wells and Marvin Ginsburg for data processing.

This work was supported by the Insurance Institute for Highway Safety.

REFERENCES

1. *The Status of the Nation's Highways: Conditions and Performance.* U.S. Department of Transportation, 1983.
2. *Fatal and Injury Accident Rates on Federal-Aid and Other Highway Systems: 1981.* U.S. Department of Transportation, March 1983.
3. D. E. Cleveland and R. Kitamura. Macroscopic Modeling of Two-Lane Rural Roadside Accidents. In *Transportation Research Record 681*, TRB, National Research Council, Washington, D.C., 1978, pp. 53–62.
4. J. A. Deacon, C. V. Zegeer, and R. C. Deen. Identification of Hazardous Rural Highway Locations. In *Transportation Research Record 543*, TRB, National Research Council, Washington, D.C., 1975, pp. 16–33.
5. T. L. Maleck. Michigan Dimensional Accident Surveillance (MIDAS) Model: Progress Report. In *Transportation Research Record 672*, TRB, National Research Council, Washington, D.C., 1978, pp. 16–19.
6. E. Hauer and B. Persaud. Common Bias in Before-and-After Accident Comparisons and Its Elimination. In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 164–174.
7. J. C. Glennon and M. C. Sharp. Methods for Improving Analysis of Roadside Safety. In *Transportation Research Record 709*, TRB, National Research Council, Washington, D.C., 1979, pp. 20–25.

8. J. C. Glennon. *NCHRP Report 148: Roadside Safety Improvement Programs on Freeways.* TRB, National Research Council, Washington, D.C., 1974.
9. J. H. Hall and T. E. Mulinazzi. Roadside Hazard Model. In *Transportation Research Record 681*, TRB, National Research Council, Washington, D.C., 1978, pp. 50–53.
10. W. W. Hunter, F. M. Council, A. K. Duit, and D. G. Cole. Methodology for Ranking Roadside Hazard Correction Programs. In *Transportation Research Record 672*, TRB, National Research Council, Washington, D.C., 1978, pp. 1–9.
11. P. H. Wright and L. S. Robertson. Priorities for Roadside Hazard Modification. *Traffic Engineering*, Vol. 46, No. 8, Aug. 1976.
12. P. H. Wright and P. Zador. Study of Fatal Rollover Crashes in Georgia. In *Transportation Research Record 819*, TRB, National Research Council, Washington, D.C., 1982, pp. 8–17.
13. J. W. Hall and P. Zador. Survey of Single-Vehicle Fatal Rollover Crashes in New Mexico. In *Transportation Research Record 819*, TRB, National Research Council, Washington, D.C., 1982, pp. 1–8.
14. H. H. Bissell, G. B. Pickington, J. M. Mason, and D. L. Woods. *Roadway Cross Section and Alignment in Synthesis of Safety Research Related to Traffic Control and Roadway Elements*, Vol. 1. Report FHWA-TS-82-232. FHWA, U.S. Department of Transportation, Dec. 1982.
15. P. H. Wright, J. Hall, and P. Zador. Low-Cost Countermeasures for Ameliorating Run-off-the-Road Crashes. In *Transportation Research Record 926*, TRB, National Research Council, Washington, D.C., 1983, pp. 1–7.
16. *Highway Performance Monitoring System: Field Manual for the Continuing Analytical and Statistical Data Base.* FHWA, U.S. Department of Transportation, Sept. 1980.
17. P. Zador, H. Stein, J. Hall, and P. H. Wright. Superelevation and Roadway Geometry: Deficiency at Crash Sites and on Grades. In *Transportation Research Record 1026*, TRB, National Research Council, Washington, D.C., 1984, pp. 43–50.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.

Evaluation of Opportunity-Based Accident Rate Expressions

MARK PLASS AND WILLIAM D. BERG

Recent development of opportunity-based accident rate expressions provides a potentially more sensitive set of indicators for use in safety studies. A comparison and evaluation of conventional versus opportunity-based accident rate expressions was undertaken for a set of 50 case study signalized intersections in Broward County, Florida. The effect of level of aggregation of the exposure data was examined, as well as differences in the rating of intersections by degree of hazard. It was found that hourly traffic volume counts may be necessary for reliable estimation of opportunity-based exposure levels and that the use of opportunity-based accident rate measures will yield significantly different hazard rankings compared with conventional accident rate expressions. Issues relating to exposure-based versus conflict-based opportunity expressions are also discussed.

The use of accident rates is a commonplace but not necessarily unbiased method of analyzing hazardous roadway locations. Typically, an accident rate is defined as either the total number of accidents per million vehicle miles or the total number of accidents per million entering vehicles. The first measure would apply to roadway segments, whereas the second is used at specific locations such as intersections.

M. Plass, City of Green Bay, City Hall, 100 N. Jefferson Street, Green Bay, Wis. 54301. W. D. Berg, Department of Civil and Environmental Engineering, 2206 Engineering Building, University of Wisconsin, Madison, Wis. 53706.

Although these rate expressions are easily calculated, because of aggregation effects, it is not clear that they accurately reflect the true degree of hazard. In both formulations, number of accidents is expressed as the sum of all accidents that have occurred at a given location over a specified time period. Locations such as intersections often have predominating types of accidents, the existence of which is not apparent because of this aggregation. In addition, the rate formula for intersections uses total entering vehicles and thus does not account for possible correlation between specific accident types and certain combinations of vehicular movements. Reality is therefore lost by the implied assumption that all entering vehicles have an equal probability of being involved in any type of accident.

Recent work by Council et al. (*1*) has resulted in the specification of a set of opportunity-based accident rate expressions that account for the correlation between accident type and vehicle movement. The opportunity-based accident rate differs from the conventional rate in that the number of opportunities for a given type of accident to occur is used as the exposure measure rather than total number of entering vehicles. An opportunity consists of the presence of certain prerequisite conditions related to vehicle speeds and relative positions. Without these conditions, the opportunity and therefore the likelihood of a given type of accident do not exist.

Although the work by Council et al. produced a complete specification of opportunity-based accident rate expressions, no evaluation was made of the impact of their application to hazardous location identification, countermeasure development, or before-and-after studies. The research reported here was undertaken with the objective of performing such an evaluation (2). In addition, a second objective was to assess the impact of the level of aggregation used in the calculation of the traffic flow parameters. This is an important issue in terms of the amount of data that is necessary to reliably estimate the exposure levels. The scope of the study was limited to signalized intersections.

REVIEW OF OPPORTUNITY-BASED ACCIDENT RATE EXPRESSIONS

The opportunity-based accident rate expressions (*1*) are derived using an assumed four-leg intersection (Figure 1). For each of four approaches ($i = A, B, C,$ and D), an entering flow rate (f_i) and an approach speed (v_i) are specified. Also recorded are the respective approach widths, W_{ac} and W_{bd} (opposite approaches are assumed to have equal widths so that $W_a = W_c$ and $W_b = W_d$), and the overall area of influence of the intersection (L).

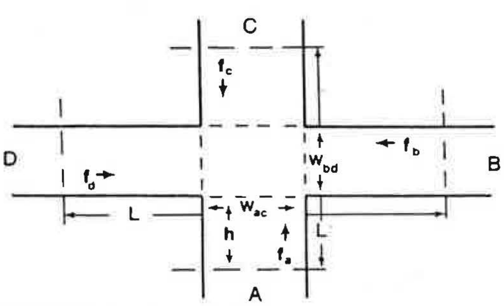


FIGURE 1 Schematic layout of intersection referred to by opportunity equations.

Single-Vehicle Accident Opportunities

A single-vehicle accident is one in which a vehicle runs off the road or strikes a fixed object, or both. The condition that constitutes the opportunity for this type of accident to occur is the presence of a single vehicle within the defined limits (L) of the intersection. The number of opportunities (O_{sv}) at a four-leg intersection during a period of time T is equal to the total number of vehicles entering the intersection. The opportunity equation takes the form

$$O_{sv} = T(f_a + f_b + f_c + f_d) \tag{1}$$

where T is the time period and f_i is the total entering flow rate on approach i.

Rear-End Accident Opportunities

A rear-end accident occurs when a moving vehicle strikes a stopped or slowed vehicle from behind. The opportunity for this type of accident consists of two conditions: two vehicles traveling in the same direction and both vehicles simultaneously within the limits of the intersection. The opportunity equation predicts the number of such pairs of vehicles during a given time period T through the use of a probability distribution function. The distribution function determines the proportion of vehicle headways less than the limits of the intersection L. The number of opportunities during time period T on approach i is equal to

$$O_R^i = Tf\{1 - \exp[-(f_i/\bar{v}_i)L]\} \tag{2}$$

where $\bar{v}_i = v_iL/(L + v_id_i)$ and d_i is the delay experienced by vehicles on approach i because of the signal.

Head-On Accident Opportunities

A head-on accident is one in which a vehicle strikes a stopped or moving vehicle that is traveling in the opposite direction, including left-turning vehicles. The opportunity for this type of accident consists of two conditions: two vehicles traveling in opposite directions and both vehicles simultaneously within the limits of the intersection. The opportunity equation predicts the number of vehicles traveling in the opposite direction met by an average vehicle on a given approach during both the red and green portions of the cycle for that approach. Opportunity equations are developed for both pairs of approaches (*AC* and *BD* as shown in Figure 1). The equation for the pair of approaches *AC* is

$$\begin{aligned} O_H = {} & (Tf_af_c/7200f_{tot})\ \{f_{bd}\ [(2cf_{bd}/f_{tot})] \\ & + (h + W_{bd})\ [(v_a^* + v_a)/v_av_a^*] + [(2h + W_{bd})/v_a] \\ & + f_{ac}\ [c(f_{bd}/f_{tot})] + 3[(2h + W_{bd})/v]\} \end{aligned} \tag{3}$$

where

$f_{tot} = f_a + d_b + f_c + f_d$,
$f_{ac} = f_a + f_c$,
$f_{bd} = f_b + f_d$,
v_a^* = average velocity of a vehicle that has accelerated from zero at the stop bar,
c = cycle length, and
h = length of intersection approach.

The opportunity equation for the pair of approaches *BD* has the same form as that for approaches *AC*. The total number of opportunities for the intersection is obtained by adding the equations for each pair of approaches.

Angle Accident Opportunities

Angle accidents involve vehicles traveling at right angles to one another that collide within that part of the intersection bounded by the stoplines of each approach. In this case there are two conditions that make up the opportunity: two vehicles traveling at right angles to one another, and both vehicles simultaneously within the area bounded by the stoplines of each approach. Flow products corresponding to perpendicular approaches ($f_a f_b$, $f_b f_c$, etc.) are used as an estimate of the number of pairs of vehicles that could be involved in an angle collision. The sum of these products, representing an estimate for the entire intersection, is multiplied by an estimate of the percentage of vehicles on each approach that simultaneously pass through the intersection on either a green or red light. This product is then multiplied by an estimate of how long vehicles remain within the area of the intersection. This estimate uses average vehicle speeds to account for those vehicles passing through the intersection on a green or yellow light and those accelerating from a stop. The longer a vehicle takes to pass through the area of angle accident opportunities, the longer it has the opportunity to be involved in an angle accident. The opportunity equation is

$$O_A = T/5280\{[(W_{ac}/v_b^*) + (W_{bd}/v_a^*)](f_a f_b + f_b f_c)(P_{g_a} P_{r_b} + P_r P_g)\} \quad (4)$$

where $v_a^* = [v_a f_a + 0.83(W_{bd} f_b)^{1/2}]/(f_a + f_b)$ and $v_b^* = [v_b f_b + 0.83(W_{ad} f_a)^{1/2}]/(f_a + f_b)$. P_{g_i} and P_{r_i} are the decimal percentages of vehicles on approach i entering the intersection during the green and red intervals, respectively.

Sideswipe Accident Opportunities

In a sideswipe accident one of two vehicles traveling in the same direction in adjacent lanes encroaches on the other vehicle's lane, which leads to a collision. The conditions of opportunity for a sideswipe accident are two vehicles in adjacent lanes simultaneously within the intersection as defined by L and portions of the vehicles being side by side. An estimate for each approach of such pairs moving through the intersection during the green phase is made, and the number of pairs that form during the red phase at each approach is then added to it.

TABLE 1 INTERSECTION CHARACTERISTICS

Characteristic	No. of Intersections	Characteristic	No. of Intersections
Geometry		Average daily traffic	
Four legs	45	Major street	
Three legs	5	0–10,000	0
Signal control		10,000–20,000	3
Two phase	6	20,000–30,000	8
Three phase	11	30,000–40,000	17
Four phase	8	40,000–50,000	12
Five phase	1	>50,000	10
Six phase	5	Minor street	
Seven phase	1	0–5,000	13
Eight phase	18	5,000–10,000	12
		10,000–15,000	7
		15,000–20,000	5
		20,000–25,000	4
		>25,000	9

The opportunity equation for a given approach is the sum of the opportunities occurring during the green and red phases:

$$O^i_{ss} = O^i_{ss,g} + O^i_{ss,r} \quad (5)$$

where

$O^i_{ss,g} = r_{bd}[Tf_1 f_2(v_1 - v_2)L/5280 v_1 v_2]$ if $L(v_1 - v_2)/v_2 > 40$ ft

$O^i_{ss,g} = r_{bd}[40Tf_1 f_2/5280 v_1]$ if $L(v_1 - v_2)/v_2 \leq 40$ ft

$O^i_{ss,r} = r_{ac}(Tf_i/2)$

$r_{ac} = (f_b + f_c)/f_{tot}$

$r_{bd} = (f_a + f_c)/f_{tot}$

The total number of opportunities for the intersection is obtained by adding the opportunities for the individual approaches.

Accident Rate Equations

Using the opportunity equations just described, two types of accident rates may be calculated. The first is the rate for a specific type of accident and has the form

$$r_i = a_i/O_i \quad (6)$$

where a_i is the number of accidents of type i, and O_i is the number of opportunities for accident type i.

The second type of rate is the total, or aggregate, rate for a given intersection. The total rate may be expressed in two ways:

$$R_1 = \sum_{i=1}^{k} a_i / \sum_{i=1}^{k} O_i \quad (7)$$

$$R_2 = \sum_{i=1}^{k} r_i \quad (8)$$

The R_2 measure only includes the opportunities for the accident types that have actually occurred. Before the accident rates are calculated, a decision must be made regarding an appropriate time period (T) during which flow rates and signal timing will be assumed to remain constant.

DATA BASE

The evaluation of the opportunity-based accident rate expressions was limited to 50 signalized intersections in Broward County, Florida. The basic characteristics of these intersections are summarized in Table 1. Accident data were obtained for each intersection from the Florida Department of Transportation for the year 1982. From these data the total number of vehicle-related accidents at each intersection was determined as well as the number of accidents by type. Conventional accident rates were also calculated for each of the intersections.

The number of opportunities per day at each intersection for

TABLE 2 RESULTS OF F-RATIO TEST FOR SIGNIFICANCE OF LEVEL OF AGGREGATION ($\alpha = 0.05$)

Accident Rate	No. of Intersections	Average Day Versus Hourly			Peak/Off-Peak Versus Hourly		
		F^a	$F_c{}^b$	Significant	F^a	$F_c{}^b$	Significant
R_1	50	28.91	4.08	Yes	16.89	4.08	Yes
R_2	50	1.04	4.08	No	1.02	4.08	No
Angle	39	3.36	4.17	No	4.59	4.17	Yes
Single-vehicle	29	Rate Values Identical Over Level of Aggregation					
Head-on	7	0.30	5.99	No	0.008	5.99	No
Rear-end	41	119.58	4.08	Yes	119.31	4.08	Yes
Sideswipe	37	36.12	4.17	Yes	50.48	4.17	Yes

[a]Calculated F-ratio.
[b]Critical F-ratio.

each accident type was computed for three levels of aggregation:

1. Average day: Accident opportunities are calculated by using average hourly flow rates for the 16-hr period from 6:00 a.m. to 10:00 p.m.
2. Peak/off-peak: Accident opportunities are calculated by using average hourly flow rates for the peak (7:00–9:00 a.m. and 4:00–6:00 p.m.) and off-peak (6:00–7:00 a.m., 9:00 a.m.–4:00 p.m., 6:00–10:00 p.m.) periods, and then the results for the two periods are weighted and summed.
3. Hourly: Accident opportunities are calculated for each of the 16 hr from 6:00 a.m. to 10:00 p.m., and then the results for the 16 periods are summed.

Opportunity-based accident rates were then calculated for each level of aggregation at each intersection. Each rate was calculated as the total number of accidents occurring during the study year divided by the annual number of opportunities (365 times the number of opportunities during the average day).

FINDINGS

Effect of Level of Aggregation

The level of aggregation used in calculating the number of opportunities for a given type of accident has the potential to significantly influence the value of the resulting accident rates. This can subsequently introduce uncertainty into the ranking of intersections on the basis of relative hazard, as well as into the evaluation of countermeasure effectiveness.

An evaluation of these impacts was made by comparing the R_1, R_2, and individual opportunity-based accident rates calculated at each of the levels of aggregation identified earlier. For each accident rate type (R_1, R_2, angle, single-vehicle, head-on, rear-end, and sideswipe), an $m \times 3$ matrix of accident rates was prepared. The number of rows (m) corresponded to the number of intersections that experienced that accident type, each row representing a specific intersection. Each of the three columns corresponded to one of the three levels of aggregation. Using the hourly level as the base, a set of F-ratio tests (*3*, pp. 383–384) was performed to determine whether accident rates calculated using a higher level of aggregation are significantly different at the 95 percent level of confidence. As summarized in Table 2, there were statistically significant differences between accident rate values calculated at a low level of aggregation (hourly) and those calculated at higher levels (average day, peak/off-peak) for the R_1 total rate and the rear-end and sideswipe individual rates. This implies that level of aggregation does have an important effect and that the opportunity expressions may need to be calculated at the hourly level to assure the most reasonable and reliable safety evaluations.

Sensitivity of Hazard Rankings to Exposure Measures

Rankings of the 10 most hazardous study intersections on the basis of accident rate were made by using both conventional (accidents per million vehicles) and opportunity-based (accidents per million opportunities) rate measures as summarized in Tables 3 and 4. The opportunity-based measures were calculated at the hourly level of aggregation. This level was selected on the basis that it would provide the most accurate estimate of the opportunities for the occurrence of accidents and, as revealed by the F-ratio tests, some accident rates calculated at this level are significantly different from those calculated at a higher level of aggregation.

As shown in Table 3, the R_1 and R_2 rankings differ from the

TABLE 3 HAZARD RANKING OF CASE STUDY INTERSECTIONS BY OVERALL ACCIDENT RATE

Conventional		R_1		R_2	
Intersection	Rate[a]	Intersection	Rate[b]	Intersection	Rate[b]
49	3.64	15	0.171	36	5.847
15	2.79	49	0.118	47	0.679
5	2.72	42	0.082	31	0.196
42	2.64	7	0.079	15	0.171
40	2.61	29	0.078	49	0.122
24	2.52	40	0.064	12	0.089
34	2.40	39	0.056	42	0.086
22	2.19	5	0.054	27	0.085
7	2.09	34	0.051	29	0.082
21	2.02	1	0.049	7	0.079

[a]Accidents per million vehicles.
[b]Accidents per million opportunities.

TABLE 4 HAZARD RANKING OF CASE STUDY INTERSECTIONS BY INDIVIDUAL ACCIDENT TYPES

R_1		Angle		Single-Vehicle		Head-On		Rear-End		Sideswipe	
Inter-section	Rate	Inter-section	Rate	Inter-section	Rate	Inter-section	Rate	Inter-section	Rate	Inter-section	Rate
15	0.171	15	4598.2	49	0.552	15	0.375	34	1.13	15	0.0400
49	0.118	49	111.3	42	0.329	7	0.327	42	1.11	49	0.0205
42	0.082	29	64.2	34	0.286	29	0.169	6	1.08	42	0.0220
7	0.079	37	39.7	22	0.233	9	0.119	11	1.06	7	0.0150
29	0.078	48	37.0	32	0.195	1	0.115	5	0.96	39	0.0140
40	0.064	31	21.7	29	0.180	5	0.106	24	0.96	28	0.0130
39	0.056	30	17.0	8	0.164	18	0.079	15	0.92	8	0.0116
5	0.054	40	16.5	43	0.154	8	0.079	8	0.85	5	0.0114
34	0.051	46	10.7	44	0.137	10	0.065	23	0.83	22	0.0112
1	0.049	39	10.0	23	0.094	19	0.050	22	0.82	23	0.0108

conventional at each position within the ranking. In addition, of the 10 most hazardous intersections according to the conventional accident rate, only 7 appear in the R_1 list and 4 in the R_2 list. This is significant only from the standpoint that it reflects the difference in what is being measured (i.e., accidents per entering vehicles as opposed to accidents per opportunities). Because different things are being measured, it would be expected that the rankings would also differ. If the rankings obtained through the opportunity-based R_1 and R_2 measures did not significantly differ from the conventional ranking, the higher level of sensitivity to hazard implied by the opportunity-based measure would be in doubt.

The differences between the R_1 and R_2 rankings is a reflection of the varying sensitivity to hazard found within the opportunity-based measures. This sensitivity is related to the level of aggregation used in establishing the total accident opportunities. In the case of the R_1 measure, the level of aggregation is high, because all possible opportunities are used in the denominator of the rate expression. In effect, the R_1 rate measure provides an indication of the overall "level of service" offered by an intersection. The inclusion of opportunities for occurring accident types gives an indication of relative hazard, whereas the additional use of opportunities for nonoccurring types allows for a reflection of the relative safety at the intersection.

The R_2 measure, on the other hand, uses only opportunities for accident types that actually occurred. In instances where only one type of accident has occurred, the R_2 measure becomes, in essence, an individual rate measure and is therefore more sensitive to the specific hazard than the R_1 measure. In cases where more than one accident type has occurred, the R_2 measure tends to mask specific hazards, as does the R_1 measure because of the increased aggregation of opportunities. However, the R_2 measure also fails to completely account for the level of safety implied by the lack of certain types of accidents. In addition, it has the potential to be biased in cases where the occurrence of a given accident type at an intersection is reduced to zero from a given year to the next. Because the R_2 rate expression uses only opportunities corresponding to occurring accident types in its denominator, a reduction to zero for a given accident type can result in a significant change in the value of the denominator, and therefore in the rate value (and implied hazard) assigned to the intersection. In the following hypothetical example, the total number of accidents at one of the study intersections has been reduced by 33 percent, resulting in an increase of approximately 11,000 percent in the R_2 rate measure (expressed as accidents per million opportunities):

	1982		*1983 (hypothetical)*	
	No. of Accidents	*R_2 Rate*	*No. of Accidents*	*R_2 Rate*
Angle	4		4	
Sideswipe	2		0	
Total	6	0.196	4	21.713

The data in Table 4 can also be used to examine the relationship between overall accident rate and individual accident types. Using the ranking of the 10 most hazardous intersections based on the R_1 rate, only 3 of these appear on the lists of the 10 intersections with the highest angle and head-on accidents, only 4 appear on the single-vehicle list, and only 6 appear on the rear-end and sideswipe lists. This demonstrates that overall accident rates are not necessarily good indicators of the existence of special types of hazardous conditions that may merit additional attention. This should not be unexpected given that collision diagrams generally reveal an accident pattern in which some, but not all, accident types dominate. This furthermore suggests that some treatable intersection problems may escape notice if overall accident rates are the only indicators used to identify hazardous locations. Upon implementation of a countermeasure to address a specific problem, a before-and-after comparison using the associated rate for that accident type and adjusted for regression to the mean would clearly be the most sensitive indicator of countermeasure effectiveness.

CONCLUSIONS

The level of aggregation used in calculating the opportunity expressions has a significant impact on the value of the R_1 total rate and the rear-end and sideswipe individual rates. This suggests that the use of hourly traffic count data for the calculation of the opportunity expressions will reduce the likelihood of creating bias in hazard rankings or error in before-and-after comparisons.

On the basis of the case study comparisons, it is clear that the conventional accident rate does not provide the same indication of hazard as the opportunity-based measures. The conventional measure is considered to be a less sensitive indicator because it assumes that all vehicles entering an intersection are equally likely to be involved in any type of accident. In addition, the predominance of certain accident types cannot be made apparent by the rate values because of the aggregation of all occurring accidents in the rate expression.

In comparing the R_1 and R_2 opportunity-based measures, the criteria for using a total accident rate measure and the degree to which each measure meets the criteria must be considered. The reason for using a total rate is simply to provide a basis for comparison of relative overall intersection hazard. The R_1 measure achieves this in its use of both opportunities corresponding to occurring accidents, which denotes relative hazard, and opportunities corresponding to accident types that did not occur, which denotes relative level of safety. The R_2 measure is sensitive to specific hazard in instances where only one type of accident has occurred, in which case it effectively becomes an individual rate. Whenever more than one type of accident occurs, the R_2 measure becomes similar to the R_1 from the standpoint that specific problems are masked. However, because opportunities for accident types that did not occur are not included in the R_2 rate expression, the relative safety of an intersection is not reflected in the rate value. Because neither measure is strictly able to identify specific hazards, the R_1 measure, which offers the most balanced appraisal of overall relative hazard, is considered the most appropriate for use as a means of overall comparison.

In considering the applicability of the various accident rate measures to the identification of hazardous locations, the development of countermeasures, and the performance of before-and-after studies, several recommendations are offered. First, the identification of hazardous locations is obviously critical and is achieved to some degree by the conventional and the R_1 and R_2 measures. Although each of these is capable of illustrating relative hazard among a group of intersections by the assignment of aggregate rate values, none is able to address the specific hazards. As discussed previously, the aggregation present in each of their measurements causes the "true" hazard at a given location either to be masked or, in the case of the R_2 measure, to be represented in a biased manner.

The use of individual opportunity-based accident rates would be the most effective means of identifying specific hazards. Rather than a single ranking of hazardous intersections whose true hazards would not be apparent if total accident rates were used, individual rankings by accident type would be more appropriate. The use of individual rate measures would not only provide an efficient and effective means of hazard identification, but would also facilitate the development of countermeasures because the hazards they are designed to alleviate would be made apparent.

For before-and-after studies, both conventional and total opportunity-based measures lack sensitivity because of the aggregation of accident types present in their rate expressions. The effect of a countermeasure will not always be apparent from these measures because it is not possible to determine whether the accident type related to the countermeasure has been reduced. Individual rates, on the other hand, address specific accident types and therefore offer the best appraisal of the effect a countermeasure has had.

FUTURE RESEARCH

The opportunity-based accident rate expressions examined in this research are now undergoing further refinement by Council and his colleagues at the University of North Carolina. Nevertheless, the general observations noted regarding the effect of the level of aggregation used in calculating opportunities and the sensitivity of various accident rate formulations to specific types of hazard are likely to remain relevant. One issue that merits additional research is the relationship between the specification of the opportunity-based accident rates and their sensitivity to changes in intersection geometry and signal timing. At the heart of this matter is the fundamental question of whether these opportunity expressions are measuring exposure to accidents or traffic conflicts that may result in an accident. An excellent discussion of this definitional problem can be found in a paper by Hauer (*4*).

If the opportunity expressions are specified to measure exposure to various accident types, then their numerical value for any given intersection should be independent of geometrics and signal timing (except where certain movements become prohibited) and should only be a function of the exposed traffic flows. The safety effectiveness of common geometric and signal timing improvements will then be measurable by using accident rates formulated with exposure-based opportunities. This is an important capability because safety is fundamentally achieved by separating traffic flows either spatially or temporally. Accordingly, safety measures of effectiveness should be sensitive to these types of countermeasures.

On the other hand, if opportunity expressions are specified to measure expected number of vehicle conflicts of various types, then their numerical value will be dependent on the geometrics and signal timing at the intersection. This means that accident rates calculated by using conflict-based opportunities will be relatively insensitive to the geometric and signal timing characteristics of the site because the effect of these elements will have already been accounted for in the denominator of the accident rate expression. However, such accident rates would presumably remain sensitive to the effects of human factors, environmental conditions, and information system design at the intersection.

The implication of these comments is that if one can predict conflicts, a certain fraction of which result in accidents, then the expected number of conflicts becomes a surrogate measure for the expected accident rate. This is analogous to the premise underlying the traffic conflicts technique (*5*). Expressions for estimating the expected number of conflicts as a function of traffic flows, intersection geometry, and signal timing would become useful planning and design tools for the engineer. They would effectively complement the delay-based evaluation techniques found in the *Highway Capacity Manual* (*6*).

Where accident data rather than conflict data are to be used in evaluating relative safety, the denominator of the accident rate expression should reflect exposure to accidents. With a reasonable formulation of exposure, it should be possible to have an indicator that can be used to evaluate the safety effectiveness of a variety of countermeasures aimed at achieving higher levels of flow separation. These would include various forms of channelization and signal timing (especially left-turn phasing alternatives).

The opportunity expressions examined in this research include several that fall in the category of a conflict measure rather than an exposure measure. It is recommended that future research closely examine the exposure versus conflict issue as well as the sensitivity of the resulting accident rate expressions to typical countermeasures.

REFERENCES

1. F. M. Council, J. R. Steward, D. W. Reinfurt, and W. W. Hunter. *Exposure Measures for Evaluating Highway Safety Issues*, Vol. 1: *Final Report*. Report FHWA/RD-83/088. University of North Carolina Highway Safety Research Center, Chapel Hill, N.C., 1983.
2. M. Plass. *Evaluation of Exposure-Based Accident Rates*. Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Jan. 1985.
3. G. Keppel. *Design and Analysis: A Researcher's Handbook*, 2nd ed. Prentice-Hall, Englewood Cliffs, N.J., 1982.
4. E. Hauer. Traffic Conflicts and Exposure. *Accident Analysis and Prevention*, Vol. 14, No. 5, 1982, pp. 359–364.
5. W. D. Glauz and D. J. Migletz. *NCHRP Report 219: Application of Traffic Control Analysis at Intersections*. TRB, National Research Council, Washington, D.C., 1980.
6. *Special Report 209: Highway Capacity Manual*. TRB, National Research Council, Washington, D.C., 1985.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.

Demonstration of Regression Analysis with Error in the Independent Variable

RICHARD M. WEED AND RICARDO T. BARROS

Regression analysis is frequently used in the engineering field to develop mathematical models for a wide variety of applications. Of the several assumptions upon which regression theory is based, one of the most fundamental is that the *X*-values are known exactly and that any error is associated only with the *Y*-measurements. Because this is not the case for many engineering applications, a study was conducted (*a*) to determine the magnitude of this problem and (*b*) to develop and test a software package that incorporates a theoretical solution found in the literature. Computer simulation is used to demonstrate both the seriousness of the problem and the effectiveness of the solution. An example based on early-strength tests of concrete is presented.

Many engineering applications require the development of a mathematical model (equation) to characterize some physical relationship. Examples include those shown in Table 1.

In the first example, the objective is a reliable early predictor of the 28-day strength of concrete, a measure upon which many acceptance procedures are based. The objective of the second example is to replace a costly and time-consuming subjective rating procedure with a simple mechanical device. In the third example, a relationship is sought that will become an integral part of a pavement management system.

New Jersey Department of Transportation, 1035 Parkway Avenue, Trenton, N.J. 08625.

The variable to be predicted or estimated is placed on the *Y*-axis and an equation of the form $y = f(x)$ is desired. The equation may be linear, quadratic, exponential, or any other appropriate form. The analyst, from his understanding of the physical process, will often know the correct form in advance. In other cases, it may be necessary to let the data dictate the form.

The desired relationship is often derived empirically from a set of *X,Y*-data values by using the technique of least squares (*1*) as shown in Figure 1. The procedure, invisible to the analyst when executed by a computer program, consists of solving for

TABLE 1 PHYSICAL RELATIONSHIPS CHARACTERIZED BY MATHEMATICAL MODELS

Characteristic of Interest	*X*-Data (Independent Variable)	*Y*-Data (Dependent Variable)
Compressive strength of concrete	Seven-day test results	Twenty-eight-day test results
Rating of highway pavement serviceability	Output of mechanical roughness-measuring device	Average rating of a team of panelists
Rating of highway pavement serviceability	Cumulative axle loads	Current rating of serviceability

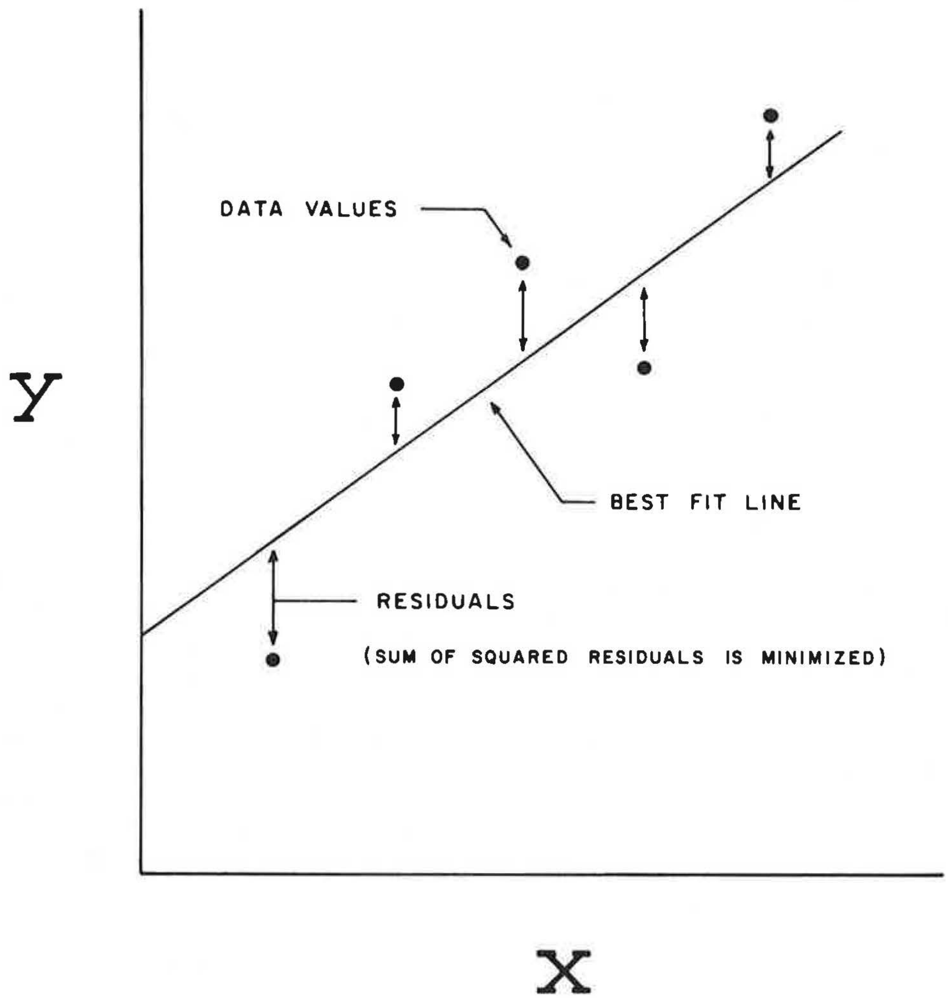


FIGURE 1 Concept of the ordinary least-squares technique.

the line that best fits the data. When ordinary least squares is used, the best fit is defined as that line of the chosen form that minimizes the sum of the squared residuals in a direction parallel to the *Y*-axis.

In carrying out this procedure, the user makes several theoretical assumptions, one of the most fundamental of which is that the *X*-values are known exactly and that any error of measurement is associated only with the *Y*-values. Because it is often impossible or impractical to achieve this idealized condition in practice, a study (*2*) was undertaken (*a*) to investigate the effect of failing to satisfy this assumption and (*b*) to develop and test a software package that incorporates a theoretical procedure for dealing with *X*-error (*3*). Linear models are addressed because only the linear solution of the *X*-error problem has been published to date.

USE OF COMPUTER SIMULATION

In order to demonstrate the extent of the *X*-error problem and the effectiveness of the solution, a method was required to observe and quantify the accuracy and precision of the regression estimates. This can readily be accomplished with computer simulation by performing the following steps:

1. Randomly generate a bivariate normal *X*,*Y*-data set with known regression (population) parameters:
   (*a*) intercept (β_0),
   (*b*) slope (β_1), and
   (*c*) residual error (σ_{yx}).
2. Include a fixed amount of *X*-error, either in absolute terms or as a percentage of σ_{yx}.
3. Use the randomly generated data to estimate the regression parameters:
   (*a*) intercept (B_0),
   (*b*) slope (B_1), and
   (*c*) residual error (S_{yx}).
4. Repeat the entire process many times in order to compare the distributions of the regression estimates with the known parameters. Ideally, the sampling distributions of the estimates should be centered on the true population parameters and have relatively narrow dispersions.

This technique can be used to provide a very dramatic demonstration of the bias introduced by error in the *X*-variable and the conditions that accentuate it. It will also be used to demonstrate the effectiveness of the procedure developed to overcome this problem.

TABLE 2 EXAMPLES OF BIAS INTRODUCED BY THE PRESENCE OF X-ERROR

Parameter	True Value	Regression Estimates Obtained by Ordinary Least Squares for Selected Levels of X-Error[a] 0	25	50	75	100
Intercept	100	100.14	108.11	129.25	160.81	200.23
Slope	10	10.00	9.84	9.41	8.78	8.00
Residual error	5	4.93	13.21	24.62	35.08	44.79

NOTE: Results obtained by computer simulation with 1,000 replications of 30 data points spanning the range between approximately $X = 30$ and $X = 70$.

[a] X-error is measured as the ratio σ_{xx}/σ_{yx}, expressed as a percentage, in which σ_{xx} represents the error in individual X-measurements.

TABLE 3 EFFECT OF SLOPE ON THE DEGREE OF BIAS INTRODUCED

Parameter	True Value	Regression Estimates Obtained by Ordinary Least Squares for Fixed X-Error[a] and Selected Levels of Slope 0.0	0.5	1.0	5.0	10.0
Intercept	100	99.94	105.14	109.95	150.26	200.23
Slope	–[b]	0.00	0.40	0.80	3.99	8.00
Residual error	5	4.95	5.44	6.69	22.58	44.79

NOTE: Results obtained by computer simulation with 1,000 replications of 30 data points spanning the range between approximately $X = 30$ and $X = 70$.

[a] The level of X-error is fixed at 100 percent ($\sigma_{xx} = \sigma_{yx}$ in which σ_{xx} represents the error in individual X-measurements).

[b] Variable (values given in column headings).

DEMONSTRATION OF THE PROBLEM

Table 2 has been prepared with dimensionless data to demonstrate the detrimental effect that even a moderate amount of X-error can have under certain conditions. It may be observed that when there is no X-error, the estimated values (averages for 1,000 replications) of all three regression parameters are extremely close to the true population values. When the amount of X-error is as little as 25 percent of the Y-error (σ_{yx}), it may be seen that a substantial amount of bias has been introduced in the estimates of both the intercept and the residual error. As the degree of X-error increases, all three regression parameters begin to show considerable bias. When the X-error and the Y-error are approximately equal—a fairly common situation in actual practice—the regression estimates differ substantially from the true population parameters.

It should be noted that this example was chosen to dramatize the potentially serious nature of the X-error problem. Although the three population parameters ($\beta_0 = 100$, $\beta_1 = 10$, $\sigma_{yx} = 5$) are not extreme in any sense, the effect is pronounced because the slope is fairly steep. Figure 2 presents a conceptual illustration of the effect of the slope in translating X-error into apparent Y-error, error that the ordinary least-squares procedure attributes solely to the Y-variable. The examples in Table 3

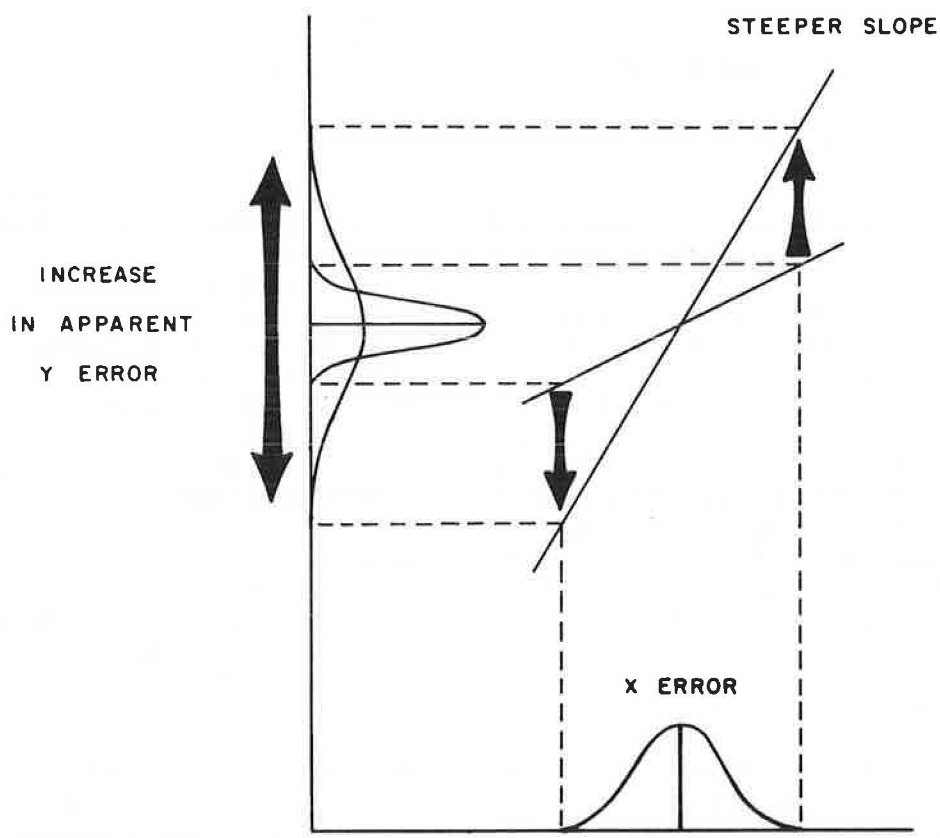


FIGURE 2 Effect of slope on the translation of X-error into apparent Y-error.

further demonstrate this effect. Viewed collectively, the examples in Tables 2 and 3 provide an empirical indication of the conditions that tend to influence the magnitude of the *X*-error problem: the ratio of *X*-error to *Y*-error and the slope of the regression line.

MANDEL'S SOLUTION

A theoretically derived procedure for avoiding the bias in the regression estimates due to *X*-error has been published by Mandel (*3*). Use of the procedure requires one additional bit of information that is usually readily available or readily obtainable: the ratio of the variances associated with the *X*- and *Y*-measurements. Although the mathematical procedure is somewhat involved, the concept is easy to visualize. Ordinary least squares minimizes the sum of the squared residuals in a direction parallel to the *Y*-axis. In the presence of *X*-error, the minimization process is performed by Mandel's procedure in a direction oblique to the *X*- and *Y*-axes, the exact angle being determined primarily by the relative magnitude of the *X*- and *Y*-error.

In order to test the effectiveness of Mandel's method, it was applied to the same data sets used to develop Table 2. The results are reported in Table 4 and the values from Table 2 are repeated for ease of comparison. It can be seen from Table 4 that Mandel's method is extremely effective in removing the bias that exists when an application with *X*-error is analyzed by ordinary least squares. Its only discernible weakness in this example is a possible small downward bias of the estimate of the intercept when the *X*-error is quite large.

To judge whether this apparent bias was real, the simulation program was modified to print out a histogram and elementary statistics for 1,000 intercept estimates. The *X*-error was held constant at 100 percent of σ_{yx}. Although not strictly applicable because the distribution of intercept estimates was somewhat skewed, a *t*-test indicated that the average intercept of 93.06 was highly significantly different ($\alpha < 0.001$) from the true value of 100.0. Although it is not obvious from the results in Table 4, a similar test suggests that the slope estimates may also be biased to a very small degree. Consequently, although Mandel's method appears to be very effective and is far superior to ordinary least squares, it must be concluded that it is not totally unbiased in all cases.

Figure 3 shows in a graphical way the effects that have been observed in Table 4. The distributions shown were drawn from histograms generated by the same simulation programs used to develop Tables 2–4. Mandel's method can be seen to be essentially unbiased in that the means of the distributions generated by that method are very close to the true population parameters. In marked contrast, the distributions produced by ordinary least squares are shifted substantially away from the true parameters. Another important observation in Figure 3 is that the distributions for the intercept and the slope obtained with Mandel's method are only slightly more dispersed than those obtained by ordinary least squares. This indicates that a substantial gain in accuracy has been achieved with only a slight loss of precision. For the residual error, Mandel's method is both more accurate and more precise.

An interesting feature of Mandel's method is that, unlike the least-squares technique, the same regression line will be obtained regardless of which variable is considered to be independent (*X*) and which is dependent (*Y*). Furthermore, the degree of uncertainty associated with predictions made by this method is the same for either choice of variables (*3*, p.9). This property conveniently avoids a controversial aspect of the calibration application, the need to work backward through the regression procedure to estimate what value of *X* gave rise to an observed value of *Y*.

Another series of computer simulation tests was performed by using the appropriate procedures for computing interval estimates for the intercept, slope, and σ_{yx}. A level of confidence of $1 - \alpha = 0.95$ was selected and the number of times that the interval estimate actually contained the true population parameter was counted. For 1,000 replications, the empirically observed results should fall within the range of approximately $0.95 \pm 2\{[(0.95)(0.05)]/1{,}000\}^{1/2} = 0.95 \pm 0.014$ when the interval estimation process is working properly. It can be seen from the results in Table 5 that, even for small amounts of *X*-error, the interval estimates computed by ordinary least squares contain the population parameters substantially less often than desired. In contrast, all of the interval estimates computed by Mandel's method are satisfactory.

TABLE 4 COMPARISON OF MANDEL'S METHOD WITH ORDINARY LEAST SQUARES

Parameter	True Value	Method	Regression Estimates Obtained for Selected Levels of *X*-Error[a] 0	25	50	75	100
Intercept	100	Mandel	100.18	100.19	98.34	96.69	93.06
		OLS	100.14	108 11	129.25	160.81	200.23
Slope	10	Mandel	10.00	9.99	10.03	10.06	10.14
		OLS	10.00	9.84	9.41	8.78	8.00
Residual error	5	Mandel	4.93	4.94	4.97	4.94	4.97
		OLS	4.93	13.21	24.62	35.08	44.79

NOTE: Results obtained by computer simulation with 1,000 replications of 30 data points spanning the range between approximately $X = 30$ and $X = 70$. OLS = ordinary least squares.

[a] *X*-error is measured as the ratio σ_{xx}/σ_{yx}, expressed as a percentage, in which σ_{xx} represents the error in individual *X*-measurements.

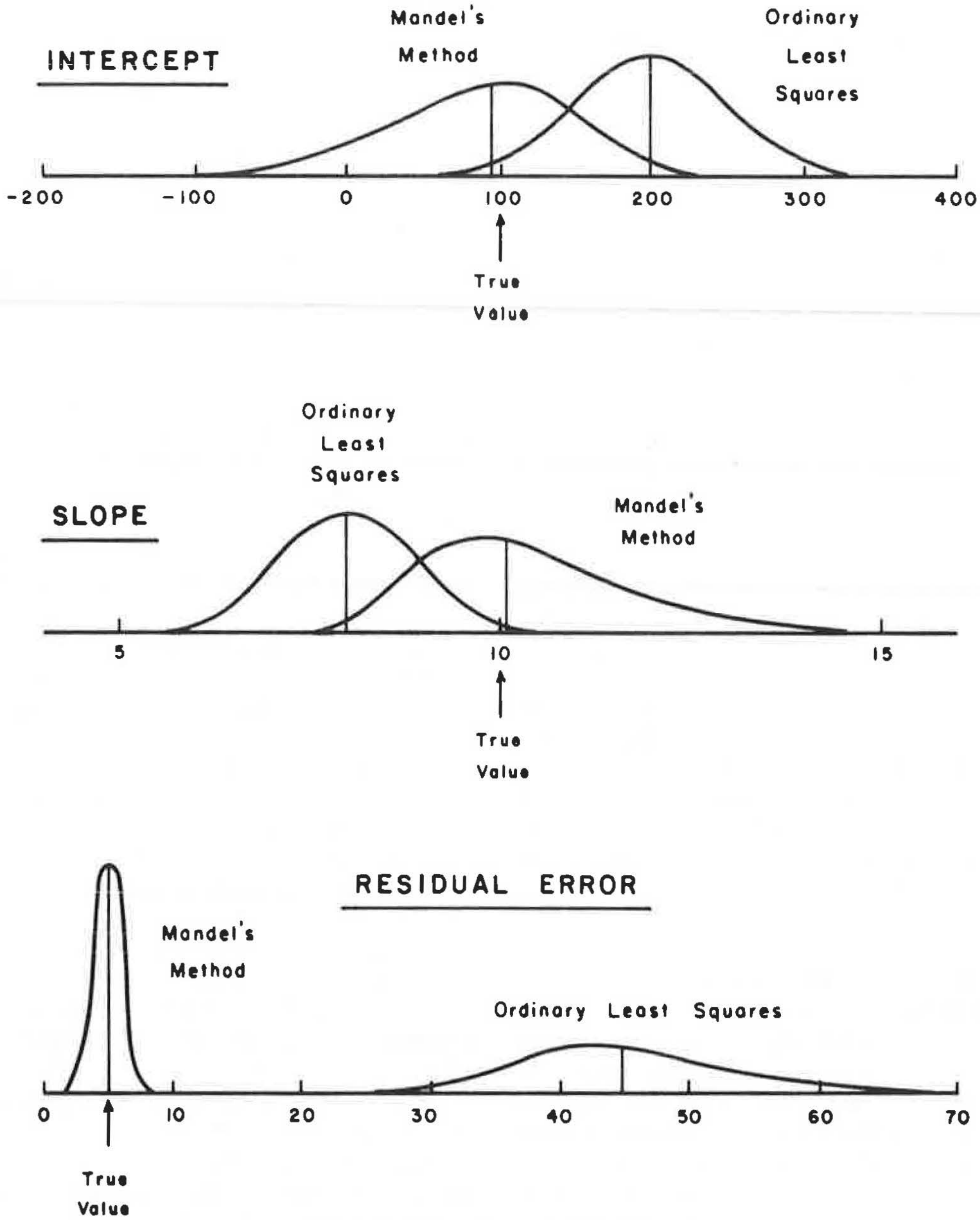


FIGURE 3 Comparison of distributions of regression estimates.

TABLE 5 EFFECT OF X-ERROR ON INTERVAL ESTIMATES

Parameter	Desired Confidence Level	Method	Empirically Observed Confidence Levels at Selected Levels of X-Error[a]				
			0	25	50	75	100
Intercept	0.95	Mandel	0.950	0.938	0.954	0.939	0.953
($\beta_0 = 100$)		OLS	0.951	0.898	0.768	0.548	0.321
Slope	0.95	Mandel	0.948	0.941	0.956	0.948	0.952
($\beta_1 = 10$)		OLS	0.946	0.889	0.755	0.534	0.301
Residual error	0.95	Mandel	0.955	0.956	0.963	0.947	0.959
($\sigma_{yx} = 5$)		OLS	0.955	0.0	0.0	0.0	0.0

NOTE: Results obtained by computer simulation with 1,000 replications of 30 data points spanning the range between approximately $X = 30$ and $X = 70$. OLS = ordinary least squares.

[a] X-error is measured as the ratio σ_{xx}/σ_{yx}, expressed as a percentage, in which σ_{xx} represents the error in individual X-measurements.

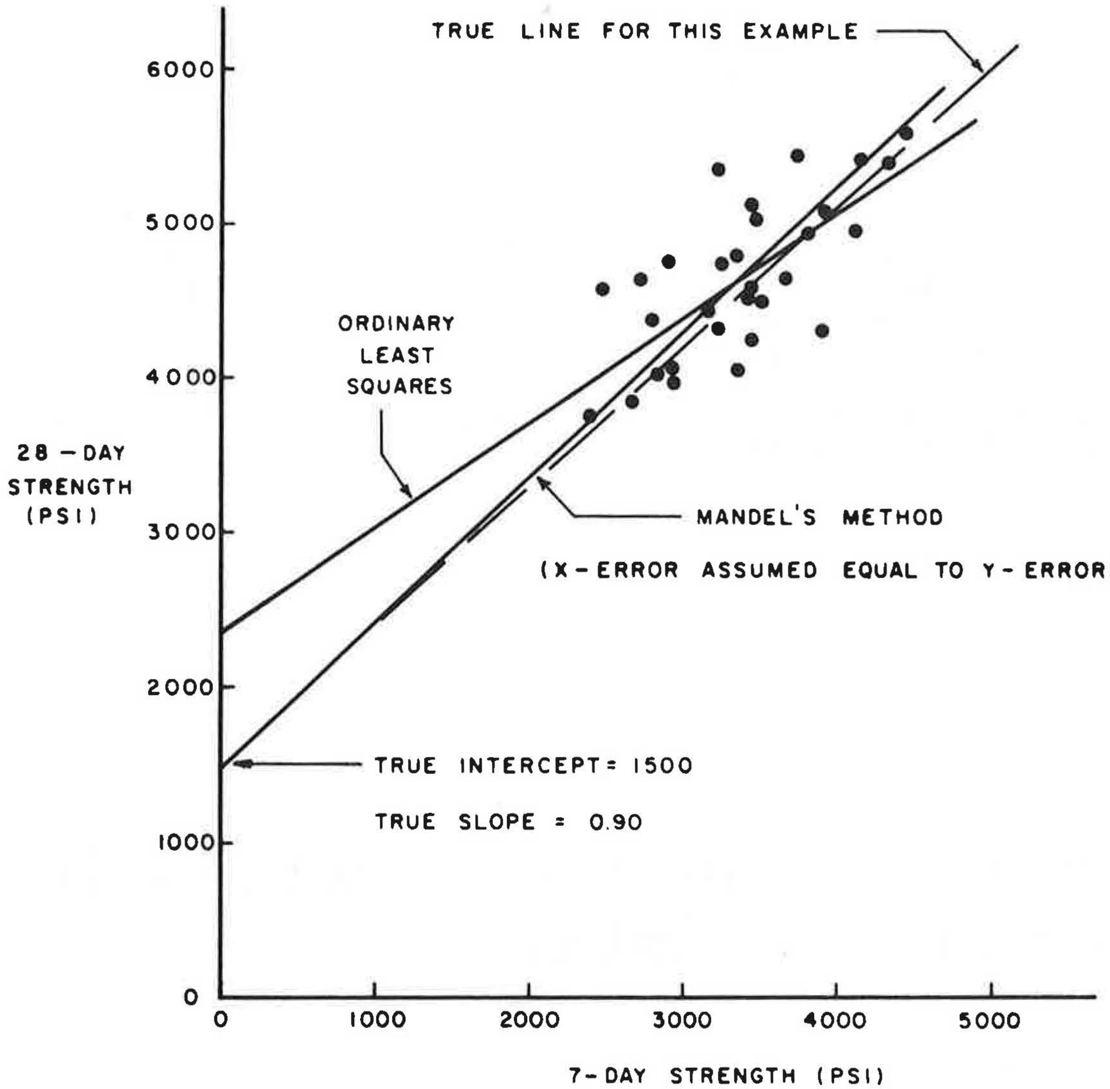


FIGURE 4 Typical regression results with concrete strength data.

EXAMPLE BASED ON CONCRETE STRENGTH DATA

The following example is based on concrete strength data collected from a construction project in New Jersey. It is a contrived example in that the data set that is used was randomly generated from a population having the same statistical parameters as those observed in the field. This approach provides a known control against which the results obtained by the two methods may be compared. Otherwise, it could only be observed that the results obtained by the two methods were distinctly different and it would not be known how close either one came to estimating the true population parameters.

In order to use Mandel's procedure, the ratio of the X- and Y-error variances must be known or assumed. Because both the 7-day and 28-day strengths are at relatively high levels, it has been assumed that the measurement error is the same for both sets of data. Therefore, the variance ratio to be entered into the Mandel procedure is 1.0.

The data points and regression results are shown in Figure 4. These results are typical in that they are examples of central values of the distributions shown in Figure 3. Like the dimensionless examples in Table 4, ordinary least squares has produced a considerably biased estimate, whereas Mandel's method has produced an estimate very close to the true location of the line.

SUMMARY AND CONCLUSIONS

Regression analysis is frequently used in the engineering profession to develop mathematical models for many different applications. In regression theory the X-values are assumed to be known without error, a requirement that often cannot be met. Computer simulation was used to demonstrate that under certain relatively common conditions, regression estimates obtained by ordinary least squares can be seriously in error.

Conditions that appear to warrant concern are (*a*) *X*-error approaching the level of *Y*-error combined with moderate degrees of slope or (*b*) lesser degrees of *X*-error combined with greater degrees of slope.

It was also demonstrated by computer simulation that Mandel's method, which might be termed "oblique least squares" because of the manner in which it minimizes the sum of squared residuals, is extremely effective at removing most of the bias introduced by error in the *X*-variable. Figure 3 and Tables 4 and 5 clearly show that, in general, Mandel's method provides substantially more accurate results than ordinary least squares and Figure 4 illustrates this fact with a specific example based on concrete strength tests. The complete theoretical development, along with a more quantitative guideline to determine when it is advisable to use it, is contained in the original source document (*3*). The FORTRAN coding necessary to apply the procedure is contained in the project report (2).

REFERENCES

1. J. Neter, W. Wasserman, and M. Kutner. *Applied Linear Regression Models*. Richard D. Irwin, Homewood, Ill., 1983.
2. R. T. Barros and R. M. Weed. *Applied Regression in the Presence of X Error*. FHWA, U.S. Department of Transportation, 1987.
3. J. Mandel. Fitting Straight Lines When Both Variables Are Subject to Error. *Journal of Quality Technology*, Vol. 16, No. 1, 1984, pp. 1–14.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.

Validation of a Nonautomated Speed Data Collection Methodology

FRED R. HANSCOM

The objective of this research was to develop a validated spot speed study procedure that does not rely on automated equipment. The field study procedure applied a variety of speed collection techniques and compared results against baseline speeds obtained with reliable pavement instrumentation. A recommended manual-timing technique was based on observed accuracies with various vehicle-selection strategies, site conditions, sample sizes, observation period lengths, and observer characteristics.

The conduct of spot speed studies with nonautomated equipment involves a variety of methodological considerations (*1*). Although such studies have long been used in traffic engineering, a number of factors have hampered their valid application (*2*). Among these factors are observer vehicle-selection bias (e.g., the human ability to select a truly random sample), impact of vantage point (e.g., cosine error associated with radar measurement), technique reliability (e.g., stopwatch timing measurement error), and observer human factors (e.g., experience, fatigue).

The objective of this research was to address the effects of the foregoing factors in order to develop a spot speed data collection procedure that does not rely on automated equipment. The field study procedure involved applying a variety of speed collection techniques and comparing results against baseline speeds obtained with reliable pavement instrumentation. A recommended manual timing technique was based on achieved accuracies with various vehicle-selection strategies, site conditions, sample sizes, observation period lengths, and observer characteristics.

Transportation Research Corporation, 2710 Ridge Road, Haymarket, Va. 22069.

VEHICLE-SELECTION STRATEGIES

Basic Application

Specific techniques were evaluated that controlled observer bias in selecting vehicles for speed measurement. Thus, two speed collection methods (radar and manual timing) were applied by using the following vehicle-selection strategies:

1. *Subjective random (all vehicles)*: The observer designates vehicles that appear to be traveling at a speed representative of overall traffic characteristics. Observer instructions are merely to collect a "random, representative" sample of vehicle speeds.

2. *Subjective random (free-flow vehicles)*: The observer designates vehicles that appear to be traveling at a speed representative of overall traffic characteristics and that appear sufficiently isolated in the traffic stream that drivers can select their desired speeds. Observer instructions are to designate "vehicles in which drivers can select their own speeds, unimpeded by other vehicles."

3. *Systematic (Nth vehicle)*: The observer designates vehicle arrivals at some predetermined interval. Example observer instructions are to collect speeds "on every tenth vehicle."

4. *Randomized (vehicle arrival time)*: The observer uses a scientific random time generator to designate times at which the next vehicle arrival will be selected for speed measurement. The applied technique was to program a hand-held computer to wait a random time (e.g., ranging from 5 to 15 sec) and then to instruct the observer to measure the speed of the next vehicle arrival.

5. *Randomized (designated vehicle)*: The observer uses a scientific random procedure (e.g., modified random number table) to designate vehicle arrival. The applied technique was to program a hand-held computer to randomly select a vehicle (e.g., ranging from the first to the fifth vehicle arrival) for speed measurement.

6. *Subjective platoon weighting*: The observer measures speed for the lead vehicle in a platoon and weights this speed by the total number of platooned vehicles. When radar is applied, this method is known as the radar-platoon technique (*3*).

Speed measurement was conducted with the foregoing sampling techniques on roadway sections instrumented with the Traffic Evaluator System (TES) as a source of baseline data against which to establish the reliability of each technique. [TES is a large-scale data acquisition system developed by FHWA. It consists of electronic roadway sensors and recording apparatus designed to retain information on all passing vehicles. Its accuracy has been established in previous research (*4*).] Sufficient samples were obtained to establish statistical confidence of 1 mph or better: sample sizes were approximately 250 vehicles over a period of 2 hr. Both experienced and inexperienced observers applied the techniques.

Average measurement error (i.e., the difference between TES baseline and sample speeds) for each tested vehicle-selection technique is shown in Table 1. The baseline is taken to be the all-vehicle population at the site during each data collection period. (An exception is made for free-flow sampling; the baseline free-flow population in this case comprises only vehicles with headways of 9 sec or greater.)

Close agreement is shown between the traffic baseline speeds and those gathered by each selection technique. No statistical differences were noted for mean, 15th-, or 85th-percentile speeds gathered either by radar or by manual timing. Little difference in accuracy was noted between radar and manual timing. Average all-vehicle error in miles per hour for these two methods for each speed parameter is as follows:

	Mean Speed	*15th Percentile*	*85th Percentile*
Radar	0.4	0.7	0.1
Manual timing	0.4	0.5	0.4

The vehicle-selection strategy for free-flow vehicles proved to be well suited for that specific application. Both manual timing and radar demonstrated very good ability to match all-vehicle free-flow samples, with the following accuracies:

	Mean Speed	*15th Percentile*	*85th Percentile*
Radar	0.4	0.4	0.1
Manual timing	0.1	0.3	0.2

The remaining five strategies were designated to estimate speeds for the all-vehicle population. In order to distinguish among these strategies in terms of accuracy, accuracies are ranked as follows (1 = most accurate to 5 = least accurate):

	Manual Timing		*Radar*	
Strategy	*Mean Speed*	*85th Percentile*	*Mean Speed*	*85th Percentile*
Subjective random	1	3	4	4
Systematic (Nth vehicle)	5	4	5	5
Randomized (vehicle arrival time)	4	2	2	1
Randomized (designated vehicle)	1	1	1	1
Subjective platoon weighting	3	5	2	1

The rankings indicate consistent superiority of the randomized (designated vehicle) strategy, which ranked first (although twice tying with others) as the most accurate to measure both mean and 85th-percentile speeds by using either radar or manual timing. Furthermore, measurement error (average of mean and 85th-percentile measurement differences for both radar and manual timing) indicated the following relative accuracy associated with each technique:

Strategy	*Avg Error (mph)*
Randomized (designated vehicle)	0.2
Randomized (vehicle arrival time)	0.3
Subjective random	0.4
Subjective platoon weighting	0.6
Systematic (Nth vehicle)	0.8

Again, the randomized (designated vehicle) strategy is seen to be slightly superior.

The data conclusively demonstrate that although all tested vehicle-selection strategies produce acceptable (e.g., not statistically different) agreement with baseline traffic speeds, the randomized (designated vehicle) strategy is preferable. The desirability of its use with either radar or manual timing will be

TABLE 1 SPEED MEASUREMENT ERROR ASSOCIATED WITH VARIOUS SAMPLING STRATEGIES, RURAL FREEWAY

	Manual Timing			Radar[a]		
Strategy	Mean Speed	15th Percentile	85th Percentile	Mean Speed	15th Percentile	85th Percentile
Subjective random (all vehicles)	−0.2	−0.5	−0.5	−0.6	−1.1	−0.4
Subjective random (free-flow vehicles)	+0.1	+0.3	−0.2	−0.4	−0.4	−0.1
Systematic (*N*th vehicle)	−0.7	−0.7	−1.1	−0.7	−0.4	−0.9
Randomized (vehicle arrival time)	−0.6	−1.1	+0.3	−0.2	+0.8	−0.2
Randomized (designated vehicle)	−0.2	+0.2	−0.2	−0.1	−1.0	−0.2
Subjective platoon weighting	−0.5	−0.4	−1.3	+0.2	−0.3	−0.2

NOTE: Measurement error is in miles per hour.

[a]Corrected for cosine error.

addressed in a subsequent section dealing with applied vehicle-selection strategies in varied highway settings.

Lane Specificity

One variation to the applied vehicle-selection strategies just discussed was to designate vehicles by lane for speed measurement. The underlying rationale for this procedure was an attempt to account for the fact that vehicles in the right-hand lane tend to travel more slowly than do vehicles in the left lane. The applied procedure involved making a lane-specific volume count immediately before commencing speed observation. Thus, a hand-held computer was programmed to randomly select vehicles by lane: the proportion of selected vehicles in each lane was based on observed lane occupancy. This lane specificity selection option was applied to both randomized (vehicle arrival time) and randomized (designated vehicle) strategies. A comparison of measurement error (baseline all-vehicle versus selected sample) is as follows:

	Lane Specific		*Not Lane Specific*	
Strategy	*Mean Speed*	*85th Percentile*	*Mean Speed*	*85th Percentile*
Vehicle arrival time	−0.3	0.3	−0.6	+0.3
Designated vehicle	−0.4	0	−0.2	−0.2

A slight overall improvement in accuracy was found with lane-specific vehicle selection for both tested strategies. Average 85th-percentile speed error was 0.15 mph (versus 0.25 mph) and average mean speed error was 0.35 mph (versus 0.40 mph) when the lane specificity option was applied. However, this improvement is so slight (and not statistically significant) that it, in and of itself, cannot constitute a basis for using lane-specific vehicle selection on an operational basis.

A recommendation regarding lane-specific vehicle selection (as opposed to random arrivals regardless of lane presence) must consider the operational application of this procedure and its trade-offs against the potential gain in accuracy. For this reason, application of lane-specific selection cannot be justified in view of the insignificant demonstrated increase in accuracy. The following operational considerations provide the basis for this recommendation.

First, the lane-specific vehicle-selection procedure is time consuming and cumbersome to initiate in the field. A volume count must first be gathered and entered into the hand-held computer. Further, operation of the hand-held computer would be encumbered by the more complex procedure and programming required to accommodate situations of varied lane number. Second, and more important, data collection with the lane-specific selection option is much more time consuming and thus reduces the overall data collection efficiency. This is especially true under low to moderate volume conditions where long intervals exist between vehicle arrivals in the left lane. Greater statistical accuracy can be expected because of the larger sample obtained, within a given time frame, when a straightforward random arrival selection technique is applied.

VARIED SITE CONDITIONS

Limited validation of speed collection techniques was conducted across a variety of site conditions. The purpose of this activity was to examine the possible effect of differing highway conditions (e.g., available vantage point) on speed observation results. Speeds were collected by using radar and manual timing at four site types: urban four lane, urban two lane, rural Interstate, and rural two lane. The vehicle-selection strategy applied at each site was subjective platoon weighting.

Measurement error (difference between platoon-weighted sample and all-vehicle population and difference between lead vehicle sample and free-flow vehicle population) obtained at each site type with both radar and manual timing is shown in Table 2. (Recall that subjective platoon weighting involves measuring lead vehicle speed and weighting this value by the number of vehicles in the platoon.)

Results shown in Table 2 indicate reasonably small measurement error for all-vehicle speed estimation when the platoon-weighting technique is applied with radar as the speed collection method. This average error, across sites, is 0.5 mph for mean speed and 0.7 mph for 85th-percentile speed: a statistical match with baseline speed was achieved under all site conditions. Another tested application of speed collection techniques across sites was to estimate free-flow speed parameters based on the lead vehicle sample used for platoon weighting. Again, when radar was applied as the speed collection method, the technique was seen to work fairly well. (Across sites, average

TABLE 2 SPEED MEASUREMENT ERROR ASSOCIATED WITH SUBJECTIVE PLATOON WEIGHTING AT FOUR SITE TYPES

	Urban, Four Lanes		Urban, Two Lanes		Rural Interstate		Rural, Two Lanes	
Sample	Mean Speed	85th Percentile	Mean Speed	85th Percentile	Mean Speed	85th Percentile	Mean Speed	85th Percentile
Platoon weighted								
Radar	−0.9	−1.0	−0.7	−1.1	+0.2	−0.2	+0.2	−0.6
Manual timing	−2.5[a]	−2.9[a]	−0.8	−1.7[a]	−0.5	−1.3	−3.8[a]	−4.0[a]
Lead vehicle								
Radar	−0.5	−1.4	0.3	−1.7[a]	−0.4	−0.9	0	−1.3
Manual timing	−1.9[a]	−2.9[a]	−0.6	−1.9[a]	−1.1	−1.4	−3.8[a]	−4.9[a]

NOTE: Measurement error is in miles per hour.
[a]Statistically significant ($\alpha = .05$).

errors were 0.3 mph for mean speed and 1.3 mph for 85th-percentile speed.) A statistically different 85th-percentile speed (1.7 mph error) at the urban two-lane site was likely due to small sample size (e.g., 137 lead vehicles).

Manual timing as a speed collection method was not shown to be reliable under all tested conditions. Statistical differences were evident at three of the four sites that did not have elevated observer vantage points. However, somewhat promising results (e.g., mean speed error of 0.7 mph) were found at the urban two-lane site, where the observer was standing at street level. As noted earlier, the significant (1.9 mph) error in 85th-percentile speed may be due to sampling conditions. A detailed evaluation (i.e., a vehicle-by-vehicle error determination) was not possible to fully assess the maintenance of manual timing speed measurement accuracy under this condition.

In summary, radar and manual speed timing methods using the platoon-weighting technique were applied at four site types: urban two- and four-lane highways, rural Interstate, and rural two-lane highways. Radar platoon weighting demonstrated good results across site conditions. All-vehicle population speeds were estimated by using this technique with the following average accuracies: mean speed, 0.5 mph; 85th-percentile speed, 0.7 mph. Radar sampling of lead vehicles was shown to estimate mean free-flow speeds with an accuracy of 0.3 mph. Manual timing was not shown to be reliable at sites without elevated observer vantage points.

RELIABILITY OF MANUAL SPEED TIMING

Stopwatch timing is a frequently applied manual method of speed measurement. In order to examine the accuracy of this technique, the following studies were conducted: (*a*) vehicle-by-vehicle comparison of manually timed speeds with those obtained from a commercial speed-monitoring device, (*b*) intercoder reliability study comparing between-observer results, and (*c*) minimum required observation period.

Vehicle-by-Vehicle Comparisons

Manually timed speeds from two coders were compared on a vehicle-specific basis with results obtained from a commercial automated speed-monitoring device. The applied manual technique used a hand-held computer configured as an electronic stopwatch to time vehicles between two pavement markers spaced 300 ft apart. One coder was relatively inexperienced (but highly motivated), with approximately 2 days of previous speed data collection experience; the second was the project principal investigator, who had considerable speed data collection experience.

The following speed parameters were compared on a vehicle-by-vehicle basis for the automated device (pavement loop) and manual collection (stopwatch timing): mean speed, 15th percentile, and 85th percentile. Average difference (i.e., error between techniques) for each observer is shown (in miles per hour) in Table 3. A minus sign indicates that speed obtained manually was slower than that obtained with the automated device. Sufficient sample sizes were obtained in each trial to establish mean speed confidence (at the 0.01 level) within 1.0 mph. Sample sizes ranged from 215 to 241 vehicles per trial.

Error associated with individual vehicle speed measurements was also examined. Results obtained for each observer are summarized in Table 4. Although individual vehicle measurement errors were shown to be surprisingly large (e.g., approximately 40 percent exceeded 1 mph), the errors were shown to be largely compensating in nature as evident from corresponding mean speed differences between techniques, which ranged from 0.1 to 0.9 mph. The resulting assessment of the manual speed-timing procedure is that the method produced mean data

TABLE 3 AVERAGE DIFFERENCE BETWEEN SPEEDS TIMED MANUALLY AND BY AUTOMATED DEVICE

	Mean Speed		15th Percentile		85th Percentile	
	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
Inexperienced	−0.1	+0.1	+0.6	+0.1	−1.0	−0.1
Experienced	−0.9	−0.3	−1.2	+0.3	−0.5	−0.2

TABLE 4 ERROR BY EXPERIENCE OF CODER

	Avg Error[a]	Maximum Error	Percent Error by Speed (mph)				
			>5	>4	>3	>2	>1
Inexperienced	1.1 ± 0.14	11.4	1.6	2.6	9.3	19.5	46.9
Experienced	1.4 ± 0.22	12.8	1.1	1.7	6.6	11.8	37.9

[a]0.05 confidence interval.

statistically equivalent to speeds gathered with the automated speed-monitoring device. However, as noted in the previous section, manual speed timing can be considered reliable only in those highway settings that provide an elevated vantage point.

Intercoder Reliability

To determine intercoder reliability, two observers manually timed speeds for specifically selected vehicles. Matched vehicle-by-vehicle speed data were then analyzed to examine agreement between coders. The operational application of the intercoder reliability study is its use as a training aid.

In this experiment, two intercoder reliability studies were undertaken. The first, conducted by the research team, consisted of a traffic engineer with considerable speed collection experience and an assistant with only one day of previous experience. The second study involved two FHWA employees, both of whom were familiar with speed collection procedures. In each study differences (in miles per hour) between observers were found not to be significant (Table 5). Error between coders in each of the studies is shown in Table 6.

Despite the relatively large magnitude of individual measurement differences, results indicate close overall between-coder agreement (approximately 0.5 mph mean speed difference in both tests). The larger sample obtained in the TRC study resulted in additional opportunity for larger individual speed measurement differences (thus explaining the maximum measurement error of 18.3 mph). However, as shown in the previous section, manual timing speed errors were seen to be compensating (i.e., approximately equal in both positive and negative directions). This is further substantiated by the fact that two-thirds of between-coder speed measurements differed by more than 1.0 mph, yet averaged speeds differed by only 0.5 mph.

These tests provide results of a procedure to assess speed measurement ability between observers of varying skill levels. Two precautions must be noted. First, observers are aware that results are being monitored and may therefore perform with more vigilance. Second, manual speed timing must be conducted from an elevated vantage point. Nevertheless, these intercoder reliability studies demonstrate comparable between-observer results for manual speed timing.

TABLE 5 MEASUREMENT DIFFERENCE BY CODING TEAM

	Observer	Sample Size	Mean Speed	Standard Deviation	85th Percentile
TRC	1	200	57.5	4.80	61.9
	2	200	57.0	4.49	61.3
FHWA	1	40	58.6	3.76	62.7
	2	40	59.1	3.53	63.0

NOTE: TRC = Transportation Research Corporation.

TABLE 6 ERROR BY CODING TEAM

	Avg Error[a]	Maximum Error	Percent Error by Speed (mph)	
			>5	>1
TRC	0.49 ± 0.78	18.3	28	66
FHWA	0.54 ± 0.64	6.5	20	68

[a]0.05 confidence interval.

Observation Period

In order to assess the suitability of manual timing to estimate speeds in a one-time study, comparisons were made between sampled speeds and the all-vehicle population for a variety of observation conditions (e.g., length of observation period, time of day, and previous experience). Period durations (10, 20, and 45 min) were randomly ordered throughout each of the two data collection days. In addition, systematic scheduling ensured that both long periods (45 min) and short periods occurred both early and late on different days as a check on coder fatigue. Ten-minute rest breaks were taken between each data collection period, and a 1-hr lunch break was taken at midday. Two observers, one experienced and one inexperienced, participated in this experiment.

Speed measurement accuracy was determined by comparison of manually timed speeds for each observer with an all-vehicle baseline consisting of TES data for each collection period. Summary results contrasting mean speed error (difference in miles per hour between TES data and manual timing speeds) for the experienced and inexperienced observers are as follows:

Observation Period (min)	*Observer*	
	Experienced	*Inexperienced*
10	0.5	2.6
20	0.9	1.3
45	0.4	0.7

Data collected by the inexperienced coder (who regrettably exhibited a lackluster motivation) demonstrated a distinct error effect associated with period duration. The results from the 10-min observation period for the experienced coder indicated surprisingly close agreement between manually coded and TES

TABLE 7 RESULTS FROM 45-MIN OBSERVATION PERIOD

Trial No. and Data Source	Sample Size	Mean Speed	SD	15th Percentile	85th Percentile	95th Percentile	Period of Day	Result
1								
Coder	94	57.4	4.0	54	61	63	1	No statistical differences
TES	171	58.0	4.3	54	62	64		
2								
Coder	81	57.9	5.5	52	63	65	2	Coder mean speed low by 1.5 mph
TES	144	59.4	4.8	54	64	66		
3								
Coder	112	58.8	6.2	54	63	67	4	Coder variance high
TES	208	58.8	4.5	55	63	66		
4								
Coder	103	57.7	4.4	53	62	65	5	Amazing
TES	166	57.5	4.5	53	62	65		
5								
Coder	88	56.3	5.3	53	61	63	5	No statistical differences
TES	168	57.0	4.9	51	62	63		
6								
Coder	91	56.7	6.3	52	62	65	8	No statistical differences
TES	179	56 7	5.7	52	62	64		
7								
Coder	144	58.1	4.7	54	62	64	8	No statistical differences
TES	440	58.1	4.1	54	62	64		
8								
Coder	115	57.7	5.4	53	62	66	10	Coder variance high
TES	323	57.7	4.2	53	61	65		

(all-vehicle) speeds. During each 10-min period, the coder measured speeds on samples ranging in size from 20 to 37 vehicles. This sample represented between 38 and 69 percent of the total vehicle population measured by TES data. Subsequent 20- and 45-min periods resulted in similar sampling percentages. The 20-min period data resulted in a lesser degree of mean speed accuracy. However, the results from the 20-min period showed improved agreement in measured speed variance (no statistical difference).

As expected, closer overall agreement was obtained between TES and coder speeds (both means and all selected percentiles) during the 45-min observation periods. Examination of results from eight individual periods (Table 7) indicates that although statistical differences were found during three trials, a minimal effect was realized in terms of measurement error magnitude. The single incidence of significantly different mean speed was 1.5 mph. The average mean speed error was 0.36 mph. An examination of the raw data indicated that the mean 85th-percentile speed was in error by only 0.45 mph.

The impact of observer fatigue was approached by using observation period duration as a surrogate. The appropriateness of this surrogate lies in the fact that tested conditions represent time requirements to gather statistically suitable samples. With this approach, the effect of fatigue was examined by two procedures. First, within-period fatigue was examined for the data from the 45-min period; yet no degradation in accuracy was found for speed measurements obtained late in any specific period. Second, mean speed error (all-vehicle versus sample difference) demonstrated a trend for less error to occur later in the day. Ranked period-specific mean measurement errors associated with time of day are as follows (Period 1 begins at 9:00 a.m.; Period 10 ends at 5:00 p.m.):

Error (mph)	*Period*
0	8
0.1	9
0.3	6
0.4	10, 1
0.8	5
0.8	3
0.9	7
1.1	4
1.3	2

Results of this experiment indicated that although 45 min is the minimum acceptable period duration, specifying period duration alone does not ensure an adequate sampling requirement. Both observation duration and sample size must be specified. Therefore, sample-size effects were studied next.

DETERMINATION OF MINIMUM SAMPLE SIZE

In this experiment the suitability of small spot speed samples to estimate all-vehicle speed populations was investigated. The objective was to determine minimum sample requirements in order to optimize manpower and financial resources without sacrificing statistical integrity of the study.

The applied procedure involved comparing results obtained with varied sample sizes versus results from the all-vehicle population. Two days of speed observation were applied at a rural Interstate site during hours of uniform traffic flow. Subsamples consisting of 10, 20, 50, 100, and 200 vehicles were randomly selected from the all-vehicle population. Five iterations (random selections) were conducted in each sample size category. Samples were extracted from specific durations (e.g.,

TABLE 8 AVERAGE AND MAXIMUM SPEED MEASUREMENT ERROR FOR VARYING SAMPLE SIZES

Sample Size	Mean Speed Day 1	Mean Speed Day 2	85th Percentile Day 1	85th Percentile Day 2
Average Error				
10	0.8	0.5	1.0	0.8
20	0.2	0.3	0.4	0.6
100	0.1	0.1	0.1	0.2
200	0.1	0.1	0.1	0.2
Worst-Case Error				
10	3.0	2.8	5.1	5.4
20	1.3	3.1	3.0	2.9
50	0.9	1.1	1.3	2.3
100	0.8	0.7	0.6	0.7
200	0.5	0.6	1.0	0.9

NOTE: N = 50 observation periods for each sample size.

a half-hour) in the database so as to represent operational data-gathering periods. A total of 50 observation trials were made for each tested sample size.

No statistical differences (α = .05) were found between samples and population mean speeds. In certain instances, standard deviations differed for samples of 10, 20, and 50 vehicles. Speed measurement error (i.e., all-vehicle populations versus sample groups) is summarized in Table 8. Average mean and 85th-percentile speed differences are shown in the top portion of the table. These averages represent magnitude of error without regard to direction (i.e., a +1.0-mph error and a −1.0-mph error would average to 1.0 mph). The results in the upper portion of the table imply that very good results were obtained with relatively small sample sizes. That is, average precision of better than 1.0 mph was achieved with sample sizes as small as 20 vehicles.

However, in order to examine the maximum sampling error likely to be associated with each sample size, the worst-case difference from all 50 trials within each size category is shown in the lower portion of the table. These results indicate that sample sizes of 10 to 50 vehicles can result in mean or 85th-percentile speed sampling errors ranging from 0.9 to 5.4 mph. However, a sharp reduction in error was noted for 85th-percentile speeds as sample size increased from 50 to 100 vehicles. A further increase to 200 vehicles did not yield any real benefit. Thus, maximum expected measurement error associated with a random sample of 100 vehicles was shown to be 0.75 mph for mean speed and 0.65 mph for 85th-percentile speed. Results indicate that under uniform flow conditions (e.g., during non-rush periods), mean and 85th-percentile speeds can be measured with an accuracy of better than 1.0 mph if two sampling minimums are met: a 45-min observation period (as seen from the previous section) and a sample of 100 vehicles.

EFFECT OF OBSERVER EXPERIENCE

Emphasis in this research was placed on the relative accuracy achieved with tested techniques used by experienced versus inexperienced observers. Each of the foregoing vehicle-selection strategies was applied by both an experienced observer (i.e., a traffic engineer with 14 years' experience) and an inexperienced observer (i.e., part-time personnel with short training session) using both radar and manual timing techniques. Four inexperienced observers were used; the same experienced observer conducted data collection for each tested technique as a basis for comparison.

No significant effect in radar application was noted as a function of observer experience. Table 9 gives the manual speed timing accuracies associated with the inexperienced observers. Two speed measurement criteria were applied: (*a*) difference between speed results coded by inexperienced observers and actual speeds of the vehicle population, and (*b*) difference between results of the experienced versus the inexperienced coders. Relative error is shown for each inexperienced coder, averaged across all trials. Three women and one man were used as the inexperienced observers; ages ranged from 25 to 39. Measurement differences for each observer are ranked in the table according to age. Results generally indicate that improved accuracy is associated with greater motivation, as was shown by the results for the two older observers. Mean speed measurement errors recorded by the inexperienced coders ranged from 0.1 to 2.0 mph; differences between the experienced and inexperienced observers ranged from 0 (exact agreement) to 1.4 mph. In one case, however, the inexperienced coder produced more accurate results than did the experienced coder.

TABLE 9 COMPARATIVE MANUAL-TIMING SPEED MEASUREMENT ACCURACIES FOR INEXPERIENCED OBSERVERS

Coder and Age	Inexperienced Coder Versus Actual: Mean Speed	Inexperienced Coder Versus Actual: 85th Percentile	Experienced Versus Inexperienced Coder: Mean Speed	Experienced Versus Inexperienced Coder: 85th Percentile
Eleanor, 25	2.0	1.5	1.4	1.0
Barbara, 27	1.5	1.9	0.9	0.5
Dave, 35	0.1	0.6	−0.5[a]	0.3
Carol, 39	0.3	0.5	0	0.4

[a]Minus sign indicates superior performance by comparison with experienced coder.

The interpretation of these results leads to the following conclusion regarding observer experience and its effect on manual speed timing accuracy. Although generally improved results were associated with age (i.e., observers in their thirties demonstrated improved results in comparison with those in their twenties), no consistent difference was noted between male and female coders. The field experience during this research demonstrated two important factors. First, motivation is more significant than specific observer characteristics in determining suitability for this task. Those personnel who demonstrated a serious attitude and who appeared to genuinely want to do the work proved to be more accurate in their results.

Second, intercoder reliability trials (i.e., vehicle-by-vehicle data comparisons between observers) are essential in order to predetermine the suitability of any employee to conduct manual speed timing. In the case of a motivated observer, one 2-hr training session is likely to be sufficient to provide needed experience. A second training session, conducted on a different day, is recommended to control for within-observer performance variation.

SUMMARY OF RESULTS

A field evaluation of speed data collection techniques was conducted by comparing actual traffic speed characteristics with those measured with the following procedural variations. Six vehicle-selection strategies were tested in order to eliminate observer bias. These strategies included subjective, systematic, and random vehicle-selection procedures. Lane-specific vehicle selection was also tested. The reliability of two methods (radar and manual timing) using the platoon-weighting technique was assessed on four highway types: rural Interstate, rural two lane, urban four lane, and urban two lane. The effect of observer experience (age and practice) was examined. Relative precision for spot speed measurement was determined for a variety of observation period effects (e.g., duration, time of day, and within-period observer fatigue). Spot speed sampling accuracies were determined for minimum cost-effective sample sizes.

Results of the series of field studies are as follows:

1. Six vehicle-selection strategies were tested in order to eliminate observer bias: subjective, systematic, computer-assisted random, and platoon-weighted procedures using both radar and manual timing methods. Lane-specific vehicle selection was also tested but was determined not to be beneficial. All the tested strategies yielded results that were statistically equivalent to real traffic speeds. However, the randomized (designated vehicle, not lane specific) strategy consistently proved best and resulted in mean and 85th-percentile speed error of 0.2 mph or less.
2. The reliability of two methods (radar and manual timing) using the platoon-weighting technique was assessed on four highway types: rural Interstate, rural two lane, urban four lane, and urban two lane. Radar produced the following accuracies: mean speed, 0.5 mph, and 85th-percentile speed, 0.7 mph. However, manual timing was not shown to be reliable in highway settings that do not afford an elevated vantage point.
3. The accuracy of manually timed speed observation was determined from vehicle-by-vehicle comparisons with an automated speed collection device. Despite considerable vehicle-specific error (i.e., approximately 40 percent of the measurements were in error by 1.0 mph or more), these errors were largely compensating in nature. Averaged trials for two observers resulted in sample means and 85th-percentile speeds within 0.5 mph accuracy.
4. Relative precision for spot speed measurement was determined for a variety of observation period effects (e.g., duration, time of day, and within-period observer fatigue). Results showed that a minimum 45-min observation period is required but that no accuracy degradation due to fatigue during this time is expected. Rest periods throughout the day resulted in no manual speed timing accuracy reduction at the end of an 8-hr day.
5. Spot speed sampling accuracies were determined for minimum cost-effective sample sizes. A minimum of 100 vehicles is required for mean speed accuracy of 0.5 mph and 85th-percentile speed accuracy of 1.0 mph.

The product of this series of field studies is a recommended manual technique for speed determination. Manual observation is suggested for applications such as assessment of traffic control device effectiveness and other uses where continuous speed monitoring with automated equipment is not feasible. Application of manual procedures developed in this series of field experiments was determined to yield mean speeds accurate to 0.5 mph at the 0.01 confidence level.

The recommended manual speed collection method consists of the following procedure:

1. Speed-timing personnel should be trained with at least two intercoder reliability trials (on separate days), requiring mean agreement between coders of 0.5 sec or better for individual speed measurements.
2. Speeds should be clocked by using an electronic stopwatch capable of measuring and displaying time to an accuracy of 0.01 sec.
3. Overhead observation points, such as overpasses, should be used.
4. Speed-timing markings should be painted on the pavement at a minimum spacing of 270 ft.
5. A minimum observation period of 45 min and total sample size of 100 vehicles should be used.
6. Observations should be conducted at times of day exhibiting stable speed conditions (e.g., only rush or only nonrush conditions).

ACKNOWLEDGMENT

The work reported in this paper was conducted under a contract with FHWA. The contract monitor was Davey Warren.

REFERENCES

1. F. R. Hanscom. *An Assessment of Speed Measurement Techniques and Practices: A Literature Review*. FHWA, U.S. Department of Transportation, 1983.
2. W. J. Homburger and J. H. Kell. *Fundamentals of Traffic Engineering*. Institute of Transportation Studies, University of California, Berkeley, 1977.
3. M. Van Aerde and S. Yagar. Radar-Platoon Technique for Recording Speeds and Volumes. In *The Effects of Speed Limits on Traffic Accidents and Transport Energy Use*, Organization for Economic Cooperation and Development, Paris, 1981.
4. G. S. Kolsrud. *Diagrammatic Guide Signs for Use on Controlled Access Highways*, Vol. 3. Report FHWA-RD-73-26. FHWA, U.S. Department of Transportation, 1972.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.

Standard Target Contrast: A Visibility Parameter Beyond Luminance To Evaluate the Quality of Roadway Lighting

F. W. Jung and A. Titishov

An additional yardstick to evaluate the quality of fixed highway lighting is proposed, specifically, the contrast of a critical-size standard target object, such as a 20 × 20-cm (8 × 8-in.) square middle-gray card placed vertically on the road surface perpendicular to the road axis. The proposed measure of standard target contrast would be used in conjunction with other currently used yardsticks, such as luminance and glare. The concept of a standard target contrast is demonstrated with reference to a test section of conventionally designed luminaires. Although luminance levels meet the standards for uniformity, spots of unsafe low contrast are clearly revealed by the new yardstick. Contrast has been defined in such a way that values for silhouette vision range from no contrast at zero to a maximum contrast of −1, representing a comfortable range for all practicable evaluation work. The prospect of more effective nonsymmetrical luminaires for one-way traffic is also investigated. It is shown that about half the light output must be widely and uniformly distributed to meet luminance requirements and that the remaining light can be effectively directed toward the driver for enhancing contrast values. Conventional symmetrical luminaires need a distinct overlap of main beams to provide sufficient contrast for silhouette vision at all points along the road.

In the field of roadway and expressway lighting, considerable research has resulted in the adoption of luminance (reflected light) as a design standard for fixed lighting systems. Before this, the methods and standards of design were based on incident light (illuminance) only. Design and evaluation of lighting systems with road surface luminance as an additional standard is just one small forward step toward a true visibility criterion. Performance evaluation of fixed highway lighting is still inadequate, as will be shown in this paper.

At present, research efforts are directed toward finding a visibility index. However, there are difficulties with such an index, because it contains too many transient quantities that are difficult to turn into standard values. In the case of luminance, there were only a few values, such as the viewing angle α, that could still be standardized (although not easily). The visibility index, as it is being discussed now, is apt to be loaded with physical and human factors, and thus it becomes much more difficult to agree on representative or standard values for system parameters.

Research and Development Branch, Ontario Ministry of Transportation and Communications, 1201 Wilson Avenue, Downsview, Ontario M3M 1J8, Canada.

PROPOSED YARDSTICK

The solution proposed here is not to try a perfect modeling or definition of such an index but rather to concentrate on a less sophisticated parameter that can easily be computed at each location on the roadway surface by using only the physical dimensions and properties of the lighting system. Such a parameter can be used in the same way as glare, illuminance, or luminance, namely, as a system evaluation yardstick revealing weaknesses of the system much more clearly than any other design parameter.

The parameter suggested for use is the contrast of a critical-size standard target object on the road surface perpendicular to the line of sight or road axis. Only two quantities of such a target must be defined:

- Reflectance properties (such as middle gray, 20 percent reflectance, and nonspecular perfectly diffusing surface), and
- Height above ground (such as 20 cm), which appears to be less important.

Note that it is possible to use several standard values of reflectivity (say, 20 percent, 50 percent, etc.).

UNDERSTANDING AND INTERPRETING TARGET CONTRAST

To understand the concept of standard target contrast, imagine a square 20 × 20 cm (8 × 8 in.) cut from a Kodak middle-gray card (which has a diffuse surface of 18 percent reflectance). This card is positioned vertically on the pavement surface perpendicular to the road axis and in this position can be moved along the surface forward and sideways. In each position, the contrast of the target object against the background of the pavement surface at a normal driver viewing angle can be determined. The target surface can be specified as a perfect diffuser with a uniform reflectance such as 20 percent or another chosen value. Further, imagine a driver who is 80 to 100 m (about 300 ft) away from the gray card, approaching it as a "critical-size" object situated in his driving path. "Critical size" means that the driver is sufficiently motivated to take evasive action when detecting such an object. Note that the capability of the driver to detect this card in time for evasive action is proportional primarily to the contrast between the card surface and the pavement background. Only secondarily does this capability depend on the level of luminance on the road surface.

In this context, contrast is defined as the difference between the object and background luminance divided by the background luminance, that is, $(L_o - L_b)/L_b$. In this way, a powerful and well-understood additional criterion can be established to measure the quality of roadway lighting. Defined in this way, the contrast can be calculated or measured for each grid point on the road surface in terms of a definite number. The meaning of such a number should be immediately clear. For instance,

- A contrast close to −1 means a strong silhouette effect. This is favorable for detecting objects but not necessarily for recognizing what they are, which fortunately is less important.
- A contrast close to zero means that the target almost disappears, which can be an unsafe condition.
- A large positive number means that the gray card appears bright against a darker background.

It should be noted that the visibility of such targets under fixed roadway lighting is determined predominantly by negative contrast or silhouette effect against the brighter background luminance of the roadway surface.

IMPLEMENTATION

With relatively little effort, the target contrast parameter can be incorporated into existing computer programs for illumination design so that values of contrast (C) can be calculated in the same manner as that currently used for illuminance, luminance, and disability veiling glare.

It is also possible to subscript the contrast parameter C (for example, C_{10} or C_{20}) to denote the percentage of reflectivity of the target object (10 percent, 20 percent, etc.). Further, it is possible to add a term to the vertical target illuminance to take into account the additional illumination of the target object by vehicle headlights. However, headlights are not very effective at 90 m viewing distance and beyond.

According to current knowledge, major highways need an average luminance of 0.5 to 1.2 cd/m^2. It is conceivable that the higher value is needed for situations of poor contrast, whereas the low values of luminance could be adequate in cases of good contrast over the whole surface.

Finally, introducing standards of contrast would promote innovation toward new luminaires that yield more visibility for each unit of consumed energy. This point needs further comment. Fixed lighting with improved visibility on high-speed expressways can be achieved economically by nonsymmetrical luminaires with main beams turned toward the driver and with good cutoff characteristics. Besides a cutoff in the driver's direction between 80 and 85 degrees to minimize glare, there should be little light in the direction of travel, opposite to the driver's view, just enough to provide a minimum level of luminance on the road surface (say, 0.5 cd/m^2), which is important to maximize (negative) contrast. Luminaires with such characteristics are not yet on the market but could well be developed after contrast standards have been introduced.

The current design standards of illuminance and luminance are blocking such development because they are unable to reveal the spots of bad visibility (i.e., of low contrast), as will be shown in the next section. On the other hand, an innovative luminaire as just described would probably violate present uniformity standards, which have little to do with visibility. Thus, introducing a standard of contrast would greatly promote the state of the art of lighting systems design.

A CASE OF CONTRAST DISTRIBUTION MEASUREMENT

Measurements of illuminance, luminance, and contrast of a standard target object were carried out at a test in Ontario with luminaires of adjustable output *(1)*. The pole distance was 59.5 m (200 ft) and the mounting height 15 m (50 ft), using two 400-W high-pressure sodium (HPS) Type III luminaires per pole with 1-m (3.3-ft) overhang. The road width was 10.5 m (34 ft); that is, there were three lanes running east to west. The installation consisted of six poles in one row, with two luminaires installed at each pole and bracket arm (to maximize the range of lighting levels).

The measurement field was chosen in the center, between the third and fourth poles. The results of the measurements, carried out in the center of each lane, are plotted in Figures 1 and 2. Luminance and contrast distributions are shown for a driver who is looking from the right, that is, approaching from the right side of the figures. Figure 1 shows distributions for full illumination and Figure 2 for an illumination output reduced to 14 percent.

The standard target was a 20 × 20-cm middle-gray card of about 20 percent reflectance in a vertical position and perpendicular to the road axis. The contrast values, C_{20} (20 stands for 20 percent), in Figures 1c and 2c have been calculated from luminance measurements of the target object (L_o) and the road surface background (L_b) using the formula $C_{20} = (L_o - L_b)/L_b$. Each figure contains three lines or curves, one for each of the three lanes.

The results shown in Figures 1 and 2 are interpreted as follows:

1. The distribution of luminance in each lane is relatively uniform. This means that the luminance yardstick cannot be used to discover a deficiency in contrast; that is, no weakness in the system can be determined.
2. The standard target contrast is basically all negative, which means that there is silhouette vision almost all along the road section. The object appears dark against a brighter background.
3. There are contrast values that are close to zero somewhere between the entrance point (the fourth pole) and the center (i.e., toward the approaching traffic), which means that critical-size objects may almost disappear at certain spots. Thus, contrast measurements reveal a weakness in the system.

If the evaluation of this fixed lighting system were based on illuminance and luminance only, it would have to be judged satisfactory or fair. Note that the luminance distribution is very uniform. However, contrast values close to zero mean that critical-size objects in the driving path (mufflers, rocks, etc.) may disappear or almost disappear for a short period of time

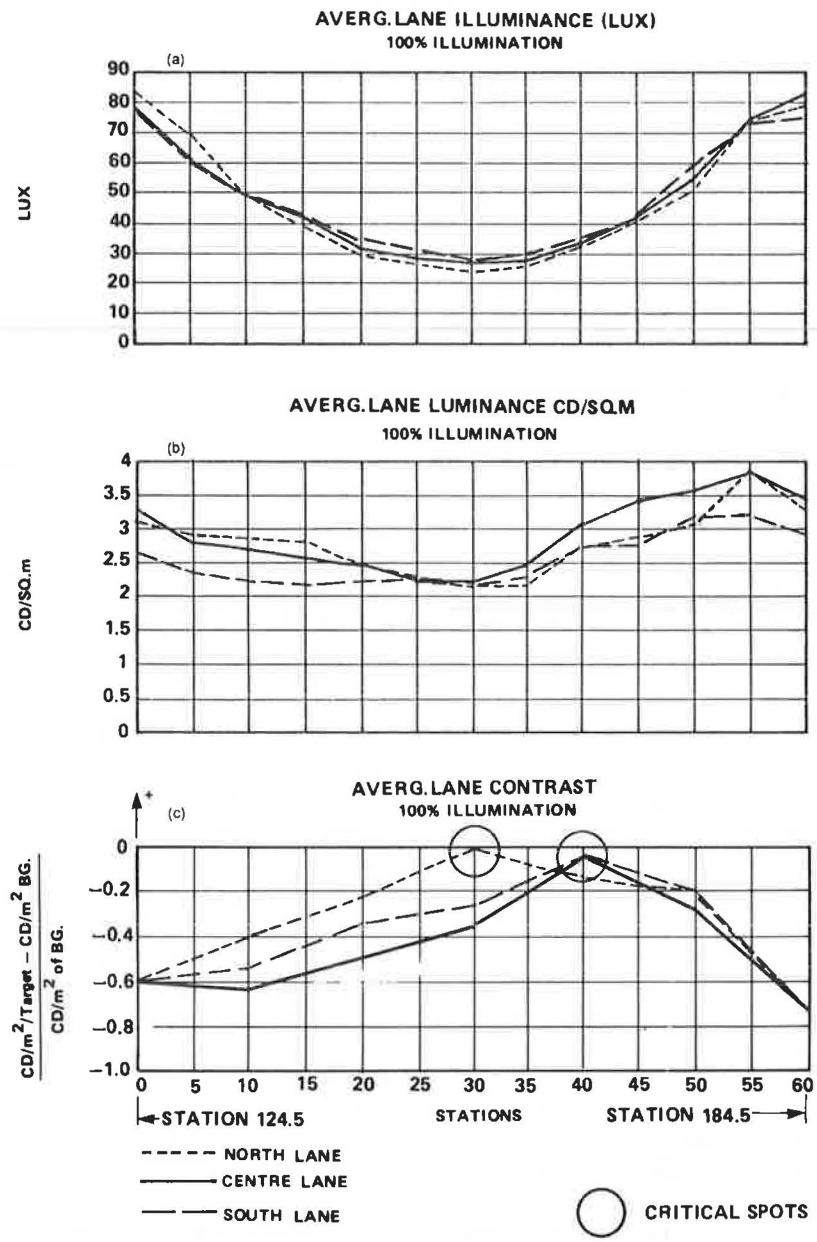


FIGURE 1 Distribution of illuminance, luminance, and standard target contrast for 100 percent output.

while one is driving along. Thus, it is important to look also at the standard target contrast distribution, which should be introduced as another important yardstick for the quality of fixed highway lighting, together with the currently used yardstick of luminance.

Negative contrast means that the target object is darker (has lower luminance) than the road surface in the background. Car headlights, which are effective at distances of less than 90 m (300 ft), increase the luminance on the target object, and thus may reduce the negative contrast. However, this requires further investigation; it is conceivable that headlights may worsen the situation.

Only the contrast diagrams in Figures 1 and 2 show the deficiency of the lighting system: between 30 and 45 m, that is, beyond the center of the section and toward the right in the direction of the approaching driver, contrast values are at or close to the zero line. The illuminance and luminance diagrams do not reveal any deficiency.

PILOT COMPUTATION AND ILLUSTRATION

System Parameters

Illumination systems for highways usually consist of rows of luminaires on poles. The relationship of each pole and luminaire with regard to the point P on the road surface as seen by an approaching driver is shown in Figure 3. Computer programs are available to calculate illuminance and luminance at predetermined grid points (P) for a standard viewing angle of

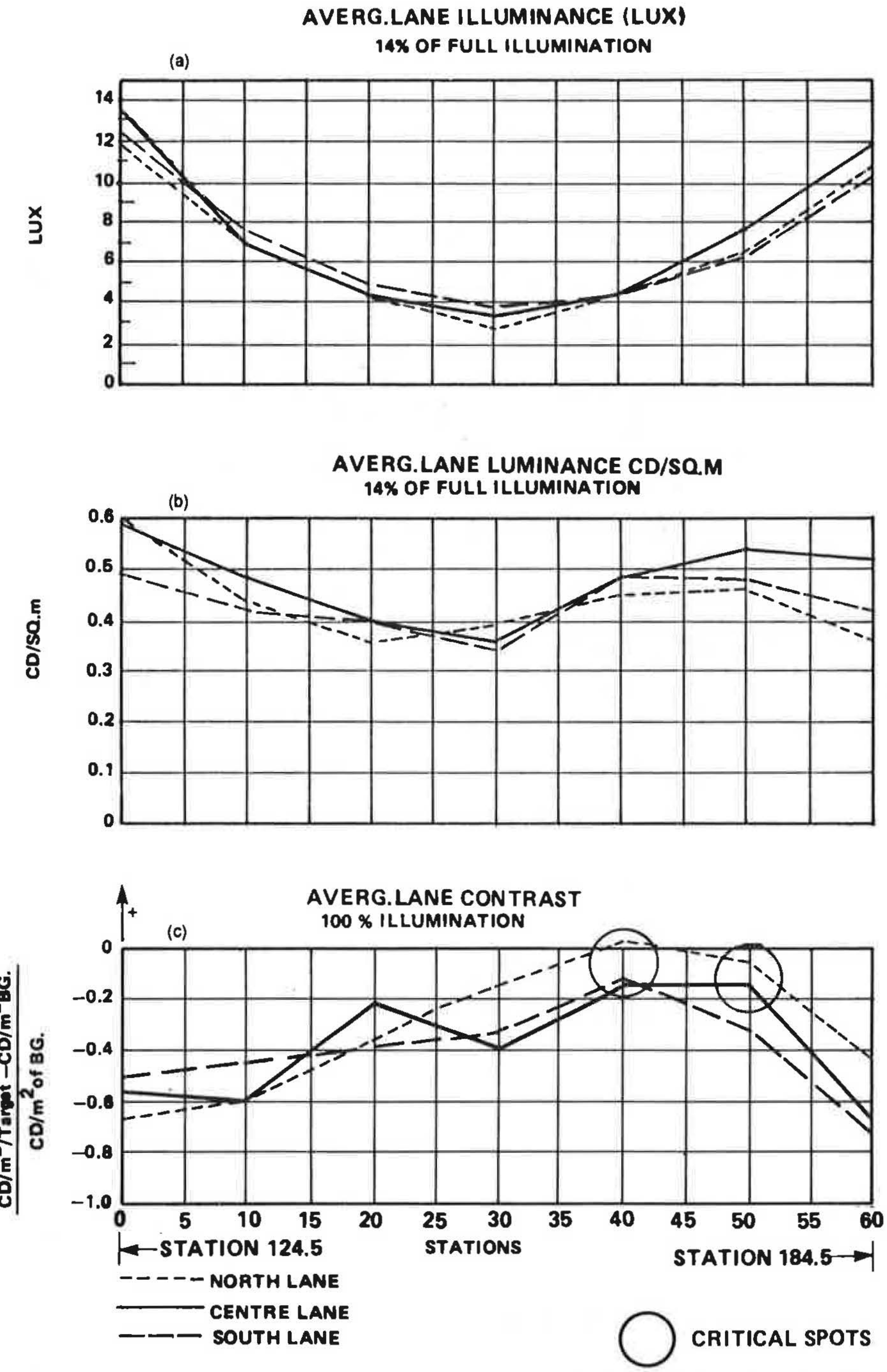


FIGURE 2 Distribution of illuminance, luminance, and standard target contrast for 14 percent output.

an approaching driver (α = 1 degree). Such programs can be extended easily to include the contrast of a standard critical-size object, a middle-gray perfect diffuser with a 20 × 20-cm surface perpendicular to the road axis. The strengths of such a new parameter can be evaluated by the following pilot study.

For illustration, the system is simplified to two dimensions longitudinally along the axis of a straight horizontal road of a few lanes. One row of luminaires is arranged in the center position at a mounting height (H) above the road surface. This system is shown in Figure 4 and may be understood as accommodating either one-directional traffic from left to right or two-way traffic.

Terms used in this discussion denote the following variables:

H = mounting height (m);
S = spacing of luminaires (m);
λ = S/H = ratio of spacing to mounting height;
ζ = fraction of distance between point P and the pole at $i = -1$ in terms of pole spacing;
i = index of location of light poles;

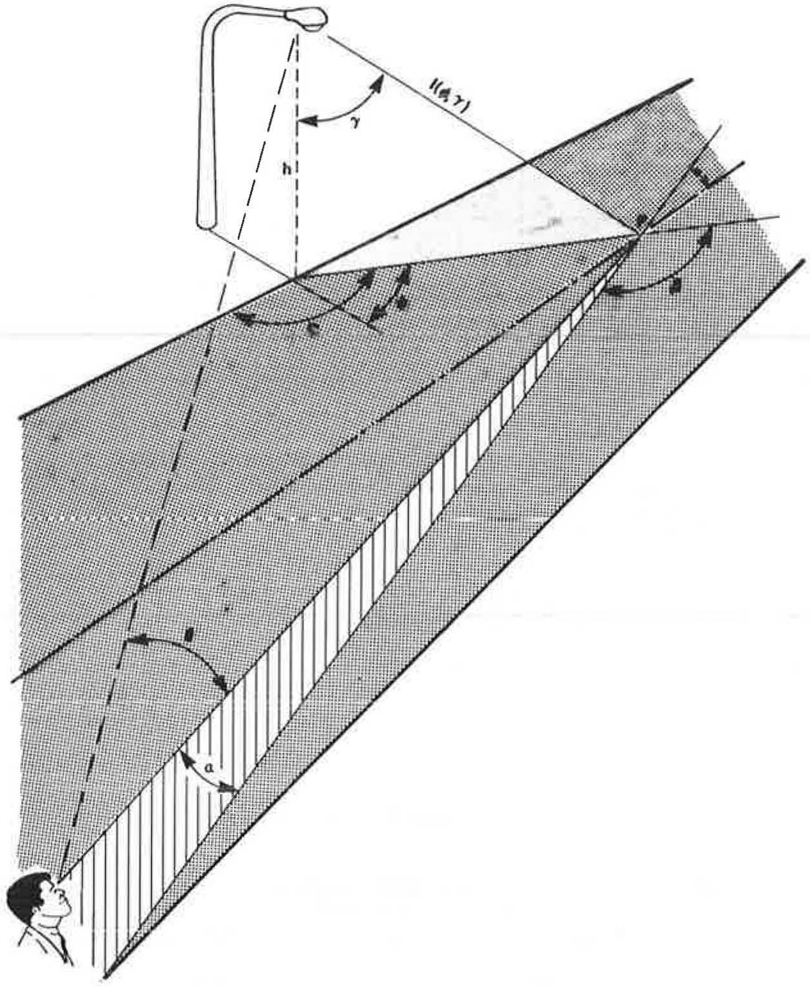
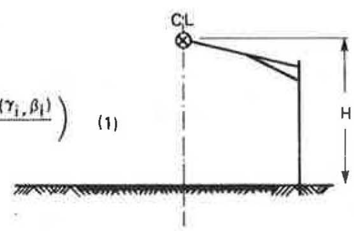


FIGURE 3 Illumination system parameters for highways.

$$L_p = \sum_{i=-2}^{i=+3} \left(\frac{I(\gamma_i, \phi_i) \cdot R(\gamma_i, \beta_i)}{H^2} \right) \quad (1)$$



$$\tan \gamma_i = \frac{(1-\xi)\,\lambda H - i\lambda H}{H} \quad (2)$$

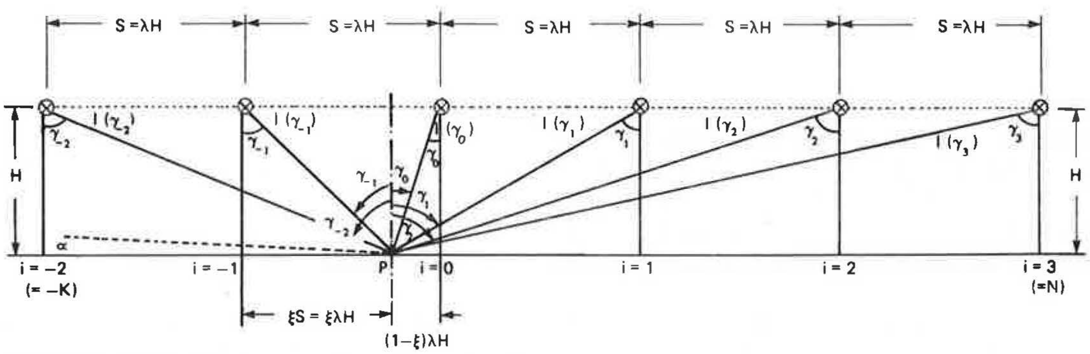


FIGURE 4 Simplified system (two-dimensional).

L_p = luminance at point P (cd/m^2), added contribution from many luminaires;
I = luminous intensity function (lumens per space angle);
γ = angle of light incidence or vertical angle of the intensity function I;
Z = size of target object above ground (m);
ϕ = angle of azimuth (i.e., horizontal angle of the intensity function I; see Figure 3);
R = reduced luminance coefficient; and
β = angle between the direction of I and the road axis.

For simplicity of demonstration, the luminous intensity function (I) is assumed to be independent of the horizontal angle ϕ. Further, the angle β is assumed to be between zero and 10 degrees, so that a simplified assumption can be made for the reduced luminance coefficient (R). These two simplifications mean that variations perpendicular to the road axis are neglected. With ϕ and β being zero, the equation to calculate the luminaires' contributions at point P is

$$L_p = \frac{I(\gamma) * R(\gamma)}{H^2} \qquad (1)$$

where

$$\tan \gamma = (1 - \xi)\,\lambda + (i * \lambda) \qquad (2)$$

Reduced Reflectance Coefficient $R(\gamma)$

The aforementioned assumption for the reflectance coefficient is shown in Figure 5. The curve shown is represented by the equation

$$R = 0.0305\,[\cos\gamma + 0.37\cos(2\gamma)] \times \frac{q_0}{0.070} \qquad (3)$$

where q_0 is the brightness parameter *(2–5)*. This curve is assumed to sufficiently approximate the function R for small angles of β derived from the standard reflectance table (*R*3). Note that there are variations in age and location of the reflectance parameter along any type of pavement, so approximations of this kind are appropriate.

Luminous Intensity Function $I(\gamma)$

In the two-dimensional simplified system here, luminous intensity functions for luminaires vary with the vertical angle only. For example, if a cutoff point is assumed at some angle between 80 and 90 degrees, the functions could be expressed as a function of γ by a polynomial such as the following:

$$I = f(d + a\gamma + m\gamma^{\mu} - t\gamma^{\nu}) \qquad > 0 \qquad (4)$$

If uneven exponents are chosen, such as $u = 3$ and $v = 5$ or 7, the intensity distribution would be nonsymmetrical. Such nonsymmetrical functions may also be expressed by two polynomials, different for positive or negative values of γ, and so forth.

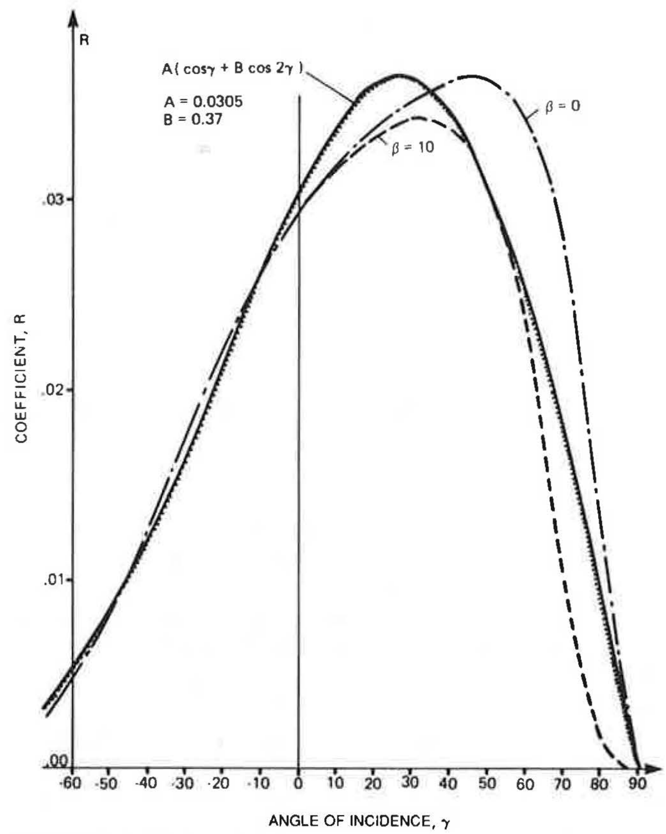


FIGURE 5 Reduced reflectance coefficient *R* for two-dimensional system.

A nonsymmetrical luminaire could be described as shining mainly backward, toward the driver. This type of luminaire would be most economical and should therefore be included in the investigations.

It should be understood that Equation 4 represents theoretically assumed intensity distributions, disregarding whether a luminaire can be built according to this or similar specifications.

At present, the luminaires of street lighting systems are symmetrical, with main beams reaching far in both directions, because this is perceived to be an economical feature. The luminous intensity function of a symmetrical luminaire can be simulated by the following function, with even values of u and v, setting $a = 0$ and $d = 1$:

$$I = 1{,}000 * (1 + m\gamma^{\mu} - t\gamma^{\nu}) \qquad > 0 \qquad (4a)$$

Modeling of Contrast C_{20}

The area of the standard target object should generally be defined as a perfect diffuser, so that deviations from a perpendicular viewing angle or variations in the angle of light incidence have no influence on the contrast calculations.

The luminance of the target surface must have sufficient contrast against the background luminance of the road surface behind the target. Details of the target geometry may be seen in Figure 6. With respect to target contrast, the vertical illumi-

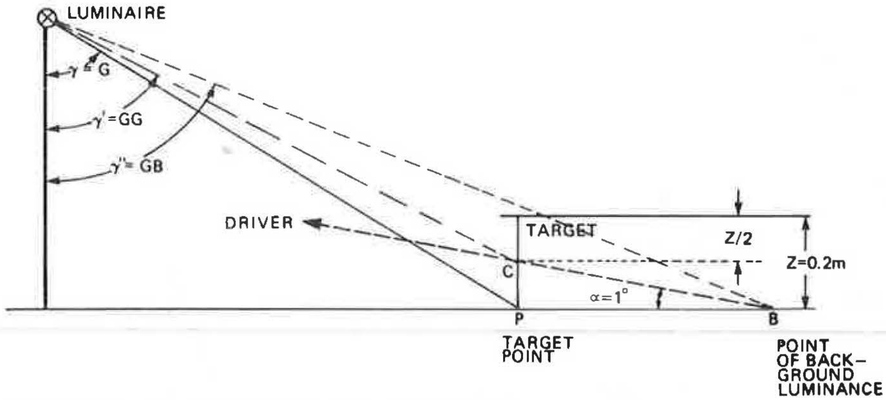


FIGURE 6 Details of critical-size target geometry (two-dimensional).

nance must be calculated at point C, the front center point of the target. The corresponding background luminance must be in the driver's line of sight.

In expressway driving, it is necessary to perceive a critical object at a distance of about 90 m. Car headlights are not very effective at that distance. Under these conditions, therefore, objects are seen predominantly by silhouette vision, that is, by the negative contrast generated by fixed roadway lighting.

Adopting a standard using the visibility of a critical object in the calculation or measurement of contrast would permit a true comparison of lighting systems based on the quality of night visibility that they provide.

As shown in Figure 4, there are three different angles of γ involved:

$$\tan \gamma = (1 - \xi) * \lambda + (i * \lambda) \quad (2)$$

$$\tan \gamma' = [(1 - \xi) * \lambda + (i * \lambda)] * \{H/[H - (Z/2)]\} \quad (5)$$

$$\tan \gamma'' = [(1 - \xi) * \lambda + (i * \lambda)] * \{Z/[2 * H * \sin(\alpha)]\} \quad (6)$$

The vertical illuminance at point C is calculated as follows, adding up contributions from luminaires in front of the target only:

$$E_{vc} = \frac{I(\gamma') \cos^2(\gamma') \sin(-\gamma')}{[H - (Z/2)]^2} > 0 \quad (7)$$

The horizontal luminance at point B is calculated by using Equation 1, except for the larger angle γ'':

$$L_b = \frac{I(\gamma'') R(\gamma'')}{H^2} \quad (8)$$

The contrast C of the target object at point P can then be computed:

$$C_{20} = \frac{E_{vc}(0.20/\pi) - L_b}{L_b} \quad (9)$$

In an actual lighting program, Equations 7 and 8 are more complex, containing expressions of the angles β and ϕ (Figure 3), whereas Equation 9 remains the same.

As shown in Figure 4, contributions from all luminaires must be added. Because the target surface is defined as a perfect diffuser, the luminance coefficient is independent from the viewing angle of the driver and from the angle of incidence; for 20 percent reflectivity, it is simply $0.20/\pi$.

Contrast Calculations of Two Types of Luminaires

In accordance with Figure 4, contrary to the results of measurements in the last section, the traffic is assumed to be moving from left to right.

Symmetrical Luminaires

The following example has been calculated for various symmetrical luminaires:

Mounting height: $H = 12.5$ m,
Pole spacing: $S = 4 * H = 50.0$ m, $\lambda = 4$.

The results of the contrast calculations are plotted in Figure 7. There is predominantly negative contrast; that is, the background luminance is higher than the target luminance. The symmetrical luminaire always throws some light on the target, although very little at zero and between 30 and 50 m near the end. At the quarter point of the distance beyond a pole, the target is relatively bright, and the contrast may be close to zero if the main beams of the luminaires do not overlap sufficiently. It should be noted that the luminance distribution of symmetrical luminaires was found to be relatively uniform for all cases (1, 2, and 3).

Thus, for conventional luminaires with symmetrical light distribution, an important design principle has been confirmed by introducing contrast as a quality measure:

The quality of fixed highway lighting using symmetrical luminaires depends on the degree of overlap of the main beams, provided that glare can be held to a minimum acceptable value.

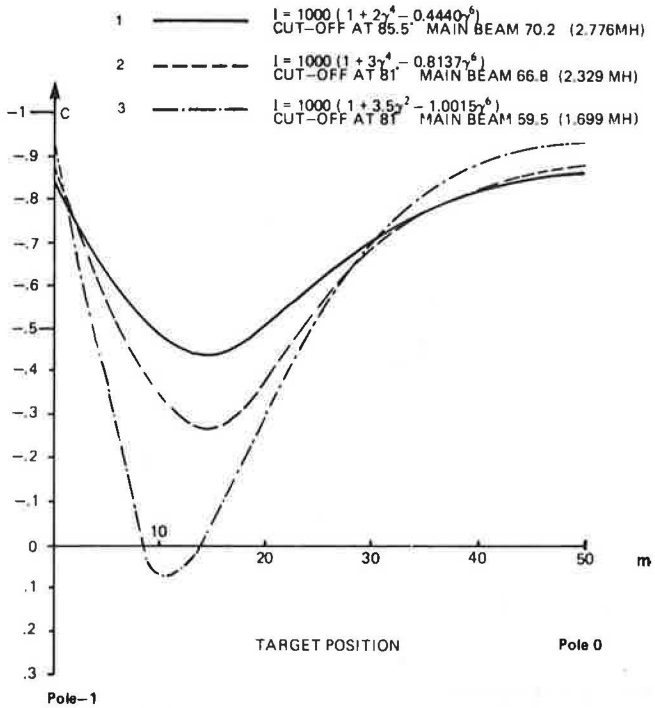


FIGURE 7 Distribution of contrast for symmetrical luminaires.

Nonsymmetrical Luminaires

Introducing contrast as a criterion for the quality of fixed highway lighting (together with luminance) would mean that new luminaire types could be developed. For separated expressways or roads with one-way traffic, it is possible to maximize the negative contrast by directing the main beam toward the driver, that is, in the direction opposite that of the traffic. However, at the same time, the (longitudinal) luminance distribution should remain fairly uniform. This is difficult to achieve with large pole distances if no light is directed away from the driver, that is, in the direction of the traffic. In other words, some light is needed in all directions. Numerous calculations of systems were carried out in order to find criteria for such an innovative luminaire, at least for the simplified system of two dimensions.

Using the same example as that used for symmetrical luminaires, the luminance intensity distributions are listed and plotted in Figure 8. The assumed cutoff angles are ±81 degrees. Between these cutoff points, there is a block of constant light output of 1,000 lumens, and for $\gamma > 0$ there is additional output directed toward the driver of approximately the same order of magnitude.

The luminance and contrast distributions are plotted in Figures 9 and 10, respectively. Both parameters have large enough values and reasonable uniformity for all three distributions of intensity: *1*, relatively narrow; *2*, medium; and *3*, wide. Narrow and medium distributions such as those labeled *1* and *2* appear feasible. The following observation can be made:

> Innovative nonsymmetrical luminaires for unidirectional traffic should have about half their output evenly distributed and the other half directed toward the driver.

This is a tentative conclusion that must be investigated further.

CONCLUSION

The contrast of a standard critical-size target object with 20 percent diffused reflectivity has been studied and may be regarded as a powerful criterion for evaluating the quality of

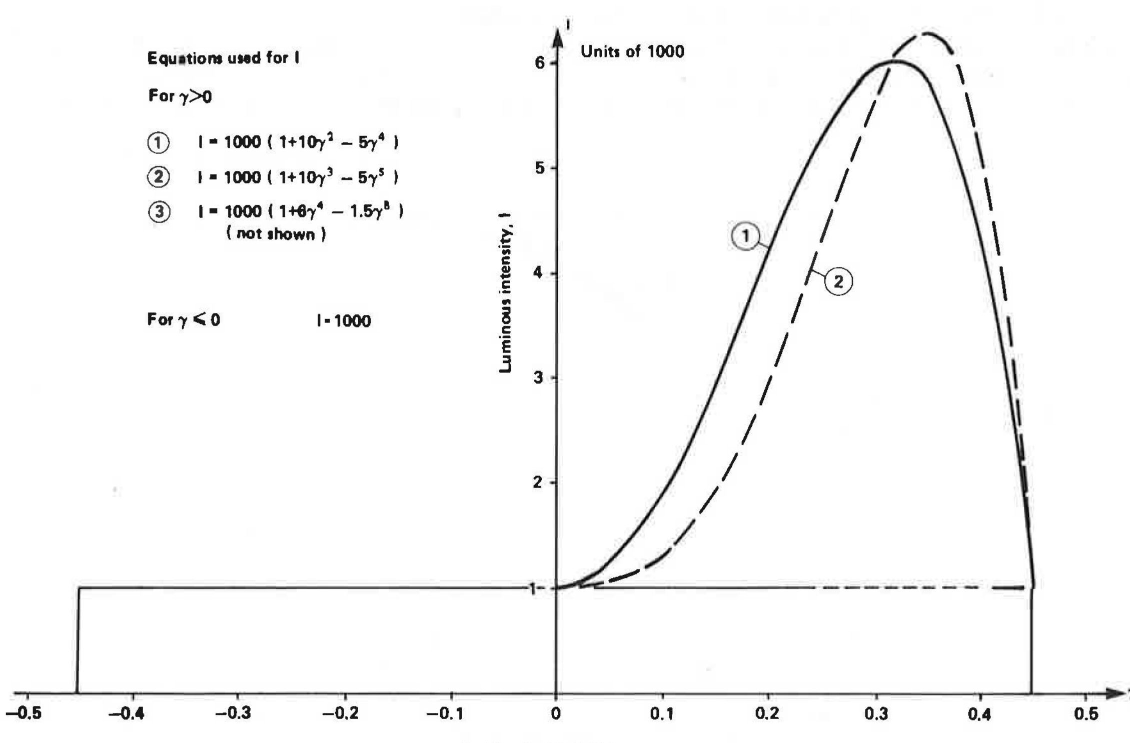


FIGURE 8 Luminance intensity for nonsymmetrical luminaires.

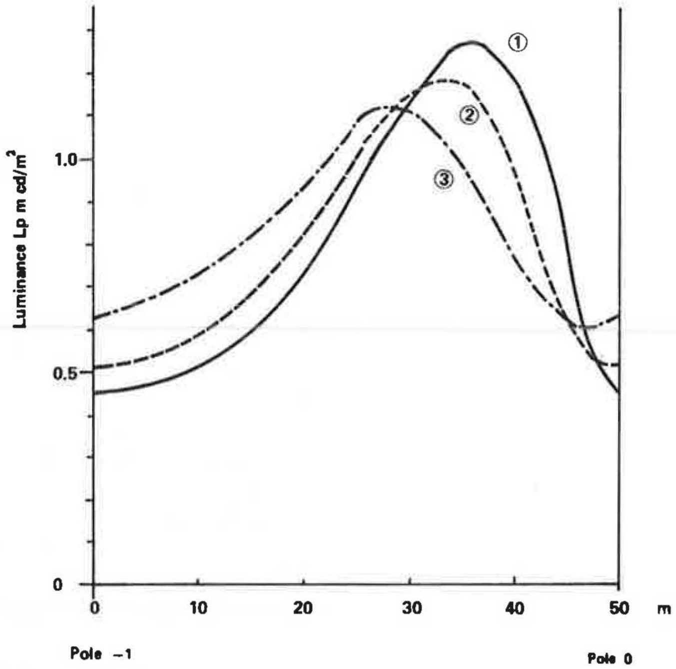


FIGURE 9 Luminance distribution for nonsymmetrical luminaires.

fixed lighting systems. Various reflectivity percentages may be chosen in a future system of standards.

This concept of contrast has been applied in a simplified systems calculation together with the usual illuminance and luminance distributions to find some basic criteria for the quality of roadway lighting with respect to night visibility and to evaluate the concept of standard target contrast.

In currently used systems with symmetrical luminaires, the overlapping of the main beams is important to avoid spots with low or zero contrast. Increased negative contrast can be achieved with nonsymmetrical luminaires directed toward the driver without having the main beams touch or overlap and with good distribution of luminance. Cutoff properties appear to be more critical in such innovative cases.

When this contrast concept is used (together with luminance and glare) to evaluate fixed roadway lighting systems, illuminance standards should be discarded.

These preliminary findings, once recognized, could lead to new and improved lighting standards and to an innovative development of new types of luminaires that are more energy efficient.

Without the introduction of additional standards of target contrast, manufacturers will have no incentive to develop innovative or improved conventional luminaire designs because the

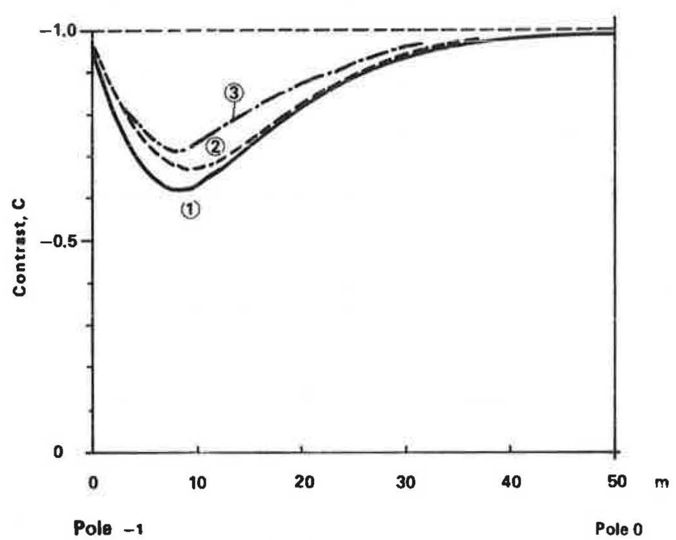


FIGURE 10 Contrast distribution for nonsymmetrical luminaires.

present standards of uniformity and level of illuminance and luminance in fixed highway lighting cannot reveal a system's weakness.

REFERENCES

1. *Determination of Minimum-Maximum Luminance Levels for Programmable Lighting Operations.* Internal Report. Ministry of Transportation and Communications, Downsview, Ontario, Canada, 1984.
2. CIE Committee TC-4.6. *Calculation and Measurement of Luminance and Illuminance in Road Lighting.* CIE Publication 30 (TC-4.6). Commission Internationale d'Eclairage, Paris, 1976.
3. M. G. Bassett, S. Dmitrevski, P. C. Kramer, and F. W. Jung. Measurements of Reflection Properties of Highway Pavement Samples. *Journal of the Engineering Society,* Sept. 1981.
4. J. B. DeBoer and D. A. Schreuder. Theoretical Basis of Road Lighting Design. In *Public Lighting,* Chap. 3, Philips Technical Library, Eindhoven, Netherlands, 1967.
5. A. Erbay. *Atlas of the Reflection Properties of Road Surfaces.* Institut fuer Lichttechnik, Technische Universitaet Berlin, Berlin, West Germany, 1974.

Publication of this paper sponsored by Committee on Visibility.

A Method of Calculating the Effective Intensity of Multiple-Flick Flashtube Signals

MARC B. MANDLER AND JOHN R. THACKER

A method of determining the effective intensity of light flashes composed of multiple pulses (flicks) of light was devised. Detection thresholds were measured for such flashes when the flick frequency and flash duration were varied. Thresholds decreased with increasing flick frequency and flash duration. At each flick frequency the relationship between threshold and flash duration was well characterized by the Blondel-Rey relation (a = 0.2), provided a multiplicative frequency-dependent fitting parameter was chosen. The fitting parameter, β, increased linearly with frequency between 5 and 20 Hz. A method of determining effective intensity was described that uses the flick frequency, number of flicks, and the calculated effective intensity of a single flick to arrive at the solution. It was concluded that this method should be used for all multiple-flick signals, provided the single-flick duration is less than 0.01 sec and the frequency is between 5 and 20 Hz. The method of Allard should not be used, because it consistently overestimates effective intensity.

U.S. Coast Guard Research and Development Center, Avery Point, Groton, Conn. 06340-6096.

A flashtube is a capacitive discharge device capable of emitting brilliant flashes of light in extremely brief time periods (on the order of microseconds). The highly intense flashtube burst can be detected at great distances and has been noted as a conspicuous signal in a typical aid-to-navigation system *(1).* Moreover, the efficiency of converting input energy to visible output is greater than that of an incandescent light *(2).* These factors make the flashtube attractive as an aid to navigation.

There are three major disadvantages associated with the use of flashtubes. The intense nature of the flick tends to momentarily blind the close observer *(3).* Also, the duration of the single flick is so brief that mariners have difficulty fixing the exact location in the visual field *(3).* Finally, mariners report difficulty judging the distance to the flashing source *(3).* The latter two difficulties can be ameliorated by presenting several flicks in rapid succession so that the appearance is not one of individual flicks, but of a longer-duration flash. Previous studies have shown that individuals can take line-of-sight bearings with greater speed and accuracy when the flash duration is increased in this way *(4).*

The detection distance of a lighted aid to navigation is valuable information because it not only allows one to calculate the range at which a light will become visible, but it is also one measure of the signal effectiveness of the aid. A typical aid-to-navigation system is composed of both steady and flashing lights. The detection distance of a steady light, where intensity does not vary with time, can be calculated from the familiar Allard's law (5, p. 62):

$$E = IT^D/D^2$$

where

E = illuminance threshold of the eye [typically 0.67 sea-mile candle *(6)*],
I = intensity of the light,
T = transmissivity of the atmosphere, and
D = distance at which the light is visible.

As the intensity of a flashing light source is a function of time, the detection distance can be calculated by using Allard's law provided that a steady-light-equivalent intensity, termed the effective intensity of the light, can be determined. Effective intensity is defined as follows *(7)*:

> If a flash is found to be just seen in conditions in which a steady light of intensity I_e is also just seen at the same distance and in the same atmospheric conditions, the flash is said to have an effective intensity I_e.

The three generally accepted methods of calculating the effective intensity of a single flash are the methods of Allard *(5)* (not to be confused with Allard's law discussed previously), Schmidt-Clausen *(8, 9)*, and Blondel-Rey-Douglas *(10, 11)*. All these methods are condensed and described by the International Association of Lighthouse Authorities (IALA) *(7)*.

As noted earlier, the increased duration of the multiple-flick flash is desirable, but to incorporate these multiple-flick flash-tube devices in an aid-to-navigation system requires a method of specifying its detection range. The IALA formally recognized that of the three single-flash methods of calculating effective intensity, only the Allard method is appropriate for calculating the effective intensity of multiple-flick flashes *(12)*:

> The reason for recommending . . . the method of Allard for trains of rapidly repeated flashes was that this method would yield an effective intensity that increased with increasing number of flashes in the train and approached asymptotically a steady-state response that for very rapid rates was identical with Talbot's Law. The other two methods could not yield any satisfactory effective intensity for trains of flashes. It cannot, however, be said that there is any direct experimental confirmation of the effective intensity obtained by the Allard method for repeated flashes.

The Allard method involves lengthy computer calculations of the explicit solution of a differential equation *(7)*. The primary purpose of this work was to find a simple, accurate method of determining the effective intensity of multiple-flick signals based on the characteristics of the signal. A secondary goal of this work was to provide the missing experimental confirmation of the Allard method.

DEFINITIONS

A single light pulse from a flashtube is called a flick. When several flicks are presented in rapid succession so that dark periods are not distinguished between the individual flicks, this is referred to as a multiple-flick flash. The rate at which the flicks are delivered in a multiple-flick flash is termed the flick frequency. The flash duration or flash length is the time between onset of the first flick and the cessation of the last flick and is a function of the flick frequency and the number of flicks. Figure 1 provides two examples of multiple-flick signals. The bottom portion of Figure 1 shows relative intensity as a function of time for a 20-Hz, 13-flick signal. The time between each flick is 1 per flick frequency, which in this case is 0.05 sec. The flash duration is 0.6 sec. In the upper portion of the figure, a 5-Hz, 4-flick signal is shown. Each flick is delivered every 0.2 sec. As with the 20-Hz signal, the flash duration is 0.6 sec, though the number of flicks and total integrated light intensities of the two signals are different.

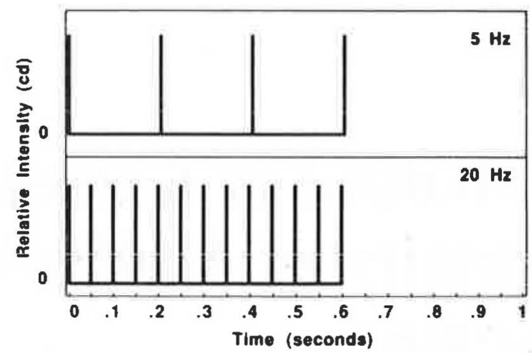


FIGURE 1 Plot of intensity versus time of multiple-flick flashes.

APPARATUS

Figure 2 shows a schematic of the apparatus used for data collection. A flashtube (Automatic Power, Inc., Part 9001-0295) was installed inside an integrating sphere. A 0.025-in. aperture was attached to the output of the integrating sphere and defined the source size. A variable neutral-density wedge provided computer control of illuminance of the signal. The signal was reflected off a mirror and superimposed on a low-luminance (0.0045-ft-lambert) background provided by a Kodak slide projector.

The signal appeared as a point source (actual angular diameter of 0.0003 mrad) when viewed from 83.25 in. The background subtended approximately 1.0 rad by 1.0 rad. To minimize the observer's uncertainty of the position of the signal, four small low-intensity fixation points were provided, each 17.5 mrad from the signal location.

The flashtube circuitry was modified so that it could be triggered by a computer pulse. A single output flick was presented for each input trigger with the maximum rate being 50 flicks per second.

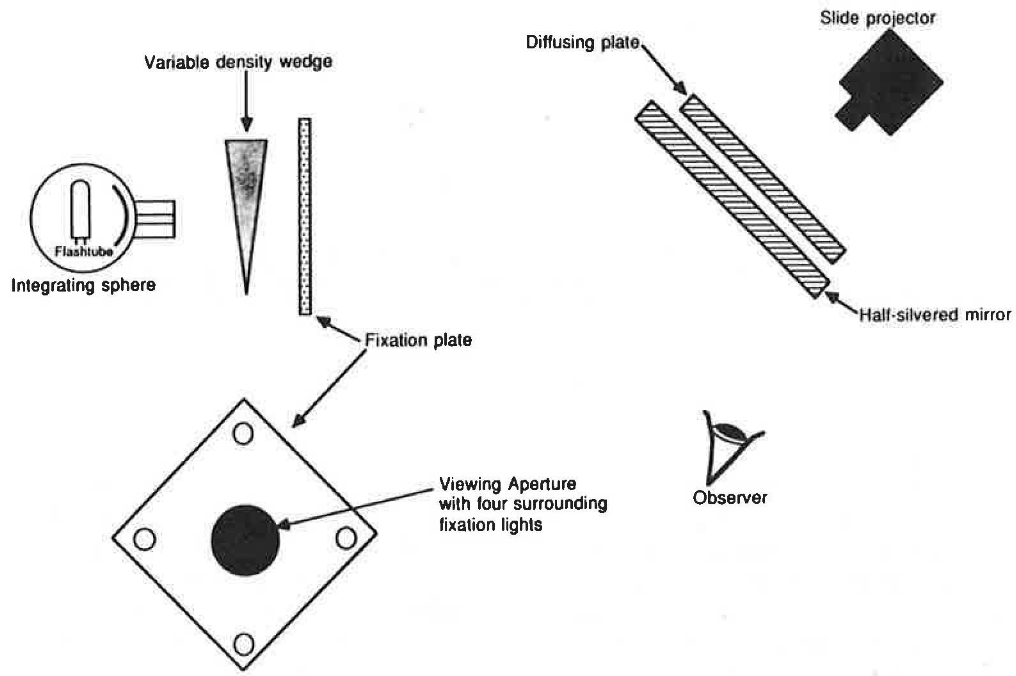


FIGURE 2 Schematic of laboratory apparatus.

CALIBRATION

Figure 3 is a plot of the intensity as a function of time for a single near-threshold flick. To obtain this curve, an EG&G photomultiplier tube (PMT) with photometric filter was positioned in the apparatus where the observer's eye was typically positioned. The PMT, which had a rise time of less than 10 nsec, was calibrated against an EG&G Model 555 photometer system using a steady light, so that the relationship between illuminance and output voltage from the PMT was known.

Figure 3 shows that the peak output occurs about 20 μsec after flick onset and the intensity decays to 10 percent of peak after 55 μsec, with negligible low-level output for as long as about 85 μsec after onset.

PROCEDURE

Detection thresholds were obtained for a total of 34 separate signals. Table 1 shows the six flick frequencies that were used,

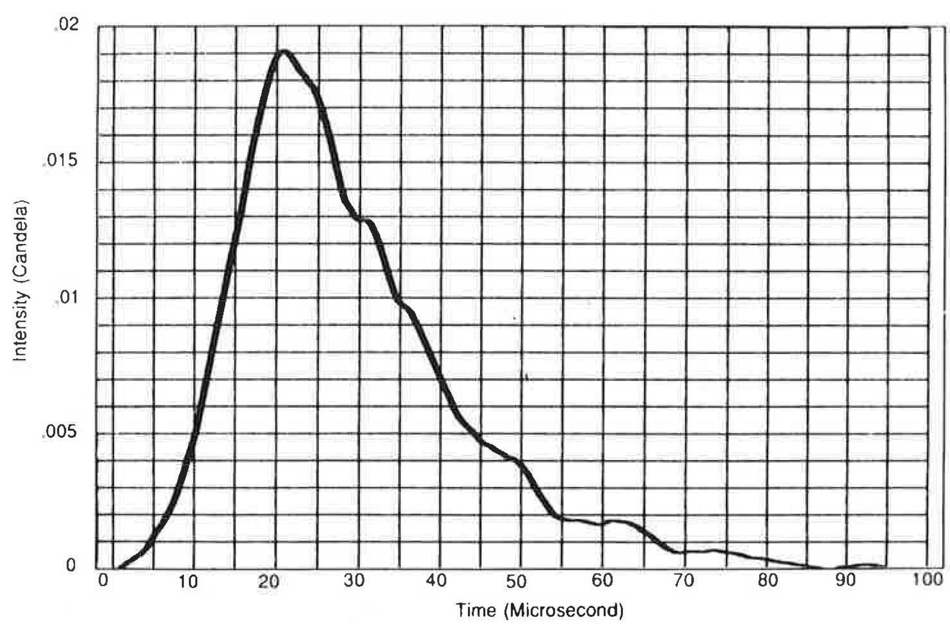


FIGURE 3 Intensity profile of a single flick.

TABLE 1 TEST SIGNALS

Flick Frequency (Hz)	No. of Flicks	Flash Duration (sec)
5	1,2,3,4	0.000,0.200,0.400,0.600
8	1,2,4,6	0.000,0.125,0.375,0.625
11	1,3,5,7,9	0.000,0.182,0.363,0.546,0.727
14	1,3,5,7,9,11	0.000,0.143,0.286,0.429,0.572,0.714
17	1,3,5,7,9,11,13	0.000,0.118,0.235,0.353,0.471,0.588,0.706
20	1,3,5,7,9,11,13,15	0.000,0.100,0.200,0.300,0.400,0.500,0.600,0.700

the number of flicks provided at each frequency, and the corresponding flash duration (see Equation 4).

A "staircase" procedure was used to measure thresholds simultaneously for all the signals of a particular frequency. On a given trial, one of the 34 signals was presented and the observers responded as to whether or not the signal was detected by pressing one of two computer-readable switches. If the signal was not detected, the illuminance was raised by 0.1 log unit (25.9 percent). When the signal was detected three consecutive times, the illuminance was decreased by 0.1 log unit. This illuminance staircase continued until the illuminance had reversed direction eight times. Threshold was taken to be the mean of the peak illuminances where the staircase reversed direction. This procedure yields a threshold that corresponds to approximately a 79 percent probability of detection *(13)*.

Four observers with 20/20 vision or better wearing spectacles participated in this experiment. Observers dark adapted for at least 20 min before data collection began. All viewing was done monocularly (with one eye), primarily because of apparatus limitations.

The test signals were always provided in the center of the four-light fixation target (see Figure 2) and observers were required to fixate at this point throughout the experiment.

RESULTS

Figure 4 shows one observer's thresholds as a function of flash duration for six flick frequencies. Flash duration is defined to be the time from the start of the first flick to the end of the last flick. Threshold is defined as the illuminance of the signal that could be detected 79 percent of the time. Because the illuminance of each flick varies with time, the peak illuminance is used as the measure of threshold.

For all frequencies, as flash duration increased, lower peak illuminance was required to detect the light, and thus the threshold is considered to have decreased. Moreover, as frequency increased, and thus the number of flicks per flash increased, thresholds also decreased.

As in any study of visual sensitivity, observers have different thresholds *(14)*. In this experiment, the thresholds of the most sensitive and least sensitive observers differed by more than a factor of 2. Because concern here is not with an absolute measure of threshold, but rather a relative measure of how threshold changed with flash duration and frequency, all observer one-flick thresholds (flash duration = 80 μsec) were normalized to 1.0. This means that thresholds at all other flash durations were proportionately less than 1.0, revealing the ex-

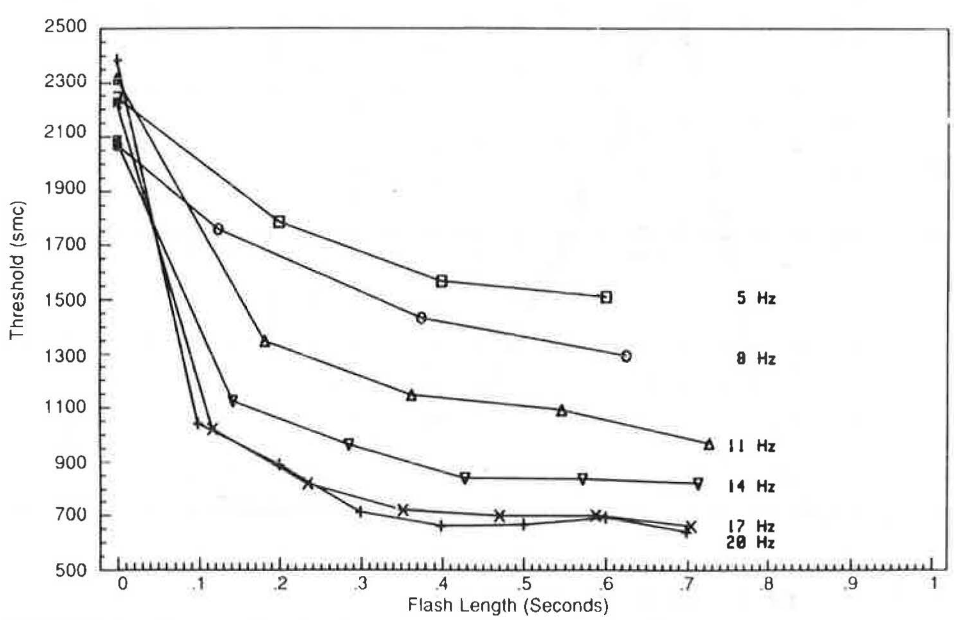


FIGURE 4 Observer thresholds as a function of flash duration.

tent to which the illuminance of these other signals could be reduced, relative to the one-flick threshold, to bring it to the threshold criterion. Moreover, because the authors' interest was with effective intensity and not threshold, the data were converted from a threshold ordinate to an effective intensity ordinate. Because two signals at threshold have identical effective intensities, the data of Figure 4 represent the peak illuminances of various test signals of equal effective intensities. The threshold ordinate of Figure 4 can be converted to an effective intensity ordinate by taking the reciprocal of threshold. That is, if one signal has a threshold that is 0.5 times that of another signal, then it has twice the effective intensity.

Figure 5 shows the relative effective intensity functions of the six test frequencies. The squares represent the mean of the four observers, and the vertical bars are ±1.0 standard error of the mean. The solid curve through the data is a theoretical function fit to the data and will be discussed in detail later. The circles show the predictions of the Allard integral, the preferred method of calculating effective intensity of multiple-flick signals *(7)*.

It is clear that the Allard method overestimates the effective intensity of the multiple-flick signals. The amount by which effective intensity is overestimated increases with frequency and flash duration. For example, at 20 Hz the Allard method overestimates effective intensity by as much as 22 percent.

The solid curves through the data are best fits of the function

$$I_e = 1.0 + \beta * [t/(a + t)] \qquad (1)$$

where

β = fitting parameter,
t = flash duration, and
a = constant of 0.2.

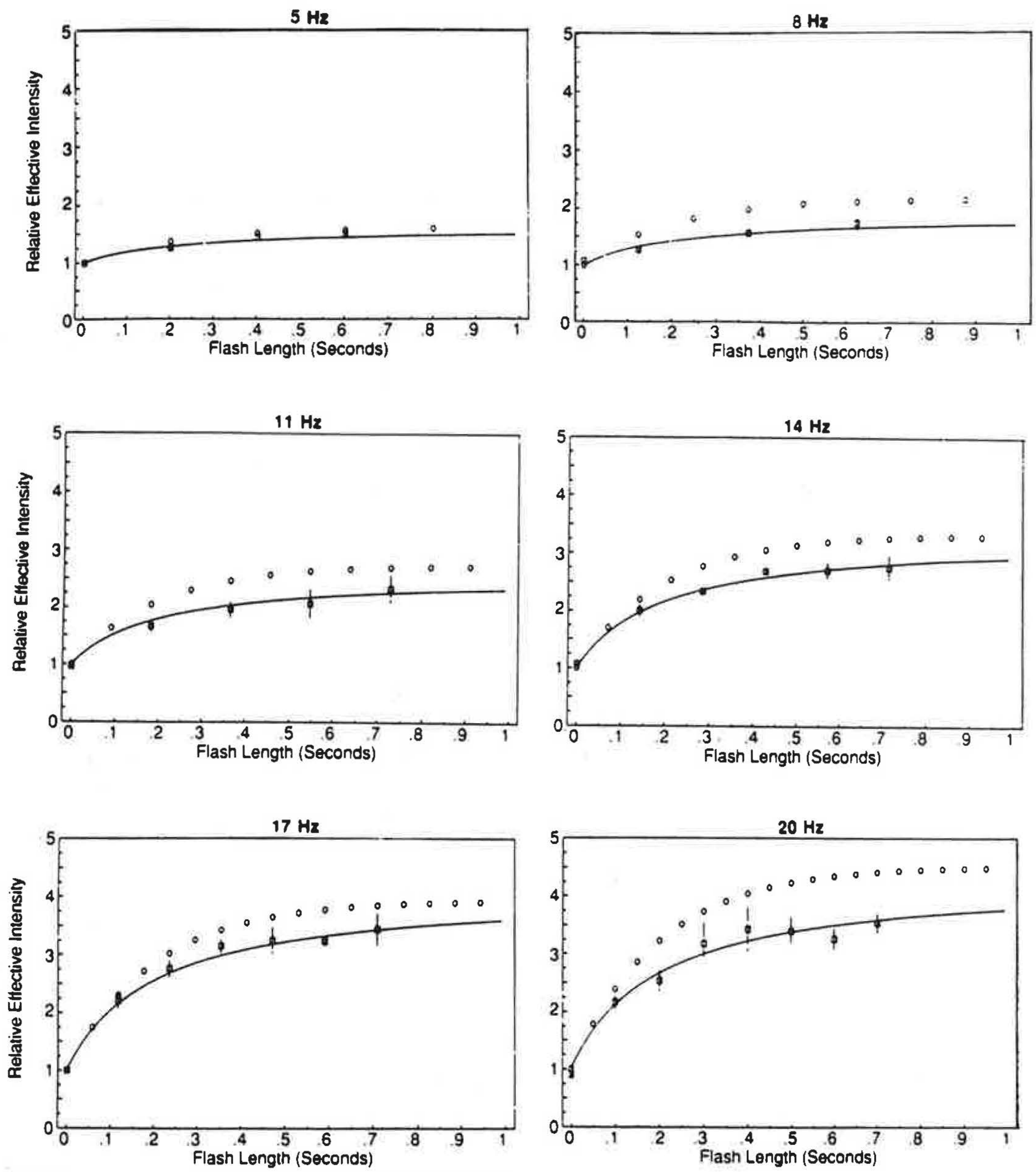


FIGURE 5 Relative effective intensity functions.

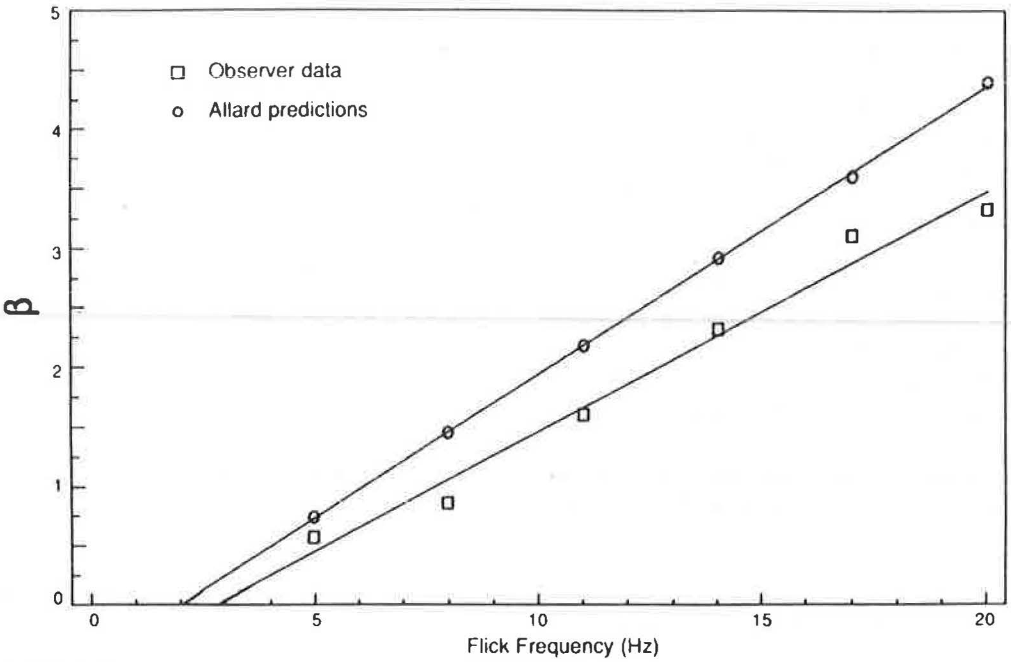


FIGURE 6 Best-fitting β, a = 0.20.

This is essentially the Blondel-Rey relation *(10)*. The constant 1.0 was added in keeping with the normalization performed on the data. By optimizing β, it was possible to fit the group data at each frequency with this function, as shown by the solid curves of Figure 5.

The β providing the best fit at each frequency is shown in Figure 6. The observer data are shown as squares and the Allard calculation as circles. The lines are least-squares fits to the data.

As frequency increases, the fitted β increases. Again, as noted earlier, the fits for the Allard predictions are greater than those for the observer data. The β fits for the Allard predictions fall along a straight line with slope of 0.243 and intercept of −0.488. The observer data can be reasonably approximated by a straight line with slope of 0.203 and intercept of −0.577.

DISCUSSION OF RESULTS

The purpose of this experiment was to establish a method of specifying the effective intensity of multiple-flick flashes. It has been shown that for each flick frequency, observer data can be fit with a Blondel-Rey equivalent function (Figure 5) having a frequency-dependent fitting parameter, β (Figure 6). Had the thresholds been independent of frequency, the ratio of the effective intensities at a particular flash duration would have been the same as the ratio of the total energy in the flashes regardless of frequency. This is not the case, as can be seen in Figure 5. A 13-flick 20-Hz signal and a 4-flick 5-Hz signal both have flash durations of 0.600 sec. The ratio of the total energies of these two signals is 3.25, yet the ratio of the effective intensities is 2.30. Consequently, a frequency-dependent relationship is required, as shown in Figure 6.

Proposed Method of Calculating Effective Intensity

For any multiple-flick signal the effective intensity can be determined by using Figure 5 as a nomogram and finding, along the ordinate, the value of the relative effective intensity for the appropriate frequency and flash duration. This value, when multiplied by the calculated effective intensity of a single flick, yields the effective intensity of the signal.

An alternative, but equivalent, approach is to derive the equations that can be used to calculate the effective intensity. Equation 1, which was used to fit the observer data of Figure 5, provides the relative effective intensity of any signal provided the fitting parameter, β, is known. The effective intensity (I_e) of a multiple-flick flash can be determined from

$$I_e = I_{e1} * 1.0 + \beta\,[t/(a + t)] \qquad (2)$$

I_{e1} is the effective intensity of a single flick. To calculate this single-flick effective intensity, any of the three methods proposed by IALA is sufficient *(7)*. In these calculations the three methods yielded nearly identical results for the flick shown in Figure 3. The calculated effective intensities from these three methods are as follows:

Method	*Effective Intensity*
Schmidt-Clausen	4.6189×10^{-7}
Blondel-Rey-Douglas	4.6170×10^{-7}
Allard	4.6800×10^{-7}

The value β can be found from the least-squares fit in Figure 6, given the flick frequency. It can be calculated by

$$\beta = (0.203 * f) - 0.577 \qquad (3)$$

TABLE 2 RELATIVE EFFECTIVE INTENSITIES OF COMMON FLASHTUBE SIGNALS

Flick Frequency (Hz)	No. of Flicks per Flash 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	1.219	1.292	1.329	1.350															
6	1.291	1.401	1.458	1.493	1.517														
7	1.352	1.497	1.575	1.625	1.659	1.684													
8	1.403	1.582	1.683	1.748	1.793	1.827	1.852												
9	1.447	1.658	1.781	1.862	1.919	1.961	1.994	2.020											
10	1.485	1.727	1.872	1.969	2.038	2.090	2.130	2.162	2.189										
11	1.518	1.789	1.956	2.068	2.150	2.212	2.260	2.299	2.331	2.357									
12	1.547	1.845	2.033	2.162	2.256	2.327	2.384	2.430	2.468	2.499	2.526								
13	1.573	1.897	2.105	2.250	2.357	2.437	2.504	2.556	2.600	2.637	2.668	2.695							
14	1.597	1.944	2.172	2.333	2.452	2.544	2.618	2.679	2.728	2.770	2.805	2.837	2.864						
15	1.612	1.989	2.234	2.411	2.543	2.645	2.728	2.795	2.851	2.899	2.939	2.974	3.005	3.033					
16	1.637	2.028	2.293	2.484	2.629	2.742	2.833	2.908	2.971	3.024	3.069	3.109	3.143	3.174	3.201				
17	1.654	2.065	2.348	2.554	2.711	2.835	2.935	3.017	3.086	3.145	3.195	3.240	3.278	3.312	3.342	3.370			
18	1.670	2.100	2.399	2.620	2.789	2.923	3.032	3.122	3.198	3.262	3.318	3.367	3.410	3.448	3.482	3.512	3.539		
19	1.684	2.132	2.448	2.682	2.864	3.008	3.126	3.224	3.306	3.377	3.438	3.491	3.538	3.580	3.617	3.651	3.681	3.708	

where f is frequency in hertz. Flash duration (t) is calculated by

$$t = (1/f) * (n - 1) + d \qquad (4)$$

where

f = frequency (Hz),
n = number of flicks in the flash, and
d = duration of a single flick.

This equation simply calculates the time between the start of the first flick and the end of the last flick.

Table 2 provides the solution of the foregoing equations for flick frequencies between 5 and 20 Hz and various numbers of flicks. The effective intensity can be determined by multiplying the appropriate tabulated value in Table 2 by the effective intensity of a single flick.

Flick Duration

The method proposed here is based on threshold measurements for signals with the 80-μsec flick profile of Figure 3. Not all flicks or light pulses have the same temporal profile, and thus the generality of this method for other flick profiles must be addressed. It is well established that the shape of a light pulse does not affect threshold or effective intensity if its duration is less than the critical duration (i.e., Bloch's law is obeyed) *(10, 15)*. Within certain time intervals, the eye acts as a perfect integrator, and thus pulse shape is irrelevant. Although the time period over which Bloch's law is obeyed varies with many stimulus conditions (*16*, p. 154), it can conservatively be estimated at 0.01 sec. It appears that as long as the flick duration is less than 0.01 sec, this method can be used. Although this has not been verified empirically in this work, two considerations support this conclusion. First, the effective intensity, as calculated by any of the three methods of IALA *(7)*, is independent of pulse shape for pulses between 0.0 and 0.01 sec. Second, the amount by which Allard's method overestimates effective intensity is constant provided the flick duration is less than 0.01 sec. It is concluded that this method of calculating effective intensity can be used for any multiple-pulse light flash as long as each pulse is less than 0.01 sec.

Visual Time Constant (a)

It has been internationally agreed *(7)* that for nighttime observation the visual time constant, a, used for calculations of effective intensity, should be equal to 0.2. Nighttime observation is assumed to be at a background luminance less than 0.1 cd/m^2 *(7)*. The background used in this experiment was 0.015 cd/m^2, considerably less than the maximum permitted. For the curve fitting performed on the data of Figure 5, a was assumed to be 0.2. It can be seen that the theoretical curve does an adequate job of fitting the observer data. It was disturbing, though, that the Allard method overpredicted I_e. As an exercise, the analysis was repeated using different values of a in search of an a that would bring the observer data and the Allard calculation into agreement. The observer data were refitted with different values of a and the Allard calculation was performed with these same values and the results were compared. Figure 7 shows the best-fitting β for the observer data and the Allard calculation assuming an a of 0.155. The least-squares fit to the observer data yields a slope of 0.189 and an intercept of −0.540, whereas the slope and intercept for the Allard calculations are 0.183 and −0.445, respectively.

Areas of Further Investigation

In this experiment only a single background luminance was used. It is of interest to determine whether the slope of the curve in Figure 6 varies with background luminance so that the generality of this approach can be assessed. Further, it was argued that the proposed method can be used with many different flick profiles and durations. This argument should be empirically verified.

Another area in which much work is needed is in determining the optimum flick frequency and flash duration. Such an approach not only must take into account the effective intensity of the signal, but also must be concerned with the speed and

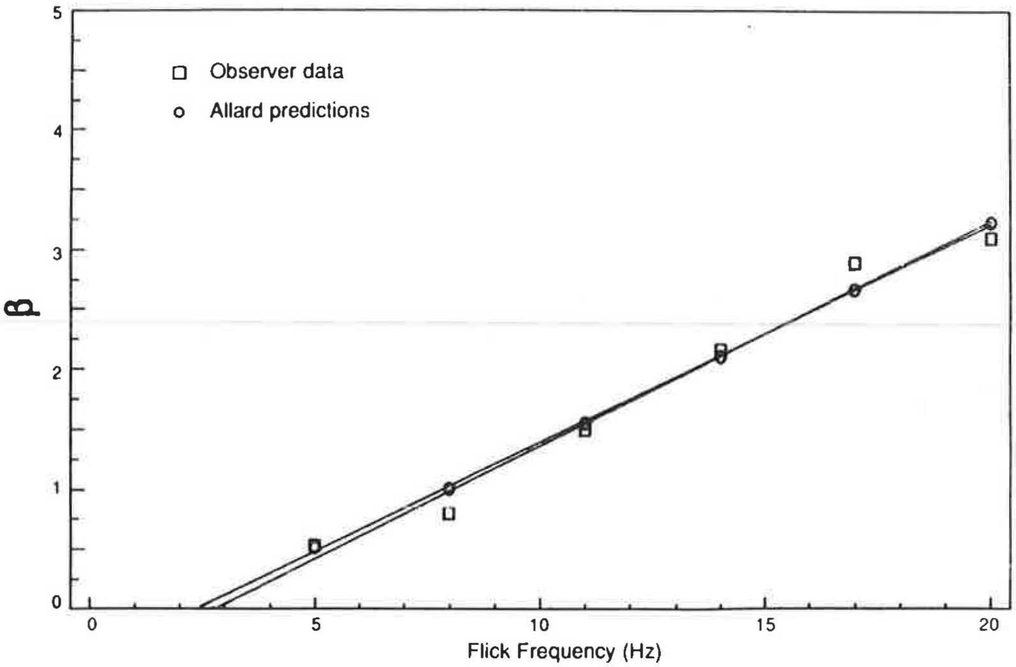


FIGURE 7 Best-fitting β, a = 0.155.

accuracy with which a bearing can be taken and must perform a comparative analysis of battery power requirements for such signals. Thacker *(4)* showed that as flash duration increases, speed and accuracy in taking a bearing improve. Montonye and Clark *(2)* performed a partial analysis of flashtube battery power requirements. Edgerton *(17)* has measured the relationships between the size of the flashtube storage capacitor, initial capacitor charging voltage, the peak output intensity, the electrical input to visual output efficiency, and the flick duration. Before flashtubes are widely deployed in the field, all these approaches to flashtube optimization must be fully analyzed in conjunction with one another.

CONCLUSIONS

The following conclusions are based on the results reported here.

1. The proposed method of calculating effective intensity (I_e) should be used for any multiple-flick signal regardless of the single-flick time-intensity profile, provided the single-flick duration is less than 0.01 sec and the flick frequency is between 5 and 20 Hz. The effective intensity can be calculated from Equations 2, 3, and 4 found in the discussion section.
2. The Allard method should not be used for calculating the effective intensity of multiple-flick flashes because it consistently overestimates the effective intensity. The greater the flick frequency, the greater the error.

RECOMMENDATIONS

The purpose of this effort was to determine a means of calculating the effective intensity of a multiple-flick signal. The results reported here do not fully address the issue of the optimum flashtube signal to incorporate into field-deployable hardware. There remain several issues that must be resolved before the optimum signal is chosen. Accordingly, the following recommendations are made:

1. Research should be conducted to specify the flashtube energy consumption as a function of number of flicks employed in a flash and the circuitry of the system. This analysis would compare and contrast the results obtained from standard incandescent sources. This work will reveal how signals are limited by energy consumption. Work should also be done to optimize the flashtube circuitry for Coast Guard applications.
2. Research should be sponsored that addresses issues of human performance with respect to flashtube signals. Specifically, problems of depth perception and difficulty in obtaining a fix on the light should be studied more carefully.
3. Once the foregoing data have been collected, an optimum range of flashtube signals should be selected based on performance measures, calculated effective intensity, and energy cónsumption considerations.

REFERENCES

1. C. B. Devoe and C. N. Abernethy. *Field Evaluation of Locomotive Conspicuity Lights.* Report FRA-OR-D-75-84. Federal Railroad Administration, U.S. Department of Transportation, 1975.
2. J. T. Montonye and G. P. Clark. *Evaluation of LS-59 Xenon Flashtube System.* U.S. Coast Guard, 1970.
3. D. F. Murphy. *A Laboratory Evaluation of the Suitability of a Xenon Flashtube Signal as an Aid-to-Navigation.* Master's thesis. Naval Postgraduate School, Monterey, Calif., 1981.
4. J. Thacker. *An Evaluation of Flashtube Signal Characteristics.* Report CG-D-26-84. U.S. Coast Guard, 1984.
5. E. Allard. *Memoire sur l'Intensité et la Portée des Phares.* Imprimerie Nationale, Paris, France, 1876.

6. *Visual Signalling, Theory and Application to Aids-to-Navigation.* Ocean Engineering Report 37, CG-250 Series. U.S. Coast Guard, 1970.
7. International Association of Lighthouse Authorities. Recommendations on the Determination of the Luminous Intensity of a Marine Aid-to-Navigation Light. *IALA Bulletin*, Vol. 3, 1978, p. 2.
8. H. J. Schmidt-Clausen. Experimental Investigations of the Validity of the Blondel-Rey Equation. Report 5-1-1. Presented at Eighth International Conference on Lighthouses and Other Aids to Navigation, Stockholm, Sweden, 1970.
9. H. J. Schmidt-Clausen. The Influence of the Angular Size, Adaptation Luminance, Pulse Shape and Light Colour on the Blondel-Rey Constant "a." In *The Perception and Application of Flashing Lights*, London, Adam Hilger, 1971.
10. A. Blondel and J. Rey. Application aux Signaux de la Loi de Perception des Lumières Brèves á la Limite de Leur Portée. *Journal de Physique*, I (5th Series), 1911, p. 643.
11. C. A. Douglas. Computation of the Effective Intensity of Flashing Lights. *Illuminating Engineering*, Vol. 52, 1957, p. 641.
12. International Association of Lighthouse Authorities. Recommendations for the Calculation of Effective Intensity of a Rhythmic Light. *IALA Bulletin*, Vol. 2, 1981, p. 27.
13. G. B. Wetherill and H. Levitt. Sequential Estimation of Points on a Psychometric Function. *British Journal of Mathematical and Statistical Psychology*, Vol. 18, 1965, p. 1.
14. S. Hecht, S. Shlaer, and M. H. Pirenne. Energy, Quanta, and Vision. *Journal of General Physiology*, Vol. 25, 1942, p. 819.
15. A. M. Bloch. Experiences sur la Vision. *Paris: Soc. Biol. Mem.*, Vol. 37, 1885, p. 493.
16. N. R. Bartlett. Thresholds as Dependent on Some Energy Relations and Characteristics of the Subject. In *Vision and Visual Perception* (C. H. Graham, ed.), John Wiley and Sons, New York, 1965.
17. H. E. Edgerton. *Electronic Flash, Strobe*, 2nd ed. MIT Press, Cambridge, Mass., 1979.

Publication of this paper sponsored by Committee on Visibility.

Evaluation of Alternative Sign-Lighting Systems To Reduce Operating and Maintenance Costs

JONATHAN E. UPCHURCH AND JEFFREY T. BORDIN

The study objective was to identify a sign-lighting system that has a lower electric power cost and reduced maintenance requirements and that provides adequately for the motorists' needs in terms of legibility and illumination level. Twenty-five candidate lighting systems were identified through a review of technical data and specifications for lamps and fixtures by an independent lighting expert. Photometric tests and computer analyses of sign illumination levels reduced the number of candidates to 10 alternative systems, which were then field tested. Each alternative lighting system was field tested for 10 to 14 months. Sign luminance was measured with a telephotometer. Power consumption was monitored. Maintenance requirements and lamp life were noted. A human factors study determined legibility distance and rated viewing comfort, lighting uniformity, and color rendition. An economic analysis was performed in which the initial cost of acquiring and installing the lighting systems and annual costs for electric power, washing, relamping, and ballast replacement were considered. A lighting system using the high-pressure sodium light source was recommended. Compared with the existing commonly used fluorescent system, it uses one-third as much electric power and has about one-third of the annual owning and operating costs. The recommended system has a satisfactory illumination level and provides the best legibility distance of the 10 systems tested.

Department of Civil Engineering, Arizona State University, Tempe, Ariz. 85287.

During the past 5 years interest has been increasing nationwide in overhead guide-sign lighting because of the increasing cost of the energy to provide illumination. In California, for example, the annual cost of electric power to illuminate overhead signs on the freeway system increased from $993,000 in FY 1977–1978 to $2,200,000 in FY 1982–1983 (*W.A.J. Hoverstern*, California Department of Transportation, unpublished data, June 1985). The nationwide cost of power for overhead sign lighting (for all overhead signs on all roadway systems) was estimated to be about $20 million annually in 1986.

In addition to the cost of electric power, highway agencies are also concerned about the maintenance costs and labor requirements for sign-lighting systems. California's annual maintenance cost for its overhead signs is $800,000 per year

(*W.A.J. Hoverstern,* California Department of Transportation, unpublished data, June 1985). Resources are scarce and the monies and manpower available to highway agencies have been declining in real terms. Thus, with increasing operating and maintenance costs and limited resources there is a need to stretch dollars further and reduce manpower needs.

Although cost reduction is important, overhead sign-lighting systems must also serve the needs of the motorist. Signing must be sufficiently visible and allow the driver adequate time to respond. These factors led the Arizona Department of Transportation (ADOT) to initiate a research project designed to identify lighting systems that would be more power efficient, require less maintenance, and, at the same time, satisfy the needs of the motorist.

RESEARCH APPROACH

Introduction

The whole issue of overhead guide-sign visibility at night is quite complex. There are a tremendous number of variables that affect what the motorist sees:

The Sign

- Type of sign materials used for the legend and background and their luminance or reflectivity
- Contrast between the legend and the background
- Color of sign background
- Age of sign material (sign material deterioration)
- Dirt, dust, and road film accumulated on the sign
- Presence of rainwater, dew, or frost on the sign
- Size of letters in legend

Illumination

- Illuminated versus nonilluminated
- Type of light source
- Illuminance level
- Color rendition
- Presence of ambient lighting (surround luminance)
- Presence of glare sources behind sign or other competitive background lighting

Environmental Factors

- Snow, rain, fog, haze, blowing dust

The Vehicle

- Headlight characteristics (e.g., photometry, aim, clean or dirty, wet with rainwater)
- Windshield characteristics (e.g., tinted glass, clean or dirty, wet with rainwater)

Roadway Geometry

- Sign orientation (Perpendicular to road? Does road have horizontally or vertically curved alignment?)

The Motorist

- Observer visual characteristics (e.g., night vision, which is a function of age)

Other Factors

- Use of high-beam or low-beam headlamps
- Traffic volume (heavy stream of traffic provides more headlight illumination than a single vehicle)
- Vehicle position (lane position and distance from sign)
- Blockage of view by other vehicles (e.g., trucks)

There are additional factors that affect sign legibility requirements:

- Time required for the driver to recognize a sign, read it, and react to it;
- Length or complexity of the message on the sign;
- Vehicle speed (determines viewing time available); and
- Kind of response required of the motorist (immediate response or delayed reaction?).

These variables greatly complicate the task of quantifying the motorists' needs and complicate the development of a research approach to solve the sign-lighting problem.

To simplify the research problem and make the research project more manageable, two basic tenets were accepted:

1. The requirement in the *Manual on Uniform Traffic Control Devices* (MUTCD) *(1)* that overhead signs on freeways be illuminated or have a reflectorized background was accepted. The study was limited to illuminated signs.
2. Existing sign-lighting standards published by the Illuminating Engineering Society (IES) *(2)* and AASHTO *(3)* were accepted. A decision was made to develop lighting systems that met these standards.

An additional issue was color rendition. The MUTCD *(1)* requires that regulatory and warning signs show the same color by day and by night when illuminated. It does not require that guide signs have good color rendition. AASHTO's guide *(3)* states that "the light source . . . [should] . . . preserve the colors on the sign." Unlike the MUTCD, the AASHTO guide is only advisory; it is not a legal requirement.

The issue of color rendition is important because some light sources (high-pressure sodium, low-pressure sodium) do not provide good color rendition. Assessment of the need to see green at night is highly subjective and there is a great diversity of opinion. A decision was made to include light sources (high-pressure sodium and low-pressure sodium) that provide poor color rendition in the alternative lighting systems considered.

Study Objectives

The principal objective of the study was to identify a lighting system that has a lower power cost and reduced maintenance requirements compared with those of currently used lighting sources and that provides adequately for the motorists' needs in terms of color rendition and illuminance level.

Each of the light sources or lighting systems was evaluated on the following bases:

1. Illuminance level: compared with AASHTO and IES guides;
2. Economics: costs of lamps, fixtures, installation, electric power, and maintenance;
3. Maintenance required: person hours for installation, washing, cleaning, lamp replacement, and other maintenance;
4. Lamp life;
5. Legibility: the distance from which a sign is legible when illuminated;

6. Color rendition: a subjective assessment;
7. Light uniformity (whether the sign is uniformly lit): an assessment (*a*) in comparison with AASHTO and IES guides and (*b*) subjectively by human observers;
8. Viewing comfort: an assessment of glare or harshness due to brightness of a sign in a dark environment.

It was recognized at the beginning of the study that, potentially, there are a very large number of alternative sign-lighting systems. A system is composed of a light source, a lamp of a given size, a fixture, the ballast, and a specific number of lamps and fixtures. With approximately six principal light sources, roughly five lamp sizes for each source, several different fixtures on the market, and various numbers of lamps and fixtures that could be used to light one sign, the potential number of lighting systems available could easily be more than 100. A summary of the choices available follows.

Light Source
- Fluorescent (the standard light source now used by ADOT)
- Mercury vapor (available in a "clear" and a "deluxe" version)
- Metal halide (available in a "clear" and a "color-improved" version)
- High-pressure sodium (available in a "clear" and a "color-improved" version)
- Low-pressure sodium

Lamp Size
- Each light source is available in several sizes (wattages) and lamp configurations

Fixture
- Various manufacturers market a variety of fixtures. Design of the fixtures varies considerably. Design of the reflector (behind the lamp) and the refractor (the glass cover or lens in front of the lamp) can have a dramatic effect on the ability of the fixture to distribute light over a sign panel. One type of fixture is used for the long, narrow fluorescent lamp. A second type generally can be used for most high-intensity discharge (mercury vapor, metal halide, high-pressure sodium) lamps.

Ballast
- A variety of ballasts are available on the market for use with specific light sources and lamp sizes. The ballasts vary in efficiency.

Number of Lamps and Fixtures
- A given size of sign panel can be illuminated by using one, two, three, or more lamps and fixtures. The choice of number of lamps and fixtures affects the level of illuminance (footcandles), light uniformity, and economics of installation and operation.

The performance of an individual sign-lighting system is dependent on the choices made in the foregoing list. There were a very large number of possible combinations of light source, lamp size, fixture, ballast, and number of lamps and fixtures that could potentially serve in a sign-lighting system.

The challenge of the study was to weed out the lesser systems and identify the best one. This was done through a three-step process. Each succeeding step was more detailed and rigorous than the previous step. The three steps were a preliminary evaluation, a laboratory evaluation, and field testing.

TESTING OF ALTERNATIVE LIGHTING SYSTEMS

Preliminary Evaluation

As noted, a large number of lighting systems were potentially available on the market. As a preliminary evaluation the whole range of alternatives was evaluated in terms of their ability to meet IES recommended illuminance levels for typical sign sizes, their ability to be competitive from the standpoint of power use, and their ability to be competitive from an overall economic standpoint. This evaluation was conducted by reviewing the technical data and specifications available for lamps and fixtures and through a subjective review. Contenders that did not meet the evaluation criteria were eliminated. This preliminary evaluation reduced the number of alternatives to 25.

Laboratory Evaluation

As a second step, each of the 25 systems underwent a laboratory evaluation. Photometric tests quantified the illuminance levels. Computer analysis of the photometric test results, using a program named SITELITE, predicted the illumination level that each alternative would provide on a typical sign face having dimensions of 8 ft high by 21 ft wide.

A review of the SITELITE computer analyses allowed a further reduction in alternatives. Alternatives were rejected if they did not provide illuminance levels as recommended by the IES, if they provided uneven light distribution, or if adequate illumination could be provided by a smaller-wattage lamp. The IES standards state that signs located in "medium" ambient light locations should have an average of 20 to 40 footcandles of illumination maintained. Alternatives were rejected if they did not provide an average of 20 footcandles of illumination maintained. In some instances it was apparent that a smaller-wattage lamp would provide adequate illumination. In these cases the smaller-wattage lamp was used in subsequent field testing.

As a result of the foregoing analysis, the number of alternatives for field testing was reduced to 10 (including the standard ADOT fluorescent lighting system). These 10 remaining alternatives were then subjected to field testing. A list of the 10 systems selected for field testing is given in Table 1.

Another element of the preliminary evaluation was an inventory of sign panel sizes. Sign panel dimensions are important in determining the performance of a sign-lighting system. A system that performs well on a small sign may perform poorly on a large sign. Conversely, a sign-lighting system that performs well on a large sign may provide illumination "overkill" and waste electric power on a small sign. For these reasons an inventory of existing sign panel sizes on the ADOT roadway network was compiled early in the study.

Great variability was discovered regarding sign dimensions. No standardization of sign panel sizes was discovered. In fact, the 355 sign panels measured represented 117 different sign panel sizes. The significant findings are as follows:

- Lengths varied from 6 to 28 ft,
- Heights varied from 5 to 14 ft,
- Length of 90 percent of all signs was between 8 and 21 ft, and
- Height of 94 percent of all signs was between 6 and 12 ft.

FIELD TESTING

Ten sites were selected for field testing on an 11-mile-long freeway segment in the Phoenix urban area. The sign sizes at each test site were representative of the total sign population and each was approximately 8 ft high and 20 ft wide. With one exception, all sites had porcelain enamel backgrounds [the tenth site (System 2) had a high-intensity reflective sheeting background]. The legends were all white porcelain enamel with reflector buttons. With one exception all signs were interchange sequence signs having three lines of legend. The lighting systems to be tested were installed, and wiring modifications were made to allow power consumption to be monitored during field testing. Following installation, each lighting system was field tested for a period of 10 to 14 months.

Sign Luminance

One aspect of field testing was the measurement of sign luminance. Photometric tests and the SITELITE computer program described previously were employed to predict sign illuminance (the amount of light shining onto the sign face). Sign luminance is the amount of light coming from the sign face. In general, luminance is related to illuminance but is also affected by the amount of light reflected by the sign material (dependent on color and surface characteristics), the angle of incidence of the illuminance on the sign face, and the position of the observer or measuring instrument with respect to the sign.

Sign luminance data were obtained for three purposes: to compare actual field performance with IES recommended luminance levels, to compare the actual performance of individual lighting systems with one another, and to compare actual performance (based on luminance) with predicted performance (based on illuminance) by the SITELITE program.

Luminance was measured by using a Spectra Pritchard photometer, Model 1980. The IES *Guide for Photometric Measurements of Roadway Sign Installations (4)* was followed. Measurements were recorded every 2 ft across the horizontal axis of the sign face, and every 1 ft on the vertical axis. Data were recorded on a segmented chart representing the sign face (see Table 3) for both the white legend and the green background.

Telephotometer readings were compared with IES standards. Each of the 10 lighting systems was designed, on the basis of SITELITE program evaluations, to provide an average of at least 20 footcandles of illuminance on an 8-ft-high by 20-ft-wide sign panel. The 20-footcandle value meets the IES standard for medium ambient lighting conditions. The IES standard also prescribes required luminance (reflected illumination) levels for the white legend. This value is 14 footlamberts and it assumes that white sign letters will reflect 70 percent of the illuminance. Therefore, field performance (telephotometer readings) was compared with the IES standard of 14 footlamberts.

Data on measured luminance for each of the 10 lighting systems are presented in Table 2. The values are estimates of the average luminance over the entire sign face based on telephotometer measurements of the legend. Measured luminance ranged from 10.6 footlamberts on System 5 to 20.9 footlamberts on System 7. If the IES luminance standard of 14.0 footlamberts is applied rigorously, three lighting systems fail to meet that standard. It is the opinion of the principal investigator that the IES standard is a broad guideline to be followed and that small deviations from that guideline have insignificant effects on visibility. As described later, two of the

TABLE 1 ALTERNATIVE SYSTEMS SELECTED FOR FIELD TESTING

System Number	Light Source	Lamp Size	Fixture (Manufacturer and Model)	Number of Fixtures	Predicted Average Footcandles *	Predicted Uniformity *
1	Fluorescent	800 milliamp	Nu-Art NAFL	3 fixtures 6 lamps	20.1	4.1:1
2	Clear Metal Halide	175 watt	Holophane Expresslite	2	28.1	6.5:1
4	Clear High Pressure Sodium	70 watt	Guth Signliter	2	33.7	7.3:1
5	Clear Metal Halide	175 watt	Guth Signliter	2	47.6	7.5:1
6	Clear High Pressure Sodium	70 watt	General Electric Versaflood II	2	26.4	7.5:1
7	Clear Metal Halide	175 watt	General Electric Versaflood II	2	47.3	6.1:1
8	Low Pressure Sodium	35 watt	Holophane Expresslite	3	27.2	3.8:1
10	Clear High Pressure Sodium	150 watt	Holophane Panel-Vue	1	23.3	5.9:1
11	Clear Metal Halide	175 watt	Holophane Panel-Vue	1	22.1	
12	Clear Mercury Vapor	250 watt	Holophane Panel-Vue	1	22.3	

*The values given are the overall average footcandles of illumination and the uniformity ratio predicted by the SITELITE program for an 8 foot high by 20 foot wide sign. Uniformity ratio is based on the maximum and minimum foot-candle values for 1 foot squares.

TABLE 2 FIELD PERFORMANCE: LUMINANCE AND UNIFORMITY RATIO

Lighting System[1]	Predicted Luminance (Foot-Lamberts)[2]	Measured Luminance (Foot-Lamberts)[3]	Uniformity Ratio[4]
1	14.0	18.8	3.3:1
2	19.7	approx. 12.	5.5:1
4	23.6	approx. 17.5	8.0:1
5	33.3	10.6	6.4:1
6	18.5	approx. 20.	3.4:1
7	33.1	20.9	5.9:1
8	20.5	20.6	6.0:1
10	16.3	11.9	3.3:1
11	15.5	14.0	5.0:1
12	15.6	16.7	4.3:1

[1]See Table 1 for a description of each lighting system.

[2]Predicted Luminance is the predicted overall luminance for a white legend. It is based on the predicted overall illuminance level from the SITELITE program multiplied by 0.7.

[3]Estimated overall luminance based on telephotometer measurements of the legend.

[4]Uniformity Ratio is based on telephotometer readings. These are estimates only.

systems that had measured luminance of less than 14 footlamberts (Systems 2 and 10) had the best legibility distances in the observer study.

Luminance measurements were also used to determine the uniformity ratio for each sign-lighting system. The uniformity ratio is the ratio of the brightest luminance to the darkest luminance on the sign face. The IES standard states that this ratio should not exceed 6:1. In Table 2 estimated uniformity ratios based on telephotometer readings are presented. The best uniformity ratio was 3.3:1, and the worst was 8.0:1. Two lighting systems exceeded the 6:1 standard.

A comparison of actual performance in the field (based on luminance) to predicted performance (based on illuminance) by the SITELITE program shows mixed results. Comparisons were made of average luminance levels as shown in Table 2 and for individual points on the sign face as shown in Table 3. Some lighting systems showed good agreement between field performance and predicted performance. Lighting Systems 8, 11, and 12 are good examples. Other lighting systems showed poor agreement, notably Systems 2, 5, and 7.

Poor agreement could result from several factors: a higher-than-expected degradation in lamp light output, a greater-than-expected accumulation of dust and dirt on the fixtures, the possibility that lamps used in the laboratory photometric tests were not ordinary lamps, instrumentation errors, sign legend materials that reflect more than or less than 70 percent of the incident illuminance, the angle of incidence of the illuminance, and others. Although any of these factors could have resulted in poor agreement, none was identified as being a definite contributor.

Table 3 shows a comparison of the field performance and predicted performance for individual points on a sign face. The data presented are for System 1.

Telephotometer readings were used to determine one other parameter—the contrast between the white legend and the green background. Luminance of the white legend was generally 10 times the luminance of the green background.

Observer Studies

An important element of field testing was the evaluation of legibility distance, viewing comfort, lighting uniformity, and color rendition provided by each sign-lighting system. These four characteristics are defined as follows:

1. *Legibility distance*: the distance from which the sign can be read.
2. *Viewing comfort*: effect of brightness of the light source; discomfort in viewing may occur as the motorist approaches the sign, because of the bright light, or just after he has passed the sign, because of the sudden change from a brightly lit to a dark environment. An analogy would be the discomfort experienced when one drives out of a dark tunnel into bright sunlight or when one drives from bright sunlight into a dark tunnel.
3. *Lighting uniformity*: the range between bright spots and dark spots on the sign.
4. *Color rendition*: the presence or absence of color distortion. With certain light sources, notably, high-pressure sodium

TABLE 3 COMPARISON OF LUMINANCE MEASUREMENTS IN FIELD WITH SITELITE PREDICTION

	COLUMN									
	1	2	3	4	5	6	7	8	9	10
1										
2	10.3 7.4	14.0 9.2	16.0 10.4	17.5 11.1	19.7 11.5				16.4 9.2	13.1 7.4
3										
4	13.1 11.6	16.7 14.8	19.7 16.5	23.0 17.4	25.0 17.9	27.0 17.9	27.0 17.4	26.0 16.5	25.0 14.8	12.3 11.6
5										
6										
7	11.5 12.6	15.7 16.8	17.4 17.9	21.0 18.3	23.0 19.0				23.0 16.8	16.5 12.6
8										

Notes: The above matrix represents an 8 foot high by 20 foot wide sign
Each row is 1 foot high; each column is 2 feet wide.
Data are presented for Lighting System 1.
The value in the upper left of each cell is the luminance (in foot-lamberts) measured in the field for the white legend.
The value in the lower right of each cell is the predicted luminance (in foot-lamberts) for a white legend.
It is based on the predicted illumination level from the SITELITE program multiplied by 0.7.

and low-pressure sodium, the sign colors appear much different in the nighttime than they do in the daytime.

The foregoing characteristics were evaluated with an observer study. Details of the methodology used in the observer study are not presented in this paper, for the purpose of brevity. The details are documented in the project final report *(5)*. The major findings of the observer study are noted as follows.

Two different groups of observers were used. The first group was composed of hired observers, subdivided into two age groups—a group of young adults ranging in age from 18 to 33 and a group of senior citizens ranging in age from 61 to 86. The second group was composed of transportation professionals. Forty-three observers participated.

The average legibility distance for all hired observers for the 10 sign-lighting systems was 862 ft. The average legibility

TABLE 4 LEGIBILITY DISTANCE

	Legibility Distance (Feet)			
	Young Adult Group		Senior Citizen Group	
Lighting System[1]	Mean	Standard Deviation	Mean	Standard Deviation
1	899	214	802	148
2	924	215	897	231
4	873	188	794	257
5	850	250	839	200
6	912	252	811	199
7	842	214	861	178
8	820	228	856	193
10	952	251	912	261
11	*		832	194
12	837	234	835	203

*System not operational during testing of this group

[1]See Table 1 for a description of lighting systems 1, 2, 4, 5, 6, 7, 8, 10, 11 and 12.

NOTE: Four different tests for statistical significance show that there is no statistically significant difference in legibility distance between any two lighting systems (95% level of confidence).

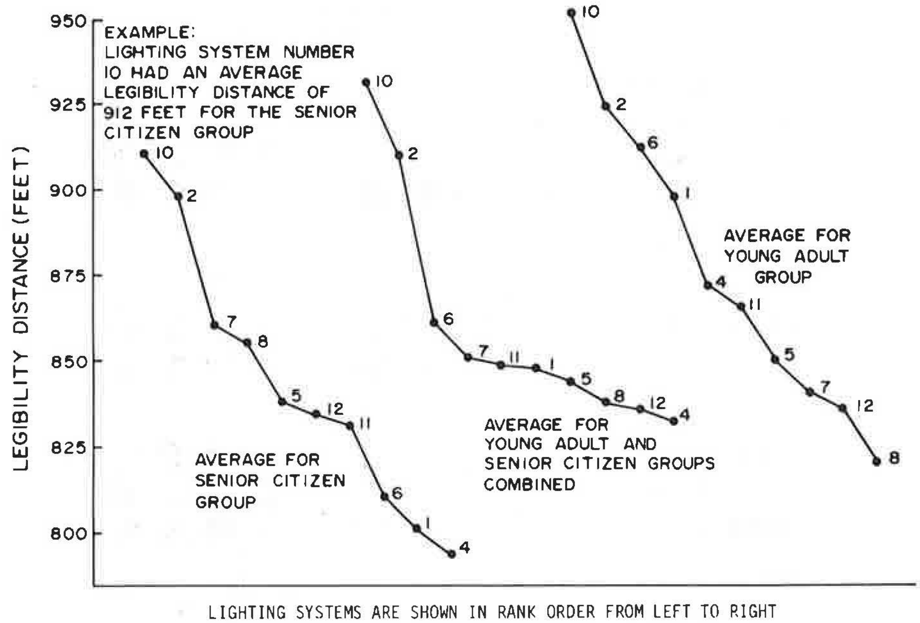


FIGURE 1 Legibility distances of lighting systems by group.

distance for the senior citizens was 844 ft; for the young adults it was 879 ft. In comparing the various lighting systems, the legibility distance ranged from a low of 794 ft for the senior citizens on System 4 to a high of 952 ft for the young adults on System 10 (see Table 4). The legibility distance of a specific lighting system generally tended to fluctuate greatly from observer to observer, as shown by the standard deviation.

As indicated in Figure 1, legibility distance varies little between lighting systems. Tests for statistical significance show that there is no statistically significant difference in legibility distance between any two lighting systems (95 percent level of confidence). The time span between the greatest and the shortest legibility distance is only 1.96 sec. Noteworthy, however, is that both Systems 10 and 2 consistently had noticeably greater legibility distances than the other lighting systems tested (1.03 sec and 0.61 sec, respectively) when compared with the standard fluorescent lighting system (System 1). System 10 had the greatest legibility distance, with an average of 932 ft.

All 10 lighting systems provided luminance levels within a relatively narrow range. Luminance levels generally meet the IES guidelines for medium ambient light conditions. Two systems with lower luminance levels were found to have the best legibility distance in the observer study. On the basis of these results, all 10 lighting systems provide satisfactory luminance levels.

For the characteristics of viewing comfort, lighting uniformity, and color rendition the observers rated individual signs as excellent, good, marginal, poor, and abysmal. These ratings were converted to a numerical scale (excellent = 5, abysmal = 1) so that a quantitative average score could be determined for each characteristic. Significant differences between lighting systems were found as shown in Figure 2.

Lamp Life

Lamp life is important because it determines how often maintenance is required. The costs of manpower and equipment (trucks) to perform maintenance is significant; the longer the time interval between routine maintenance visits, the less the maintenance costs will be.

The 1-year field test period used in this study was not long enough to make conclusions about lamp life, because the life of all lamps tested exceeded 1 year. As a result, comparisons of lamp life can be based only on manufacturer claims. The values for lamps tested in this study are as follows:

Lamp	*Size*	*Life (hr)*
Fluorescent	800 mAmp	18,000
Clear mercury vapor	250 W	~28,000
Clear metal halide	175 W	10,000
Clear high-pressure sodium	70, 150 W	~28,000
Low-pressure sodium	35 W	18,000

The lamp-life values represent the average life for a random sample of lamps. Fifty percent will fail in less than the lamp-life values given.

ADOT's practice is to use a group replacement program with a replacement period short enough so that nearly all lamps are replaced before they fail. Sign-lighting lamps are lit for about 4,000 hr a year. ADOT uses a 2-year replacement period for fluorescent lamps, which results in an age of about 8,000 hr when lamps are replaced (compared with an 18,000-hr average life).

On the basis of the manufacturer claims of lamp life, the following intervals between group relamping were established for use in an economic analysis of each lighting system:

Lamp	*Interval (years)*
Fluorescent	2
Clear mercury vapor	3
Clear metal halide	1
Clear high-pressure sodium	3
Low-pressure sodium	2

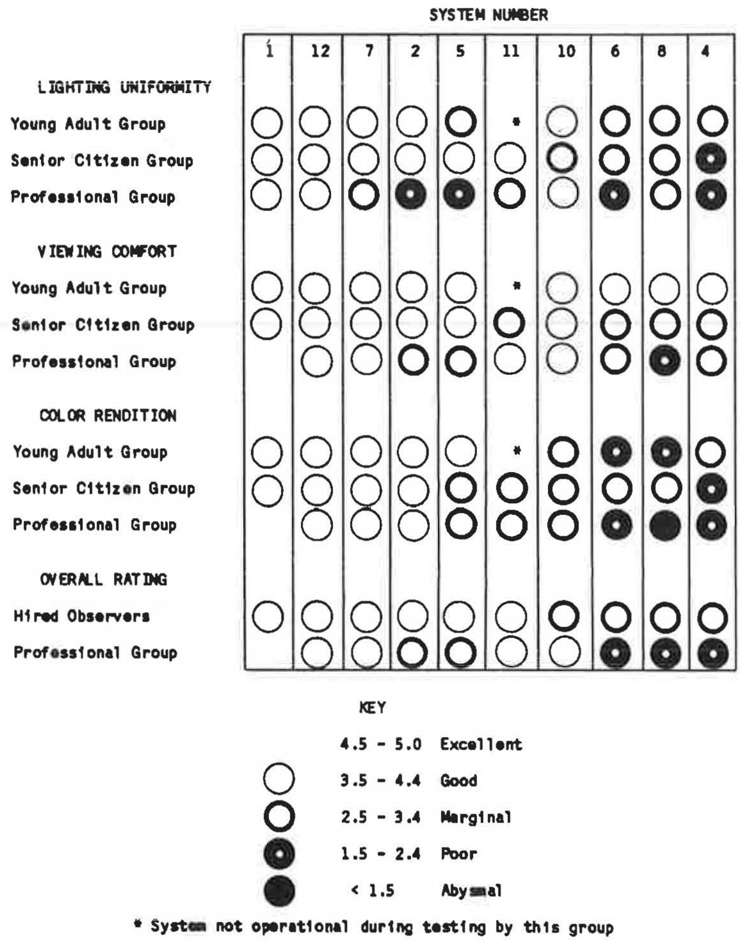


FIGURE 2 Observer ratings of lighting uniformity, viewing comfort, and color rendition.

Maintenance

During field testing ADOT's personnel kept detailed records of any maintenance required at the 10 field test sites. Maintenance was required at some test sites, but a careful review showed that, in each case, it was required by a malfunction external to the lighting system. All 10 systems performed equally well in that they did not require maintenance.

Power Consumption

During field testing, power consumption for each of the lighting systems was monitored monthly by using a wattmeter. The levels of energy consumption by the various lighting systems demonstrated little fluctuation over time. Although low-pressure sodium lamps are characterized by a gradual increase in power consumption over time, no trend was shown by the data.

Table 5 and Figure 3 present data on the power consumption by each lighting system. The current ADOT lighting system, which uses a fluorescent lamp, is represented as System 1. It had the highest level of energy consumption with an overall average of 531 W. This is in sharp contrast with the three most energy-efficient lighting systems (System 4, 183 W; System 10, 158 W; and System 6, 148 W). Each of these systems used a high-pressure sodium lamp.

ECONOMIC ANALYSIS

An economic analysis was performed to compare the 10 sign-lighting systems. The initial costs for fixtures and lamps and the labor and equipment for installation as well as the annual operating costs for electricity, washing, relamping, and ballast replacement were considered. The cost information used in the economic analysis is given in Table 6. The following points describe various inputs to the economic analysis.

- Prices for fixtures, lamps, and ballast replacement were obtained from local suppliers for purchases in both large and small quantities. (The values in Table 6 are for large quantities.)
- Installation cost was based on an ADOT estimate of the amount of time required to install fixtures. An ADOT labor rate of \$17.86/hr and an equipment rate (for a truck) of \$10.50/hr were used to calculate cost.
- A 10 percent interest rate was used.
- On the basis of ADOT experience with fluorescent lighting systems, all lighting systems were estimated to have a

TABLE 5 POWER CONSUMPTION

System Number	Light Source	Lamp Size	Fixture	Power Consumption (watts)	Power Consumption Compared to Standard Fluorescent System (Percent)
1	Fluorescent	800 milli-amp	Nu-Art NAFL	531	100
2	Clear Metal Halide	175 watt	Holophane Express-lite	385	73
4	Clear High Pressure Sodium	70 watt	Guth Sign-liter	183	35
5	Clear Metal Halide	175 watt	Guth Sign-liter	376	71
6	Clear High Pressure Sodium	70 watt	General Electric Versaflood II	148	28
7	Clear Metal Halide	175 watt	General Electric Versaflood II	432	81
8	Low Pressure Sodium	35 watt	Holophane Expresslite	289	54
10	Clear High Pressure Sodium	150 watt	Holophane Panel-Vue	158	30
11	Clear Metal Halide	175 watt	Holophane Panel-Vue	262	49
12	Clear Mercury Vapor	250 watt	Holophane Panel-Vue	282	53

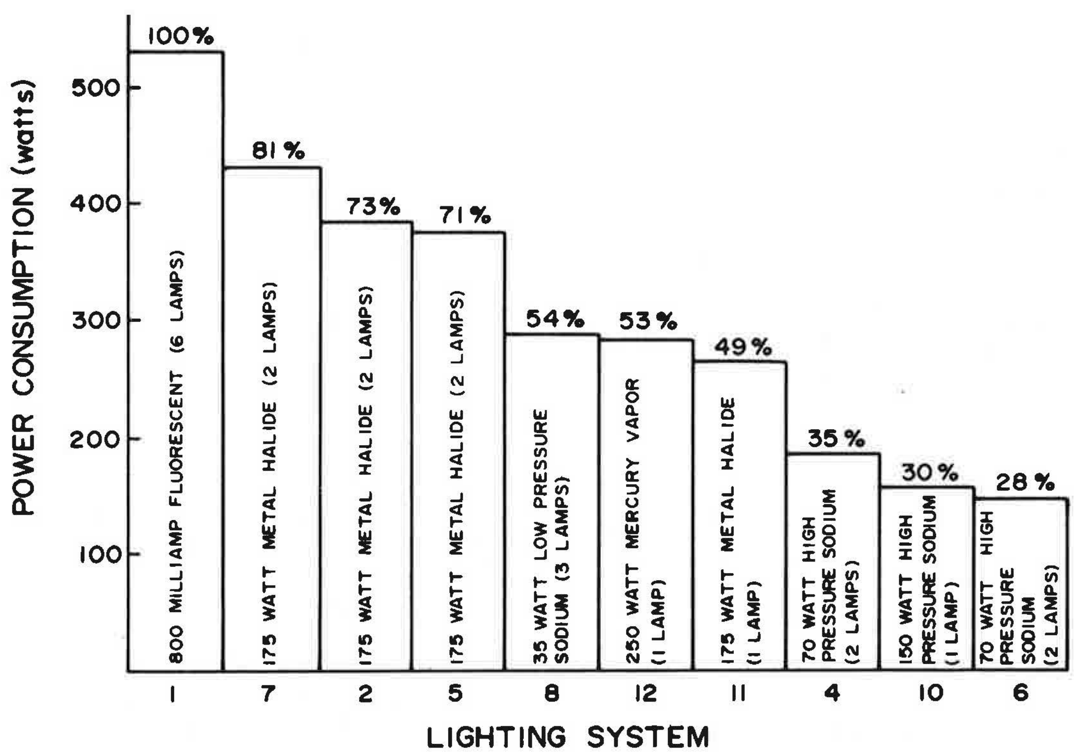


FIGURE 3 Relative power consumption.

TABLE 6 COST INFORMATION USED IN ECONOMIC ANALYSIS

System Number	10	12	11	6	4	8	5	2	7	1
Number of Fixtures	1	1	1	2	2	3	2	2	2	3
Cost per Fixture (includes ballast) ($)	175.00	175.00	175.00	233.33	185.00	150.00	175.00	150.00	206.67	330.00
Installation Cost per Fixture ($)	46.22	46.22	46.22	46.22	46.22	46.22	46.22	46.22	46.22	46.22
Number of Lamps per Fixture	1	1	1	1	1	1	1	1	1	2
Total Number of Lamps	1	1	1	2	2	3	2	2	2	6
Cost per Lamp ($)	33.55	20.79	27.23	31.08	31.08	10.80	27.23	27.23	27.23	5.06
Interest Rate (%)	10.	10.	10.	10.	10.	10.	10.	10.	10.	10.
System Life (Years)	20	20	20	20	20	20	20	20	20	20
Salvage Value (% of Initial Cost)	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
Power Consumption (Watts per Fixture)	158.	282.	262.	74.	92.	96.	188.	193.	216.	177.
Annual Operating Hours	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000
Power Price per Kilowatt-Hour (¢)	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5
Energy Cost Escalator (percent per year)	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
Maintenance Labor Rate ($ per Hour)	17.86	17.86	17.86	17.86	17.86	17.86	17.86	17.86	17.86	17.86
Time Required to Wash Lamp and Fixture or to Replace Lamp and Wash Fixture (Hours/Fixture)*	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Equipment Rate (Truck for Crew)($ per Hour)	10.50	10.50	10.50	10.50	10.50	10.50	10.50	10.50	10.50	10.50
Time Between Washings (Years)	1.5	1.5	1.0	1.5	1.5	1.0	1.0	1.0	1.0	1.0
Time Between Group Relamping (Years)	3.0	3.0	1.0	3.0	3.0	2.0	1.0	1.0	1.0	2.0
Number of Ballasts per Fixture	1	1	1	1	1	1	1	1	1	2
Estimated Ballast Life (Years)	12	12	12	12	12	12	12	12	12	12
Ballast Material Replacement Cost ($)	76.00	72.00	73.00	30.75	50.00	82.96	50.00	80.00	39.09	45.00
Time Required to Replace Ballast (Hours/Fixture)*	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8

*2 person crew for washing and relamping
1 person crew for ballast replacement

useful life of 20 years. The salvage value was assumed to be nil.

- Power consumption was based on actual experience during field testing. Annual operating time was 4,000 hr.
- ADOT currently purchases electric power at a weighted average cost of 8.5 cents/kW-hr.
- It was assumed that the cost of electric power would escalate no faster than the cost of labor and replacement parts.
- Current ADOT labor and equipment rates were also used for washing, relamping, and ballast replacement functions.
- The time required to wash fixtures and to relamp was estimated on the basis of ADOT's experience with fluorescent lighting systems.
- On the basis of ADOT's past practice of group relamping, this same practice was applied to all 10 lighting systems. The frequency of relamping was based on lamp life.
- The frequency of washing was based on the frequency of relamping and ADOT's past experience with dirt accumulation and washing needs.
- Based on manufacturer claims, a 12-year ballast life was

established for all systems. The time for ballast replacement was based on an ADOT estimate.

The computer program COSTLITE was used to calculate annual costs. The results for each of the 10 lighting systems are given in Table 7. COSTLITE calculates costs as follows:

Initial cost: Costs for a system's fixtures and lamps and their installation are determined.

Annual owning cost: A capital recovery factor for a 10 percent interest rate and a 20-year lifetime is applied to the initial cost.

Annual power cost: Power cost is based on consumption, hours of operation, and power price.

Annual washing cost: Washing cost is the time required multiplied by the labor and equipment rates and divided by the washing frequency.

Annual lamp replacement cost: Lamp replacement is lamp cost divided by the replacement period. Labor and equipment costs for lamp replacement are included in washing cost.

Annual ballast replacement cost: Ballast replacement is the time required multiplied by the labor and equipment rates. Ballast material replacement cost is added. The total is divided by the estimated ballast life.

Annual operating cost: Annual operating cost is the sum of the four preceding items. For none of these four items are increases in costs of labor, equipment, lamps, and ballast in future years considered. All annual costs are based on current prices.

Total annual owning and operating cost is the sum of *annual owning cost* and *annual operating cost*.

Review of Table 7 shows great differences in the annual costs of the 10 lighting systems (they are ranked in order of total annual costs). Total annual costs range from $115 per year to $423 per year. The following observations explain some of the dramatic differences in annual cost:

- Systems 10, 11, and 12 use only one fixture to illuminate a sign 8 ft high by 20 ft wide. Initial cost is considerably less than that for other systems. Conversely, System 1 requires three lighting fixtures and has a high initial cost.
- Systems 8, 5, 2, 7, and 1 have much higher annual operating costs. Four factors contribute to this: (*a*) these systems have higher power consumption, (*b*) they all require annual washing, (*c*) they have shorter lamp life than most of the other systems, and (*d*) the annual ballast replacement cost tends to be higher than that of the other systems.

It is emphasized that the cost information presented in Table 7 is for lighting a sign 20 ft wide. Systems 6, 4, 8, 5, 2, and 7 use two fixtures to light a sign of this width. For narrow signs these systems would be adequate with one fixture, and annual cost would be cut in half. For Systems 6 and 4 this would mean that the annual cost (approximately $90) would be even less than that of System 10.

The discussion thus far has compared the annual cost of 10 different lighting systems for new installations. The existing fluorescent system is inferior to all of the other nine alternatives, but it is also important to evaluate the economics of allowing the existing fluorescent lighting systems to remain in place versus replacing them with a different system. The last column in Table 7 shows the annual cost of operating an

TABLE 7 INITIAL COST, ANNUAL OWNING COST, AND ANNUAL OPERATING COST FOR EACH LIGHTING SYSTEM

System Number	10	12	11	6	4	8	5	2	7	1	**
Initial Cost	$254.77	$242.01	$248.45	$621.26	$524.60	$621.06	$496.90	$446.90	$560.24	$1159.02	$0.00
Annual Owning Cost	29.93	28.43	29.18	72.97	61.62	72.95	58.37	52.49	65.81	136.14	0.00
Annual Power Cost	50.56	90.24	83.84	47.36	58.88	92.16	120.32	123.52	138.24	169.92	169.92
Annual Washing Cost	15.13	15.13	22.69	30.25	30.25	68.06	45.38	45.38	45.38	68.06	68.06
Annual Lamp Replacement Cost	11.18	6.93	27.23	20.72	20.72	16.20	54.46	54.46	54.46	15.18	15.18
Annual Ballast Replacement Cost	8.22	7.89	7.97	8.91	12.11	26.41	12.11	17.11	10.30	33.84	40.61
Annual Operating Cost	85.09	120.19	141.73	107.24	121.97	202.84	232.27	240.47	248.37	287.01	293.78
Total Annual Owning and Operating Cost	$115.02	$148.61	$170.91	$180.21	$183.58	$275.79	$290.64	$292.96	$314.18	$423.15	$293.78

The costs shown are those for illuminating an 8 foot high by 20 foot wide sign. Systems are ranked in order of Total Annual Owning and Operating Cost.

**See Text

existing fluorescent system. It treats the initial cost of the system as a sunk cost that has already been expended and for which there is no annual owning cost. On the basis of information provided by ADOT, an average age of 10 years and a remaining useful life of 10 years is assumed. An annual operating cost of $294 is shown, a value nearly three times as large as the annual owning and operating cost of the most cost-effective system.

SELECTION OF A RECOMMENDED SIGN-LIGHTING SYSTEM

The rationale used to select a recommended sign-lighting system for use by ADOT is described. The various factors considered in the selection process are summarized in Table 8.

Many factors were evaluated in this study and considered in selecting a recommended system. Color rendition, lighting uniformity, and viewing comfort were evaluated by two observer groups. As shown in Table 8, three systems received overall ratings of marginal to poor by both the hired observers and the professional group. All other systems received an overall rating of either good or excellent from one or both of the two groups.

An important decision in the selection process is whether the high-pressure sodium light source has acceptable color rendition. In the observer study, a low relative importance was placed on color rendition. On the basis of the finding that lack of evidence that color rendition is important for overhead guide signs and the significant economic savings that can be achieved with high-pressure sodium, it was decided that this system does have acceptable color rendition. The research team also noted that four other states—Nebraska, Tennessee, Utah, and Virginia—are using high-pressure sodium for sign lighting.

All 10 lighting systems were about equal in legibility distance. Systems 2 and 10 had a slightly greater legibility distance. All 10 systems had satisfactory luminance levels.

Lamp life and maintenance requirements were considered as part of the economic analysis.

From an economic standpoint, it appeared that five systems should be considered—Systems 10, 12, 11, 6, and 4.

Considering all of the factors described earlier and summarized in Table 8, the following observations led to the selection of a recommended system:

- Systems 11 and 12 were very comparable in terms of observer group rating, legibility, and illumination level. System 12 was preferred due to its lower cost.
- Systems 4, 6, and 10 all use a high-pressure sodium lamp. Systems 4 and 6 have significantly higher annual costs than does System 10. They also received poorer ratings from the observer groups. Therefore, System 10 was preferred.
- A comparison of the two remaining systems showed that System 12 provided better color rendition and System 10 offered slightly more legibility distance. In view of the substantially lower annual cost, System 10 was selected as the preferred system.

Therefore, System 10 is recommended as the best overall lighting system.

ESTIMATED SAVINGS IN ARIZONA

It is estimated that the 699 existing illuminated signs (virtually all using fluorescent lighting) use 1,546 fluorescent fixtures. The annual operating cost for these 1,546 fixtures is $151,400. If they were converted to the recommended lighting system, the annual owning and operating cost would be $86,380. In addition to a lower annual cost, ADOT would have a lighting system in place with a 20-year life as compared with a remaining life of approximately 10 years for the fixtures now in place.

The initial investment for a conversion would be significant

TABLE 8 EVALUATION OF LIGHTING SYSTEMS

	LIGHTING SYSTEMS										
System Number	1	12	7	2	5	11	10	6	8	4	
Overall Rating by Observer Groups											
Hired Observers	○	○	○	○	○	○	**○**	**○**	**○**	**○**	
Professional Group		○	○	**○**	**○**	○	○	◉	◉	◉	
Legibility	○	○	○		○	○		○	○	○	Systems 2 and 10 had greater legibility distances
Illumination Level	○	○	○	○	○	○	○	○	○	○	All systems had satisfactory illumination levels
Total Annual Owning and Operating Cost*		$114,603**				$128,353	$86,380	$120,477		$122,723	

*Total Annual Owning and Operating Cost is for the 699 signs currently illuminated on the ADOT system. Costs are shown only for the five less costly systems.

**System 12, as field tested, used a 250 watt lamp. For the 699 illuminated ADOT signs it was found that use of a 175 watt lamp would be more economical. This size lamp would still provide adequate sign luminance. For definitions of the symbols used in Table 8, see Figure 2.

but would result in a relatively short payback period. Initial cost for fixtures, lamps, and installation for 699 signs would be $191,332. The annual savings in operating costs would be $87,497. Thus, the investment would pay for itself in less than 2½ years.

CONCLUSIONS

The study conclusions are as follows:

- ADTO spends annually about $87,500 in electric power costs, $106,000 for washing, and $23,500 for lamps for illuminating 699 overhead guide signs on freeways.
- There is no standard sign size on the Arizona freeway system. The great variety in sizes is a challenge in selecting the best sign-lighting system.
- All 10 sign-lighting systems tested provided satisfactory luminance. Only one of the systems had unsatisfactory lighting uniformity. All 10 systems had about the same legibility distance.
- Power consumption can be greatly reduced by using high-pressure sodium as a light source.
- All nine of the alternative lighting systems tested have substantially lower owning and operating costs than the standard fluorescent system.
- Conversion of existing sign-lighting systems from fluorescent lighting to System 10 would reduce annual operating cost from $151,400 to $63,903. The initial investment to conduct the conversion would be $191,332.
- Use of the recommended lighting system on future installations would save an average of $189 per sign in annual owning and operating costs for the state of Arizona.
- It should be noted that the results of this sign-lighting research were influenced by the needs and requirements of the state of Arizona and the particular methodology and techniques of this research. Other states and operating agencies may find different results if other requirements and research methodologies are selected.

ACKNOWLEDGMENT

This research was sponsored by the Arizona Department of Transportation in cooperation with FHWA. More detailed documentation on this research project may be found in the project final report *(5)*.

REFERENCES

1. *Manual on Uniform Traffic Control Devices for Streets and Highways.* FHWA, U.S. Department of Transportation, 1978.
2. *Recommended Practice for Roadway Sign Lighting.* Illuminating Engineering Society, New York, 1982.
3. *An Informational Guide for Roadway Lighting.* AASHTO, Washington, D.C., March 1976.
4. *Guide for Photometric Measurements of Roadway Sign Installations.* Illuminating Engineering Society, New York, 1975.
5. J. E. Upchurch and J. T. Bordin. *Evaluation of Alternative Lighting Systems for Guide Sign Illumination.* Report FHWA/AZ-86/200. FHWA, U.S. Department of Transportation, April 1986.

Publication of this paper sponsored by Committee on Visibility.

DISCUSSION

MICHAEL S. JANOFF
JMJ Research, P.O. Box 144, Newtown, Pa. 18940.

This is a very interesting paper that combines a somewhat novel experimental test method with an economic analysis based on field data to determine a preferred lighting system for highway signs.

However, the results stated by the authors concerning their choice of "best" lighting system and their recommendations for its proposed use by Arizona and other states are, I believe, based on both insufficient evidence and potential problems in the experimental research.

My comments are primarily addressed to two aspects of the paper: (*a*) the economic analysis and (*b*) the field experiment.

The economic analysis, which is used to support the far-ranging recommendations, is based on test data from only one lighting system of each type, and furthermore, such lighting systems were only evaluated for one size sign.

The authors' measurements of power consumption differ from manufacturer's published specifications for many of the studied lamps. Are we to believe that one field measurement is more valid than extensive laboratory testing? Clearly, more field testing was required.

The number of lamps in each lighting system is the single most influential factor in defining the initial operating and maintenance costs of each system, but the number is dependent on the size of sign selected for study.

If a sign of a different size were studied, the costs would change radically. For example, a slightly larger sign would require two 150-W high-pressure sodium (HPS) lamps in the "best" lighting system, significantly increasing the costs of this system but having only a marginal effect on the costs of the system employing 35-W low-pressure sodium (LPS) lamps. Similarly, a smaller sign would still require one 150-W HPS lamp but fewer 35-W LPS lamps. Such changes would drastically alter the economic results and hence the choice of "best" system. To suggest that the 150-W HPS system is preferred for all applications is thus unsupported by the research, especially because the range of sign sizes in Arizona, and most states, is quite large, and not limited to the size of sign selected for study.

The experimental design includes a number of facets that are either invalid or at least unexplained in both the paper and the referenced final report.

The 10 sign locations were not described, other than their location on the freeway. If the backgrounds against which the signs were viewed by the test subjects differed in visual complexity, the experimental results could have been confounded by these differences and this would invalidate the selection of "best" lighting system. The lack of significant differences in legibility distances reported by the authors could have resulted from this problem (and the ones discussed subsequently). If, for example, the LPS system—a poor performer—was viewed against a complex background and the 150-W HPS system—the "best"—was viewed against a simple background, the background itself could have dominated the subjective ratings of legibility.

The authors state that all lighting was evaluated in the preliminary studies against a "medium" ambient light background but were all field locations carefully checked for similarity? Can the authors support the necessary similarity of backgrounds in any objective manner?

The subjective, and subject-controlled, method of measuring legibility distances that the authors employed is somewhat novel and may have induced potential problems related to the accuracy and repeatability of the legibility distances. Classic sign legibility research by Forbes and others employed test methods that were quite different than the one reported here. Did the authors investigate alternative methods or attempt to test their methodology to determine its repeatability, accuracy, and validity? What led the authors to select such a method?

The test subjects all viewed the 10 signs in the same order of presentation, which could have resulted in a learning curve that biased the results. Counterbalancing the order of presentation would have been preferred.

The rating scale resulted in a narrow range of subjective ratings (e.g., as low as 2.8 to 4.4), indicating possible central tendency effects that might have been eliminated by better instructions (which were not described), a better rating scale, and other, better psychophysical testing procedures. No statistics are presented to support conclusions or indicate the significance of the differences in lighting system performance.

The authors state that field experiments began about 1/2 hr after sunset and continued thereafter. It is my experience—and published sky luminance values support it—that the sky on a clear day has considerable brightness at 1/2 hr after sunset, not reaching full darkness until at least 1 hr after sunset. Such sky luminance differences could have influenced the results.

Other comments include the lack of any descriptive material supporting the reduction in number from 100 to 25 lighting systems in the preliminary analysis; the lack of information describing Lewin's analysis methodology or the SITELITE results; the choice of only one ambient lighting background (medium) rather than many; the choice of only one sign size, resulting in the bias described previously; and the use of a lighting fixture not designed for the LPS lamp in a sign-lighting application.

In conclusion, it is my belief that the potential problems just described may have invalidated the results of this research, and that the recommendation that Arizona and other states use such a sign-lighting system is unsupported.

Either the present paper should be rewritten to limit the results to those obtained under the exact study conditions (which still need better explanation) and should exclude the far-ranging recommendation for use of such a lighting system for all highway signs, or my comments should be included with the published paper to provide the prospective user of these results with a very different interpretation of the research and its implications.

AUTHORS' CLOSURE

The comments by Michael Janoff are greatly appreciated. His comments stimulate discussion on this important topic and offer the opportunity to present additional information on this research project.

Janoff notes that in the economic analysis, the cost data presented are based on only one lighting system of each type and implies that a larger sample size should have been used for determining system cost. Table 6 presents all the cost factors that went into the economic analysis. Of those several factors, the only one that would change if a larger number of lighting systems of a given type were evaluated is power consumption.

Janoff states that the measured power consumption differs from manufacturer's published specifications. There are differences between the measured power consumption and the rated lamp wattage. Considering ballast losses, however, only one system is substantially different. The power consumption for System 8 (low-pressure sodium) is much greater than expected and this discrepancy cannot be explained.

A sensitivity analysis shows that major changes in power consumption would be required to change the rank order of the 10 systems in the overall economic analysis. If the power consumption of the low-pressure sodium system is, in fact, about 135 W (approximately the expected value based on rated wattage), there would still be five other systems that had lower annual owning and operating costs.

We agree with Janoff's statement that there is great variety in sign sizes and this is supported by an inventory of sign sizes conducted in the study. As a part of the economic analysis, we did consider the fact that the lighting system that is most economical for a sign 20 ft wide would not necessarily be most economical for a different size of sign. Systems 4 and 6 would be more economical for a sign less than 10 ft wide because only one fixture would be required. The Arizona Department of Transportation preferred to use one lighting system for all signs to simplify parts inventory and maintenance. Based upon the mixture of sign sizes in Arizona, System 10 provided the lowest overall cost.

Janoff notes that differences in background complexity at the 10 test sites could have affected the evaluation of the systems by the observers. Every possible attempt was made in this study to have the 10 test sites identical in terms of approach geometry, mounting height, size of sign, amount of legend, and ambient illumination. We believed that the most important characteristic was the size of the sign face. All other factors being equal, a change in the size of the sign face can result in significant changes in overall luminance levels and in lighting uniformity. We believed that it was desirable to have 10 test sites where the signs were close to the same size, where all signs had three lines of legend, where signs were mounted individually (rather than in pairs), and where test site locations were relatively close together for convenience in observer studies. Ambient illumination levels were comparable at all locations. Although it cannot be proven on an objective basis, backgrounds were similar at the 10 test sites.

The method of measuring legibility distances was selected for simplicity. The test method employed by Forbes (sign-reading errors) may be more rigorous. One advantage of the method employed in this project is that the observers were approaching the signs at highway speed—a more realistic condition. The stopwatch method used in this study has also been used in similar signing studies conducted by the Texas Department of Highways and Public Transportation.

We agree that counterbalancing the order of presentation of the 10 test sites would have been a more rigorous approach.

It is our opinion that twilight sky luminance had no impact on the observer studies. All observations at actual test sites were made more than 1 hr after sunset. In the urban area test site environment skyglow caused by urban lighting over-powered any twilight sky luminance at 1 hr after sunset.

Evaluation of the Effectiveness of Crash Cushion Delineation

F. THOMAS CREASEY, CONRAD L. DUDEK, AND R. DALE HUCHINGSON

The objective of this study was to evaluate the effectiveness of a limited number of crash cushion delineation techniques in the field. Three candidate treatments were selected for field testing: *(a)* a yellow diamond-shaped object marker, *(b)* a yellow-and-black chevron-patterned nose panel, and *(c)* yellow-and-black chevron-patterned nose and back panels. Because accidents involving crash cushions are relatively rare events, it is difficult to make statistically valid comparisons. In this study vehicle encroachments into the gore area were considered to be indicators of the potential for accidents with crash cushions. Studies were conducted at three sites in El Paso, Texas. A low-light-level camera and time-lapse video recorder were used to collect continuous encroachment and traffic volume data at the sites. Three candidate delineation treatments and the existing delineation treatment were tested at each of the study sites. A classification system was developed to differentiate among the gore sites on the basis of the geometrics of the gore approach. Data were collected over a 3-day period for each of the candidate treatments and for the existing treatment at the three sites. Crossover rates were used to compare the effectiveness of the delineation treatments. Analysis of the data indicated no difference in crossover rates among the treatments. The results, based on a limited sample, suggest that added delineation did not reduce crossover rates at locations where sight distance was not a critical factor and that accident problems at these sites may not be related to poor conspicuity alone, but instead may have also been influenced by informational deficiencies in signing and markings.

The use of crash cushions (impact attenuators) to protect vehicles from crashes with fixed objects in freeway gore areas has become a widespread practice. Use of crash cushions has been shown to reduce impact severity *(1)*. However, crash cushions increase the frequency of accidents. This increase may result from reducing the area of the recovery zone, reducing decision or reaction time or both, or simply adding another fixed object in the roadway environment for vehicles to strike. Although crash cushions reduce fatalities and injury severity, collisions with crash cushions may lead to serious secondary accidents or disruptions in traffic flow. There is also a risk to maintenance personnel who are exposed to traffic during repair operations. Thus, the safety benefits derived from crash cushion use are offset to some degree by increased maintenance, labor, and operational costs.

A possible reason that some impact attenuators are more frequently struck than others may be a lack of conspicuity in gore areas. Drivers having to simultaneously process complex information inputs from geometric features, signing, and markings and from other vehicles in the traffic stream may fail to distinguish a gore area or crash cushion embedded in the visual

F. T. Creasey and C. L. Dudek, Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843. R. D. Huchingson, Department of Industrial Engineering, Texas A&M University, College Station, Tex. 77843.

field and may strike the cushion while entering, exiting, or making evasive maneuvers. Thus, improving crash cushion conspicuity in gore areas by providing effective delineation may be helpful in reducing certain accidents in which drivers fail to perceive the presence of the crash cushion.

OBJECTIVE

The objective of the field studies was to evaluate the effectiveness of a limited number of crash cushion delineation techniques. The purpose of the field evaluation was not to determine the best treatment, but rather to determine those treatments resulting from the laboratory studies *(2)* that are effective in the field. Thus, it was not necessary to test differences between treatments, but rather to test operational differences resulting from a candidate delineation treatment. Three candidate delineation treatments—a yellow diamond-shaped object marker, a yellow-and-black nose panel, and yellow-and-black nose and back panels—were selected for field testing.

Both short- and long-term analyses were to be conducted. The short-term analysis included a study of driver performance. The long-term analysis involved visual field inspections of the delineation treatments after 4 to 6 months. Limited funding prevented traffic stream measurements during the long-term evaluations.

DATA COLLECTION

Selection and Classification of Study Sites

El Paso, Texas, was selected as the location for the study. The El Paso District Office of the Texas State Department of Highways and Public Transportation (SDHPT) identified three sites for the study as the most frequently hit crash cushions. All three sites were located within the Interstate 10–US-54 interchange near downtown El Paso. The study sites and existing delineation treatments before the installation of the test treatments are described in the following paragraphs.

Site 1: I-10 Westbound at US-54

This location, referred to as Interchange Ramp A, is the exit ramp for all US-54 traffic from westbound I-10. The existing delineation treatment consisted of a black-and-white chevron-patterned wraparound nose panel and a Type 1 diamond-shaped object marker as specified by the *Manual on Uniform Traffic Control Devices* (MUTCD) *(3)* (Figure 1). There were three crashes involving repairs at this location between May 1983 and July 1985.

Site 2: I-10 Westbound at US-54 East-West Split

Referred to as Interchange Ramp A-F, this site is located at the split of US-54 immediately downstream from Site 1 (Interchange Ramp A), with the eastbound (right-hand) split heading

FIGURE 1 Existing delineation treatment at Site 1 (Interchange Ramp A).

toward New Mexico and the westbound (left-hand) split heading toward Juarez, Mexico. The existing delineation treatment consisted of a black-and-white chevron-patterned wraparound nose panel, an MUTCD Type 1 object marker in the front, and two MUTCD Type 2 object markers, one vertical and one horizontal, in the rear (Figure 2). The crash cushion installation at this site was struck twice in 1985 (records of repairs that may have been made before then were not available).

Site 3: I-10 Eastbound Entrance Ramp (Gateway Boulevard East) at Copia Street

This location, referred to as the Copia Street Ramp site, is a left-hand entrance ramp from the frontage road onto I-10 eastbound. The existing delineation consisted of a black-and-white chevron-patterned wraparound nose panel, an MUTCD Type 1 diamond-shaped object marker, and two rows of small rectangular yellow reflective-tape sections arranged in a checkerboard pattern (Figure 3). The crash barrels at this site were repaired five times between July 1982 and July 1985.

Delineation Treatments

Four crash cushion delineation treatments were studied at each site:

FIGURE 2 Existing delineation treatment at Site 2 (Interchange Ramp A-F).

FIGURE 3 Existing delineation treatment at Site 3 (Copia Street).

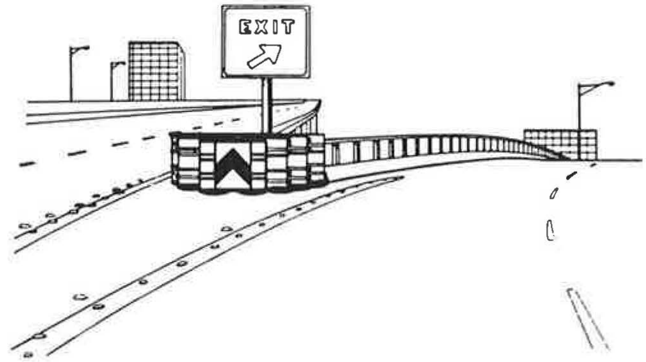


FIGURE 5 Nose panel delineation treatment.

1. Existing,
2. Object marker,
3. Nose panel, and
4. Nose panel and back panel.

The first test treatment was the existing treatment previously discussed. The second test treatment consisted only of an all-yellow, diamond-shaped 18- × 18-in. MUTCD Type 1 object marker. The marker was mounted on a small sign post with its bottom tip located at the top surface of the front crash barrels (Figure 4).

The third test treatment was a 2- × 3-ft yellow-and-black chevron nose panel (Figure 5). High-intensity reflective sheeting was used for the yellow portions of the panel.

The fourth experimental treatment combined the nose panel mentioned earlier with an 8- × 8-ft yellow-and-black chevron-patterned back panel (Figure 6). The back panel also utilized high-intensity reflective sheeting and was mounted behind the back barrels of the crash cushion with its bottom edge flush with the top of the barrels.

Measures of Effectiveness (MOEs)

The effects of the delineation treatments should be evaluated either directly in terms of accident reduction or indirectly through surrogate measures. From a literature review to examine MOEs used in past gore area studies to quantify driver behavior and traffic performance, the following four MOEs were identified:

1. Accidents,
2. Repair history,
3. Erratic maneuvers, and
4. Gore intrusions (encroachments).

The most direct MOE is the number of vehicles colliding with the crash cushion during a specified period. However, there were some practical limitations to using accidents as an MOE in this study. First of all, the number of vehicular collisions with a crash cushion at a given gore area within the 4- to 6-month field test period available in this study was expected to be too small for statistical testing. Second, gore area accident records are not always available. Although the literature did not provide any definitive answers as to the most effective MOE, crash cushion repair history and gore intrusions (encroachments) were initially selected as the MOEs in this study because they appeared to be the most promising alternatives.

Encroachments were classified as either crossover or sideswipe. Four types of crossovers and two types of sideswipe encroachments considered in this study are shown in Figures 7 and 8.

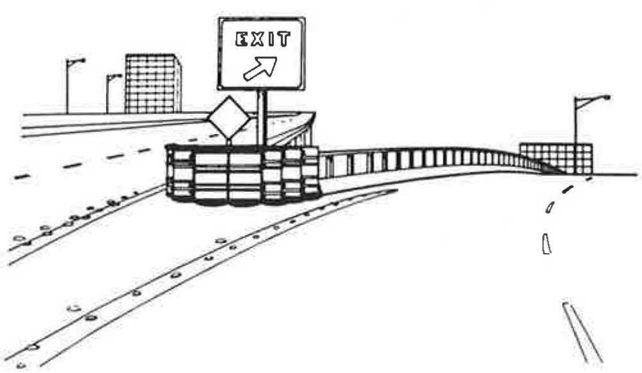


FIGURE 4 Object marker delineation treatment.

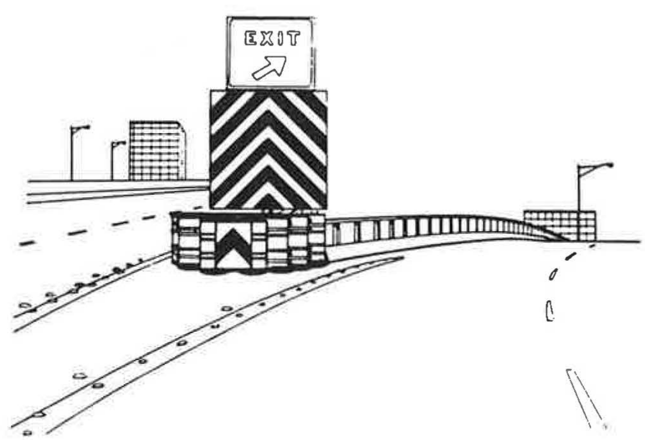


FIGURE 6 Nose and back panel delineation treatment.

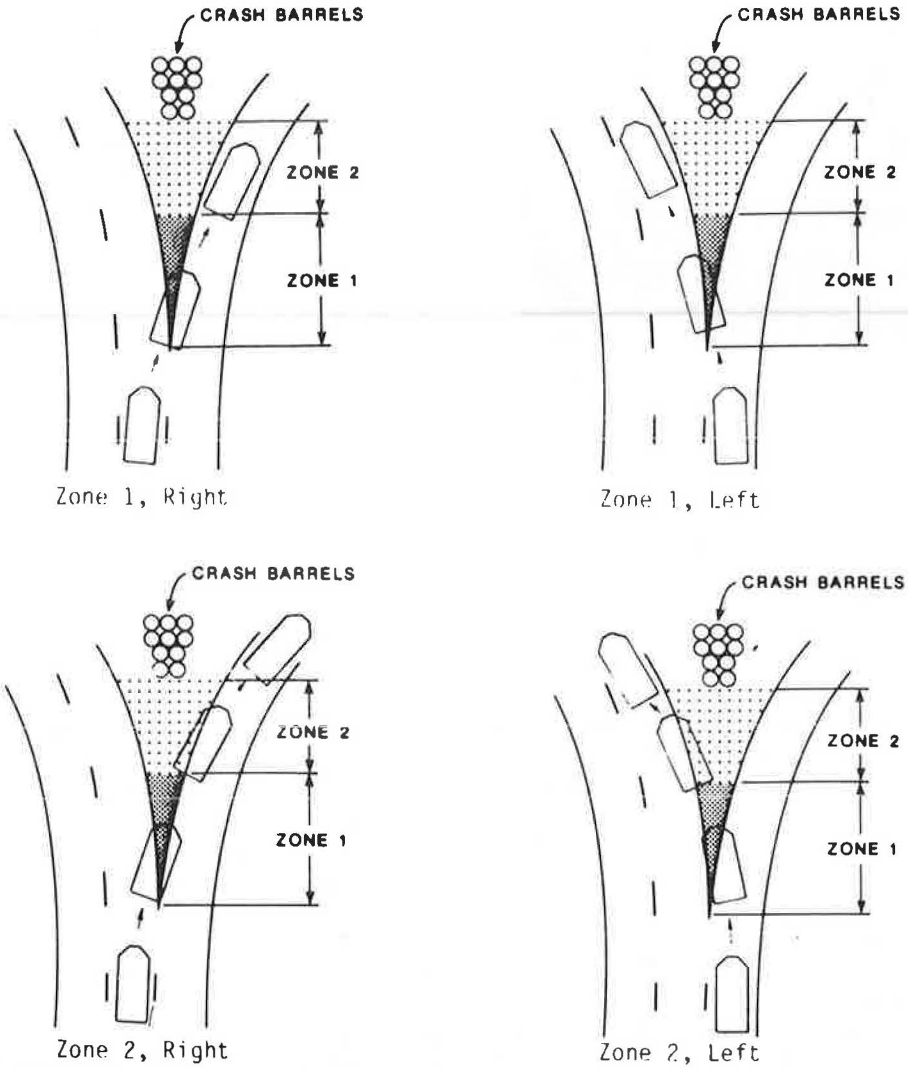


FIGURE 7 Crossover encroachments.

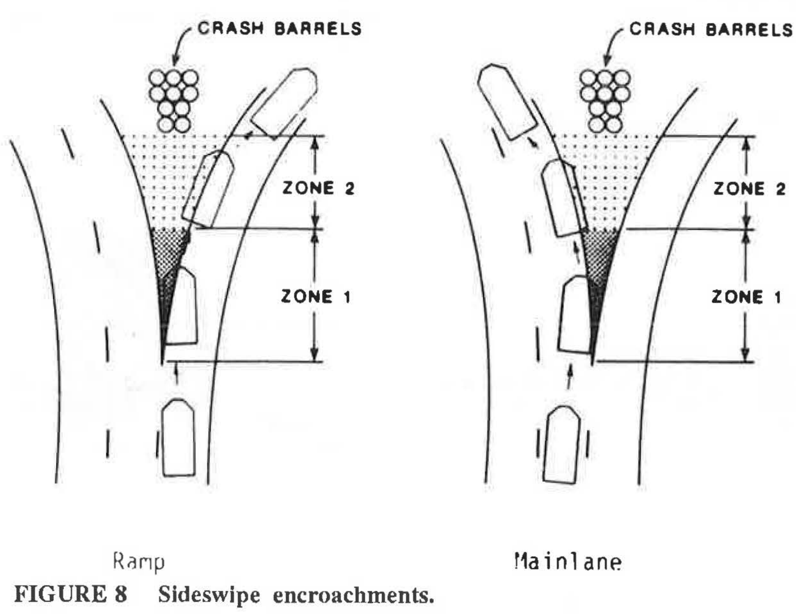


FIGURE 8 Sideswipe encroachments.

Experimental Design

The willingness of the El Paso District to install more than one treatment at each site, with certain restrictions, prompted an experimental design that allowed each delineation treatment to be studied at each of three different sites. A major restriction was that the District was not receptive to leaving the object marker treatment at any of the gore areas for prolonged time periods (more than one week) because of a concern for safety. In addition, the District did not want the object marker installed after either of the two experimental treatments. It was the opinion of District engineers that the object marker treatment was a step down from the existing treatments.

The experimental design is shown in Table 1. The fourth (last) treatment for each site was scheduled to remain at the site for approximately 4 to 6 months in order to conduct long-term visual evaluations. The insistence by the District that the object marker not remain at a site longer than one week or that the object marker not be installed after either of the other two experimental treatments required another revision to the experimental design. Note in Table 1 that only the nose panel alone and nose panel plus back panel were varied in order from site to site.

TABLE 1 EXPERIMENTAL DESIGN

	Treatment Order by Site		
	Site 1 (Ramp A)	Site 2 (Ramp A-F)	Site 3 (Copia St. Ramp)
1	Existing	Existing	Existing
2	Object marker	Object marker	Object marker
3	Nose panel	Nose and back panel	Nose and back panel
4	Nose and back panel	Nose panel	Nose panel

Data Collection

Equipment and Installation

A low-light-level video camera and time-lapse recorder were used to collect the data. The only available camera mounting location for the Ramp A and A-F studies was on a traffic light mast-arm at an intersection southeast of the ramps. Unfortunately, this location was to the side of the gore areas and, as discussed later, this presented some problems with respect to determining sideswipe encroachments.

Scheduling

At each study site, data were collected for four delineation treatments: existing, object marker, nose panel, and nose and back panel. It was desirable to collect data on nights with the highest traffic volumes (to obtain the largest possible sample size). Thus, data were collected on from Wednesday through Friday each week, beginning at approximately 9:00 a.m. on Wednesday morning and continuing until approximately 9:00 a.m. on Saturday morning.

Due to project time constraints, only one full week was devoted to data collection for each candidate treatment at each site. Approximately 72 hr of continuous time-lapse data were collected for each treatment at each site.

CLASSIFICATION OF GORE AREAS

Before the research conducted in El Paso, the Texas Transportation Institute (TTI) completed studies in Houston and Fort Worth *(4, 5)* in which the short-term effects of alternative delineation treatments were evaluated. These studies produced inconsistent results, which prompted TTI to evaluate other factors that might in some way have affected the consistency of the results. It was hypothesized that two major factors could be influential:

1. Total driver information, and
2. Geometrics of approach to the gore area.

Drivers are guided in large part by the formal information (i.e., information provided by signs and markings and by the location and positioning of signs and markings) provided on a highway. Poor information or poorly placed information can have a detrimental effect on driver behavior and could lead to erratic behavior (encroachments) at gore areas. Adequate delineation of gore areas may not be able to offset the erratic behavior caused by insufficient advance information. Study of the total driver information system is outside the scope of the research reported here.

Geometrics also play an important role in driver behavior and, alone or in combination with inadequate driver information, can lead to erratic driving behavior at gore areas. In further analyzing the results of the Houston and Fort Worth gore area studies, it became apparent that delineation requirements may not be the same at all gore areas. Because of geometrics and inadequate sight distances, certain types of gore areas may require extensive delineation, whereas locations with adequate sight distance may require lower levels of delineation. This hypothesis prompted TTI to develop a classification system for gore areas. The classification for right-hand exits is shown in Figure 9. A similar classification could be developed for left-hand exits.

Type I gore area represents a typical gore location with tangent alignment of the main roadway and a well-designed exit ramp. There are no unusual geometric features (e.g., lane drops) and sight distance to the gore area is 1,500 ft or greater. Sight distances of 1,500 ft have been found to provide adequate response time on high-speed facilities *(6, 7)*. Sight distances less than 1,500 ft could result in operational problems.

Type II gore area represents similar conditions to Type I with the exception that sight distance is restricted (e.g., by an overpass). Type IIa represents gore areas in which the sight distance is between 800 and 1,500 ft. Type IIb gore areas have sight distances less than 800 ft. Type II gore areas are more critical than Type I because of the more restricted sight distances. It is likely that Type II gore areas will require more extensive

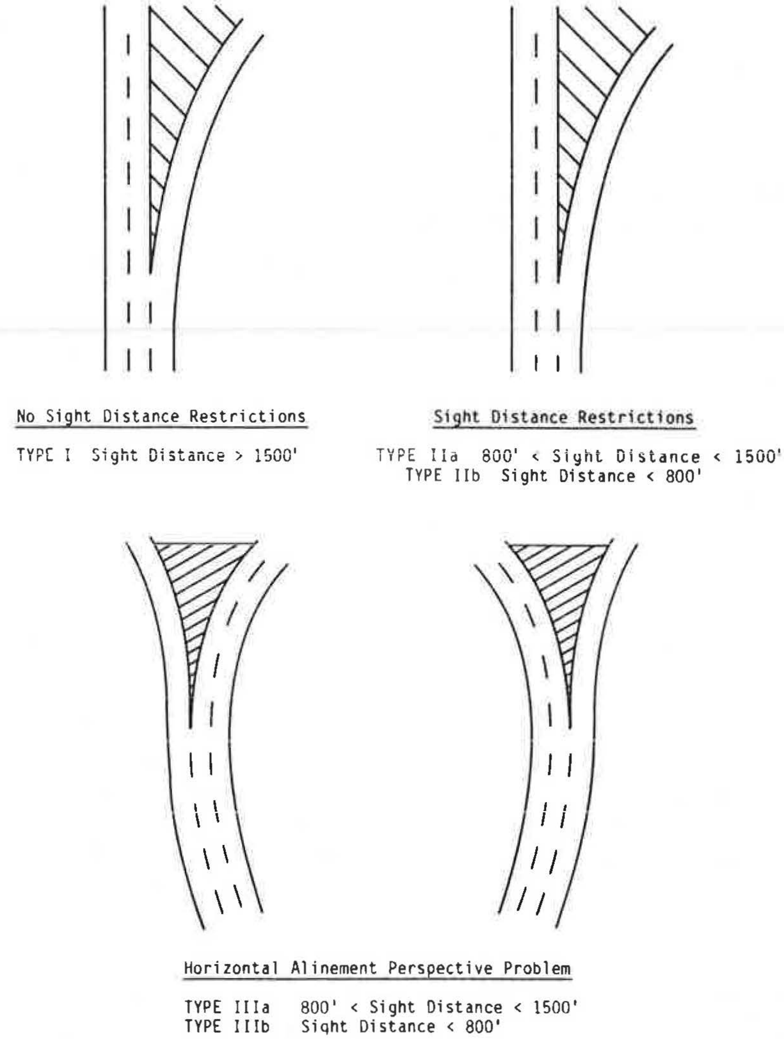


FIGURE 9 Gore area classification system.

delineation treatments than Type I. For example, a delineated back panel may be required to increase the effective sight distance to the gore area for Type II, whereas sight distance is not a problem for Type I and therefore a back panel may not be necessary.

Type III gore areas introduce another geometric feature—curvature—which, in combination with lane drops, lane additions, and so on, results in a visual perspective that may be confusing to the driver. Although Type I and Type II direct the driver past the gore area (either to the left or the right), Type III directs the driver, for a period of time, into the gore area (either into the nose or the side of the crash cushions). The roadway abruptly changes to move the driver away from the gore. However, the perspective problem in combination with inadequate (less than 1,500 ft) sight distance often lead to gore area accidents. It is possible that the perspective and sight distance problems cannot be solved by increased gore area delineation alone. Improvements to the communication system or in some cases improvements in geometrics may be necessary.

Type IIIa gore area contains the characteristics noted earlier with sight distance between 800 and 1,500 ft. The sight distance to Type IIIb gore area is less than 800 ft.

An examination of the conditions in El Paso indicates that the three gore area study sites may be classified as follows.

Site 1, Ramp A—Type IIa;
Site 2, Ramp A-F—Type IIIa; and
Site 3, Copia Street Ramp—Type IIa (left-hand "exit").

A driver's perspective while approaching Site 1 is shown in Figure 10. Similar perspectives of approaches to Sites 2 and 3 are shown in Figures 11 and 12, respectively.

ANALYSIS OF DATA

Sample Periods

From the data collected for each day, a 7-hr nighttime sample period (9:00 p.m. to 4:00 a.m.) and a 7-hr daytime sample period (9:00 a.m. to 4:00 p.m.) were selected for the purpose of analysis. These periods were selected for two reasons. First, peak traffic periods were not included, eliminating the effects of heavy traffic volumes (e.g., close following, abrupt slowing

FIGURE 10 Driver perspective while approaching Site 1.

FIGURE 12 Driver perspective while approaching Site 3.

or stopping, and swerving). Second, the selected periods excluded the transition in lighting conditions that occurs during dawn and dusk hours.

Encroachments

The number of crossovers and sideswipes was totaled to determine the number of encroachments during the nighttime and daytime data collection periods. An analysis of the data, however, revealed serious inconsistencies in the sideswipe data, which prompted close scrutiny of the data reduction process.

It became apparent that the side viewing angle of Site 1 (Ramp A) and Site 2 (Ramp A-F) coupled with video pictures of less than top quality made it difficult to consistently identify sideswipes, particularly when the right tires encroached into the gore area. The video camera was mounted at the best possible locations for the field studies. Unfortunately, the only practical camera location for Sites 1 and 2 was to the side of the sites. Field inspections before the field studies indicated that sideswipes could be identified in spite of the viewing angle. However, losing the three-dimensional perspective while viewing the scenes on a monitor that had a picture of less than high quality made it extremely difficult to identify sideswipes. Consequently, a decision was made to remove the sideswipe data from further analysis and to focus entirely on crossover encroachments. The loss of sideswipe data was considered to be less important than the loss of crossover data because with

FIGURE 11 Driver perspective while approaching Site 2.

sideswipes the driver was likely to be in the correct lane, whereas with crossovers the driver was more likely to be confused, leading to a late lane change.

Crossover rates were calculated by dividing the sum of all crossovers during the time period (nighttime or daytime) by the sum of the traffic volumes in the two lanes bordering each side of the gore area. The assumption was made that vehicles traveling in the lanes bordering the gore area would be more likely to cross the gore area.

Statistical Tests

A gore area crossover is a relatively rare event. In general, rates of relatively rare events can be assumed to follow a Poisson distribution. Under the assumption that a vehicle crossover is a Poisson random variable, the crossover rate at a particular site can be considered to be a measure of the average rate of occurrence. The fact that a Poisson distribution has equal mean and variance allows for use of the chi-square test statistic in testing for significant differences among crossover rates for different delineation treatments.

At each of the three study sites, the first step was to test the hypothesis that the average crossover rates for Wednesday, Thursday, and Friday (nighttime and daytime periods) were not significantly different for the delineation treatment in question. If the crossover rates for all three nights or days were not significantly different from each other, then the overall crossover rate for the delineation treatment could be compared with the overall rates for the other treatments (meeting the same criteria) to determine whether any particular delineation treatment was better than the others from a statistical standpoint.

However, if the overall crossover rate chi-square value for a specific delineation treatment exceeded the critical value for the appropriate level of confidence (95 percent) and degrees of freedom, it indicated that one of the nightly or daily rates was drawn from a different population than the other samples. Thus, the overall rate for the treatment could not be considered a good estimate of the crossover rate for that treatment and any comparisons using that overall rate would not be statistically valid. For example, if the Friday night crossover rate for the object marker treatment at one of the study sites was drastically different from the Wednesday and Thursday night rates, the average overall rate for all three nights would not be a good estimate of the crossover rate for that treatment, and would not be valid for statistical comparison with other treatments.

RESULTS

Total Crossovers

Site 1: Ramp A

A chi-square test on the crossover rates for individual nights and days was performed to test the consistency of the individual rates. There was no significant difference among the nighttime or daytime rates within any of the treatments. Therefore the data for individual nights and days were combined to obtain overall rates.

A summary of Site 1 data is shown in Table 2. A malfunction in the video system (assumed to have been caused by a power outage) caused the daytime data sample for the existing treatment on Thursday to be reduced by about 50 percent and the Thursday nighttime data to be totally lost. Overall nighttime crossover encroachment rates were calculated to be 1.1, 0.7, 0.6, and 0.6 crossovers per 1,000 vehicles for existing, object marker, nose panel, and nose and back panel treatments, respectively. Overall daytime crossover rates were calculated to be 1.2, 0.4, 0.6, and 0.5 crossovers per 1,000 vehicles for existing, object marker, nose panel, and nose and back panel treatments, respectively.

TABLE 2 SUMMARY OF SITE 1 (Ramp A) CROSSOVER ENCROACHMENTS

Treatment	Total No. of Crossovers	Sample Period Volumes Combined	Rate (cross/ 1,000 vehicles)
Nighttime (9:00 p.m. to 4:00 a.m.)[a]			
Existing	5[b]	4,403[b]	1.1
Object marker	5	7,058	0.7
Nose panel	4	6,544	0.6
Nose and back panel	5	7,953	0.6
Daytime (9:00 a.m. to 4:00 p.m.)[c]			
Existing	20[d]	17,363[d]	1.2
Object marker	9	21,476	0.4
Nose panel	10	18,141	0.6
Nose and back panel	10	20,122	0.5

[a]No test could be performed.
[b]Thursday night data not available because of video system malfunction.
[c]$\chi^2 = 9.69$;
[d]Thursday data sample size reduced because of video system malfunction.

The nighttime crossover rates could not be compared among treatments because of the low crossover frequencies. There was a significant difference in the daytime rates among treatments [$\chi^2 = 9.68$ ($p = .02$)], with the existing treatment having a higher crossover rate than the other treatments. However, there was no significant difference among the other three treatment rates (object marker, nose panel, and the nose and back panel).

Site 2: Ramp A-F

A summary of Site 2 crossover data is shown in Table 3. Wednesday nighttime existing treatment data were excluded because of rain and fog. Overall nighttime crossover rates were calculated to be 1.0, 0.7, 2.0, and 2.6 crossovers per 1,000 vehicles for existing, object marker, nose panel, and nose and back panel treatments, respectively. Overall daytime crossover rates were calculated to be 2.1, 2.3, 2.8, and 3.2 crossovers per 1,000 vehicles for existing, object marker, nose panel, and nose and back panel treatments, respectively. A chi-square test on the crossover rates for individual nights and days during each treatment showed the data to be consistent in each situation. A comparison was made among the four treatments and no statistically significant difference was found among them for nighttime or daytime conditions [$\chi^2 = 7.00$ ($p = .07$) and 4.12 ($p = .25$)].

TABLE 3 SUMMARY OF SITE 2 (Ramp A-F) CROSSOVER ENCROACHMENTS

Treatment	Total No. of Crossovers	Sample Period Volumes Combined	Rate (cross/ 1,000 vehicles)
Nighttime (9:00 p.m. to 4:00 a.m.)[a]			
Existing	4[b]	4,119[b]	1.0
Object marker	4	5,354	0.7
Nose panel	10	4,927	2.0
Nose and back panel	13	5,072	2.6
Daytime (9:00 a.m. to 4:00 p.m.)[c]			
Existing	30	14,475	2.1
Object marker	33	14,299	2.3
Nose panel	41	14,630	2.8
Nose and back panel	44	13,798	3.2

[a]$\chi^2 = 7.00$; $p = .07$.
[b]Wednesday night data excluded because of rain and fog.
[c]$\chi^2 = 4.12$; $p = .25$.

Site 3: Copia Street Ramp

A summary of Site 3 crossover data are shown in Table 4. The existing-treatment data for Wednesday nighttime and daytime periods were not available because of technical difficulties, and the nose panel treatment data for the Wednesday daytime period could not be used because of rain. Overall nighttime crossover rates were calculated to be 4.2, 3.7, 2.9, and 3.9 crossovers per 1,000 vehicles for existing, object marker, nose panel, and nose and back panel treatments, respectively. Overall daytime crossover rates were calculated to be 4.3, 3.1, 3.0, and 4.2 crossovers per 1,000 vehicles for existing, object marker, nose panel, and nose and back panel treatments, respectively. A chi-square test on the crossover rates for individual nights and days during each treatment showed the data to be consistent in each situation, meaning that all four overall rates were considered to be good estimates of crossover rates for the four delineation treatments. A comparison was made among the four treatments and no statistically significant difference

TABLE 4 SUMMARY OF SITE 3 (Copia Street) CROSSOVER ENCROACHMENTS

Treatment	Total No. of Crossovers	Sample Period Volumes Combined	Rate (cross/ 1,000 vehicles)
Nighttime (9:00 p.m. to 4:00 a.m.)[a]			
Existing	8[b]	1,924[b]	4.2
Object marker	10	2,719	3.7
Nose panel	8	2,740	2.9
Nose and back panel	11	2,852	3.9
Daytime (9:00 a.m. to 4:00 p.m.)[c]			
Existing	41[b]	9,637[b]	4.3
Object marker	43	13,838	3.1
Nose panel	29[d]	9,632[d]	3.0
Nose and back panel	64	15,282	4.2

[a] $\chi^2 = 0.58$; $p = .90$.
[b] Wednesday nighttime and daytime data not available because of technical difficulties.
[c] $\chi^2 = 4.38$; $p = .22$.
[d] Wednesday daytime data not available because of rain.

was found among them for nighttime or daytime conditions [$\chi^2 = 0.58$ ($p = .90$) and 4.38 ($p = .22$)].

Crossovers by Type of Gore Area

As previously noted, Sites 1 and 3 were classified as Type IIa gore areas and Site 2 as Type IIIa. Consequently, assuming that the motorist information (signing, lane markings, etc.) upstream of the gore is adequate at Sites 1 and 3, one would expect a random distribution of crossovers across type of crossovers and gore area treatments. Higher frequencies of crossovers during specific gore area treatments would be attributed to the differences between the treatments.

In contrast, one would expect a specific pattern (type) of crossovers at Site 2 (Type IIIa) regardless of treatment. Geometrics plays a significant role in the type of crossovers at Type IIIa gore areas.

In order to further evaluate the four gore area treatments, the data were classified and analyzed by crossover type. The results of this analysis are presented in the following sections.

Site 1: Type IIa

Table 5 compares crossover encroachments by summarizing frequency totals across sites and lighting conditions, and also presents the crossover rates for each condition. The results indicate that there were not discernible nighttime crossover patterns for this site (row 1, Table 5). Crossover frequencies appeared to be randomly distributed by type of crossover and across gore area treatments.

Site 2: Type IIIa

For both nighttime and daytime data, it is evident from Table 5 that significantly more crossovers were occurring in a left direction than in a right direction. Referring back to Figure 11, it may be noted that the left fork leads to Juarez, Mexico, a large traffic generator, and that vehicles in the right lane at Site 1 have only about 1,500 ft to move into the center or left lanes. Unless adequate advance signing exists, the drivers may be trapped in the right lane headed for New Mexico. It is surmised that a large number of drivers (146) made a crossover to the left because of *(a)* the Type IIIa geometrics, *(b)* the congestion, *(c)* inadequate advanced lane directions, or *(d)* a combination thereof. Number of lanes available in a left-hand exit may also be a factor in some applications.

Site 3: Type IIa

The Copia Street entrance to I-10 is again a left-hand entrance so it is not too surprising that crossovers were predominantly in a right-to-left direction, both night and day (Table 5, rows 5 and 6). Note that at Site 1, which was a right-hand exit, the daytime data showed twice as many right crossovers.

Day Versus Night

Frequency of crossovers at all sites was greater for day than night. This was expected because of the much greater traffic volumes and the frequent problem of getting into the exit lane in heavy traffic. The rate data, which correct for volume, show less difference between day and night. At Sites 1 and 3, there was very little difference in day and night rates.

Long-Term Evaluation

One of the objectives of the field studies was to conduct an on-site inspection of the gore area treatments to subjectively assess the quality of the treatments after prolonged use (4 to 6 months' duration).

TABLE 5 TOTAL CROSSOVER FREQUENCIES AND RATES (all treatments)

			Left					Right					
			Zone 1		Zone 2			Zone 1		Zone 2			
Site	Type	Time	Frequency	Rate	Frequency	Rate	Total Frequency	Frequency	Rate	Frequency	Rate	Total Frequency	Left Versus Right Frequency
1	IIa	Night	6	0.2	2	0.1	8	6	0.2	5	0.2	11	No difference
1	IIa	Day	9	0.1	6	0.1	15	20	0.2	14	0.2	34	Difference
2	IIIa	Night	17	0.9	12	0.6	39	1	0.1	1	0.0	2	Difference
2	IIIa	Day	88	1.5	58	1.0	146	0	0.0	1	0.0	1	Difference
3	IIa	Night	24	2.4	6	0.6	30	3	0.3	4	0.4	7	Difference
3	IIa	Day	115	2.4	22	0.5	137	10	0.2	30	0.5	40	Difference

The last treatment studied at each site was to remain at the site for at least 4 months before the field inspections. As indicated by the field study experimental design (Table 1), the nose and back panel treatment was to be left at Site 1 (Ramp A) and the nose panel treatment at Sites 2 and 3 (Ramp A-F and Copia Street Ramp). However, three accidents resulting in crash cushion repairs occurred after the completion of the short-term data collection and before the 4-month long-term period, which ruled out any long-term field inspections.

The nose panel treatment left in place at Site 2 was hit sometime in January 1986, requiring the treatment to be replaced. During the Easter weekend (March 28–30), the Site 2 crash cushion again was struck. The Site 1 crash cushion (nose and back panel treatment) also was struck. Both of these sites had new crash cushions and nose panels installed. However, the new nose panels were different from the original nose panels used for the short-term data collection. The chevron patterns were accidentally reversed by the El Paso District maintenance personnel. The new nose panels had a yellow chevron in the center with black corners, while the original nose panels had a black chevron in the center with yellow corners. This was not discovered by the research staff of the El Paso District contact person until the final inspection of the study sites was made in April. Therefore, no long-term comparative assessment could be made.

SUMMARY AND DISCUSSION OF RESULTS

Three candidate gore area delineation treatments were selected for field evaluations: *(a)* object marker, *(b)* yellow-and-black nose panel, and *(c)* yellow-and-black nose and back panel. The I-10–US-54 interchange in El Paso was selected as the study area. Three specific gore area sites at the interchange were identified by the El Paso District of the SDHPT as being problem gore areas. All three sites had an existing treatment that became part of the field evaluation studies. Although it would have been more desirable to study the candidate treatments at sites without an existing treatment in order to have a more suitable base condition, most gore areas exhibiting accident problems will have some type of delineation treatment in place.

The El Paso District of the SDHPT agreed to install all three candidate treatments at each of the three sites. This allowed the opportunity for a much stronger experimental design in terms of evaluating differences between candidate treatments than that specified in the research contract. However, the District would not agree to leaving the object marker in place for longer than one week.

One major problem, discovered after the data had been collected, was that the video camera location for two of the gore area sites (1 and 2) made it difficult to accurately identify all gore area encroachments. Encroachments were identified as either sideswipes or crossovers. The side viewing angle, contrary to expectations based on actual field assessments, made it difficult to identify sideswipes at Sites 1 and 2. Therefore, only crossover encroachments were used in the analysis. Because the number of crossovers was relatively small, the data base was consequently smaller than expected.

However, in rationalizing between crossovers and sidewipes, crossover data would appear to be more relevant. Crossovers can be interpreted to mean that the driver was in the wrong lane and made a late decision to change lanes (or was restricted by traffic from lane changing until it was almost too late). With sideswipes, the driver is in the correct lane and for some reason swerves into the gore area—forced by traffic or wind gusts, not paying attention, and so on. It appears that it is less likely that he was confused by the delineation treatment (or lack thereof) or by advanced signing or geometrics.

A basic gore area classification system was developed and proposed as part of the research study. It was hypothesized that safety problems are more prominent with certain classes of gore areas and that delineation treatments can enhance safety for these classes. Also, there are classes of gore areas for which it may be difficult, if not impossible, to solve the safety problems by increased gore area delineation alone because of less-than-desirable geometrics, sight distance, or advanced driver information or all three. Improvements to the information system, or in some cases improvements in geometrics, may be necessary. The classification system was developed during the course of the research contract and was not used in this research to select gore area study sites. However, it does help to explain the results of the field studies. Perhaps further development and use may lead to a better systematic evaluation of solutions to gore area problems.

The three El Paso study sites were classified by using the scheme proposed by TTI. An evaluation of the crossover data indicated that the results were consistent with expectations based on the classification of the gore area study sites. For example, as expected, Site 2, classified as a Type IIIa gore area, exhibited a very high rate of crossovers from right to left in comparison with Sites 1 and 3, which were classified as Type IIa.

An analysis of the crossover data at each site indicated no difference in crossover rates among the four treatments: existing, object marker, nose panel, and nose and back panel. These results were consistent with the gore area classification concepts. Sites 1 and 3 were classified as Type IIa. Sight distance to the gore areas was not a problem. The results indicated that added delineation, based on a limited sample, did not reduce crossover rates. In particular, the back panel, designed to increase the effective sight distance to a gore area, apparently was not warranted for Sites 1 and 3. Site 2 (Ramp A-F) was classified as a Type IIIa gore area. Sight distance did not appear to be a problem. However, adverse geometrics and inadequate or confusing signing, or both, resulted in a relatively high rate of crossovers. It is hypothesized that additional gore area delineation would not alleviate the crossover problem and that improvements in geometrics or signing or both may be more effective.

The results of this and previous TTI studies indicate that increased delineation can reduce encroachments and accidents at some gore areas where sight distance is restricted *(4, 5)*. However, when sight distance to the gore is not a critical factor, encroachments and accidents may be less affected by increased delineation of crash cushions.

It was assumed that by evaluating all of the delineation treatments at each of the three study sites, a stronger conclusion could be drawn based on the redundancy of the experiment.

However, this procedure reduced the amount of data collected because of time restrictions on the study. Because encroachments are a relatively rare event, the encroachment rates used in the analysis were typically small. Although the chi-square test is sensitive to small sample sizes, the fact that no significant differences were found between treatments indicates that encroachments may be somewhat insensitive when used as MOEs. Because of their rarity of occurrence, simply collecting more encroachment data for the same type of analysis may not provide any different results. It is recommended that further research be performed in this area that will utilize more sensitive MOEs and will also expand that scope of the study to include geometrics of the gore area and the overall information system (signing, markings, delineation, etc.) associated with the gore.

ACKNOWLEDGMENTS

The research reported here was conducted as part of a project entitled "Evaluation of the Effectiveness of Crash Cushion Delineation," sponsored by FHWA. Truman Mast served as the contracting officer's technical representative. COMSIS Corporation, Wheaton, Maryland, was the prime contractor, with Neil D. Lerner serving as principal investigator. The Texas Transportation Institute, Texas A&M University, was a subcontractor to COMSIS.

REFERENCES

1. L. I. Griffin. How Effective Are Crash Cushions in Reducing Deaths and Injuries? *Public Roads,* March 1984.
2. N. D. Lerner and C. L. Dudek. *Evaluation of the Effectiveness of Crash Cushion Delineation.* Draft Report. FHWA, U.S. Department of Transportation, Sept. 1986.
3. *Manual on Uniform Traffic Control Devices.* FHWA, U.S. Department of Transportation, 1986.
4. R. C. Wunderlich. *Evaluation of Gore Area Crash Cushion Delineation in Houston.* Report 0937-1. Texas Transportation Institute, Texas A&M University, College Station, July 1984.
5. R. C. Wunderlich and C. L. Dudek. *Evaluation of the Short-Term Effect of Gore Area Crash Cushion Delineation in Ft. Worth.* Report 0102-1. Texas Transportation Institute, Texas A&M University, College Station, Aug. 1984.
6. C. L. Dudek, S. H. Richards, and M. J. S. Faulkner. *Traffic Management During Urban Freeway Maintenance Operations.* Research Report FHWA/TX-82/2+228-10F. Texas Transportation Institute, Texas A&M University, College Station, Jan. 1982.
7. H. W. McGee, W. Moore, B. G. Knapp, and J. H. Sanders. *Decision Sight Distance for Highway Design and Traffic Control Requirements.* Report FHWA-RD-78-78. BioTechnology, Inc., Falls Church, Va., Feb. 1978.

The contents of this paper reflect the views of the authors, who are responsible for the accuracy of the material presented herein. The contents do not necessarily reflect the official views or policy of the Federal Highway Administration.

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

Special Traffic Control To Meet Motorist Information Needs on Long, Steep Grades

JOSEPH L. HENDERSON, EUGENE M. WILSON, GEORGE A. DALE, AND CLAUDIA L. SCRIVANICH

A new concept in traffic control devices is evaluated. The subject of the study is a special message information sign on US-14A near Burgess Junction, Wyoming. The sign presents motorists with percent and length of grade information in response to a long sustained 10 percent downgrade. This diagrammatic sign is located in a turnout. Vehicles such as tractor semitrailers, single-unit trucks, motor homes, and vehicles pulling trailers are required to turn out and read the sign. This method for presenting complex information to motorists is effective and the compliance with the required turnout is high.

Informing motorists of roadway conditions can normally be accomplished by using traditional methods of traffic control. Quite often, warning signs must be used to alert motorists to geometric and alignment conditions that may violate driver expectancy or create a confusing situation. Presented in the following sections is a case study describing a new concept in the use of traffic control devices. The study was conducted on a mountainous section of US-14A in northern Wyoming. A special information warning sign located at a turnout requires certain motorists to turn out and stop. Motorist compliance with this requirement and message comprehension were the focus of this study.

PREVIOUS RESEARCH

The state of the art in driver message signing has been advancing in the areas of directional and guide signing, variable message signs, and positive guidance. Driver responses to the message content, response time, and route choice have been analyzed by many researchers.

Previous studies have covered subjects such as motorist route selection criteria *(1)*, diversionary signing and driver behavior *(2)*, tourist information systems *(3)*, positive guidance *(4)*, and traffic operations, safety, and positive guidance projects *(5)*. Studies related to guide signs are directed toward new applications of directional and guide signs *(6–9)*. Research concerning driver systems addresses the need for coordinating signs, maps, and tourist information to reduce user costs and traffic congestion *(10, 11)*. These studies are designed to evaluate specific route and grade information that is provided to selected motorists.

A method for providing grade and length of grade information to truck drivers is now being researched. A grade severity rating system *(12)* has been designed to designate a specific speed for a tractor semitrailer driver to use when driving on a downgrade. As the results were first envisioned, each downgrade on every road would be rated by using this system, and drivers would be expected to reference previous driving experiences to take advantage of the system. This idea gave way to a weight-specific speed sign that would show a specific speed for different weight classes. The Colorado Department of Highways *(13)*, and highway departments from nine other states participated in a program to test this signing concept. The results have been inconclusive to date, and observations have shown that a high percentage of drivers do not recognize the intent of the signs.

In the following section the study is discussed including a description of the study site and the traffic control used.

THE STUDY

Turnout signing is a unique concept because motorists stop to read these signs, whereas the information on other signs is conveyed to drivers as they travel along the road. Two signs located at turnout areas were designed to alert drivers to a long, steep downgrade that exists on US-14A between Burgess Junction and Lovell, Wyoming. Drivers who turn out are separated from through traffic and have an unlimited amount of time to study these signs. The primary study objectives were to determine the ability of the sign to convey grade data, the motorists' comprehension of this information, motorists' compliance with the required turnouts, and driver expectancy of grades.

The locations of the study sites are shown in Figure 1. Highways of interest are US-14 and US-14A between Burgess Junction and Cody, Wyoming. These are both two-line primary highways with a similar type design and are located in rugged mountainous terrain. US-14A contains a section of road located 20 mi west of Burgess Junction that has a great deal of 10 percent grade. Most motorists have never experienced a long sustained 10 percent grade. The grades are deceptive because of the high quality of the road. A driver who is unaware of or unfamiliar with such grades may tend to be accelerating more than he perceives. Also, some of these motorists are driving

J. L. Henderson, Centennial Engineering, P.O. Drawer 1307, Arvada, Colo. 80001. E. M. Wilson, University of Wyoming, P.O. Box 3295, University Station, Laramie, Wyo. 82071. G. A. Dale, Wyoming Highway Department, 5300 Bishop Boulevard, Cheyenne, Wyo. 82002. C. L. Scrivanich, Washington State Highway Department, Olympia, Wash. 98504.

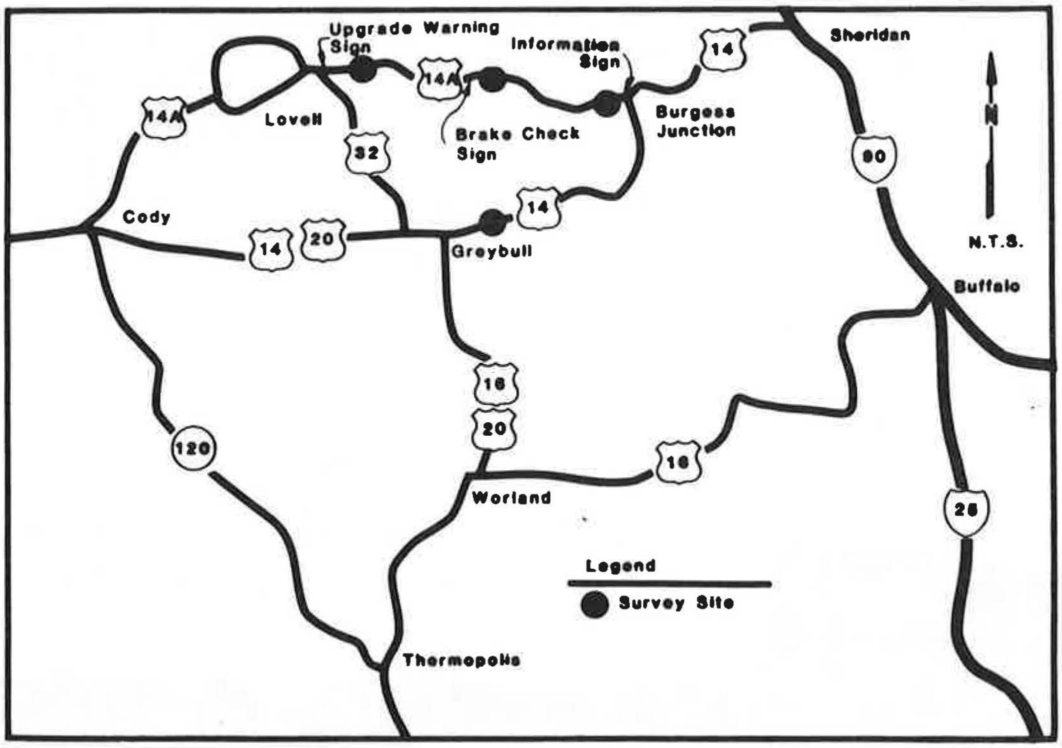


FIGURE 1 General study area.

some type of recreational vehicle or a vehicle with a trailer. More traditional grades of 5 to 7 percent exist on US-14. Special attention has been given to these highways because they are major routes to the east entrance of Yellowstone National Park and are traveled by nonresidents of Wyoming. The majority of westbound traffic at Burgess Junction uses US-14. During the study period, 62 percent of the westbound traffic used US-14. Out-of-state vehicles accounted for 86 percent of the traffic on US-14 and 57 percent on US-14A.

An information sign (Figure 2) is located on US-14A about 1 mi west of Burgess Junction. This sign, which measures about 10 ft by 30 ft, has been designed to show that the distance from Burgess Junction to Cody is approximately the same by way of US-14A or US-14, but that the grades are much more severe on US-14A. Advance signing requires all semitrailer trucks, buses, recreational vehicles, and vehicles pulling trailers to stop and read the information sign (Figure 3).

Another turnout warning sign is located about 25 mi west of Burgess Junction on US-14A at a brake check safety area. This sign, shown in Figure 4, provides more detailed information about the 10 percent downgrade, including the location of a brake cooling turnout and the three truck runaway ramps. The same vehicle classifications are required to turn out.

Signing associated with the uphill travel direction is shown in Figure 5. This sign is located just east of Lovell, Wyoming, approximately 20 mi before the steep upgrade section. This is the only grade signing in the uphill direction.

DATA COLLECTION

Data were collected manually and by using driver interviews during the summers of 1984 and 1985. The interview sites were located at the Burgess Junction information sign turnout west of Burgess Junction on both US-14 and on US-14A at the bottom of the grade and on US-14A at the top of the grade (see Figure 1). Sample sizes for the various surveys are given in Table 1. The surveys were designed to gather information concerning comprehension of the information signs, driver's understanding of percent and length of grade, and route selection criteria. Also of interest were any problems encountered on the grades in the study area and whether drivers diverted to US-14 after stopping at the Burgess Junction information sign. Vehicle classification, origin of license, and stopped time were obtained.

Major findings of this study are presented in the following sections. The major findings discussed in this paper are only those that were statistically significant at the 0.05 level. Topics of discussion are turnout signing, downgrade analysis, and upgrade analysis.

TURNOUT SIGNING

Compliance with the required turnout was essential for this signing technique to be effective. Of all the vehicles observed at the sign, only about 32 percent stopped. However, the rate of compliance for vehicles required to stop was quite high. More than 62 percent of the vehicles required to stop did stop. Vehicles required to stop are trucks, recreational vehicles, and vehicles with trailers. Out-of-state motorist compliance with the required turnout was almost 80 percent, whereas only about 20 percent of local Wyoming motorists complied. More first-time users of US-14A stopped at the sign than any other group (80 percent). The overall rate of compliance was slightly higher in 1985 and local Wyoming compliance decreased slightly.

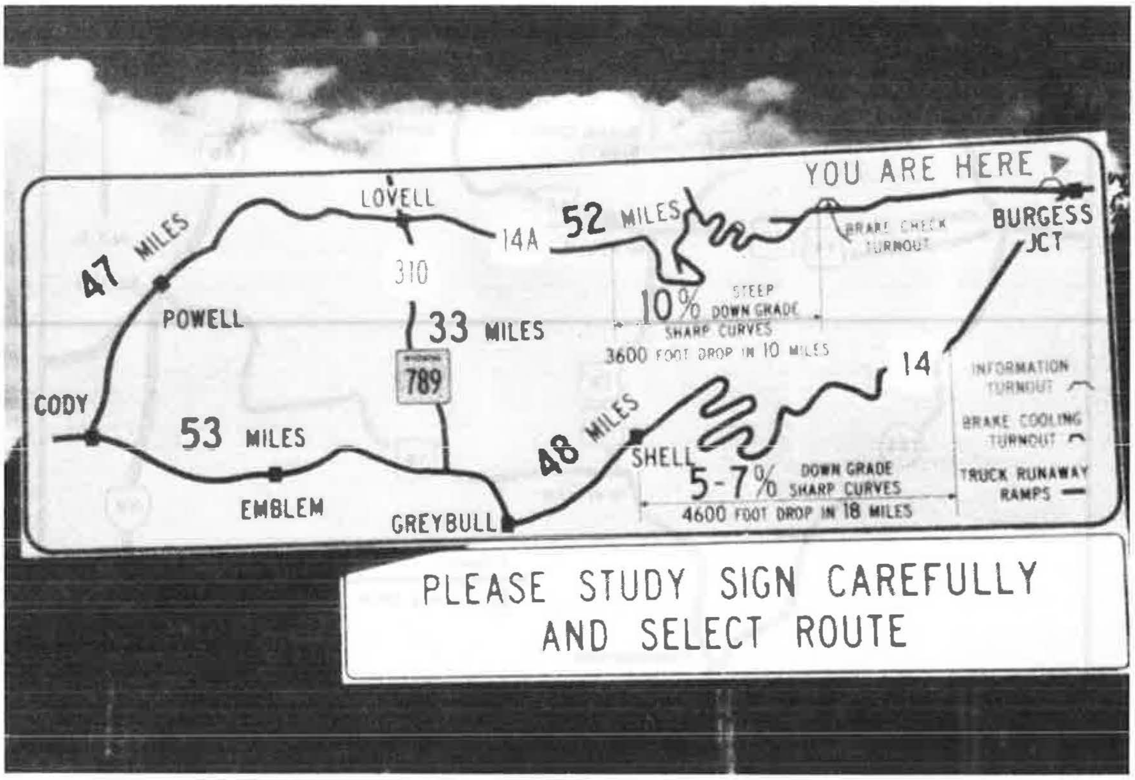


FIGURE 2 Burgess Junction information sign.

Driver behavior was observed for vehicles that stopped at the sign. The amount of time that drivers took to read the sign was recorded. Wyoming motorists generally stopped for less than 30 sec, whereas out-of-state drivers generally stopped for more than 30 sec. Drivers of vehicles required to stop viewed the sign longer than drivers of other vehicles. Average times were 37 sec and 29 sec, respectively. The vehicles that stopped at the sign and then diverted to US-14 were also observed. During the data collection period, 110 vehicles were observed that diverted to US-14. Of this number, 80 percent were from out of state and 67 percent were required to turn out. Of the 4,558 vehicles observed, 7.6 percent of the total vehicles using the turnout diverted to US-14. This was 2.4 percent of the total US-14A traffic observed. No route diversion signing existed before the turnout signing. This additional signing was incorporated into the design because of the reopening of US-14A on a new steeper alignment.

Driver behavior at the brake check sign turnout was also examined. Less than 50 percent of the vehicles required to turn out at the sign did so; however, 56 percent of the out-of-state vehicles stopped that were required to do so. Fifty-nine percent of the drivers who stopped at the Burgess Junction information sign also stopped at the brake check sign.

Drivers' actions at the brake check sign, including stopped time, were observed. Of the vehicles that stopped at the sign, average viewing time for vehicles required to stop was longer than for vehicles not required to stop (53 sec versus 28 sec). This difference was attributed to the message content of the sign, which is targeted toward vehicles required to turn out. The sign provides more specific information than does the Burgess Junction information sign concerning the percent grade, overall length of grade, location of three truck escape ramps, and location of the brake cooling area. The difference in average viewing times indicated that motorists were spending the additional time necessary to assimilate this detailed information on the brake check sign.

DOWNGRADE ANALYSIS

The downgrade of US-14A was examined to determine the driver expectancy of the length and percent of grade and to

TABLE 1 NUMBERS OF SURVEYS AND OBSERVATIONS BY LOCATION AND YEAR

	No. of Driver Surveys			No. of Driver Observations	
Year	US-14A Downhill	US-14 Downhill	Burgess Junction	Burgess Junction	Brake Check Area
1984	186	204	0	3,635	0
1985	573	198	211	923	574
Total	759	402	211	4,558	574

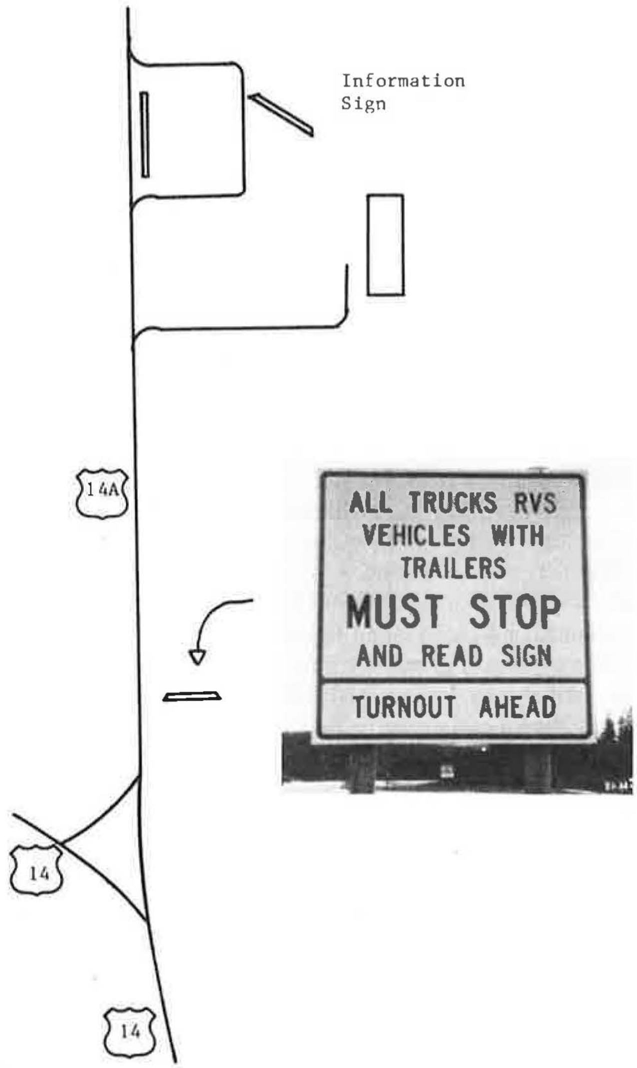


FIGURE 3 Advance signing at Burgess Junction.

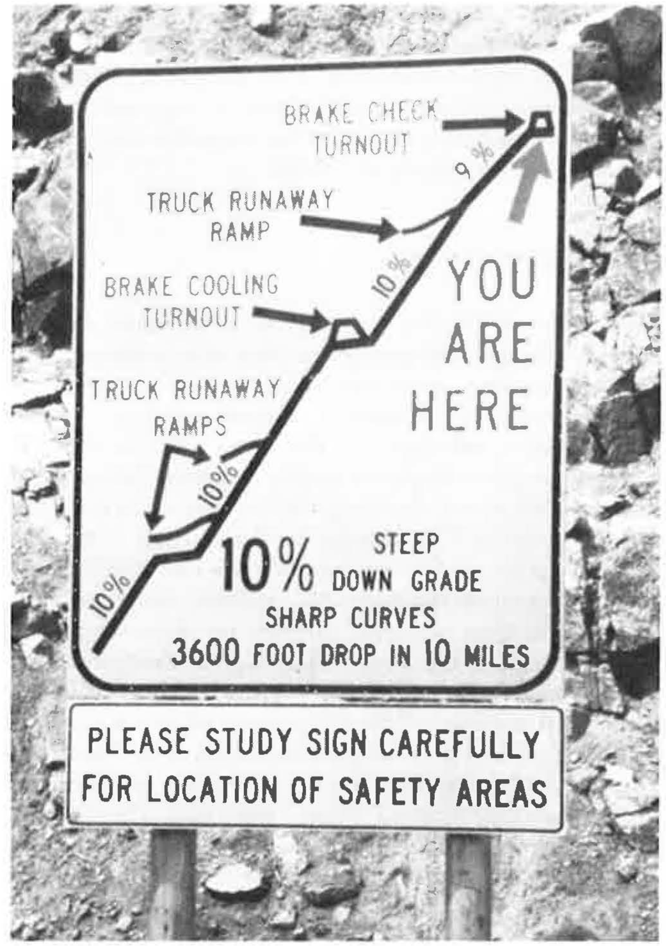


FIGURE 4 Brake check sign.

determine whether operational problems existed. Driver expectation of percent of grade on US-14A was affected by driver's domicile and frequency of travel on this road. The local Wyoming motorists and regular users of the road felt more comfortable driving on the road than other motorists. Only about 40 percent of the motorists interviewed indicated that they had used a lower gear to descend the grade. In many cases, not using a lower gear resulted in brakes on vehicles descending the grade becoming extremely hot. This not only occurred on vehicles required to stop at the signs, but also on some passenger vehicles. One passenger vehicle lost its brakes during the study period in 1985. This vehicle, with a heavily loaded trailer, used the middle truck escape ramp. Quite a few motorists stopped while descending the grade on US-14A. Some of the motorists who stopped did so more than once. A few motorists even stopped at truck escape ramps, which creates a hazard in the event that the ramp is needed by a vehicle that has lost its brakes.

Results of the downgrade driver interviews indicated that most motorists did not understand the concept of percent grade. After the 1984 study three additional signs with the text All Vehicles Use Lower Gear were added. This substantially reduced the problem of hot smoking brakes. In 1985, 573 interviews were conducted below the 10 percent grade. There were 51 vehicles (8.9 percent) that had noticeably hot brakes. Additional signing, Save Your Brakes, Use Lower Gear, has been recommended.

A comparison was made between the motorists' perception of the downgrade on US-14A and that on US-14 by surveying motorists on both routes. Most motorists thought that the

FIGURE 5 Upgrade advance-warning sign.

grades on US-14A were steeper than expected and that the grades on US-14 were not as steep as expected. This result depended on the domicile of the driver. In general, Wyoming motorists responded that the grades were as expected. Out-of-state motorists generally responded that the grades were steeper than expected, particularly on US-14A.

UPGRADE ANALYSIS

The upgrade on US-14A was analyzed to determine characteristics of the uphill motorist and to learn what problems these motorists may have experienced driving up the mountain. More than 80 percent of the vehicles observed traveling upgrade were passenger vehicles: very few heavy vehicles such as tractor-semitrailers, single-unit trucks, and motor homes were observed. Only one of the heavy vehicles observed was a local Wyoming vehicle. This indicated local awareness to the difference in grade severity between US-14A and US-14. Many vehicles stopped on their way up the mountain. Some overheating problems were observed, but most motorists stopped to take advantage of the scenic vistas. Uphill travel problems were observed for bicyclists, older vehicles, and heavily loaded vehicles. A number of bicyclists were observed in 1985 attempting to ride up the steep portion of US-14A. Two cross-country cyclists spent an entire day attempting to ascend the grade before they received motorist assistance. The rest that were observed turned around.

Motorists indicated that the uphill advance warning sign, located about 1 mi east of Lovell, was understandable (see Figure 5). This sign is located on a relatively flat tangent section about 20 mi before the steep uphill section. Some motorists thought that the sign should indicate more detail concerning the grade on US-14A. This sign is the only one that provides information to the motorist about the grade before or during their trip up the mountain. Quite a few vehicles were observed traveling part of the way up the mountain and then turning around. This was probably because the drivers did not know how long the grade was.

The foregoing discussion contains the major results of the study conducted in 1984 and 1985. On the basis of these major study findings, the following conclusions are made:

1. The use of turnout signing is an effective method of traffic control. Motorists will comply with the requirement to turn out and stop.
2. Signing in a turnout should be considered when information content is long or complex or when messages are appropriate for only a select group of motorists.
3. Motorist comprehension of percentage of grade is poor. Diagrammatic signing or word messages explaining this concept are necessary to help motorists when grades are severe.
4. Providing uphill and downhill turnout areas on long, steep grades is necessary to help lessen the effects of brake and vehicle overheating.
5. Special uphill signing is necessary to inform motorists of length and percentage of long, steep grades.
6. Consideration should be given to prohibiting bicycle traffic on long, steep grades unless special provisions have been made for this type of traffic.

RECOMMENDATIONS

The use of turnout signing has been shown in this study to be an effective method. It should only be considered in situations where low traffic volumes exist.

Most of the drivers who were interviewed at the Burgess Junction information sign did not understand percentage of grade, and some suggested that the sign should also attempt to educate about it. To do so, diagrammatic signs could be used to show a comparison between a possible known or alternative grade and the one about to be traversed. Another possible solution would be to design a sign that would be placed in a turnout or rest area that explains grade.

Quite a few motorists stopped while descending the grade on US-14A. An adequate number of brake-cooling areas should be provided so that drivers can stop as often as desired. Operational reviews should be conducted periodically to determine whether the number of brake-cooling areas is adequate and whether they are being used.

Many motorists were observed traveling part of the way up the mountain and then turning around. This behavior could be a result of not knowing actually how steep the road was or how much farther they had to drive before reaching the summit. Signing showing distance to turnout areas as well as to the summit is important on long steep upgrades. Of most benefit to the driver would be information on percentage and length of grade *(14)*.

Special warning signs alerting bicyclists to long steep upgrades could be beneficial to this group. Consideration should also be given to prohibiting bicycle traffic on grades of this severity.

ACKNOWLEDGMENTS

The authors wish to express sincere thanks to the University of Wyoming, Wyoming Highway Department, and the Federal Highway Administration. Two of the authors, Joe Henderson and Claudia Scrivanich, were graduate students at the University of Wyoming when this research project was conducted. The funding of this project through the Highway Research Program is greatly appreciated.

REFERENCES

1. R. D. Huchingson, R. W. McNees, and C. L. Dudek. Survey of Motorist Route Selection Criteria. In *Transportation Research Record 643*, TRB, National Research Council, Washington, D.C., 1977, pp. 45–48.
2. T. M. Mast and J. A. Ballas. Diversionary Signing Content and Driver Behavior. In *Transportation Research Record 600*, TRB, National Research Council, Washington, D.C., 1976, pp. 14–19.
3. P. Arthur. *Highway Research Circular 152: An Exploratory View of Tourist Information Systems*. HRB, National Research Council, Washington, D.C., Dec. 1973, pp. 15–19.
4. T. J. Post, G. J. Alexander, and H. Lunenfeld. *A User's Guide to Positive Guidance*, 2nd ed. Report FHWA-TO-81-1. FHWA, U.S. Department of Transportation, 1981.
5. H. Lunenfeld. *Evaluation of traffic Operations, Safety, and Positive Guidance Projects*. Report FHWA-TO-80-1. FHWA, U.S. Department of Transportation, 1980.

6. P. J. Dean, Highway Guide Signs: A Framework for Design and Evaluation. In *Transportation Research Record 737*, TRB, National Research Council, Washington, D.C., 1979, pp. 72–81.
7. F. R. Hanscom and W. G. Berger. *NCHRP Research Results Digest 91: Motorist Response to Guide Signing*. TRB, National Research Council, Washington, D.C., 1977.
8. G. C. Hoff. Development and Evaluation of Experimental Information Signs. In *Highway Research Record 279*, HRB, National Research Council, Washington, D.C., 1969, p. 159.
9. H. J. Wootton and R. S. Burton. Improving the Accuracy of Information on Direction Signs. In *Transportation Research Record 782*, TRB, National Research Council, Washington, D.C., 1980, pp. 30–36.
10. W. Hughes. Recreational Traffic Management Studies. *ITE Journal*, Sept. 1982, pp. 16–21.
11. S. Hulbert. *Human Factors in Highway Traffic Safety Research*. Wiley and Sons, Inc., New York, 1972.
12. P. Abbott. Grade Severity Rating System. *Public Roads*, Vol. 46, June 1982, pp. 30–34.
13. R. Erickson, Jr. *A Field Test of a Grade Severity Rating System*. Colorado Department of Highways, Denver, 1985.
14. E. M. Wilson. Total Safety Design for Long-Steep Grades. In *Compendium of Technical Papers for the ITE District 6-District 7 Joint Annual Meeting*. ITE, Vancouver, B.C., Canada, 1986.

The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the Wyoming Highway Department or the Federal Highway Administration.

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

DISCUSSION

Andrew G. Macbeth
Transportation and Traffic Engineering Program, Department of Civil Engineering, University of Toronto, Toronto, Ontario M5S 1A4, Canada.

First, I argue against the principle of compelling drivers to turn off the road to read information signs. Second, I suggest that the prohibition of bicycle traffic recommended by the authors is an inappropriate conclusion to draw from a study on road signing.

The authors' first conclusion states: "The use of turnout signing is an effective method of traffic control. Motorists will comply with the requirement to turn out and stop." The evidence, however, is not quite so convincing. The authors report that 62 percent of vehicles required to turn out (trucks, recreational vehicles, and vehicles with trailers) did, in fact, do so. But although 80 percent of out-of-state drivers complied, only 20 percent of local Wyoming drivers did so, and this proportion declined in the second year of the study.

It may be more important to target the out-of-state drivers, who are less likely to be familiar with the unusually steep grades in the area. In this light, the concept of turnout signing may be of some benefit. But is it really necessary that truck drivers (or drivers of other affected vehicles) who frequently drive the route turn out every time? A large majority of local drivers of affected vehicles ignored the sign with its "must turn out" instruction. This suggests that they believe that they know the turnout sign's contents and that consequently they do not need to turn out. If the sign is legally enforceable, would all noncomplying drivers be equally culpable? Or would enforcement be concentrated on drivers of vehicles bearing out-of-state license plates? Whether the sign is legally enforceable or not, it is likely to be disregarded by a significant proportion of drivers. The erection of traffic signs that are habitually ignored by many drivers is likely to bring traffic laws (and their makers) into disrepute.

I appreciate the difficulty in conveying complex information to drivers, but believe that compulsory turnouts do not provide a satisfactory solution. Somehow, drivers should receive all the information they need without being compelled to turn out and stop.

Creative intersection design at Burgess Junction and destination signing may be effective in diverting out-of-state drivers away from US-14A to the less steep US-14.

Second, I note that the purpose of the research described by the authors was the evaluation of a new concept in traffic control devices for motorists on long, steep grades. The authors' sixth conclusion ("Consideration should be given to prohibiting bicycle traffic on long, steep grades unless special provisions have been made for this type of traffic") and a similar recommendation would appear to fall outside the scope of the study. This conclusion and recommendation, I believe, are inappropriate and unnecessarily discriminatory. None of the authors' research described here has shown the need for (or legality of) such a measure.

The further recommendation that "special warning signs alerting bicyclists to long, steep upgrades could be beneficial to this group," although still outside the research objective, may be quite adequate for the authors' purposes of discouraging cyclists, but any stronger conclusions should in my opinion be based on specific research.

AUTHORS' CLOSURE

Mr. Macbeth's discussion was interesting, and in total considers two points. The first was his argument against a required turnout and the enforcement of the requirement to turn out. In his discussion, he indicates that turnout signing may be of benefit. The study showed that this was in fact the case. In a state with an economy of many cities and that is largely dependent on tourist traffic, it is not feasible to sign a route to create a competitive advantage for one city. Creating a motorist who is informed of the uniqueness of this long, steep grade was the goal of the turnout design. Enforcement is not really at issue. Local motorist behavior did not create a follow-the-leader effect. With an average daily traffic of 400 vehicles per day, this is not a problem.

The second comment concerning bicycle travel was due in part to our omission. The research study itself did not investigate both uphill and downhill travel behavior. Although bicycle travel uphill was not specifically detailed in this paper, we wanted to include this finding in the hope that it might be useful for designers in the future.

Advisory Speed Signs and Curve Signs and Their Effect on Driver Eye Scanning and Driving Performance

HELMUT T. ZWAHLEN

The objective of this study was to determine the effectiveness of advisory speed signs used in conjunction with curve warning signs in Ohio. A total of 40 test drivers were used to drive an unfamiliar test route on a two-lane rural road that included two typical curves equipped with curve warning signs. Curve A was a left curve with a determined advisory speed of 40 mph and Curve C was a right curve with a determined advisory speed of 25 mph. The results of the test-driver study indicate that drivers, on the average, look about two times at a warning sign (fixation duration 0.5 to 0.6 sec). There are few consistent statistically significant differences in driver eye-scanning behavior and driver control behavior (velocity, lateral acceleration, gas pedal deflection, lane position, brake activation) between Run 1 and Run 2, between inexperienced and experienced drivers, between the presence and absence of advisory speed signs, and between day and night. The daytime velocities are in general somewhat higher than the nighttime velocities. It may be concluded that advisory speed signs are not more effective in causing drivers to reduce their speeds through curves than curve and turn signs alone. It appears that the bent black arrow in the yellow diamond of the curve or turn warning sign represents such a strong and primary visual stimulus that an advisory speed sign adds very little additional information for the driver. Therefore, it is recommended that advisory speed sign maintenance and especially new installations be given a low priority.

The state of Ohio has a highway network containing more than 19,000 of four-lane and two-lane highways. On the basis of the Ohio Department of Transportation's (ODOT) curve inventory, the two-lane rural system alone contains 18,093 curves. These curves have a median curvature between 10 and 11 degrees, an average curvature of 12 degrees with a standard deviation of 12.5 degrees and a mode of curvature of 9 degrees (2,517 curves, or 13.9 percent, have a 9-degree curvature). On many roads it is impossible to drive safely at 55 mph because of the curvature, and some are considered safe at speeds of only 25 mph or less. It will probably require decades before sufficient funding is available to eliminate these sharp and dangerous curves and to rehabilitate the older highways that have substandard alignment. Therefore, drivers must be alerted to upcoming hazardous curves through the use of warning signs.

ODOT now uses curve warning signs with or without advisory speed signs to warn drivers of an upcoming curve. (The Ohio *Manual of Uniform Traffic Control Devices* defines a curve sign as a warning sign with a curved arrow, intended for use on curves with recommended speeds between 30 and 50 mph, and a turn sign as a warning sign with an arrow bent at a right angle, intended for use on curves where the maximum safe speed is 30 mph or less. Both signs will be referred to as curve signs throughout this paper.) These signs are intended to give drivers adequate time to prepare to safely traverse an upcoming hazardous curve. Because of the frequency of such hazards, the use of curve warning signs with advisory speed signs is expensive and requires the efforts of traffic engineers, maintenance crews, and others who could be utilized in other areas.

Questions may then be raised with regard to the effectiveness of such practices, some of which are

Do motorists look at the advisory speed signs?

At what distances do they look and for how long?

How effective are curve warning signs in causing motorists to decrease speeds to safe levels throughout the curve?

Does the presence of advisory speed signs increase the effectiveness of the curve warning sign in bringing about adequate speed reductions throughout the curve?

LITERATURE REVIEW

Few prior studies were found that have been devoted to the effectiveness of curve warning signs with or without advisory speed signs. Ritchie *(1)*, by recording the lateral acceleration and forward velocity measured at the time of peak lateral acceleration, investigated uninformed subjects' responses to curves with or without curve warning signs during daylight driving conditions only. The curve warning signs were presented either by themselves or in conjunction with an advisory speed sign on which the advisory speed ranged from 15 to 50 mph in increments of 5 mph (advisory speed signs with values higher than 50 mph were not investigated because the driver's responses could have been influenced by the state speed limit of 60 mph that existed at this time). Ritchie *(1)* found that drivers choose faster speeds on curves with curve warning signs than curves without curve warning signs and even faster speeds when advisory speed signs were presented with the curve warning sign. The speeds recommended on the advisory speed signs were lower than those chosen for negotiating the curves except in the case of the 45 and 50 mph signs, where the subjects chose speeds almost exactly the same as the advisory speeds. Lateral acceleration appeared to be a key variable in the driver's decision-making process. When a driver approached a curve that required a large speed reduction, he accepted lateral accelerations closer to the maximum than those accepted for curves that required smaller speed reductions.

Department of Industrial and Systems Engineering, 283 Stocker Center, Ohio University, Athens, Ohio 45701-2979.

Kneebone *(2)* reported on the effect of advisory speed signs used in conjunction with curve warning signs. He found that advisory speeds determined with ball-bank indicators were very close to the 85th-percentile speeds. Institution of curve warning signs with advisory speed signs in Australia was accompanied by a marked reduction in accidents and a reduction in the approach speeds of 2 to 3 mph. However, the average speed in the curves actually increased slightly.

In a study of minor highway improvements that contradicts the findings by Ritchie *(1)*, Hammer *(3)* found that the placement of curve warning signs by themselves in advance of curves failed to produce significant accident reductions (five curves that were not previously equipped with curve warning signs were studied). However, when curve warning signs were used in conjunction with advisory speed signs, the results were different. (In 13 of the 15 curves studied, the curve warning and advisory speed signs were erected simultaneously, whereas in the two remaining cases the advisory speed sign was added to an already existing curve warning sign.) Significant reductions in accidents, especially ran-off-road accidents at night, did occur when the two signs were used together. It should be noted that the sample sizes in Hammer's study were rather small and the experimental design did not indicate any randomization scheme or controls. Further, no information was given about the environment-accident interaction, about other minor road improvements that may have been completed at the same time or immediately following the erection of the curve warning sign, about possible changes in average daily traffic volumes or about changes in the driver population, or both, that could have been primarily or partly responsible for Hammer's results.

Shinar et al. *(4)* found in a study on driver eye-scanning behavior that as drivers approached a curve they began concentrating their eye fixations less around the focus of expansion (the area of highest concentration for straight section driving) and more on the edge lines and the roadway close to the car. The authors reached the conclusion that warning signs should be placed before the beginning of the curve approach because near that point the driver is concentrating mainly on the roadway for directional and lateral placement cues rather than on the road surroundings.

The objective of this study was to determine the effectiveness of advisory speed signs used in conjunction with curve warning signs in Ohio on typical sharp and moderate curves for both inexperienced and experienced test drivers under both daytime and nighttime conditions.

METHOD

Subjects

A total of 40 subjects took part in the experiment and were divided into one of two groups on the basis of their driving experience (either experienced or inexperienced drivers). The 21 experienced licensed drivers (12 men, 9 women) had an average age of 22 years and had driven an average of 44,000 mi during an average of 6 years. The 19 inexperienced licensed drivers (11 men, 8 women) had an average age of 17 years and had driven an average of 4,000 mi during an average of 2 years.

All subjects were initially interviewed and required to fill out a biographical and driving questionnaire. Each subject was tested in the laboratory for *(a)* foveal vision (Bausch and Lomb vision tester) and peripheral vision (Landolt rings, 10 degrees horizontal, presented left or right) and *(b)* for simple (1 choice, 0 bits) and choice (8 choices, all equally likely, 3 bits) reaction times using a CR-200 Information Response Instrument (response uncertainty mode). The subjects also underwent a limited health evaluation. The results of these tests indicated that all subjects had normal visual acuity and reaction times and were in good health. None of the subjects were familiar with the road or the experimental vehicle. All subjects were paid and told only that the study involved driving on two-lane rural roads. They were not told the actual aim of the experiment.

Apparatus

An instrumented 1973 Volkswagen 412 with an automatic transmission and type 4000 low beams was used as the experimental vehicle in this study. This vehicle contains more than 30 instruments and mechanisms that are combined into a system allowing the experimenter to monitor and record a driver's eye movements while he or she is driving the car as well as time, distance, speed, lateral lane position, steering wheel position, gas pedal deflection, brake activation, and vertical, horizontal, and lateral accelerations of the car (sampling rate of 60 Hz). A further description of the experimental car and equipment has been published by Zwahlen *(5)*.

Experimental Test Sites

In order to make the results of this study widely applicable it was necessary to choose curves representative of the rural two-lane system in Ohio with fairly low average daily traffic (ADT). The two curves chosen (Curves A and C) had an approach speed of 55 mph and were equipped with only curve warning signs (without advisory speed signs) and no raised reflective pavement markers or post delineators. Curve A required a small speed reduction for safe negotiation, whereas Curve C required a moderate to large speed reduction.

The westbound approaches to these two selected curves, which were located on SR-180 east of Laurelville, Ohio, were used. With the ball-bank indicator the advisory speed was determined to be 40 mph for Curve A (a 12.3-degree left curve with a radius of 465 ft and a superelevation of 8.6 percent) and 25 mph for Curve C (a 26-degree right curve with a radius of 220 ft and a superelevation of 9 percent). ODOT records place the ADT at 1,440 for Curve A (total count in both directions) and 930 for Curve C. Two accidents occurred on each curve between 1975 and 1981.

For this experiment the curve warning signs (both directions) were equipped either with or without an advisory speed sign. The specific intensity for an entrance angle of –4 degrees and an observation angle of 0.2 degrees was recorded for each of the curve warning and advisory speed signs. The W1-2L curve warning sign on Curve A (for westbound traffic) was 30 in. square, had an average specific intensity of 35.5 cd/(ft-candle·ft^2) and could first be seen at 1,036 ft (measured from the curve warning sign). The W1-1R curve warning sign on Curve C (for westbound traffic) was 30 in. square, had an average specific intensity of 67.6 cd/(ft-candle·ft^2) and could first be seen at 953 ft. The advisory speed signs were 18 in. square with 8-in. numbers and had an average specific intensity of 61.4 cd/(ft-candle·ft^2) for Curve A and 62.2 cd/(ft-candle·ft^2) for Curve C (for westbound traffic).

Experimental Procedure and Design

Before the test-driver study began, a local familiar-driver study was completed that involved the inconspicuous videotaping of 339 vehicles on Curve A and 312 vehicles on Curve C as they approached the two curves of interest during both daytime and nighttime conditions. During this study the two curve approaches had only the existing curve warning signs without advisory speed signs. Time and distance data and points of brake light activation were recorded. Calculations were made to determine velocities and accelerations at various distances from the curves. These data gave the experimenters a standard with which to gauge the validity of the test-driver study results.

The test-driver study involved the continuous recording of the subject's eye-scanning behavior and vehicle measures as the subject drove for 30 to 45 min along a typical rural two-lane highway that included the two curves of interest. Subjects were randomly assigned to one of eight groups such that each group had either experienced or inexperienced subjects and as close to a half-men and half-women composition as possible. Although it was originally planned for a group of six to be tested under each condition, some nighttime conditions were tested only with fewer subjects because of frequent ground fog at the test locations. Each of the eight groups was subjected to one of the different conditions; that is, they would be experienced or inexperienced drivers, drive during the day or the night (using low beams), and drive through curves that had a curve warning sign either with or without an advisory speed sign. Both curves were equipped either with or without the advisory speed sign so that no subject was exposed to one curve with the advisory speed sign and one without the advisory speed sign. Also no subject drove the test route under more than one of the eight conditions. The subjects were asked to follow the test route twice to allow the experimenters to evaluate the effects of short-term familiarity on driver performance.

The independent variables are as follows: *(a)* time of day (level of illumination, day versus night), *(b)* driver capability (inexperienced versus experienced), *(c)* presence or absence of advisory speed sign, *(d)* degree of speed reduction required in curve (moderate to large = 30 mph or more; small = 10 to 15 mph), and *(e)* familiarity (Run 1 versus Run 2 or completely unfamiliar versus somewhat familiar).

The effects of the independent variables were measured using the following dependent variables *(a)* speed (mph), *(b)* accelerator pedal position (0–7, idle; 69–73, fully deflected), *(c)* brake pedal activation (on or off), *(d)* lateral acceleration (g), *(e)* lateral lane position, and *(f)* eye movement measures (foveal and near foveal or slightly peripheral eye fixations on curve signs and advisory speed signs).

The design variables that might influence performance measures and are beyond the control of the experimenter include *(a)* traffic (ahead in opposite or in the same direction, or both), *(b)* background luminance during nighttime, *(c)* road surface condition (debris, potholes, etc.), *(d)* condition of edge lines and center lines, *(e)* visibility (haze, dust, and light fog), *(f)* environment (foliage, height of crops, and grass along the highway), *(g)* temperature and humidity, and *(h)* position of the sun, level of daytime illumination, glare, and cloud cover.

RESULTS AND DISCUSSION

Vehicle Measures

Detailed vehicle-measure and eye-scanning results for individual subjects and groups have been given by Zwahlen *(5)*. Certain points along the curves were selected for analysis in order to compare the vehicle measures for the different conditions. It was important that these points represented a balanced cross section throughout the approach and curve in order to obtain meaningful results. For this reason, the vehicle measures were analyzed at 500 ft before the curve warning sign, at the position of the sign, at the beginning of the curve, at the center of the curve, at the end of the curve, and 150 ft beyond the end

TABLE 1 SPEED AT SELECTED DISTANCES FROM CURVE SIGN FOR MODERATE CURVE (CURVE A): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

		WITH ADV. SPEED		WITHOUT ADV. SPEED	
		DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS		24	12	24	19
-500.0	AVERAGE	51.23	48.21	51.48	47.29
	STD.DEV.	3.11	3.42	2.23	2.61
0.0	AVERAGE	50.69	48.13	51.02	46.26
	STD.DEV.	2.54	3.08	2.04	2.80
422.0	AVERAGE	47.80	44.13	47.77	44.34
	STD.DEV.	3.45	2.40	2.16	3.21
883.0	AVERAGE	43.67	42.04	43.63	41.19
	STD.DEV.	2.88	2.24	1.78	2.86
1344.0	AVERAGE	48.13	44.08	47.61	43.89
	STD.DEV.	2.90	3.26	2.68	2.42
1494.0	AVERAGE	49.46	45.29	48.69	44.82
	STD.DEV.	3.05	3.63	2.93	2.65

NOTE: VELOCITY IN MILES PER HOUR
-500.0 - 500 FEET BEFORE CURVE SIGN
0.0 - AT THE CURVE SIGN
422.0 - BEGINNING OF THE CURVE
883.0 - CENTER OF THE CURVE
1344.0 - END OF THE CURVE
1494.0 - 150 FEET AFTER THE END OF THE CURVE

of the curve. *F*- and *t*-tests were performed on the vehicle measures (including velocity, lateral acceleration, lateral lane position, and gas pedal deflection) at the 0.05 level at each of the selected distance points. These tests showed very few statistically significant differences between the first and second runs and between the experienced and inexperienced subjects. These data were then combined in order to achieve larger sample sizes and therefore more sensitive statistical tests.

Tables 1 and 2 show combined group speeds (averages and standard deviations) of the experimental vehicle at selected distance points from the curve warning sign for Curves A and C, respectively, during both nighttime and daytime and with and without the advisory speed sign. The tables indicate that the speed of the vehicle decreased about 3 mph for Curve A and 8 mph for Curve C from the beginning to the center of the curve and then increased rather quickly and consistently until the end of the curve for each of the four conditions.

Table 1 shows that the average speeds at the center of Curve A (883 ft), which were the minimum speeds for the entire curve, were 43.7, 42.0, 43.6, and 41.2 mph. Table 2 shows the average speeds at the center of Curve C (642 ft), the minimum speeds for the curve, were 32.7, 29.5, 33.2, and 32.3 mph. Note that the minimum speeds for Curves A and C are higher than their respective advisory speeds of 40 and 25 mph. In fact, these average speeds and their corresponding standard deviations (between 1.6 and 4.4 mph) indicate that the use of the ball-bank indicator results in advisory speeds that are well below the 85th-percentile speeds, as discussed by Kneebone *(2)*. Tables 1 and 2 also indicate that the speeds recorded during the day were always a few miles per hour higher than those recorded at night (statistically significant at the 0.05 level in 21 of the 24 cases), when the only source of illumination was the experimental car's low beams.

There appears to be little difference between the average speeds when the curve warning sign is presented by itself and when it is presented in conjunction with an advisory speed sign. There are no statistically significant differences (at the 0.05 level) between these two experimental conditions on Curve A (the moderate curve). However, a statistically significant speed difference does exist between these two conditions at the end of Curve C (the sharp curve) during the day (38.11 mph versus 39.52 mph) and also at the center of Curve C at night (29.51 mph versus 32.28 mph). In both of these instances the drivers maintained a slightly lower average speed when the advisory speed sign was present than when the advisory speed sign was not present. However, the average lateral acceleration values for the two conditions at the center of Curve C at night are not statistically significant (0.182 g versus 0.233 g). Consid-

TABLE 2 SPEED AT SELECTED DISTANCES FROM CURVE SIGN FOR SHARP CURVE (CURVE C): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

		WITH ADV. SPEED		WITHOUT ADV. SPEED	
		DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS		24	12	24	20
-500.0	AVERAGE	48.19	42.96	48.44	43.29
	STD.DEV.	2.93	3.47	4.00	3.68
0.0	AVERAGE	42.17	38.08	42.44	39.38
	STD.DEV.	2.36	2.78	1.96	3.14
270.0	AVERAGE	41.86	37.33	42.50	39.40
	STD.DEV.	2.50	3.01	2.76	2.90
642.3	AVERAGE	32.73	29.51	33.15	32.28
	STD.DEV.	1.62	2.71	1.75	3.19
1014.6	AVERAGE	38.11	35.67	39.52	37.25
	STD.DEV.	2.22	3.98	1.85	2.46
1154.0	AVERAGE	41.57	38.38	42.79	39.75
	STD.DEV.	2.26	4.40	1.95	2.65

NOTE: VELOCITY IN MILES PER HOUR

-500.0 - 500 FEET BEFORE CURVE SIGN

0.0 - AT THE CURVE SIGN

270.0 - BEGINNING OF THE CURVE

642.3 - CENTER OF THE CURVE

1014.6 - END OF THE CURVE

1154.6 - 140 FEET AFTER THE END OF THE CURVE

TABLE 3 LATERAL ACCELERATION AT SELECTED DISTANCES FROM CURVE SIGN FOR MODERATE CURVE (CURVE A): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

		WITH ADV. SPEED		WITHOUT ADV. SPEED	
		DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS		24	12	24	19
-500.0	AVERAGE	-0.008	0.021	-0.019	0.018
	STD.DEV.	0.033	0.021	0.055	0.019
0.0	AVERAGE	0.010	0.005	0.002	0.015
	STD.DEV.	0.033	0.020	0.053	0.016
422.0	AVERAGE	0.042	0.057	0.038	0.068
	STD.DEV.	0.033	0.014	0.051	0.017
833.0	AVERAGE	0.190	0.179	0.169	0.161
	STD.DEV.	0.048	0.026	0.059	0.056
1344.0	AVERAGE	0.011	0.024	0.008	0.024
	STD.DEV.	0.026	0.022	0.057	0.023
1494.0	AVERAGE	0.021	0.031	-0.005	0.017
	STD.DEV.	0.024	0.019	0.045	0.015

NOTE: LATERAL ACCELERATION IN g'S

POSITIVE ACCEL. INDICATES LEFT CURVE

-500.0 - 500 FEET BEFORE CURVE SIGN

0.0 - AT THE CURVE SIGN

422.0 - BEGINNING OF THE CURVE

883.0 - CENTER OF THE CURVE

1344.0 - END OF THE CURVE

1494.0 - 150 FEET AFTER THE END OF THE CURVE

ering the rather small magnitude of the average speed decrease and the accompanying average lateral acceleration decrease at the center of the curve, the effect of the advisory speed sign appears to be of a rather small practical importance.

Tables 3 and 4 show the lateral accelerations (averages and standard deviations) for the selected distance points on Curves A and C, respectively. On the basis of the instrumentation, left curves result in positive lateral acceleration values, whereas right curves result in negative lateral acceleration values. The highest average lateral accelerations (0.190 g for Curve A and 0.274 g for Curve C) were obtained at about the center of each of the curves and the average acceleration values for the other distance points were close to zero. If one computes the coefficient of variation (COV = standard deviation divided by average) for the lateral accelerations, it can be seen that the COV varies from 15 to 35 percent in the center of Curve A and from 24 to 40 percent in the enter of Curve C. This variability is quite a bit higher than that found for the velocities where the COV was between 4 and 7 percent in the center of Curve A and between 5 and 10 percent in the center of Curve C. This then indicates that the drivers were able to markedly vary their lateral accelerations through fairly small steering wheel and lateral vehicle position changes without significantly varying their speed at the center of the curve. Comparing the data in Tables 3 and 4, it can be seen that the average lateral accelerations on Curve C, which requires a larger speed reduction, were higher than they were on Curve A. Because of the differences in speed between day and night conditions, the average maximum accelerations were consistently slightly lower at night than during the day.

Tables 5 and 6 show the gas pedal deflection for the selected distance points for Curves A and C, respectively. Table 5 indicates that for Curve A the average gas pedal deflection was low as the subjects entered the curve; however, by the time they reached the center of the curve the subjects began to press the gas pedal down further. At night the gas pedal was deflected further when the subjects entered Curve A than it was during the day (statistically significant); however the gas pedal was deflected further at the center of Curve A during the day than it was at night (also statistically significant).

Table 6 shows that the subjects deflected the gas pedal only slightly when entering Curve C but then very slightly increased this deflection at the center of the curve during both the daytime and nighttime. The subjects then increased the gas

TABLE 4 LATERAL ACCELERATION AT SELECTED DISTANCES FROM CURVE SIGN FOR SHARP CURVE (CURVE C): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

		WITH ADV. SPEED		WITHOUT ADV. SPEED	
		DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS		24	12	24	19
-500.0	AVERAGE	0.001	0.012	-0.007	0.003
	STD.DEV.	0.026	0.015	0.037	0.013
0.0	AVERAGE	0.068	0.035	0.037	0.034
	STD.DEV.	0.033	0.032	0.044	0.043
270.0	AVERAGE	-0.010	-0.012	-0.025	-0.010
	STD.DEV.	0.027	0.009	0.038	0.015
642.3	AVERAGE	-0.262	-0.182	-0.274	-0.233
	STD.DEV.	0.064	0.064	0.071	0.093
1014.6	AVERAGE	-0.039	-0.011	-0.037	-0.026
	STD.DEV.	0.055	0.022	0.052	0.025
1154.6	AVERAGE	0.016	0.019	0.010	0.014
	STD.DEV.	0.033	0.025	0.044	0.016

NOTE: LATERAL ACCELERATION IN g'S
NEGATIVE ACCEL. INDICATES RIGHT CURVE
-500.0 - 500 FEET BEFORE CURVE SIGN
0.0 - AT THE CURVE SIGN
270.0 - BEGINNING OF THE CURVE
642.3 - CENTER OF THE CURVE
1014.6 - END OF THE CURVE
1154.6 - 150 FEET AFTER THE END OF THE CURVE

TABLE 5 GAS PEDAL DEFLECTION AT SELECTED DISTANCES FROM CURVE SIGN FOR MODERATE CURVE (CURVE A): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

		WITH ADV. SPEED		WITHOUT ADV. SPEED	
		DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS		24	12	24	19
-500.0	AVERAGE	32.88	34.08	33.01	36.32
	STD.DEV.	8.97	8.83	7.95	8.54
0.0	AVERAGE	16.40	6.22	18.53	14.23
	STD.DEV.	10.69	3.19	9.14	9.16
422.0	AVERAGE	6.25	16.50	6.47	18.39
	STD.DEV.	1.57	11.63	1.50	11.88
883.0	AVERAGE	42.81	22.63	39.29	29.17
	STD.DEV.	7.60	12.87	12.06	11.20
1344.0	AVERAGE	31.96	24.67	30.83	27.72
	STD.DEV.	10.63	8.90	6.73	9.33
1494.0	AVERAGE	26.92	21.55	25.84	21.56
	STD.DEV.	9.32	6.26	9.24	9.86

NOTE: GAS PEDAL DEFLECTION: IDLE 1-7, FULLY DEFLECTED POSITION 69-73
-500.0 - 500 FEET BEFORE CURVE SIGN
0.0 - AT THE CURVE SIGN
422.0 - BEGINNING OF THE CURVE
883.0 - CENTER OF THE CURVE
1344.0 - END OF THE CURVE
1494.0 - 150 FEET AFTER THE END OF THE CURVE

TABLE 6 GAS PEDAL DEFLECTION AT SELECTED DISTANCES FROM CURVE SIGN FOR SHARP CURVE (CURVE C): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

		WITH ADV. SPEED		WITHOUT ADV. SPEED	
		DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS		24	12	24	20
-500.0	AVERAGE	14.46	20.34	10.65	15.35
	STD.DEV.	8.95	11.83	6.27	9.53
0.0	AVERAGE	26.92	24.09	31.24	26.78
	STD.DEV.	13.42	14.05	14.39	13.40
270.0	AVERAGE	12.07	13.75	10.96	22.20
	STD.DEV.	7.96	9.53	7.86	12.60
642.3	AVERAGE	17.08	18.63	15.43	17.58
	STD.DEV.	12.85	13.23	9.41	15.66
1014.6	AVERAGE	47.52	39.92	49.24	35.18
	STD.DEV.	8.48	17.54	11.17	11.53
1154.0	AVERAGE	44.08	27.55	36.59	26.22
	STD.DEV.	9.71	16.64	12.12	11.09

NOTE: GAS PEDAL DEFLECTION: IDLE 1-7, FULLY DEFLECTED POSITION 69-73

-500.0 - 500 FEET BEFORE CURVE SIGN

0.0 - AT THE CURVE SIGN

270.0 - BEGINNING OF THE CURVE

642.3 - CENTER OF THE CURVE

1014.6 - END OF THE CURVE

1154.6 - 140 FEET AFTER THE END OF THE CURVE

TABLE 7 LANE TRACK POSITION AT SELECTED DISTANCES FROM CURVE SIGN FOR MODERATE CURVE (CURVE A): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

		WITH ADV. SPEED		WITHOUT ADV. SPEED	
		DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS		24	12	23	20
-500.0	AVERAGE	4.85	5.16	4.68	5.15
	STD.DEV.	.70	.59	.62	.56
0.0	AVERAGE	5.06	5.89	4.78	5.71
	STD.DEV.	.73	.26	.58	.66
422.0	AVERAGE	4.44	4.28	4.15	4.63
	STD.DEV.	.55	.41	.51	.47
883.0	AVERAGE	5.35	4.03	5.37	4.52
	STD.DEV.	.63	.74	.91	.91
1344.0	AVERAGE	4.92	4.74	4.56	4.80
	STD.DEV.	.50	.66	.54	.64
1494.0	AVERAGE	4.48	4.35	4.25	4.71
	STD.DEV.	.43	.49	.60	.35

NOTE: LANE TRACKER POSITION IN FEET

-500.0 - 500 FEET BEFORE CURVE SIGN

0.0 - AT THE CURVE SIGN

422.0 - BEGINNING OF THE CURVE

883.0 - CENTER OF THE CURVE

1344.0 - END OF THE CURVE

1494.0 - 150 FEET AFTER THE END OF THE CURVE

pedal deflection as they reached the end of the curve. The drivers entered Curve C with gas pedal deflections that were lower at night than they were during the day (statistically significant only for the beginning of the curve without the advisory speed sign), but once they had passed the center of the curve, they began to press the gas pedal down further at night than they did during the day (statistically significant for all conditions at the end of the curve and 150 ft after the end of the curve except for the end of the curve with the advisory speed sign).

Many of the subjects used their brakes in addition to the gas pedal to control their speed. During the daytime, 15 of the 24 drivers used their brakes in Curve A, whereas 22 of the 24 drivers used their brakes in Curve C. At night 8 of the 16 drivers used their brakes in Curve A, whereas 12 of the 16 drivers used their brakes in Curve C. For both curves all average brake application distances for both daytime and nighttime were between the curve sign and the center of the curve.

Tables 7 and 8 show averages and standard deviations of the lateral lane position of the car measured in feet from the inside of the edge line to the longitudinal center of the car. Table 7 shows that for Curve A, the lateral lane position was larger in the center of the curve than at either end of the curve. It can also be seen that the higher speeds and lateral accelerations that were accepted by the subjects during the day were accompanied by increases in lateral lane position at the center of Curve A (statistically significant at the 0.05 level). Table 8 shows that, on the average, when the subjects entered Curve C, they positioned the experimental car left of the imaginary center of the right lane until they reached the center of the curve, at which point they swung toward the right and remained to the right of the imaginary center of the right lane throughout the remainder of the curve. However, considering the 9.2-ft lane width and 4.7-ft outside tire track width of the experimental car, the subjects were within their lane for all average values on both curves.

The average speeds of the vehicle when driven by test drivers and "local familiar" drivers on Curve A are compared in Figures 1 and 2. Because there were very few differences between the speeds of the test drivers exposed to the advisory speed sign and those who were not, the data for these two groups were combined for comparison with the data for the local familiar drivers. Figure 1 shows that both the test drivers and local familiar drivers decreased their speed by about 2 to 3

TABLE 8 LANE TRACK POSITION AT SELECTED DISTANCES FROM CURVE SIGN FOR SHARP CURVE (CURVE C): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

		WITH ADV. SPEED		WITHOUT ADV. SPEED	
		DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS		24	12	24	20
-500.0	AVERAGE	4.32	4.28	4.12	4.71
	STD.DEV.	.56	.52	.50	.57
0.0	AVERAGE	5.44	4.47	5.13	5.31
	STD.DEV.	1.03	1.03	.58	.60
270.0	AVERAGE	5.49	5.33	5.48	5.45
	STD.DEV.	.63	.57	.69	.53
642.3	AVERAGE	4.60	4.96	4.27	5.46
	STD.DEV.	.84	.86	.79	.60
1014.6	AVERAGE	4.10	4.25	4.00	3.99
	STD.DEV.	.62	.48	.82	.76
1154.0	AVERAGE	4.24	4.08	4.13	4.77
	STD.DEV.	.42	.58	.71	.67

NOTE: LANE TRACKER POSITION IN FEET
- -500.0 - 500 FEET BEFORE CURVE SIGN
- 0.0 - AT THE CURVE SIGN
- 270.0 - BEGINNING OF THE CURVE
- 642.3 - CENTER OF THE CURVE
- 1014.6 - END OF THE CURVE
- 1154.6 - 140 FEET AFTER THE END OF THE CURVE

mph before they reached the curve warning sign. However, once past the curve warning sign, the local familiar drivers, who had been going 50 mph 70 ft before the sign, reduced their speed to about 42 mph at the beginning of the curve, whereas the test drivers, who had been going 50 mph 70 ft before the sign, decreased their speed to about 48 mph at the beginning of the curve. This resulted in statistically significant speed differences for all distances after the curve warning sign had been passed. Figure 2 indicates that both groups of subjects maintained approximately the same average speeds at night until about 180 ft after they had passed the curve warning sign. From this point, the test drivers' speeds gradually increased, whereas the local drivers' speeds slowly decreased until the beginning of the curve, where their speeds were about 2 mph slower (statistically not significant) than the test drivers' speed.

The average speeds of the vehicle driven by the test drivers and the local familiar drivers on Curve C are compared in Figures 3 and 4. Figure 3 shows that both groups of drivers decreased their speed until about 185 ft before the curve warning sign. The data also show that during the day the local familiar drivers drove slightly faster than the test drivers on Curve C from 77 ft before the curve warning sign to 245 ft (statistically significant for all values between 77 and 185 ft) past the curve warning sign. Figure 4 shows that both groups of drivers decreased their speed at about the same rate when approaching the curve warning sign on Curve C at night to a point about 185 ft before the curve warning sign. Then the test drivers maintained a speed of about 37 mph, whereas the local familiar drivers actually increased their speed, reaching a maximum of about 43 mph, and then decreased their speed more rapidly heading into the curve. This resulted in speed differences that are statistically significant at the 0.05 level for only 121 and 185 ft past the curve warning sign. The results also indicate that the local familiar drivers tend to drive faster

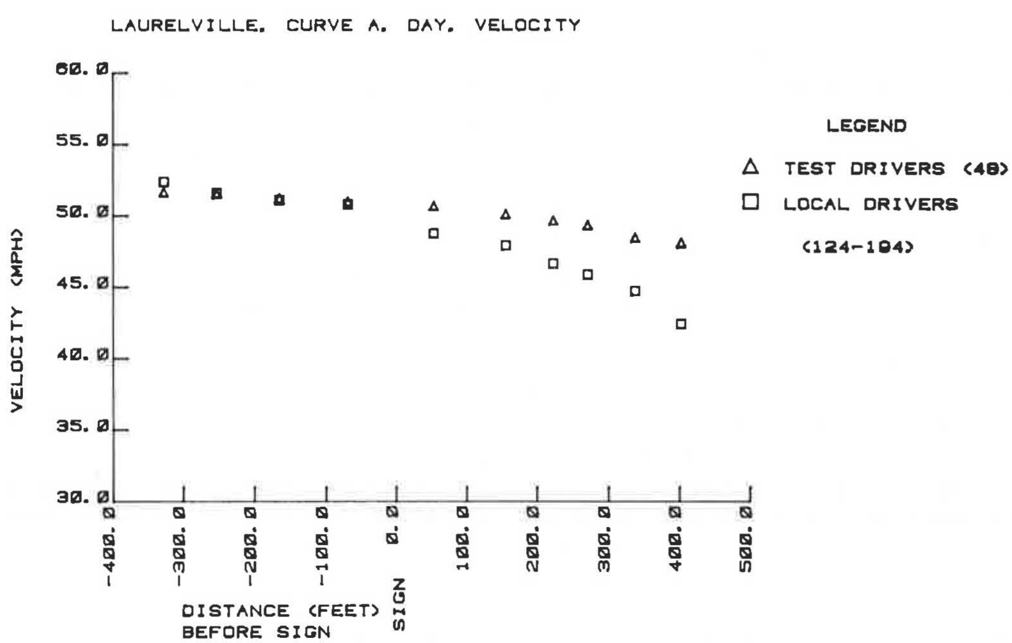


FIGURE 1 Comparison between local familiar and test drivers for Curve A during the daytime.

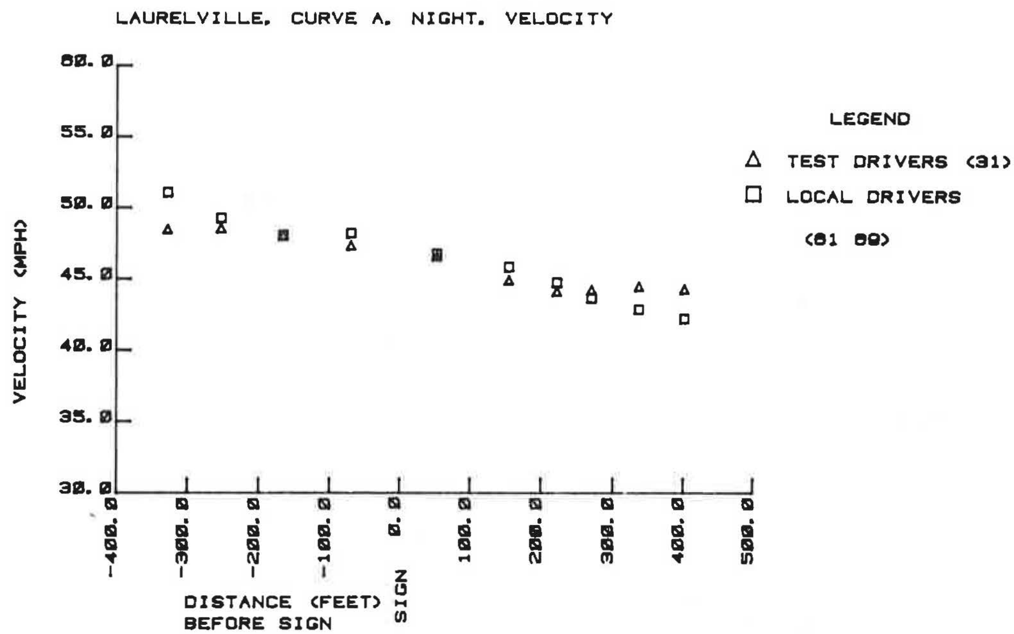


FIGURE 2 Comparison between local familiar and test drivers for Curve A during the nighttime.

around curves during the day than they do at night (statistically significant in 14 of the 19 cases tested).

Eye Scanning

Tables 9 and 10 present eye-scanning data for the curve signs on Curves A and C with and without the advisory speed sign for both daytime and nighttime conditions. These tables show that the subjects looked at the curve warning sign an average of between 1.6 and 3.5 times, with an average look duration of 0.51 to 0.62 sec. The average fixation durations when the advisory speed sign was present were equal or slightly larger (statistically not significant at the 0.05 level) than when the advisory speed sign was not present for each of the four conditions.

Tables 9 and 10 also show the first- and last-look distances (averages and standard deviations). First-look distances were

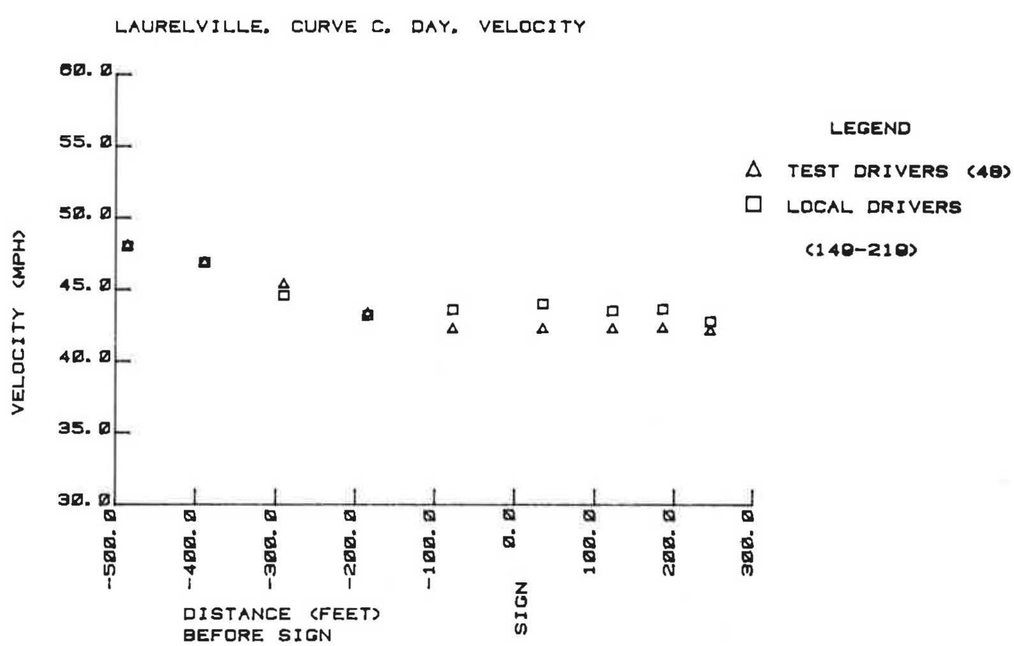


FIGURE 3 Comparison between local familiar and test drivers for Curve C during the daytime.

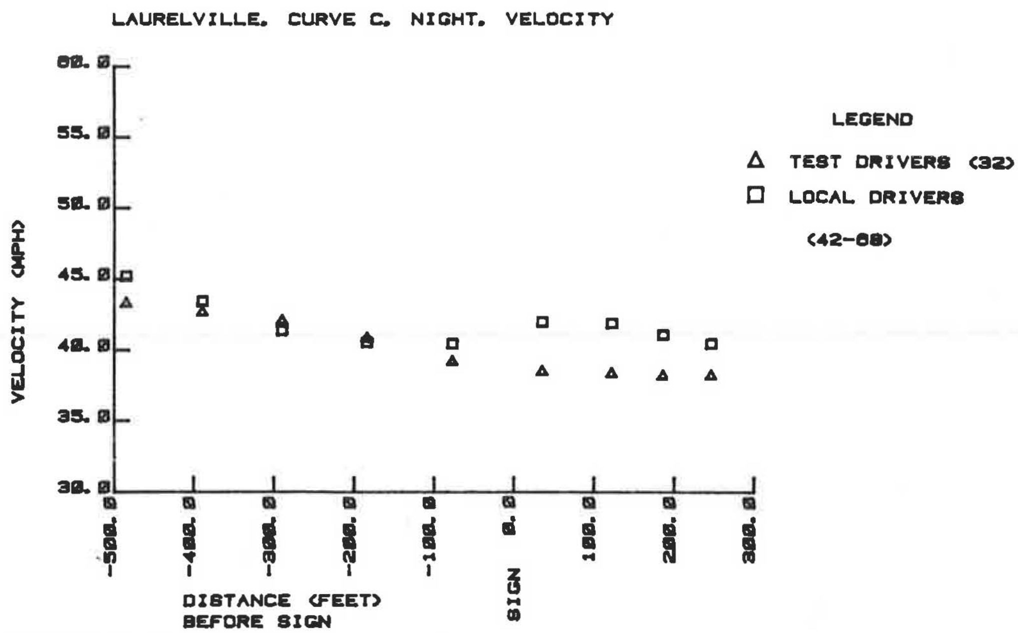


FIGURE 4 Comparison between local familiar and test drivers for Curve C during the nighttime.

TABLE 9 EYE-SCANNING SUMMARY RESULTS FOR MODERATE CURVE (CURVE A): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

	WITH ADV. SPEED		WITHOUT ADV. SPEED	
	DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS	24	12	24	19
TOTAL NO. OF LOOKS	59	29	37	67
LOOKS/SUBJECT-AVERAGE	2.45	2.42	1.55	3.53
-STD.DEVIATION	1.59	0.90	1.06	2.39
LOOK DURATION - AVERAGE	0.58	0.61	0.51	0.58
-STD.DEVIATION	0.45	0.37	0.36	0.43
FIRST LOOK DIST.-AVERAGE	490.	502.	456.	563.
-STD.DEVIATION	157.	136.	158.	160.
FL VISUAL ANGLE (ARROW)	11.7	11.4	12.6	10.2
FL VISUAL ANGLE (NUMBERS)	4.7	4.6		
AVERAGE FIRST LOOK TIME	6.5	6.3	6.1	7.6
LAST LOOK DIST. -AVERAGE	232.	180.	291.	215.
-STD.DEVIATION	96.	87.	149.	117.
LL VISUAL ANGLE (ARROW)	24.7	31.8	19.7	26.6
LL VISUAL ANGLE (NUMBERS)	9.9	12.7		
AVERAGE LAST LOOK TIME	3.1	2.5	3.8	3.1

NOTE: ALL THE UNITS OF TIME ARE IN SECS., ALL DISTANCES ARE IN FEET AND ALL VISUAL ANGLES ARE IN MINUTES OF ARC.

FL = FIRST LOOK

LL = LAST LOOK

TABLE 10 EYE-SCANNING SUMMARY RESULTS FOR SHARP CURVE (CURVE C): EXPERIENCED AND INEXPERIENCED SUBJECTS AND RUNS 1 AND 2

	WITH ADV. SPEED		WITHOUT ADV. SPEED	
	DAY	NIGHT	DAY	NIGHT
NO. OF SUBJECTS	24	12	24	20
TOTAL NO. OF LOOKS	50	22	47	31
LOOKS/SUBJECT-AVERAGE	2.08	1.83	1.96	1.55
-STD.DEVIATION	1.35	0.85	1.67	1.10
LOOK DURATION - AVERAGE	0.62	0.51	0.48	0.51
-STD.DEVIATION	0.48	0.43	0.31	0.24
FIRST LOOK DIST.-AVERAGE	406.	248.	390.	265.
-STD.DEVIATION	209.	60.	192.	62.
FL VISUAL ANGLE (ARROW)	14.1	23.1	14.7	21.6
FL VISUAL ANGLE (NUMBERS)	5.6	9.2		
AVERAGE FIRST LOOK TIME	6.1	5.1	6.0	5.5
LAST LOOK DIST. -AVERAGE	212.	140.	201.	186.
-STD.DEVIATION	198.	46.	117.	50.
LL VISUAL ANGLE (ARROW)	27.0	40.9	28.5	30.8
LL VISUAL ANGLE (NUMBERS)	10.8	16.4		
AVERAGE LAST LOOK TIME	3.3	2.4	3.1	3.5

NOTE: ALL THE UNITS OF TIME ARE IN SECS., ALL DISTANCES ARE IN FEET AND ALL VISUAL ANGLES ARE IN MINUTES OF ARC.

FL = FIRST LOOK

LL = LAST LOOK

defined as the distances measured from the curve warning sign to the point at which the driver begins to fixate his or her eyes (foveally, near foveally, or slightly peripherally) for the first time on the sign. Last-look distances were defined as the distances measured from the curve warning sign to the position where the driver moves his eyes away from the sign and does not look at it again. The visual angle when the driver was looking at either the approximately 20-in.-high black arrow on the curve warning sign or the 8-in.-high numbers on the advisory speed sign is given for both the average first-look distance and the average last-look distance in these tables. Also shown is the time it took for the experimental vehicle to pass the curve warning sign from the position of the average first-look distance and the average last-look distance.

From Tables 9 and 10 it can be seen that during the daytime, regardless of whether the advisory speed sign was present or not, the average first-look distances were considerably smaller than the maximum distances at which the signs were visible (1,036 ft for Curve A and 953 ft for Curve C). It was determined that the direction and angle of the 4-in.-wide black arrow (maximum symbol height about 20 in.) on the yellow curve warning sign could be distinguished during the daytime by a driver with 20/20 vision or better from 1,000 to 1,500 ft (visual angles 5.7 and 3.8 min of arc), and when the advisory speed sign is present, the 8-in. black numbers on the 18 × 18-in. advisory speed sign can be read by a driver with 20/20 vision or better from about 500 to 600 ft away from the curve warning sign during the day. Therefore the drivers should have been able to clearly see the shape and direction of the arrow on the curve warning sign at these average first-look distances, although they may have had some slight difficulties in reading the numbers on the advisory speed signs when they were present.

Tables 9 and 10 show very different first-look distances for the two curves at night regardless of whether the advisory speed sign was used. This may be due to the different road geometries and the different low-beam illumination conditions that could have been present on the two curves. For Curve A one can see that the average first-look distances were actually higher at night than they were during the day, whereas the first-look distances for Curve C were somewhat smaller at night than they were during the day. Even if one assumed that the maximum distances at which the arrow symbol on the curve sign could be detected at night were about 50 percent shorter than during the daytime (due to less favorable illumination conditions), the drivers should still have had little difficulty in distinguishing the shape and the direction of the arrow at the average first-look distances. However, considering the visual angles that apply to these average first-look distances during the night, it is rather doubtful that the drivers would have been able to read the advisory speed numbers shown on the advisory speed sign on Curve A but may have been able to read those displayed on the advisory speed sign on Curve C.

During the day when the curve warning sign was displayed alone, the average last-look distances were about 200 to 300 ft. Therefore, the visual angles were fairly large and the test drivers should have been able to perceive the shape and direction of the arrow on the curve warning sign very easily. These average last-look distances represent average driving times to the curve sign from 3.1 to 3.8 sec. When the curve warning sign was presented with the advisory speed sign, the average last-look distances during the daytime for the two curves were between 212 and 230 ft. The subjects should have been able to read the 8-in.-high numbers on the advisory speed signs easily, because at these distances the visual angles were between 9.9 and 10.0 min of arc. Again, these average last-look distances represent average driving times to the curve sign from 3.1 to 3.3 sec.

The average last-look distances at night, when the curve warning sign was displayed by itself, were 201 to 291 ft, with visual angles between 29 and 27 min of arc. Therefore the shape and direction of the arrow on the curve warning sign should be easily distinguishable by a driver with 20/20 vision or better. When the advisory speed sign was present, the last-look distances were 140 and 180 ft and their visual angles for the 8-in.-high numbers were 12.7 to 16.4 min of arc, which should be sufficient for the drivers to read the numbers on the advisory speed sign rather easily. Regardless of whether the advisory speed sign was present or not, the average last-look distances represent average driving times to the curve sign of between 2.4 and 3.5 sec.

In summary the results of this study indicate that the drivers first look at a curve warning sign with or without the advisory speed sign at a distance where they are close enough to the sign to be able to distinguish the shape and direction of the black arrow on the yellow background but may have some difficulty in reading the numbers on the advisory speed sign, if one is present. It appears that drivers prefer to acquire information from the road scene rather than from the warning signs when they are within about 2.5 to 3.5 sec from the warning sign. Although it would appear that the drivers should have acquired the information displayed on the signs, they still failed to slow down enough to reach the determined safe advisory speed regardless of whether the advisory speed sign was present.

CONCLUSIONS

From the results of this study, it may be concluded that advisory speed signs are not more effective in causing drivers to reduce their speeds through curves than the curve signs alone. It appears that the bent black arrow in the yellow diamond of the curve sign represents such a strong and primary visual stimulus that an advisory speed plate adds very little additional information. Therefore, it is recommended that advisory speed sign maintenance and especially new installations be given a low priority.

ACKNOWLEDGMENTS

This study was sponsored by the Ohio Department of Transportation in cooperation with FHWA, U.S. Department of Transportation.

REFERENCES

1. M. L. Ritchie. Choice of Speed Driving Through Curves as a Function of Advisory Speed and Curve Signs. *Human Factors*, Vol. 14, No. 6, 1972, pp. 533–538.
2. D. C. Kneebone. *Advisory Speed Signs and Their Effect on Traffic*. Paper 150. Department of Main Roads, New South Wales, Australia.
3. C. G. Hammer. Evaluation of Minor Improvements. In *Highway Research Record 286*, HRB, National Research Council, Washington, D.C., 1969, pp. 33–45.
4. D. Shinar, E. McDowell, and T. Rockwell. Eye Movements in Curve Negotiation. *Human Factors*, Vol. 19, No. 1, Feb. 1977, pp. 63–71.
5. H. T. Zwahlen. *Warning Signs and Advisory Speed Signs—Reevaluation of Practice*. Report FHWA/OH-84/003. Ohio Department of Transportation, Columbus, June 1983, 185 pp.

This paper does not necessarily reflect the views or policies of the Ohio Department of Transportation or FHWA. The contents of this paper reflect the views of the author, who assumes responsibility for the facts and accuracy of the data presented here. This paper in no part constitutes a standard, specification, or regulation.

Publication of this paper sponsored by Committee on Vehicle User Characteristics.

Alternatives to Enforcement in Modifying the Speeding Behavior of Drivers

STEPHEN MARONEY AND ROBERT DEWAR

Two experiments were conducted to examine alternatives to enforcement as a means of reducing speeding behavior. The first employed transverse lines painted on the roadway at progressively diminishing distances apart to produce an alerting response and an illusion of vehicle acceleration. Data were collected on 247,036 vehicles during a period of several weeks. The number of drivers that were exceeding the recommended speed by more than 30 km/hr was reduced by 25 percent, but this effect on speed began to disappear after 3 weeks. The second experiment provided drivers with feedback (using a traffic sign) about the percentage of drivers who were not speeding on the previous day. Data gathered on approximately 690,000 vehicles during 3.5 months indicated that excessive speeding could be reduced by 40 percent. This speed reduction was maintained for weeks after the sign was removed. The implications of these findings for alerting drivers and reducing speeding behavior are discussed.

It is undeniable that the traffic safety problem is very diverse and immense in scope. The high annual cost in human life and property damage also gives it priority. Although roads, vehicles, and traffic control devices are constantly being improved and made safer, there has been little advancement in the safety features of the human operator.

The traditional method of driver control is the imposition of safety rules and the use of police enforcement to obtain compliance. The presence of enforcement usually results in an immediate reduction in offenses *(1–3)*. There is also a reduction in compliance after the police have left (known as the residual impact), depending on the duration and regularity of enforcement and the frequency with which the same traffic uses the same roadway at the same time each day *(4, 5)*.

The major drawback to enforcement as a means of controlling behavior is that it does not appear to be effective over the long term. Police personnel are in short supply and very expensive, making enforcement programs haphazard and short term. Because detection rates are low, compliance is also low as enforcement does not affect the intentions of the violator.

In-depth collision analyses *(6)* have indicated that a significant proportion of traffic collisions is due to failure of drivers to pay proper attention. Although violations such as speeding are very often committed intentionally, it may well be that excessive speeding (e.g., 25 to 35 km/hr over the limit) is a result of driver inattention. Drivers may miss seeing traffic signs that indicate speed limit changes or may not be paying sufficient attention to their own vehicle speed. Increased enforcement will not necessarily increase driver attention.

By decreasing the overall demand of enforcement programs, police personnel can be directed specifically to those locations where there is the greatest need with the enforcement consistency required to make a long-term impact. It is necessary, therefore, to develop programs that encourage drivers to comply with traffic regulations regardless of police presence. The following experiments were performed to investigate two possible approaches.

S. Maroney, Calgary Police Service, 316 7th Avenue, S.E., Calgary, Alberta T2G 0J2, Canada. R. Dewar, Psychology Department, University of Calgary, Calgary, Alberta T2N 1N4, Canada.

EXPERIMENT 1: DIMINISHING LINES

Experiment 1 evaluates the impact of a strong visual stimulus on speeding. It is posited that the stimulus will increase driver attentiveness and thereby reduce speeding. The stimulus used was a series of reflective white lines, similar to those used by Helliar-Symons *(7)*, painted transversely across an exit ramp of a freeway in the city of Calgary. It has been demonstrated that the use of these lines has been quite effective in reducing speed and collisions *(7, 8)*.

Method

Subjects

Speed measurements were collected over a 5.5-week period for a total of 752 hr of data. The speeds were collected from motorists traveling the 900-m southbound exit ramp from the Deerfoot Trail freeway to the 16th Avenue split diamond interchange. A total of 247,036 vehicle speeds were recorded. Given the nature of the roadway, a considerable number of vehicles were repeat users of the ramp; however, there was also a large number of infrequent or novice users at any given time. Exact proportions were not recorded. The drivers involved were unaware that they were participating in a speed study and that their speeds were being recorded.

Apparatus

Speed and volume were measured by using a Stevens PPR II Print-Punch Traffic Classifier manufactured by Leupold and Stevens Inc. This instrument is a microprocessor-controlled, electromechanical recorder that converts output pneumatic pulses from rubber hoses to a printed record. Two ¼-in. hoses were laid perpendicular to the traffic flow 4.9 m apart. Impulses were stored in memory and then printed as a summary for a specific time interval in standard ASCII punched format on a 2.5-cm heat-sensitive paper tape. For this study the time interval was 1 hr, meaning that all vehicle speeds were averaged over a 1-hr period.

Procedure

The site for the experiment was chosen for several reasons. First, the ramp had a posted advisory speed of 50 km/hr, which was considerably less than the speed limit of 100 km/hr on the freeway. Because this was an advised speed, there was no legal compulsion for drivers to slow down. The intersection was a split diamond with straight off ramps so that drivers could operate at any speed desired given free-flow conditions. Any reduction in speed should therefore reflect a change in driver decision making. Second, the ramp approached a signal-controlled intersection that had a high collision rate. Third, a speed problem on the ramp had been independently identified through collision investigation and direct observation. Finally, the ramp was a straight, long, single-lane road, which made it ideal for painting the lines and installing the data collection equipment.

On April 28, 1982, the classifier was installed 130 m upstream of the traffic control light and 770 m downstream of the beginning of the ramp. Data were collected until May 16 that formed the pretest baseline phase of the study. Vehicle speeds in each hour were averaged by the classifer to obtain a score for that hour. The number of vehicles during each hour ranged from 4 to 865, with an average of 328 vehicles. Although this technique does not differentiate between free-flow and platooned traffic, this was not considered a problem because the average platoon speed would approximate that of the platoon leader. Also, speeding was not a particular problem when traffic was platooned. There were 336 hr of data collection (106,444 vehicles or 318 per hour) in the pretest baseline phase.

The ramp was then closed for approximately 8 hr while 90 fluorescent white lines were painted transversely on the ramp over a 404-m distance (Figure 1). Each line was 60 cm wide and 4 m long (curb to curb). The distance between the lines gradually decreased from 7.7 m at the start to 2.75 m, according to the specification outlined by Helliar-Symons *(7)*. The lines started 100 m after the ramp began and finished 400 m from the traffic light. After the lines were installed, data collection continued until June 5. There were 416 hr of data collection (140,622 vehicles or 338 per hour) in this phase.

In both phases of the experiment, data collection was not continuous. The most common reason for lost data was either a break in one of the hoses or condensation following rain or snowfall. Several hours of data recorded during snow or rain storms were also omitted, because they were unrepresentative of the driving conditions. In Phase 1, 68 hr from a total of 404 (16.8 percent) was lost. In Phase 2, 60 hr from a total of 466 (12.6 percent) was not reported. Breaks (due to street cleaning) were more of a problem in Phase 1 than in Phase 2, where condensation appeared more prevalent. The hoses were checked every morning and if necessary replaced promptly so that only small blocks of time were lost. The damage was primarily confined to early morning hours or weekends. Overall the weather was relatively consistent throughout the study. Although there was slightly more snowfall during the pretest phase, the ramp itself did not ice up and there was no apparent weather-related effect on speed.

Results

Two dependent variables were measured in this study, mean speed and the percentage of vehicles travelling faster than 80 km/hr, which is 30 km/hr over the advised speed. The latter was of interest because this speed was considered a high risk for collisions and thus a desirable target for speed reduction. Using a one-tailed *t*-test and the Mann-Whitney *U*-test, significant decreases were found in both as a result of the diminishing lines ($t = -2.64$ for mean speed and -3.73 for percentage faster than 80 km/hr, $df = 285$, $p < 0.001$). There were also decreases

FIGURE 1 Transverse pavement markings used in Experiment 1.

TABLE 1 EFFECT OF DIMINISHING LINES ON SPEED (Experiment 1)

	Mean Speed (km/hr)	Standard Deviation (km/hr)	Percent >80 km/hr	Standard Deviation (km/hr)
Before	63.46	2.96	5.45	3.78
After	61.44	2.81	4.06	3.07
Percent difference	−3.3		−25.5	
t-value	−2.64[a]	–	−3.73[a]	–

[a] $p < 0.001$.

in the standard deviations within both measures. The results are detailed in Table 1. The effects decayed over time, as shown in Table 2.

The pretest and posttest data were also compared by day of the week and hour of the day to determine whether the effects were consistent. The results for each day of the week for which there was a complete 24 hr of data were averaged to obtain a daily score. An hourly score was obtained by averaging all like hours regardless of day of the week. This analysis indicated that the decrease in the dependent variables was remarkably consistent; there were decreases in almost every instance.

TABLE 2 COMPARISON OF POSTTEST CONDITIONS

Week	Hours of Data	Mean Speed (km/hr)	Percent >80 km/hr
First	148	60.43	3.19
Second	130	61.89	4.55
Third	138	62.09	4.50
Total	416	61.44	4.06

Discussion

The results support the hypothesis that a strong visual stimulus can result in a reduction in speed. The lines create a startle effect and an illusion of acceleration, which causes drivers to pay more attention to their surroundings. Drivers responded by reducing their speed to comply more closely with the advised speed limit. The data also show that this reaction was consistent across each hour of the day and each day of the week.

It is also assumed that drivers who are the most inattentive would not decrease their speed until forced to by the traffic light. This group would most likely be exceeding 80 km/hr and an increase in attentiveness should be highly pronounced. The data support this conclusion, as shown by the 25.5 percent decrease in those exceeding 80 km/hr, which is a decrease of 764 fewer drivers a week, given an average daily volume of 8,000 vehicles.

The major failing of this manipulation is indicated in the decay of the speed reduction effect. The lines performed extremely well in the first week and then the effect dropped off. Continued monitoring of speed might have revealed a return to the baseline rates or a plateau somewhere in the middle. As mentioned, monitoring was discontinued in June because of an anticipated change in the traffic mix during the summer months, when holidays and other factors bring an increase in out-of-town drivers. It is suggested that this decay occurred because drivers learned that the lines were there and anticipated them. This reduced the startle effect so that it could be ignored. There was also no compulsion for motorists to drive 50 km/hr, because the speed posted is advisory only.

Another confounding factor was the presence of the traffic lights at the top of the ramp. Many motorists would likely ignore the lines if slowing down would prevent them from reaching the intersection on a green light. Alternatively, the lights may have slowed down some drivers who otherwise would have continued their high speed. Although the traffic light does act as a speed controller, its performance was consistent across both conditions. The light was also 130 m from the traffic classifier, so it is unlikely that its impact had a significant influence on recorded speeds.

In conclusion, the results of this study support the use of these lines to decrease speeding, particularly excessive speeding. It could be validly questioned whether a mean speed reduction of 1.5 to 2 km/hr is of practical importance. This measure simply indicates the direction of the speed change. The actual impact of the lines is apparent in the considerable reductions in the number of drivers who were traveling well above the safe speed. Reduction of their speed reduced the overall speed variance, making the ramp that much safer. Further work will have to be done to validate this effect, particularly in areas where drivers are required to slow down, but frequently do not. It would also be useful to select an area that is not affected by other traffic control devices. An ideal location for further testing would be playground and school zones, where there is a high offense rate but also strong legal and social pressures to comply with the speed limits. The fact that drivers speed in such areas despite these pressures could indicate a high level of inattention.

EXPERIMENT 2: PUBLIC POSTING

The previous experiment demonstrated that increased attentiveness can result in a decrease in speeding. VanHouten et al. *(9)* employed a method of associating the threat of enforcement with an attentiveness-increasing stimulus (a sign indicating the percentage of drivers not speeding) and obtained very positive results. The present study was performed to replicate VanHouten's results by using more sophisticated data collection techniques and a larger sample size.

Method

Subjects

Speed data were collected over a 3.5-month period from motorists traveling southbound on a two-lane arterial road in the northwest part of Calgary. A total of 1,722 hours of speed data, representing the speeds of 690,614 vehicles, were collected. The drivers in the sample were unaware of participating in a speed study.

FIGURE 2 Speed feedback sign used in Experiment 2.

Apparatus

Speed and volume were recorded with the Stevens PPR II Print-Punch Traffic Classifier described in the first experiment. The recording interval was 1 hr. Each hour of data is an average of the speeds of all vehicles in that hour. The number of vehicles during any hour ranged from 6 to 982. The independent variable was a large metal sign measuring 1.3 by 2.3 m. The sign was green with white lettering that read Drivers Not Speeding Yesterday and Best Record. After each of these phrases was a space to insert a percentage. The sign was mounted 2 m above the ground on aluminum posts that were anchored in cement blocks (Figure 2).

Procedure

The sign was installed 30 m south of a sign indicating a speed reduction ahead from 70 to 50 km/hr. A sign for the 50-km/hr speed limit was located another 20 m further south. Approximately 150 m south of the message sign the road narrowed from two lanes southbound to one lane. Speed data were measured 500 m south of the message sign approximately halfway down a 200-m downgrade. This location was selected for several reasons. First, previous enforcement efforts had shown that this particular roadway had a high incidence of speeders and that speeding was a major contributing factor to collisions in the area. Second, the speed limit reduction from 70 to 50 km/hr was similar to that used by VanHouten et al. *(9)* and was replicated for comparison purposes. Third, the narrowing of traffic from two lanes to one in each direction made speed reduction even more important to safety. Finally, the area was totally devoid of any structures or artifacts that might obscure the sign from view or distract the drivers. The sign was clearly visible for almost 800 m.

In the pretest phase, data collection began June 12, 1982, and continued until October 2, 1982. A total of 417 hr of data were collected to determine the baseline behavior. On July 8, Phase 2, the first experimental phase, started, in which the sign was in place, but no percentages were indicated. In Phase 3 the sign contained percentages indicating speeding. This phase continued until August 17. Phases 2 and 3 covered a total of 712 hr. The final phase was a posttest measurement period during which the sign was removed. This continued for 593 hr until October 2. The hours of data collected and the number of vehicles in each phase are summarized in Table 3.

During Phase 2, the sign was present without any percentages displayed on it, to determine how much impact a large sign that mentioned speeding would have on drivers' speed. Then the percentages were changed regularly to reflect average vehicle speed based on a daily readout of the Leupold classifier data. The "yesterday" percentage on the sign increased or decreased depending on the speed reading and ranged from 79 to 94 percent. The "best record" percentage started at 79 percent and increased monotonically until it peaked at 94 percent. This number represented the best percentage that drivers had reached to that date.

TABLE 3 SUMMARY OF PUBLIC POSTING EXPERIMENT

Phase	Date	Hours of Data	No. of Vehicles
1 (pretest)	June 12–July 8	417	174,900
2 (sign alone)	July 8–21	149	59,150
3 (sign with percentages)	July 21–Aug. 17	563	214,120
4 (posttest)	Aug. 17–Oct. 2	593	242,444
Total	3.5 months	1,722	690,614

Results

The results support the use of this technique for reducing speed. The sign produced significant reductions across all three dependent variables: mean speed ($F = 24.5$, $df = 3/43$, $p < 0.001$), percentage of drivers traveling faster than 65 km/hr ($F = 53.4$, $df = 3/43$, $p < 0.001$), and percentage of drivers traveling faster than 80 km/hr ($F = 30.7$, $df = 3/43$, $p < 0.001$), as indicated by one-way analyses of variance. There were also reductions in the standard deviation within each measure. Although the posted speed limit was 50 km/hr, 65 km/hr was used as the cutoff between speeding and not speeding. This speed is specified by the Alberta Highway Traffic Act as the boundary between a \$20 and \$30 summons and is commonly used as a definition of speeding. The next cutoff, between a \$30 and a \$75 fine, is 80 km/hr, which is considered to be a high-risk speed for this location. The results of the data analysis are summarized in Table 4, which also shows the percentage of difference between this phase and the pretest. The percentage of drivers traveling faster than 50 km/hr is also shown; this was not used as a criterion in this study, although the difference was significant.

The data were also tested by using a Student Newman Keuls procedure to determine differences between phases. The pretest data were significantly greater than those for any of the other three phases across all measures. The only significant difference among Phases 2–4 was between Phases 2 and 3 in the percentage over 65 km/hr conditions. Neither of these was significantly different from Phase 4.

Discussion

The sign appears to have made a significant impact on drivers' speeds. During the pretest phase, 25.2 percent of the traffic was exceeding the speed limit by at least 16 km/hr. This represents an average of more than 2,400 vehicles every day. Of this group, 336 vehicles were traveling in excess of 80 km/hr. Yet this area was well enforced by police. During Phase 3, the percentage of speeders fell to 15 percent, a reduction of almost 980 vehicles daily, or 40 percent. The high-risk group also fell by 40 percent, a reduction of 134 vehicles daily.

Another interesting result was the duration of the effect. A substantial decrease across all dependent variables was still observed during October, 4 weeks after the sign had been removed. Although it is possible that some unaccounted-for secondary variable was at work keeping the speeds down, it is also possible that the sign made a lasting behavioral change in the regular users of that route. It is also possible that the removal of the sign was in itself a sufficient change in the environment to increase attention levels and cause awareness of speeds and speed limit signs.

Obviously there are still questions to be answered regarding the use of this tool. On the basis of work by VanHouten and Nau *(10)*, it seemed that there were certain factors that could influence the sign's effectiveness. It appeared to lose its impact if the "yesterday" percentage did not change at least a few times a week. There was also a drop if that percentage and the "best record" fell below 70 to 80 percent. It appears that the larger the percentage not speeding, the larger the impact of the sign. It may also be that if the compliance rate is very low initially, the speed limit may be improper for that area and the sign will not overcome the resistance. Further work with the sign will have to be completed before these variables can be properly evaluated.

As an aid to enforcement, however, this sign has shown potential for being very cost-effective. The sign cost approximately \$500 to make and required 1 hr in installation time. By setting up this device upstream of an area marked for special attention by radar patrols, enforcement personnel can be fairly confident that those drivers who do violate the speed limit are doing so intentionally. VanHouten and Nau *(10)* have already demonstrated that when the sign is coupled with enforcement, the impact is longer lasting. If the impact does decay at any time, the sign can simply be moved to a new location for a few months and perhaps be replaced by some other device. It is not necessary to tie up expensive data collection equipment to use this device. Once this study was completed, data for each hour were correlated with daily averages and it was determined which hours of the day were the most representative of the day as a total. This allowed an unmarked radar car to sample vehicle speeds during these times over 2 to 3 hr a week and to collect sufficient speed data to change the percentage figures on the sign with a high degree of validity.

CONCLUSIONS

Both of the foregoing approaches were effective in reducing speeding behavior through an environmental cue. The public posting was more effective and longer lasting. In retrospect this may have been because the diminishing lines were not used at

TABLE 4 RESULTS OF PUBLIC POSTING EXPERIMENT

				>50 km/hr		>65 km/hr		>80 km/hr	
Phase	Mean Speed (km/hr)	Percent Difference	Standard Deviation	Percent	Percent Difference	Percent	Percent Difference	Percent	Percent Difference
1	61.5	–	3.1	89.8	–	25.2	–	3.5	–
2	59.4	3.4	–	86.2	4.0	18.7	25.8	2.1	40.0
3	58.7	4.0	–	83.4	7.1	15.0	40.5	2.1	40.0
4	59.1	25.8	–	84.7	5.7	16.4	34.9	2.2	37.1
F-test	24.5[a]	40.0	–	10.4[a]	–	53.4[a]	–	30.7[a]	–

NOTE: F critical = 4.31, $a = 0.01$, $df = 3/43$.
[a]Significant $p < 0.001$.

an appropriate location. The diminishing lines appear very effective in stimulating driver awareness to the surroundings and creating an illusion of acceleration. There will be no long-term effect on speeding, however, unless there are valid and obvious reasons for the driver to slow in that particular area. The stronger implied surveillance and threat of enforcement may have contributed to the greater impact of the feedback sign, although more work will have to be done to factor out the respective impact of each.

Once the proper location has been determined for these techniques, they can be complementary to an enforcement program. Both are relatively permanent fixtures and remain active 24 hr a day, therefore being visible to all road users. They are also very effective in reducing inattention, thereby allowing the driver an opportunity to comply with the speed limit on his own accord. Drivers who ignore these devices and speed anyway usually do so intentionally and become suitable candidates for behavior modification through police enforcement. This allows police to concentrate their efforts on those locations and times when drivers are most likely to respond to their influence.

REFERENCES

1. P. J. Cooper. Effects of Increased Enforcement at Urban Intersections on Driver Behaviour and Safety. In J. Cooper *Effectiveness of Traffic Law Enforcement,* Delevin, Cather, and Co. of Canada, Toronto, 1974.
2. F. M. Council. *A Study of the Immediate Effects of Enforcement on Vehicular Speeds.* Highway Safety Research Center, University of North Carolina, Chapel Hill, 1970.
3. K. B. Joscelyn, T. H. Bryan, D. H. Golden-Baum. *A Study of the Effects of Law Enforcement on Traffic Flow Behavior.* National Highway Traffic Safety Administration, U.S. Department of Transportation, 1971.
4. W. L. Booth. *Police Management of Traffic Accident Prevention Programs.* Charles C Thomas, Springfield, Ill., 1980.
5. O. K. Dart and W. W. Hunter. Evaluation of the Halo Effect in Speed Detection and Enforcement. In *Transportation Research Record 609,* TRB, National Research Council, Washington, D.C., 1975, pp. 31–33.
6. J. Treat, N. Tumbas, S. McDonald, D. Shinar, R. Hume, R. Mayer, R. Stansifer, and N. Castellan. *Tri-Level Study of the Causes of Traffic Accidents.* Report DOT-HS-034-3-535-77 TAC. Indiana University, Indianapolis, 1977.
7. R. D. Helliar-Symons. Yellow Bar Experimental Carriageway Markings—Accident Study. TRRL Report 1010. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1975.
8. G. G. Denton. The Influence of Visual Pattern on Perceived Speed. TRRL Report 531. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1973.
9. R. VanHouten, P. Nau, and Z. Marini. An Analysis of Public Posting in Reducing Speeding Behaviour on an Urban Highway. *Journal of Applied Behavior Analysis,* Vol. 13, 1980, pp. 383–395.
10. R. VanHouten and P. Nau. *A Study To Assess the Effects of Feedback Signs on Driving Speed.* Technical Report. Transport Canada, Ottawa, Ontario, 1984.

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

Excess Travel: Causes, Extent, and Consequences

GERHART F. KING AND TRUMAN M. MAST

The amount of excess travel in the United States is estimated on the basis of past research and new empirical studies. Excess travel is defined as the arithmetic difference between total actual highway use, exclusive of destination-free "pleasure" driving, and the use that would have resulted if all such travel had been made by using the optimum route connecting each individual origin-destination pair. Excess travel is shown to be caused by a number of different factors acting singly or in combination. These include route selection criteria and efficiencies in necessary route-planning information, in the highway information system, and in both route-planning and route-following skills. The synthesis of all available data indicates that excess travel contributes 4 percent of all vehicle miles of travel and 7 percent of all travel time for work-related trips. Corresponding figures for non-work-related trips are 20 and 40 percent, respectively. Applying these proportions to total U.S. travel results in a total of excess travel amounting to 83.5 billion mi and 914,000 person-years per year at a total estimated cost of more than $45 billion.

G. F. King, KLD Associates, Inc., 300 Broadway, Huntington Station, N.Y. 11746. T. M. Mast, Federal Highway Administration, 6300 Georgetown Pike, HSR-10, McLean, Va. 22101.

It has been estimated that transportation in the United States consumes approximately 27 percent of all energy used and 65 percent of all petroleum products *(1)*. Highway users account for 80 percent of all transportation-related use of petroleum products at an annual cost to the consumer of $40 billion. The total economic cost of highway accidents has been variably estimated to range from $40 billion to $75 billion per year *(2–5)*.

Although the relationships are somewhat complex and not necessarily linear, it has been shown that both fuel consumption and accident frequency are directly related to total vehicle mileage *(6, 7)*. Other adverse effects of personal mobility, including air and noise pollution and wear and tear on the highway system, can also be represented as functions of highway use.

Furthermore, annual total U.S. vehicle miles of travel (VMT) of nearly 1.75 trillion mi (2.8×10^{12} km) *(8)*, at an average occupancy of 1.65 [Nationwide Personal Transportation Study (NPTS) data], implies that a total of approximately 7.5 million person-years, or an average of almost 11.5 days per person, of automobile travel are accumulated each year.

Minimizing energy use, accident involvement, air and noise pollution, wear and tear on the highway system, and unproductive use of time are desirable social, as well as individual, objectives. Without extensive changes in modal split or trip making, or both, this minimization can be approached by optimizing route choice, that is, by ensuring that each motor vehicle trip made uses the optimum route in terms of both distance and time.

For any given trip, many different factors may contribute to a departure from such an optimum routing, and many different remedial measures have been tried or advocated to overcome such departures. In order to select and implement any combination of these and thus optimize highway travel, it is first necessary to develop

- An estimate, disaggregated by driver and trip attributes, of the proportion of all highway travel that is "excessive" or "wasted," and
- An estimate of the economic costs generated by such excess travel.

A recent FHWA research project designed to develop these estimates is summarized here.

EXCESS TRAVEL

The concept of excess travel rests on two assumptions:

1. Highway travel, with minor exceptions, is purposeful. Vehicle trips represent an attempt to go from one point on the highway system to another.
2. Wardrop's first law of traffic *(9)* is valid. Drivers will, if given perfect information, select the route that minimizes travel time (or costs) in traversing a network.

Excess travel is thus defined as the arithmetic difference between total actual highway use, exclusive of destination-free "pleasure" driving, and the use that would have resulted if all such travel had been made by using the optimum route connecting each individual origin-destination pair. In this definition, highway use as well as route optimization, and therefore excess travel, can be based alternatively on distance, time, or some cost function that combines these parameters.

Excess travel can be due to any of the following factors or to several acting jointly:

- Use of a route-selection criterion set that does not emphasize route optimization.
- Failure to consider an adequate number of alternative routes.
- Inadequate skills to identify optimum routes.
- Unavailability, inadequacy, or inaccuracy of the information necessary for route selection.
- Failure to follow a planned route because of deficiencies in formulating a route description or in the storage of that description.
- Failure to follow a planned route because of lack of adequate skills or of required a priori knowledge.
- Failure to follow a planned route because of deficiencies in the highway information system.
- Incorrect evaluation of real-time route-choice alternatives.
- Voluntary diversion from a planned route.
- Forced diversion from a planned route followed by selection of a suboptimum detour route.

In addition to the foregoing factors, which affect individual trips, excess travel can also occur as the result of inefficient sequencing of multilink trip chains or failure to aggregate individual trips into such trip chains.

Both the individual causes of excess travel and the excess itself have been the subject of prior studies. These studies, however, did not address all potential instances of excess travel. Furthermore, although these studies clearly indicate the existence of an excess-travel problem in terms of the proportion of times that suboptimum routes were used, existing data are inadequate to quantify the excess so generated or to stratify this excess by driver or trip attributes. Finally, a considerable proportion of the more comprehensive and more recent studies have been made abroad and the generalizability of the quantitative results obtained to U.S. conditions is somewhat questionable. To overcome these problems, a number of empirical studies were carried out. The methodology and results of these studies have been reported elsewhere *(10–12)*.

Considerable research has been done on identifying and, to a lesser extent, quantifying the distribution of route-choice criteria used by motorists. In many instances, motorists were found to apply more than one criterion to a specific route plan. The criteria selected are applied either simultaneously, in the form of an implicitly weighted linear function, or hierarchically. These studies have been summarized elsewhere *(13)*. A taxonomy of more than 50 distinct criteria is included in the summary. The choice of desired route attributes (route-selection criteria) represents largely subjective decisions by individual drivers. Optimization of these decisions probably requires long-term changes in attitudes and in individual value systems. This aspect of the problem will therefore not be considered further.

The results of some major studies oriented specifically toward quantifying the excess-travel problem are given in Table 1 *(14–26)*. Examination of this table shows that

- For nonwork, non-CBD trips, excess travel, in every study, amounted to more than 10 percent of optimum, and
- In every instance in which data on both parameters were obtained, the proportion of excess time is considerably larger than the proportion of excess distance.

No study systematically examined the causality for this excess travel. The literature, however, presents some indication of potential causes without allowing for the evaluation of the quantitative contribution of each of these. For instance, Benshoof *(27)* compared the actual route selected according to a stated criterion with the optimum route for the same criterion for relatively short trips (1.4 to 4.0 mi) to a city center. The results were as follows ($N \approx 1{,}300$):

Criterion	*Proportion Selecting Optimum Route*
Quickest	0.823
Shortest	0.598
Least stops	0.771
Least traffic	0.713

Benshoof explains this relatively poor performance by postulating that

1. Route selection, for many motorists, is a largely irrational process, and
2. Many motorists do not actually measure certain characteristics of their routes.

This second conclusion was also reached in a Swedish study of route choice *(28)*:

> The reason drivers choose different routes is not only that they ascribe different values to the road characteristics but also, to a great extent, that they simply do not accurately measure the characteristics of the routes. The capability of accurate measurement seems to decrease when this length increases.

Similar results and conclusions have been reported from other studies *(16, 29)*.

Quantitative data on failures in route following as contrasted with route planning are less common. It should be pointed out, however, that the relative contributions of route selection and route implementation for some of the studies summarized in Table 1 cannot be disaggregated. Data from a number of studies *(30–32)* show that a significant proportion of drivers reported that they lost their way or were observed at a location or traveling in a direction that could not be part of an optimum route trip plan. None of these studies, however, quantified the excess travel so generated.

Only one empirically based overall estimate of total excess travel could be located in the literature. Jeffery *(33)* synthesized the results of all studies made in Great Britain and concluded that 6.9 percent of all driver costs are due to excess travel. Using conservative estimates that route changes for repeat trips of less than 5 km (3.1 mi) are unlikely and that for a substantial proportion of non-work-related travel drivers do not seek to optimize their routes, he concluded that approximately 2.2 percent of all journey costs represent recoverable excess costs.

All studies discussed so far dealt with single trips from one origin to one destination. However, multistop, multipurpose travel is frequently undertaken. Such a trip chain, or tour, can introduce considerable excess travel if the sequencing of the individual trip segments is not optimized insofar as permitted by external constraints.

In a Canadian study *(34)* the sequencing of stops on a tour and the time used as a function of the possible minimum for tours of from one to five stops were analyzed. For tours with three or more stops the aggregate excess time consumed amounted to 7.5 percent of the total time. Optimality of individual trips within the tour was not investigated.

Excess travel is also generated by accessing activities in single trips rather than combining these into complex automobile tours. In a study in Detroit *(35)* it was estimated that 67.4 percent of all activities were accessed by complex tours and that a net saving of 7 percent of total VMT could be achieved if this proportion was raised to 83.7 percent.

TABLE 1 RESULTS OF PREVIOUS STUDIES

Reference and Year	Location	Method	Trip Purpose	No. of Subjects and Trips
14, 1975	Central London, England	Car following	Not determined	853
15, 1969	Suburban Washington, D.C.	Staged trips	Not applicable	20
16, 1970	Rome, Italy	Route mapping	Not applicable	82, 384
17, 1967	Vasteras, Sweden	Actual trips	Various	878, 1,235
18, 1981	Eindhoven, Netherlands	Actual trips	Work to home	232
19, 1966	San Francisco, Calif.	Actual trips	Work	574
20, 1966	San Francisco	Actual trips	Work	343
21, 1975–1976	England	Staged trips	Not applicable	70
22, 1974	England	Actual trips, en route interviews	Various	4,915
23, 1979	Tokyo, Japan	Simulation	All	–
24, 1972	England	Route mapping	Not applicable	337, 490
25, 1971	England	Route mapping	Not applicable	128
26, 1984	Suburban Washington, D.C.	Induced error	Not applicable	11

[a]Mean value.

FACTORS AFFECTING NAVIGATIONAL WASTE

The proportion of navigational waste is not uniform across all trips made. For any one trip the existence and magnitude of navigational waste is a function of many interrelated factors, including

- Route planner or driver attributes, or both;
- Trip purpose;
- Trip length;
- Highway systems used;
- Destination, route, and area familiarity; and
- Environmental conditions.

Trip-planning and route-following efficiency, like all cognitive activities, is a function of intelligence, skills, and experience. These items are correlated, though not perfectly, with such attributes as age, sex, education, and driving experience. Empirical evidence from past studies indicates that in most cases, the effect of demographic variables on the amount of navigational waste is small, inconsistent, and, usually, insignificant. However, the extent to which these variables have been studied systematically cannot be determined. One major exception is the correlation of ability to use a map with education, training, and especially spatial visualization. The population distribution of these characteristics is, however, not known.

The empirical portions of this research indicated a significant correlation between subject's sex and navigational efficiency. These results were, however, based on small sample sizes and furthermore were extremely inconsistent between the different levels of trip planning (i.e., the relative contributions to navigational waste of route-selection and of route-following failures). No other significant demographic effects were found.

Another class of driver attributes should be mentioned even though its effect cannot be easily quantified. The efficiency of route selection is clearly related to the amount of effort devoted to that task and to the absence of distracting influences. Similarly, the probability of error or of suboptimum decisions in route following is a function of the driver's momentary physiological and psychological state and of the presence of internal or external distracting factors. No empirical studies could be located that address this topic although the adverse effects of fatigue, alcohol, and psychological factors such as preoccupation on other aspects of the driving task have been well documented [e.g., by Shinar *(36)*].

Trip purpose is closely related to trip length and to familiarity, both of which are discussed in the following paragraphs. The major attribute of trip purpose, applying especially to the distinction between work-related trips and other trips, is the frequency with which trips are repeated. The more often a trip is repeated, the greater is the investment it represents in terms of driving costs and time, and therefore the greater is the probability that the trip-planning effort is enhanced and a greater number of alternative routes tried. However, evidence from past studies as well as questionnaire data collected as part of this project *(37)* indicate that, even for frequent trips, an inadequate number of alternative routes is tried. Furthermore, researchers have concluded that many drivers do not properly evaluate the data they obtain by trying alternative routes.

Trip length is obviously correlated with the use of different highway systems and with area familiarity. More directly, the length of a trip is an indication of the number of decision points encountered and hence of the number of error possibilities. The length of a trip is usually also directly correlated with the area covered. Increasing the size of that area increases the demands placed on cognitive mapping ability, which have been shown to be closely connected with navigational efficiency. It should, however, be pointed out that past research results that indicate that a large proportion of total excess travel occurs in the terminal phase of a trip appear to indicate that the total length of the trip may be of less importance.

The type of highway system, or highway functional classification, used for a trip exerts an effect on navigational waste through a number of separate mechanisms.

- The frequency of decision points (e.g., intersections, bifurcations, and interchanges) per unit distance is usually related to highway functional classification.
- The quality and adequacy of the highway information system, including signing, delineation, and other information

	Mean Excess (%)		Extreme Value (%)			
Distance	Distance	Time	Distance	Time	Percentile	Notes
≤7 km	+ 5.5		≅75		95	13.1 percent excess crossings
10.8, 14.3 mi	+ 47.0	+135.2	+89.3	+187.3	90	Two routes
6.25 km	+ 13.2	+ 33.0	+33.2		95	Six routes
Variable, 3.1 km[a]	+ 6.4		+30.0		95	–
4.7 km[a]	+ 7.2	+ 13.2				Least-time criterion
–	+ 3.1	+ 5.3	+10	+ 32	95	–
–	+ 5.7	+ 9.8	+21	+ 36	95	–
24.4, 107.7 km	+ 23.7	+ 29.4				Two routes
18.6 km[a]						Plus 7.2 percent generalized cost
–		+ 6.0				Improvement
–		+ 10.3				–
18 km	+ 18.9			+ 41.4	Max.	–
6.8 km	+150.0					–

sources, are usually better for the higher functional highway classification.

- The penalties associated with navigational errors—that is, excess time and distance to return to the proper route—are different, especially between conventional and limited-access facilities.
- The probability that a given highway link is properly shown on a map, especially one with a relatively large scale, is directly related to that link's classification.
- Choices between alternatives during trip planning often depend on subjectively assigned attributes of different functional classifications. These attributes may not correspond to objective reality.

Little quantitative, empirical data on these items are available except for some rather dated studies on diversion to newly opened freeways and on free versus toll road use. Some data on time-distance tradeoffs were derived from the questionnaire survey *(37)*.

Familiarity with the destination, route, and area obviously has a major effect on the probability that an optimum route will be both selected and followed. Past studies of work-trip routing, however, indicate that familiarity by itself is not a sufficient condition for minimizing navigational waste. Most empirical studies have deliberately excluded subjects familiar with the routes, destinations, or areas used in the experiments. No quantitative data on this topic are therefore available.

Different routes connecting the same O-D pair may be optimum depending on environmental conditions. If these conditions are not properly considered by the driver, if available alternative routes are not known or are avoided for other reasons, or if the environmental conditions are not anticipated, excess travel will be generated.

The preceding brief discussion has indicated the mechanisms by which individual factors may affect the extent of navigational waste. The discussion has also shown that available data are inadequate to completely disaggregate the relative contribution of each of these or assess their interactions. The following discussion of the proportion of total travel that is wasted will therefore be presented in aggregate terms except for those cases (e.g., trip purpose) in which disaggregation is possible.

PROPORTION OF TRAVEL THAT IS WASTED

The two major inputs into a determination of the proportion of all travel that is wasted are the synthesis of prior research studies, as summarized in Table 1, and the empirical data collected for this project. Insofar as past research results are concerned, major emphasis is placed on U.S. studies because the generalizability of foreign quantitative data to U.S. conditions is not known.

Work Trips

Two studies of work-trip routing *(19, 20)* have indicated excess travel of 3.1 and 5.7 percent of distance and 5.3 and 9.8 percent of time, respectively. In both cases the results were obtained by comparing self-reported actual to optimum routes. The subjects also indicated an extremely high rate of repetition of the same route. Under these conditions, it can safely be assumed that the entire excess is due to suboptimum route selection rather than to failures in route following. On the basis of these studies and European studies that indicate results of a similar order of magnitude, an estimate of 4 percent excess distance and 7 percent excess time for work trips can be supported.

Other Trips

On the basis of newly collected empirical data, trips to unfamiliar destinations average 126.5 percent of the optimum distance and 169.6 percent of the optimum time. These data thus indicate that 21 percent (26.5/126.5) of all distance traveled and 41 percent (69.6/169.6) of all time used for these trips represents navigational waste. These figures are somewhat lower than those reported from previous U.S. studies but are comparable to some more recent European results (see Table 1).

Detailed analysis of the empirical data indicates that the contributions of faulty trip planning and faulty trip plan execution to the proportion of excess travel are almost exactly equal.

Summary

On the basis of the preceding discussion, the following estimate of percent excess travel due to navigational waste can be made.

		Excess (%)	
Trip Purpose	*Causality*	*Distance*	*Time*
Work trips	Route planning	4	7
	Route following	–	–
Other trips (unfamiliar destination)	Route planning	10	10
	Route following	10	30

Further disaggregation of these results would be highly desirable but, as previously indicated, an adequate data base for that purpose is not available.

Because different proportions of excess travel apply to different trip purposes, an assumption concerning the distribution of total travel by purpose must be made. On the basis of an analysis of NPTS data *(38)* made by KLD Associates, Inc., it can be assumed that 40 percent of all VMT is work related. Three other assumptions will be made.

- Automobile travel to unfamiliar destinations, or using unfamiliar routes, amounts to 25 percent of all nonwork travel.
- Nonwork trips to familiar destinations, or using familiar routes, have the same characteristics as work trips.
- The probability of occurrence and magnitude of the consequences of an error in either trip planning or route following are independent of the criteria used for route selection. In other words, there will be a recoverable navigational waste component even in trips previously defined as incurring deliberate waste.

Given these assumptions, the proportion of all automobile travel that falls in the category of recoverable navigational waste can easily be calculated by evaluating

Proportion work travel × proportion waste in work travel + proportion other travel × (proportion unfamiliar × proportion waste in unfamiliar travel + proportion familiar × proportion waste in work travel).

Thus,

Proportion excess distance:
$0.4 \times 0.04 + 0.6 \times (0.25 \times 0.20 + 0.75 \times 0.04) = 0.064$,

Proportion excess time:
$0.4 \times 0.07 + 0.6 \times (0.25 \times 0.40 + 0.75 \times 0.07) = 0.120$.

It must be emphasized that this estimate is restricted to trip planning and route following under essentially steady-state conditions. Excess time occasioned by failures in real-time trip planning—that is, the failure to adjust trip plans to changes in traffic, highway, or environmental conditions—cannot be quantitatively estimated because of the absence of applicable empirical data. Past analyses of this problem, including some simulation studies, indicate that this component of excess travel can be substantial.

No attempt has been made to estimate the proportion of all travel that can be considered deliberate waste because little data are available that would permit evaluating the quantitative effects on travel time and travel distance of the use of "nonoptimizing" criteria in route planning.

ESTIMATED TOTAL EXCESS TRAVEL DUE TO NAVIGATIONAL WASTE IN THE UNITED STATES

Total automobile travel in the United States occurs at an annual rate of 1744.9×10^9 mi *(8)*. The latest available (1983) FHWA data *(39)* indicate that 74.8 percent of this travel is accumulated by personal passenger vehicles. Assuming that this percentage remains unchanged and applying the proportions computed in the preceding section, the total excess travel by noncommercial vehicles in the United States can be estimated as totaling 83.5 billion mi per year.

Estimating the excess time consumed by this navigational waste is somewhat more difficult because reliable data on the amount of time spent in automobile travel are not available. Using NPTS data on vehicle occupancy and representative values of average travel speed for various highway systems, the total time in automobile travel can be estimated as follows.

Work trips:
24.0×10^9 hr = 2.74×10^6 years,

Nonwork trips:
41.5×10^9 hr = 4.73×10^6 years,

All trips:
65.5×10^9 hr = 7.47×10^6 years.

Using previously derived proportions, the excess time due to recoverable navigational waste is as follows.

Work trips:
1.68×10^9 hr = 192,000 years,

Nonwork trips:
6.33×10^9 hr = 723,000 years,

All trips:
8.01×10^9 hr = 914,000 years,

or approximately 34 hr per year for every single person in the United States.

The calculations developed in this section thus indicate that 6.4 percent of all miles driven and 12.2 percent of all time spent can be conservatively assumed to represent recoverable navigational waste.

These estimates apply only to navigational waste under essentially steady-state conditions. The excess travel due to suboptimum real-time route planning—that is, adapting routes to actual traffic, highway, and climatological conditions—has not been addressed.

COSTS ASSOCIATED WITH EXCESS TRAVEL

The total costs that can be attributed to excess travel are made up of a number of component parts as follows:

- Vehicle operations,
- Accidents,
- Vehicle occupancy time,
- Maintenance and operation of the highway system, and
- Miscellaneous external costs.

Vehicle Operations

On the basis of a synthesis of published data and estimates *(6, 39, 40)* of vehicle operating costs, a figure of \$0.12/mi for variable costs was derived and is used in subsequent calculations. No change in fixed costs is assumed to occur as a consequence of excess travel.

By adjusting published data *(41)* on average vehicle fuel consumption for intervening changes in the composition of the U.S. vehicle fleet, the average energy use for the 1985 fleet can be computed as 0.0400 gal/mi. If it is now further assumed that there is no difference in fuel consumption variables between excess driving and total driving, the net energy impact of nondeliberate excess driving is 3.3 billion gal of gasoline per year.

Accidents

The total cost of motor vehicle accidents to the individuals involved as well as to society as a whole has been estimated as \$43.3 billion and \$57.2 billion, respectively, by the National Safety Council *(2)* and by NHTSA *(5)*. In a more recent study *(3)* total cost is disaggregated by accident severity. Using these figures and data on the distribution of accidents by severity *(3, 42)* yields a total societal accident cost of \$83.3 billion or \$0.0484/mi in 1985 dollars.

The accident consequences of excess driving consist of two components. The first of these is the additional exposure due to the additional vehicle miles of travel. To evaluate this component it can be assumed that the distribution of excess travel by highway type is the same as the distribution of total travel.

The second component is much more difficult to quantify. It is the possible additional accident potential due to the joint effect of route unfamiliarity and directional uncertainty. No direct data on these items are available. However, past studies of the relative odds of accident culpability as a function of route familiarity *(43, 44)* and analyses of the driving task *(7)* indicate that it appears safe and conservative to assume that the accident potential during the excess portion of driving is at least 10 percent higher and can thus be estimated to be $0.053/mi. Furthermore, on the basis of the stated assumptions, it can be estimated that excess travel is responsible for 7 percent of all traffic fatalities, or about 3,000 per year.

Time

Quantification of the value of time represents one of the most controversial and conceptually difficult aspects of highway economic analysis. Yucel has made an excellent review of these problems *(45)*.

In the major U.S. research effort on this topic, done by the Stanford Research Institute in 1967 *(46)*, the value of time was found to be highly correlated with gross hourly earnings for private nonagricultural employment. Maintaining this proportion and using the 1985 earning figure of $8.57 *(47)* yields a value of time of $8.72. This is almost identical to the value computed for work trips from project questionnaire data *(37)*.

In deriving an estimate for the value of excess time due to navigational waste, the following, mostly conservative, estimates were made based on past research *(46)* and on the analysis of the questionnaire data *(37)*.

- The unit value per hour of time is $8.50 for work trips and $6.50 for nonwork trips.
- Twenty-five percent of all travel time is accrued on trips that are so short that the excess time per trip is less than 5 min. No value is assigned to this excess.
- One-third of total vehicle occupancy for nonwork trips is contributed by children, whose time has no monetary value.

On the basis of these assumptions, the estimated cost of excess time due to recoverable navigational waste is

Trip Type	*Cost ($ billions)*
Work	10.7
Nonwork	20.6
Total	31.3

As with previous estimates, the excess time effects of deliberate waste have not been estimated. Apart from the impossibility of quantifying the net effects of such deliberate waste, the valuation of the time so consumed introduces a conceptual problem. Because this excess time is the normal result of deliberate action on the part of the driver, it may not be considered "wasted."

Air and Noise Pollution

Air pollution due to vehicular emissions is directly, although not linearly, correlated with VMT. Methodologies exist by which the increase in the amount of pollutants can be approximated. However, no reliable methodology exists by which this pollution level can be translated into incremental costs.

The quantification of noise pollution and especially its conversion into monetary terms represents an even more indeterminate subject. For these reasons, costs associated with increased pollution levels have not been considered in this evaluation.

Highway System Maintenance and Operations

The rate of deterioration of the highway system and the consequent need for maintenance, rehabilitation, or reconstruction are direct results of the physical demands placed on pavement and bridge structures. These demands are a function of axle loading and distance traveled. The relationships between axle loading and wear and tear are not linear. A moderate increase in axle loadings can lead to rapid increases in the rate of deterioration. Because commercial truck traffic was specifically excluded from the scope of this research effort, this topic will not be addressed.

The cost of highway system operations, and especially traffic control, can be directly related to traffic volumes and to their spatial and temporal distribution. Here again the relationship is not linear. A 10 percent increase in traffic volume that raises the volume-capacity ratio from 0.82 to 0.90 can have an enormous effect on the need for traffic management or even on the need for new or improved highway facilities.

However, a valid quantification of these effects would require fine-grained disaggregation of excess travel in terms of its spatial and temporal distribution and the highway systems affected. Such disaggregation is not possible with existing data.

Summary

The total annual cost of navigational waste in noncommercial travel can be estimated, on the basis of the preceding discussions and computations, as follows:

1. Vehicle operating costs, $83.5 \times 10^9 \times \$0.12 = \$10.0 \times 10^9$;
2. Accident costs, $83.5 \times 10^9 \times \$0.053 = \$4.4 \times 10^9$;
3. Cost of time, $\$31.3 \times 10^9$;
4. Total, $\$45.7 \times 10^9$.

These cost figures do not include possibly significant but unquantifiable costs due to air and noise pollution and increased highway maintenance and operations requirements. Furthermore, these costs only cover the quantifiable effects of inadvertent route-planning and route-following failure under steady-state conditions. Costs, especially those associated with excess time due to congestion, occasioned by failures in real-time route planning are not included. Also not included are all costs associated with deliberate waste, that is, excess costs accrued on trips planned with other than optimizing criteria.

CONCLUSIONS

The following conclusions are indicated on the basis of the work accomplished under this project combined with previous analyses and empirical investigations:

- Recoverable navigational waste is made up of 6.4 percent of all miles driven and 12.0 percent of all time spent in noncommercial travel.
- This excess travel accrues costs to individual drivers and to society as a whole that exceed $45 billion per year, not including costs due to increased levels of air and noise pollution or increased demands for highway maintenance and operations.
- Additional costs, unquantifiable with available data but likely to be substantial, are accrued because of failures in real-time trip planning and deliberate waste.
- Approximately half of all recoverable navigational waste can be attributed to deficiencies in trip planning. The other half can be attributed to deficiencies in route following.
- There are no significant differences in the proportion of excess travel, by trip purpose, between day and night driving.
- There are no consistent significant differences in the proportion of excess travel based on driving experience or major demographic variables except that there are some indications that male drivers may perform somewhat better than female drivers.

REFERENCES

1. *Statistical Abstract of the United States, 1985.* Bureau of the Census, U.S. Department of Commerce, 1985.
2. *Accident Facts, 1983.* National Safety Council, Chicago, Ill., 1983.
3. T. R. Miller, K. A. Reinert, and B. E. Whiting. *Alternate Approaches to Accident Cost Concepts: State of the Art.* Report FHWA/RD-83/079. FHWA, U.S. Department of Transportation, Jan. 1984.
4. *National Accident Sampling System, 1983.* Report DOT HS806 699. NHTSA, U.S. Department of Transportation, 1985.
5. *The Economic Cost to Society of Motor Vehicle Accidents.* Report DOT HS 806 342. NHTSA, U.S. Department of Transportation, Jan. 1983.
6. *Motor Vehicles Facts and Figures '85.* Motor Vehicle Manufacturers Association of the United States, Detroit, Mich., 1985.
7. P. P. Jovanis and H. L. Chang. Modeling the Relationship of Accidents to Miles Traveled. In *Transportation Research Record 1068,* TRB, National Research Council, Washington, D.C., 1986, pp. 42–51.
8. U.S. Department of Transportation, Federal Highway Administration, *Traffic Volume Trends.* FHWA, U.S. Department of Transportation, 1985.
9. J. G. Wardrop. Some Theoretical Aspects of Road Traffic Research. *Proceedings, Institution of Civil Engineers, Part 2: Research and Theory,* 1952.
10. *Economic Assessment of Potential Solutions for Improving Motorist Route Following: Interim Report, Task B—Experimental Design.* Report KLD-TR-157. KLD Associates, Inc., Huntington Station, N.Y., June 1984.
11. G. F. King. Driver Performance in Highway Navigation Tasks. In *Transportation Research Record 1093,* TRB, National Research Council, Washington, D.C., 1986, pp. 1–11.
12. G. F. King. *Economic Assessment of Potential Solutions for Improving Motorist Route Following: Final Report.* Report KLD-TR-172. KLD Associates, Inc., Huntington Station, N.Y., Aug. 1985.
13. *Economic Assessment of Potential Solutions for Improving Motorist Route Following: Interim Report, Task A—Literature Synthesis.* Report KLD-TR-146. KLD Associates, Inc., Huntington Station, N.Y., March 1984.
14. C. C. Wright. *Westminster Route Choice Survey: 2. Some Characteristics of Drivers' Route Choice in Westminster.* Transport Research Note 9. City Engineers Department, City of Westminster, London, England, Feb. 1976.
15. D. A. Gordon and H. C. Wood. How Drivers Locate Unfamiliar Addresses—An Experiment in Route Finding. *Public Roads,* Vol. 36, No. 2, June 1970.
16. F. Tagliacozzo and F. Pirzio. Assignment Models and Urban Path Selection: Results of a Survey of the Behavior of Road Users. *Transportation Research,* Vol. 7, No. 3, Sept. 1973.
17. B. Dalborg and M. Wallder. Bilisters vagval—en studie i Vasteras (Motorists' Choice of Route—A Study Conducted in Vasteras) (Swedish with English summary). Report R13:1972. Statens Institute for Byggnadsforskning, Stockholm, Sweden, 1972.
18. P. H. L. Bovy. Het korste-tijd routekeuze principze (The Minimum Time Route Choice Principle) (in Dutch). *Verkeerskunde,* Vol. 32, No. 6, 1981.
19. J. C. Orman. *Study of Driver Route Choice.* Graduate Report 1976-6. Institute of Transportation Studies, University of California, Berkeley, 1967.
20. G. R. Jansen. *A Pilot Study in Trip Assignment.* Graduate Report. University of California, Berkeley, Aug. 1966.
21. S. E. Lunn. Route Choice by Drivers. Supplementary Report 374. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1978.
22. D. J. Jeffery and N. B. Taylor. The Potential of Improved Route Guidance. In *Proceedings of the 1979 Summer Annual Meeting,* Planning and Transport Research and Computation Co., Ltd., London, England, 1979.
23. F. Kobayashi. Feasibility Study of Route Guidance System. In *Transportation Research Record 737,* TRB, National Research Council, Washington, D.C., pp. 107–112.
24. A. Morrison. Testing the Effectiveness of Road Speed Maps and Conventional Road Maps. *The Cartographic Journal,* Vol. 11, 1977, pp. 102–116.
25. D. Sheppard and J. M. Adams. A Survey of Drivers' Opinions on Maps for Route Finding. *The Cartographic Journal,* Vol. 8, No. 2, 1971, pp. 105–114.
26. T. M. Mast and W. Lareau. *Motorist Direction Finding Aids: Recovery from Freeway Exiting Errors.* Report FHWA/RD-82/098. FHWA, U.S. Department of Transportation, 1984.
27. J. A. Benshoof. Characteristics of Drivers' Route Selection Behavior. *Traffic Engineering and Control,* Vol. 11/12, 1970.
28. S. O. Gunnarson and J. Korner. *Studier av Bilisters Vagval* (Studies in Driver's Route Choice Behavior). (Swedish with English summary). National Swedish Road and Traffic Research Institute, Stockholm, Sweden, 1972.
29. E. P. Ratcliffe. A Comparison of Drivers' Route Choice Criteria and Those Used in Current Assignment Processes. *Traffic Engineering and Control,* April 1972.
30. K. D. Cross and J. J. McGrath. *A Study of Trip Planning and Map Use by American Motorists.* Final Report 170-1. Anacapa Sciences, Inc., Santa Barbara, Calif., July 1977.
31. G. F. King and H. Lunenfeld. Urban Area Highway Guide Signing. AIL Report 6566. Airborne Instruments Laboratory, Farmingdale, N.Y., May 1972.
32. D. W. Schoppert et al. Some Principles of Freeway Directional Signing Based on Motorists' Experiences. *Bulletin 244.* HRB, National Research Council, Washington, D.C., 1960.
33. D. J. Jeffery. *The Potential Benefits of Route Guidance.* Laboratory Report 997. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
34. M. E. O'Kelly and E. J. Miller. Multiple Multipurpose Travel Characteristics: An Empirical Study of Trip Length. In *Transportation Research Record 976,* TRB, National Research Council, Washington, D.C., 1984, pp. 33–39.
35. N. P. Hummon and L. D. Burns. *Automobile Tours: How We Use Our Cars.* Research Publication GMR-3771. General Motors Laboratories, Warren, Mich., June 1981.

36. D. Shinar. *Psychology on the Road: The Human Factor in Traffic Safety.* John Wiley and Sons, New York, 1978.
37. G. F. King. Driver Attitudes Concerning Aspects of Highway Navigation. In *Transportation Research Record 1093*, TRB, National Research Council, Washington, D.C., 1987, pp. 11–21.
38. R. H. Asin. *1977 Nationwide Personal Transportation Study: User's Guide for the Public Use Tapes.* FHWA, U.S. Department of Transportation, April 1980.
39. *Highway Statistics, 1983.* FHWA, U.S. Department of Transportation, 1984.
40. Hertz Explains How To Calculate Ownership and Operating Costs. *Automotive Fleet*, Vol. 23, No. 5, March 1984.
41. *Special Report 204: 55: A Decade of Experience.* TRB, National Research Council, Washington, D.C., 1984.
42. *Highway Safety Facts: Monthly Fatality Report—December.* NHTSA, U.S. Department of Transportation, Jan. 1985.
43. F. D. Bents and J. H. Hedlund. Some Characteristics of Culpable Drivers in Rural Accidents. In *Proceedings, 22nd Conference of American Association for Automotive Medicine*, 1978.
44. S. T. McDonald. Characteristics of Culpable Accident-Involved Drivers. In *Proceedings, 20th Conference of American Association for Automotive Medicine*, 1976.
45. N. C. Yucel. *A Survey of the Theories and Empirical Investigations of the Value of Travel Time Savings.* Bank Staff Working Paper 199. International Bank for Reconstruction and Development, Washington, D.C., Feb. 1975.
46. T. C. Thomas. *The Value of Time for Passenger Cars: An Experimental Study of Commuters' Values.* Stanford Research Institute, Menlo Park, Calif., May 1967.
47. U.S. Congress, Joint Economic Committee. *Economic Indicators, July 1985.* 90th Congress, 1st Session, 1985.

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

Abridgment

A Study of Route Selection from Highway Maps

GERHART F. KING AND AJAY K. RATHI

An experiment designed to assess the ability of subjects to plan long trips in unfamiliar areas by using only maps is described. The experiment was part of a larger study intended to describe and quantify the excess-travel problem in the United States. Subjects were asked to plan relatively long trips in unfamiliar areas by using only a road atlas. The sample was designed to represent the age and sex distribution of the U.S. driving population. The routes selected by the subjects were compared with the routes recommended by the American Automobile Association (postulated to be "optimum") for both distance and approximate driving time. Analyses of the data indicated that the excess distance of the routes selected by the subjects, on average, increased trip length by 12.1 percent. Age, sex, and geographic location of subjects had little effect on their performance.

Research *(1–3)* has shown that drivers face considerable difficulties in achieving optimum (i.e., in terms of minimum distance or time or both) routes from their origin to their destination. These travel inefficiencies have been shown to generate a considerable aggregate amount of excess travel.

KLD Associates, Inc., 300 Broadway, Huntington Station, N.Y. 11746.

A comprehensive literature search *(4)* indicated that excess travel may be attributed to one or more of the following trip-making activities:

1. Use of route-selection criteria that do not lead to an optimum route;
2. Trip planning (i.e., application of criteria to route selection), including inadequate trip-planning skills or unavailable, insufficient, or inaccurate information for optimum trip planning;
3. Route following (i.e., implementation of a trip plan), including all aspects of response to, reliance on, and anticipation of highway information systems; and
4. Trip chain sequencing (i.e., ordering of multiple destinations in the absence of sequential or time constraints).

As part of a major FHWA-sponsored study *(5)* of the excess-travel problem and of potential remedial measures, a series of empirical studies of trip planning and route following were implemented. The procedures used and the results obtained for an experiment on trip planning for long trips in unfamiliar areas are described. The purpose of this experiment was to assess the ability of drivers to plan trips in unfamiliar areas by using only maps and to estimate the amount of excess travel generated by trip-planning inefficiencies.

The experiment was implemented in two separate geographical locations that had previously been selected as the site for other investigations *(5)*:

- Western Fairfield County, Connecticut; and
- Western suburbs of Milwaukee, Wisconsin.

The subjects were recruited primarily through newspaper advertisements and were paid for their participation in the experiments. They represented the age and sex distribution of the U.S. driving population weighted for actual miles driven.

TEST PROCEDURES

Subjects were given the following instructions:

> You are to plan a route which minimizes both travel time and travel distance. You may use the atlas in front of you to determine your route. Record on the enclosed trip plan sheet the route number, direction of travel (e.g. North, South, East, West), the place where you get on the indicated route (from), the place where you get off the indicated route (destination), estimated travel time, estimated miles and estimated cost for car. Do not consider local travel within each of your origin cities to the route which you have outlined.

The atlas used was Rand McNally's *Road Atlas: United States—Canada—Mexico*. Subjects were under no time pressure.

Test routes were selected in accordance with the following criteria:

- Relatively long distance for each intermediate destination point or segment,
- Alternative routes possible,
- No direct Interstate route connection from origin to destination.

Route 1, which was given to all subjects, was San Antonio, Texas, to Shreveport, Louisiana, to Natchez, Mississippi, to Mobile, Alabama.

Route 2, which was given to all subjects in Wisconsin, was Norwalk, Connecticut, to Hershey, Pennsylvania, to Gettysburg, Pennsylvania, to Washington, D.C., to Norwalk, Connecticut.

Route 3, which was given to all subjects in Connecticut, was Milwaukee, Wisconsin, to Des Moines, Iowa, to St. Louis, Missouri, to Little Rock, Arkansas.

DATA REDUCTION

Data reduction consisted of determining, for each segment, the number of different routes used and the number of subjects selecting each route. Total distance, disaggregated by highway type, was determined for each route. These distances were converted into anticipated driving time by assigning representative speeds (using average U.S. values) to each link according to its highway type.

It is obvious that these average speeds do not represent, except for coincidence, the actual prevailing average route speeds on the specific highway links included in the various trip plans. However, these speeds can be considered those that may be anticipated, on the basis of highway classification only, by subjects unfamiliar with the routes. These speeds therefore approximate the inputs used by the subjects in making route-selection and distance-time trade-offs.

SUMMARY OF RESULTS

Preliminary analyses of the data indicated that there were no significant differences between Connecticut and Wisconsin subjects. All data were therefore merged and analyzed on an aggregate basis.

For 5 of the 10 segments, the minimum-distance route was also the minimum-time route. For one segment, the minimum-time route was only 1 mi longer than the minimum-distance route. For three other segments, the difference in anticipated travel time between the minimum-distance and minimum-time routes was 2 min or less.

The routes selected were compared with a "best route" as planned by the Connecticut Motor Club using normal American Automobile Association (AAA) route-planning procedures.

For four segments the AAA recommended the minimum-distance and minimum-time route; in two others the AAA selected a longer but faster route. It must be remembered that these comparisons are very sensitive to the assumed average travel speeds and that AAA probably has more accurate local travel time data.

All results obtained are summarized in Table 1. With the exception of one very short segment, the excess length of the planned route over the minimum route ranged from 6.8 to 19.7 percent for the entire sample and averaged 12.1 percent. The excess distance over AAA-recommended routes averaged 4.9 percent. The detailed discussion on selection of routes for individual segments has been presented elsewhere *(5)*. A paired *t*-test showed no significant difference due to subject sex.

Table 2 shows the distribution of routes selected. A number of different parameters indicating fractions of the sample selecting certain routes and the frequency with which routes were selected are shown. Examination of these data shows the following:

- For only 3 of the 10 segments did 50 percent or more of the sample select the minimum-distance route.
- For half of all the segments, 50 percent or more of all subjects selected a route that was more than 5 percent longer than the optimum.
- For 5 of the 10 segments, the modal route was also the minimum-distance route. For three of the other five segments, the modal route was a longer but all-Interstate route.
- Circuity of the route (i.e., the ratio of airline distance to minimum-route distance) did not correlate with any of the route-selection parameters shown.

For the 10 segments the excess-time over minimum-time routes varied from 0.6 to 17.0 percent and averaged 9.7 percent.

TABLE 1 SUMMARY OF RESULTS

		SEGMENTS										
		1	2	3	4	5	6	7	8	9	10	TOTAL
ALL	(a) TIME (HR., MIN.)											
MALE	NUMBER	61	59	57	36	36	37	36	19	20	19	
	AVERAGE	7 25	3 55	4 33	4 2	0 53	1 35	4 47	6 58	7 12	6 16	
	STD.DEV	0 26	0 27	0 12	0 29	0 3	0 23	0 45	0 36	0 17	0 1	
	% OVER MIN	10.4	13.9	6.2	16.8	1.8	17.3	7.2	10.2	7.1	0.1	9.4
	% OVER AAA	6.7	-3.4	-3.2	18.5	1.8	17.3	7.2	2.9	8.7	0.1	4.2
FEMALE	NUMBER	27	25	25	14	18	15	13	9	9	7	
	AVERAGE	7 34	3 55	4 33	4 4	0 53	1 28	5 0	7 1	7 17	6 24	
	STD.DEV	1 3	0 32	0 15	0 26	0 1	0 19	1 4	0 33	0 20	0 17	
	% OVER MIN	12.7	13.9	6.3	17.6	1.1	8.7	11.8	11.2	8.4	2.1	10.5
	% OVER AAA	9.0	-3.5	-3.1	19.4	1.1	8.7	11.8	3.8	10.1	2.1	5.2
TOTAL	NUMBER	88	84	82	50	54	52	49	28	29	26	
	AVERAGE	7 28	3 55	4 33	4 2	0 53	1 33	4 51	6 59	7 13	6 18	
	STD.DEV	0 41	0 28	0 13	0 28	0 2	0 22	0 50	0 35	0 18	0 9	
	% OVER MIN	11.1	13.9	6.2	17.0	1.5	14.8	8.5	10.5	7.5	0.6	9.7
	% OVER AAA	7.4	-3.5	-3.2	18.8	1.5	14.8	8.5	3.2	9.1	0.6	4.5
ALL	(b) DISTANCE (MILES)											
MALE	NUMBER	61	59	57	36	36	37	36	19	20	19	
	AVERAGE	421.2	215.8	245.2	230.3	49.4	87.6	271.1	404.1	408.8	365.4	
	STD.DEV	34.2	33.0	18.9	25.5	2.5	20.4	36.6	40.9	32.0	10.9	
	% OVER MIN	12.3	19.9	8.5	14.0	0.9	15.2	6.3	15.8	12.6	7.2	12.0
	% OVER AAA	12.0	-6.2	-8.2	16.9	0.9	13.7	6.3	7.7	16.8	-1.2	4.8
FEMALE	NUMBER	27	25	25	14	18	15	13	9	9	7	
	AVERAGE	423.2	214.8	244.6	230.0	49.0	80.3	283.4	408.7	399.8	360.9	
	STD.DEV	64.9	36.5	20.2	26.2	0.	13.2	63.9	36.9	32.7	19.9	
	% OVER MIN	14.2	19.3	8.2	13.9	0.	5.7	11.1	17.1	10.1	5.8	12.4
	% OVER AAA	13.9	-6.6	-8.4	16.8	0.	4.3	11.1	9.0	14.2	-2.5	5.1
TOTAL	NUMBER	88	84	82	50	54	52	49	28	29	26	
	AVERAGE	423.3	215.5	245.0	230.2	49.3	85.5	274.4	405.5	406.0	364.2	
	STD.DEV	45.5	33.9	19.2	25.4	2.0	18.9	45.0	39.0	31.9	13.6	
	% OVER MIN	12.9	19.7	8.4	14.0	0.6	12.5	7.6	16.2	11.8	6.8	12.1
	% OVER AAA	12.6	-6.3	-8.2	16.9	0.6	11.0	7.6	8.1	16.0	-1.6	4.9

TABLE 2 DISTRIBUTION BY ROUTES

		Percent Selecting					No. of Different Routes		
Segment	n	AAA Route	Minimum Route	Minimum Route +5 Percent	Modal Route	No. of Different Routes	For 50 Percent of Sample	For 85 Percent of Sample	Modal Route as Percent of Minimum
1	88	1.1	21.6	36.4	37.5	14	2	6	116.8
2	84	25.0	36.9	36.9	36.9	10	2	3	100.0
3	82	12.2	1.2	62.2	61.0	10	1	3	118.1
4	50	0.0	0.0	6.0	16.0	26	5	19	107.1/113.2
5	54	50.0	50.0	98.1	50.0	3	1	2	100.0
6	52	69.2	69.2	82.7	69.2	6	1	3	100.0
7	49	65.3	65.3	75.5	65.3	14	1	7	100.0
8	28	14.3	25.0	50.0	25.0	11	3	7	100.0
9	29	0.0	0.0	6.9	34.5	12	2	8	125.7
10	26	73.1	19.2	23.1	73.1	4	1	2	108.8

CONCLUSIONS

The data collected demonstrate that considerable excess travel occurs as a result of a driver's inability to plan trips by using maps as the primary source of information.

The data indicate that the contribution of trip-planning deficiencies to excess travel amounts to approximately 10 percent of vehicle miles of travel for the types of trips investigated. This figure is comparable with that obtained in empirical investigations of driver navigation performance for different trip types as part of this research *(6)* and during previous research efforts *(4)*. It is also comparable to, although somewhat lower than, the results obtained in previous research studies that used route mapping as the principal methodology *(7–9)*.

The significant effects of deficiencies in map-based trip planning indicate that concerted efforts to raise population skill levels in this area and to improve the clarity and legibility of maps used for that purpose may well be cost-effective.

ACKNOWLEDGMENTS

This research was conducted under a contract entitled "Economic Assessment of Potential Solutions for Improving Motorist Route Following" with FHWA, U.S. Department of Transportation. Truman Mast of FHWA served as the Contracting Officer's Technical Representative.

The contributions of Dunlap and Associates East, Inc., of Norwalk, Connecticut, subcontractor for experimental design and data collection, and of Robert Dewar of Calgary, Alberta, Canada, and Slade Hulbert, of San Ramon, California, who served as project consultants, are gratefully acknowledged.

REFERENCES

1. O. C. Orman. *Study of Driver Route Choice.* Graduate Report 1976-6. Institute of Transportation Studies, University of California, Berkeley, 1967.
2. D. A. Gordon and H. C. Wood. How Drivers Locate Unfamiliar Addresses—An Experiment in Route Following. *Public Roads,* Vol. 36, No. 2, June 1970.
3. S. E. Lunn. *Route Choice by Drivers.* Supplementary Report 374. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1978.
4. *Economic Assessment of Potential Solutions for Improving Motorist Route Following: Interim Report, Task A—Literature Synthesis.* Report KLD-TR-146. KLD Associates, Inc., Huntington Station, N.Y., March 1984.
5. *Economic Assessment of Potential Solutions for Improving Motorist Route Following: Final Report.* Report KLD-TR-172B. KLD Associates, Inc., Huntington Station, N.Y., April 1986.
6. G. F. King. Driver Performance in Highway Navigation Tasks. In *Transportation Research Record 1093,* TRB, National Research Council, Washington, D.C., 1987, pp. 1–11.
7. A. Morrison. Testing the Effectiveness of Road Speed Maps and Conventional Road Maps. *The Cartographic Journal,* Vol. 11, 1977, pp. 102–116.
8. D. Sheppard and J. M. Adams. A Survey of Drivers' Opinions on Maps for Route Finding. *The Cartographic Journal,* Vol. 8, No. 2, 1971, pp. 105–114.
9. F. Tagliacozzo and F. Pirzio. Assignment Models and Urban Path Selection: Results of a Survey of the Behavior of Road Users. *Transportation Research,* Vol. 7, No. 3, Sept. 1973.

The results and views expressed are those of the authors and do not necessarily reflect the views of FHWA.

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

CRASH Revisited: Additions to Its Clarity, Generality, and Utility

ALBERT G. FONDA

A complete rederivation of the computer program CRASH is presented, with confirmation of its theoretical basis, elimination of many of its actual or supposed restrictions, and additions to its useful outputs. The physics and algebra, although clarified, are for the most part unchanged. However, in the trajectory solution, a closed-loop iteration replaces a best-fit form of solution. In the impulse solution, the physical basis of the common velocity check is clarified, the check is revised so that more cases can be treated, and the coefficient of restitution is found. In the damage solution, delta-V accuracy is improved by better fits to crash test data, corrected treatment of oblique impact, and inclusion of the energies of restitution and intervehicular sliding. Yaw rates and impact forces are found from the impulse solution and again from the damage solution; these and other paired output comparisons indicate the quality of the reconstruction and facilitate its refinement.

In this paper the program Calspan Reconstruction of Accident Speeds on the Highway (CRASH) is reviewed, rederived, and extended, with further commentary on published criticisms of CRASH and on alternative assumptions in published reconstruction treatments.

CRASH was developed under funding from the U.S. Department of Transportation (DOT) and was based by McHenry on the spin analysis of Marquard *(1)* and the deformation energy analysis of Campbell *(2)*. Initially it was a simulation setup routine for SMAC *(3)*, but later was an independent, user-friendly digital computer program for accident reconstruction *(4, 5)* specifically for evaluation of speeds at impact and speed changes during impact from the postimpact information for two vehicles colliding on a flat surface.

DOT accident reconstruction has emphasized the systematic, standardized evaluation of the speed changes of each vehicle in impact as a measure of occupant injury exposure, with the objective of evaluating the effectiveness of various safety measures for which DOT is responsible *(6)*. For these purposes, CRASH has become a highly respected functional standard.

However, the published derivations were unduly lengthy, restrictive, and obscure. Some reservations have been expressed as to their validity *(7)*, and the necessary programmed solution has been readable only as more than 130 pages of FORTRAN. As a result, many experts have avoided using CRASH as a basis for testimony in lawsuits, and litigants have been deprived of a useful analytic tool.

Precisely such reservations were the impetus for this paper. It will be shown that the CRASH equations are fundamentally correct, as far as they go, within (and sometimes beyond) their stated restrictions. Certain assumptions will be avoided; others will be clarified. All the CRASH equations will be derived, with some revisions; then the derivations will be extended to obtain results beyond those offered by CRASH. Thus in this paper CRASH will be confirmed and its clarity, generality, and utility will be enlarged.

As shown in a companion paper *(8)*, the results have been programmed in BASIC and for use in desktop and hand-held computers, making the improvements widely available and facilitating forensic use.

DATA INPUTS

CRASH uses data gathered at the site of the accident and from both of the vehicles involved in each impact. In some cases, the tire marks or the vehicles, or both, no longer exist and the immediate sources must be photographs and police reports. General vehicle data and tire-road friction data may be obtained from published tables rather than from the vehicles, tires, and highway involved.

Vehicle speed changes (as desired by DOT) can only be determined from knowledge of the damage to the vehicles, but evaluation of the speeds before impact requires knowledge of their travel after impact. Axial impacts are indeterminate without damage data, but intersection impacts, despite the possible complication of vehicle spin after impact, can be fully reconstructed without knowing the severity of vehicle damage. This analysis is considered first.

SPINOUT ANALYSIS

For uniformly decelerated motion, CRASH utilizes the usual expression for initial speed, from the double integral with respect to time of $F = ma$,

$$U = \sqrt{2\theta\mu g s}] \qquad (1.1)$$

where] in this paper closes any expression opened by √ or Σ, and the rate of deceleration is expressed as $\theta\mu g$, where θ is either the fraction of the available friction-limited lateral force applied as braking force as the vehicle slides endwise, or unity for the laterally sliding vehicle. Fonda *(8)* furnishes auxiliary equations (from both CRASH and SMAC) for traversing differing surfaces, for speed-dependent friction, and for non-separation of the vehicles.

For the considerably more complex case of the spinning vehicle, the deceleration rate is nonuniform but may be treated by the approach of Marquard *(1)*, which considers alternating periods of predominantly angular and predominantly translational deceleration, with partial or full braking. This method was augmented by McHenry and incorporated in the START, CRASH, CRASH2, and CRASH3 programs *(3–5)*.

Consider a vehicle initially translating at rate U_s and spinning at rate $\dot{\psi}_s$, losing all of its angular momentum in time T

Fonda Engineering Associates, 558 Susan Drive, King of Prussia, Pa. 19406.

while losing only part of its linear momentum. The speed U_f before final translation for a distance s_f after the end of rotation can be found from Equation 1.1. The angle and the distance traveled while spinning are

$$\psi_s = \dot{\psi}_s T/\alpha_1 \quad \text{and} \quad s_s = (U_s + U_f)T/\alpha_5 \qquad (1.2,1.3)$$

where α_1 and α_5 are empirical coefficients that would each have the value 2 if the decelerations either were uniform or fluctuated symmetrically about fixed average values.

The actual declerations alternate between periods of predominantly angular deceleration while endwise for a total time t_1, with a total angular impulse

$$mk^2 \; \dot{\psi}_s = \alpha_2 \, (0.5l\mu mg)t_1 \qquad (1.4)$$

where $0.5l$ ($\cong\sqrt{ab}$]) is the numerical (or geometric) average of the lever arms of the tangential forces and α_2 is nominally unity, and periods of predominantly translational deceleration while broadside for the remaining time t_2 plus some deceleration during t_1 due to braking, with a total linear impulse

$$m(U_s - U_f) = (\alpha_3 \theta t_1 + \alpha_4 t_2)\mu mg \qquad (1.5)$$

where α_1 and α_4 account at least for the angularity between the instantaneous force and the average direction of motion.

Accordingly, the time duration of the spin motion is given by

$$T = t_1 + t_2 = t_1 \, (1 - \alpha_3\theta/\alpha_4) + (\alpha_3\theta t_1 + \alpha_4 t_2)/\alpha_4$$
$$= [k^2 \, (\alpha_1\psi_s/T)/0.5\alpha_2 l\mu g][1 - \alpha_3\theta/\alpha_4][(\alpha_5 s_s/T) - 2U_f]/\alpha_4\mu g$$

whence $T^2 + 2BT - C = 0$ for a quadratic solution for T (Equation 1.6c).

From Equation 1.3, the translatory and angular speeds at separation are

$$U_s = (\alpha_5 s_s/T) - U_f; \quad \dot{\psi}_s = \alpha_1 \, \psi_s/T \qquad (1.6a,b)$$

where

$$T = \sqrt{C + B^2]} - B; \quad B = U_f/\alpha_4 \, \mu g \qquad (1.6c,d)$$
$$C = \{[2k^2\alpha_1\psi_s \, (1 - \alpha_1\theta/\alpha_4)/\alpha_2 l] + \alpha_5 s_s/\alpha_4\}/\mu g \qquad (1.6e)$$

These results can be verified against the spinout equations of CRASH (5, Section 9.2.a), noting the following:

1. The CRASH treatment first derives the equations for spin without braking per Marquard, second rederives with the partial braking effect per Marquard, third rederives with McHenry's contribution of the residual velocity and generalized empirical coefficients, and fourth recapitulates while specifying polynomial coefficients and certain computation routines. The present work instead proceeds directly to the final solution.

2. The present solution embodies the solution of a quadratic in T that is the separately stated CRASH quadratic in $\dot{\psi}_s$ transformed by use of Equation 1.2 after a publishing error in the CRASH equations has been corrected by reversing the sign of its unity coefficient. The quadratic coefficients become simpler because a lengthy expression in the denominators occurs now in only one term of the numerator.

3. Once the quadratic has been solved, U_f is found from the sum $U_s + U_f$ of Equation 1.3 rather than from the more complex difference $U_s - U_f$ of Equation 1.5.

Equations 1.6 provide a complete solution for the linear and angular speeds at the start of the spinout, given appropriate expressions for the empirical coefficients. The CRASH expressions were based on an analysis of a set of 18 SMAC runs [see User's Guide *(5)*, Section 10.4.d], not known to have been published, so neither the data nor the fitting technique is known. [Fonda *(8)* shows a revaluation technique; for brevity those two equations (1.7, 1.8) are omitted here.] CRASH uses the polynomials

$$\alpha_1 = 2.6 - 7.5\rho^* + 15\rho^{*2} \qquad (1.9a)$$

$$\alpha_2 = 0.9 - 4\rho^* + 8\rho^{*2} \qquad (1.9b)$$

$$\alpha_3 = 0.23 + 5\rho^* - 10\rho^{*2} \qquad (1.9c)$$

$$\alpha_4 = 0.67 + 1.6\rho^* - 5\rho^{*2} + 6\rho^{*3} \qquad (1.9d)$$

$$\alpha_5 = 1.2 + 17\rho^* - 99\rho^{*2} + 181\rho^{*3} \qquad (1.9e)$$

as functions of the initial radius to the instant center of rotation, $\rho = U_s/\dot{\psi}_s$, divided here for convenience by 1,000; ρ^* is the radius in kinches (thousands of inches).

The CRASH routine evaluates the five polynomials from a trial value of ρ flanked by trial variations of plus and minus 15 and 30 percent, the best value then being selected by a test for a minimum value of the radius error $|\rho\dot{\psi}_s/U_s| - 1$. To the same effect, setting that error algebraically to zero and iterating may be done for simplification. From Equations 1.2 and 1.3, the radius to the instant center is

$$\rho^* = 0.001 \, (\alpha_5 s_s - TU_f)/\alpha_1\psi_s \qquad (1.10)$$

in thousands of inches, where α_1, α_5, and T are initialized at 2.0 and iterated. Convergence is strong; in the programmed version, with reasonable error criteria, the residual errors of iteration and the errors of the optional rounding in Equation 1.9 (which shortens the hand-held program) are insignificant, as are those of CRASH where the residual error in ρ varies randomly from zero to 7.5 percent.

The empirical curve fit of the CRASH polynomials extends to ρ = 250 in. This represents separation at instantaneous velocities, which, if continued in the same proportion, would give 180 degrees of rotation of the vehicle in $250\pi/12 = 65.4$ ft. This is a particular value of the distance between the readily identifiable cusps seen in typical spinout tire tracks, the successive points at which the vehicle is broadside to its new course and one axle has much more velocity than the other. CRASH locks the five polynomials at the respective values that they achieve (1.66, 0.40, 0.85, 0.85, 2.08) at ρ = 250; to the same effect ρ^* can simply be limited to 0.25.

A lightly braked vehicle with a small initial yaw rate will stop spinning and begin to roll endwise when it first rotates into alignment with its course, so large values of ρ will not persist. However, a heavily or fully braked vehicle can maintain a small yaw rate, giving slow spin to rest. So for large θ and zero s_f there is some interest in the case of large ρ. At and above

ρ = 250, with s_f = 0, $\alpha_3 = \alpha_4 = 0.85$, and either $\psi_s = 0$ or θ = 1, from Equations 1.6 the whole solution reduces to

$$U_s = \alpha_5 s_s/\sqrt{C}] = \sqrt{\alpha_4\alpha_5\mu g s_s}] = \sqrt{0.85(2.08)\mu g s_s}]$$
$$= 0.94\sqrt{2\mu g s_s}]$$

This is 6 percent less than the correct result for this case. CRASH handles this anomaly by means of a discontinuity, switching to the skid solution, $\sqrt{2\mu g s}]$, for $\rho \geq 500$ in. or $\psi_s \leq$ 20 degrees. To much the same effect, but avoiding the discontinuity, more simply the ρ^* limit is raised to 0.30 (for these polynomial coefficients), allowing U_s to reach $\sqrt{2\mu g s}]$ at ρ = 300 in.

ANGULAR IMPACT ANALYSIS

So far as possible, the aircraft notation introduced to the automotive industry by Calspan personnel in 1956 *(9)* is used, as formalized in the vehicle dynamics terminology of SAE J670 (1967ff). As in the spin analysis, the heading of a vehicle is designated as ψ (psi) and its sideslip (attitude) angle as β (beta); as shown in Figure 1a, their sum, the course angle (direction of motion) is ν (nu). With various subscripts, U will denote the course speed of either vehicle before or after impact and differences thereof, and ξ (xi) will denote various angles in the horizontal plane. The magnitude and direction of a vector U may be found from its orthogonal components as shown in Figure 1b.

The slide or spin analyses provide two speeds at separation, each in a known direction, hence two mass center velocity vectors at separation. For angular impact only, if further the vehicle weights and directions of approach are known, and horizontal tire forces for the period of the impact (even those due to underride or override) are neglected, their impact speeds can be found by impulse analysis, that is, by using the principle of conservation of momentum.

Conservation of momentum requires only that the intervehicular force act equally and oppositely on both masses, so that one mass gains as much momentum (the time integral of the force) as the other loses. Because this does not assume any commonality of velocity between the vehicles, contrary to statements in the CRASH literature, sideswipes can be treated correctly by impulse analysis, as can underride or override impacts, provided horizontal tire forces remain trivial.

With normalization to the mass of Vehicle 1 ($R = M_2/M_1$), Figure 1c shows the equivalence of the momentum at separation (from O to A via S) to the momentum at impact (from O to A via I). The direction of the mutual speed change (I to S) is necessarily the direction of the "principal" intervehicle force (the DOPF or, interchangeably, PDOF). The inset shows the vectors for a case of sideswipe; the total velocities are continuingly opposed, but their differential normal to their sides is reversed because of structural rebound.

From the linear momentum along the normal to the initial course of the respectively opposite vehicle, the vehicle speeds at impact are

$$U_{o1} = [U_{s1} \sin(\nu_{s1} - \nu_{o2}) + RU_{s2} \sin(\nu_{s2} - \nu_{o2})]/\sin(\nu_{o1} - \nu_{o2}) \quad (2.1a)$$

$$U_{o2} = [U_{s1} \sin(\nu_{s1} - \nu_{o1}) + RU_{s2} \sin(\nu_{s2} - \nu_{o1})]/ R \sin(\nu_{o2} - \nu_{o1}) \quad (2.1b)$$

The inputs are the two separation speeds from Equation 1.1 or (with spin) Equation 1.6, the four vector directions, and the mass ratio R. Each equation expresses a rear view of the impact vectors as seen along one of the approach paths.

The same vectors should be found (given the same data) by any method of reconstruction; in CRASH (as published only in 1974 in the START routine of SMAC) and in the various publications by or based on Brach, such as CARR1 *(10,* Equations 72 and 73), they are written as the two unknowns in two equations, algebraically soluble for the individual speeds. Such indirect solutions are correct, but confusing, hindering insight in use. The matrix methods of Brach are, as he notes *(11,* p. 33), neither necessary nor preferable when sufficient data are available. But as shown by the original CRASH treatment by McHenry, and contrary to Brach *(11,* p. 33), zero rebound is no bar to the solution; up to this point impact has not even been assumed to be the source of the intervehicle force, much less impact with rebound.

The approach directions must differ appreciably, preferably by more than 20 degrees, lest the vector components viewable from the rear become too small; at lesser angles the solution becomes a ratio of small quantities, unduly sensitive to usual

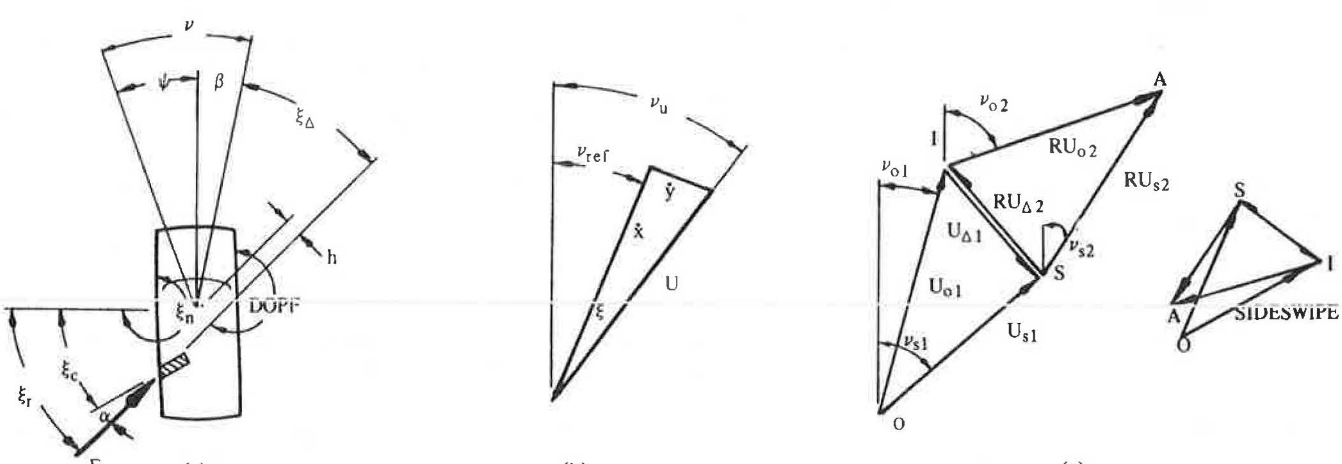


FIGURE 1 Vector relationships: *(a)* vector directions, *(b)* vector resolution, *(c)* impact vectors.

errors, and becomes indeterminate for axial impact. [Although Wooley et al. *(7)* appear to dispute this, Wooley *(12)* presents only the damage-based solution for the axial case, as in CRASH.

Once both approach velocities have been found, the velocity difference at impact, the closing velocity, may be found as a polar vector from its components,

$$U_{\Delta o},\ \xi_{\Delta o} = Pol[U_{o2}\cos(\nu_{o1} - \nu_{o2}) - U_{o1},\ U_{o2}\sin(\nu_{o2} - \nu_{o1})] \tag{2.2}$$

referred angularly to the original course heading of Vehicle 1. This is the apparent velocity of Vehicle 2 as seen from Vehicle 1 before impact. Similarly, the vector change of velocity of each vehicle during impact, from its components, is

$$U_{\Delta i},\ \xi_{\Delta i} = Pol\ [U_{si}\cos(\nu_{si} - \nu_{oi}) - U_{oi},\ U_{si}\sin(\nu_{si} - \nu_{oi})] \tag{2.3}$$

($i = 1, 2$). This gives the magnitude of the speed change of each vehicle and its direction relative to the original course heading of that vehicle. As demonstrated in Figure 1c, inherently these vectors will be 180 degrees apart in space.

The angle $\beta_i + \xi_{\Delta i}$ (with β_i usually zero) is the body-axis direction of the force of impact, hence the direction in which the vehicle is moved toward the unrestrained occupant or any other free object. It is also the direction in which the vehicle structure is deformed if isotropic ($\xi_c = \xi_r$), although allowance is made for reduced compliance in shear ($\xi_c < \xi_r$). The computed angle should be checked against the aforementioned physical evidence, often recited as a PDOF, a clock direction, which should equal $(\beta_i + \xi_{\Delta i} + 180)/30$.

The direction of the speed change and (hence) the intervehicle force relative to the normal to the surface along which sliding occurs is

$$\begin{aligned}\xi_{ri} &= \xi_{ni} - (\beta_i + \xi_{\Delta i} + 180)\\ &= \xi_{ni} - 30\ (\text{DOPF})\end{aligned} \tag{2.4}$$

which as shown by Figure 1a is positive when the shear force exerts a clockwise moment on the vehicle. By Coulomb's law this angle must not, for either vehicle, exceed in magnitude a reasonable intervehicle friction angle, on the order of arctan $0.55 \cong 30$ degrees. This limit becomes a probable value for the larger of the two angles if there is visible evidence of "scraping" (a convenient term to distinguish intervehicular sliding from tire-to-road sliding). Absent scraping, any angle between the friction limits is reasonable.

Pocketing, snagging (which is extreme pocketing, possibly with shear failure), and corner impact can change the surface orientation as the deformation proceeds. Therefore the inputted value of ξ_{ni} may not be the orientation of the original surface but that of the developed surface; this requires careful vehicle examination and visualization of the impact process. At a corner there is initially a 90 degree range of possible normals until the corner flattens to a new surface with a new normal.

Unreasonably large angles between the computed DOPF and the developed normal require reconsideration of the input data. This is not emphasized in CRASH, which is weighted toward damage-only evaluation and will not override the user choice of PDOF (except for an adjustment of not more than 7.5 degrees to obtain colinearity of the two PDOFs). Contrary to SMAC, the frictional limit on shear force has been consistently overlooked in CRASH from CRASH1 to the present. There has been no admonition against excessive angles, the user estimate of the DOPF has been checked against neither the trajectory/impulse-based DOPF nor Coulomb's law before its use in the damage calculations, and gross violations of Coulomb's law have been specifically permitted (*5*, Section 9.1.f, arctan $\mu = 75$ degrees, $\mu = 3.73$) and algebraically "corrected for" in damage evaluation (*5*, Section 9.1.f). Such angles can never occur in practice; whatever the method of reconstruction, no input data should be accepted that imply that they have occurred. Similarly there is a danger of misapplication of the matrix methods of Brach *(10, 13)*, apparently also used in CARR1 *(11)*, in which the relative shear force μ may be naively assigned its positive limit value even when it should be negative or small.

It is preferable to adopt the PDOF computed from the velocity vectors in the case of angular impact, if credible, or to adjust the input data to obtain a credible value. If this fails, and always in axial impact, the PDOF indicated by the physical evidence, subject to the limitations imposed by Coulomb's law, is used in the further calculations.

It is implied in the CRASH treatment (*5*, Section 10.5) that the condition of common velocity is fundamental to the principle of conservation of momentum; it is not. A bullet passing a mutually magnetic target interacts without impact and never a common velocity; yet equal impulses occur, so momentum is exchanged. It might better have been stated that when bodies interact by means of a structural collision, there will be both impulse and impact, with an instant of common velocity; in that instance, the subsequent rebound velocity evaluated from the trajectory data should not be excessive. This is not a simplifying assumption, it is a physical observation, and it is not imposed on the dynamics involved in impulse but deduced from the structural mechanics involved in impact.

It is the rebound velocity that is found in and limited by the common velocity check—which might better have been called the rebound velocity check. Unfortunately, the limit adopted is overly severe. The common velocity check of CRASH3 will abort the trajectory/momentum solution if the speeds of rebound from the mutual mass center differ by more than about 4 mph. In a moderately severe impact, with between 1 and 2 ft of crush, this condition is usually not met. For this reason, many reasonable solutions are aborted, frustrating the intent of CRASH and discouraging the user. If the limit had been set at about 12 to 15 mph, the common velocity check might have served its intended purpose of excluding unreasonable inputs. As it is, CRASH3 gives no momentum solution in many instances of reasonable data inputs. Either its elimination (at line 860 of SPIN2) or revision of its limit value (at Line 350 of VELCHK) is recommended. More useful calculations to the same end will be suggested.

DAMAGE ASSESSMENT

It is often necessary to infer the speed changes from the damage done to the vehicles. This can be done from measurements of the location and depths of the vertically averaged residual deformation of both vehicles combined with empirically assigned structural parameters of the vehicles.

If, following Campbell (2), it is assumed that the test speed involved in perpendicular barrier crash tests varies linearly with the resulting final crush depth while the force during impact increases as the crush depth, each from a threshold value, when the kinetic energy of approach is equated to the potential energy of crush (conservation of energy), it is found that

$$\begin{aligned} E_k &= 0.5M(V_o + CdV/dC)^2 = E_c \\ &= w_t \int (A + BC)dC + \text{constant of integration} \\ &= M[0.5V_o^2 + V_o\, CdV/dC + 0.5C^2\, (dV/dC)^2] \\ &= w_t(G + AC + 0.5BC^2) \end{aligned} \quad (3.1a)$$

whence

$$G = 0.5V_o^2\, M/w_t \text{ (inferred structural damping energy per unit width)} \quad (3.1b)$$

$$A = V_o\,(dV/dC)M/w_t \text{ (inferred threshold force per unit width)} \quad (3.1c)$$

$$B = (dV/dC)^2 M/w_t \text{ (inferred structural stiffness per unit width)} \quad (3.1d)$$

for a vehicle of mass M and involved width w_t. This establishes a data-fitting technique whereby, for given M/w_t, the test velocity intercept V_o solely determines G, the test velocity slope dV/dC solely determines B, and the threshold force $A = \sqrt{2BG}$] is a jointly dependent parameter completing the square of the binomial (that is, sized for linearity of V with crush depth).

These relationships, adapted from Campbell (2) but not shown in the User's Manual, are the source of the CRASH3 table of structural data (5) and the earlier CRASH data (14). Campbell's 1974 data (2) did show linearity of crush with test velocity for tests between 15 and 60 mph (24 and 97 km/hr). When the process is reversed, the intercept and slope data corresponding to the current CRASH3 structural tables are as shown in Table 1 for the first five vehicle classes.

Any velocity intercept and slope data, including those of Campbell (2, Table 1) and Wooley (12), may be recast into the form used in both CRASH and SMAC. This can be especially useful if specific full-width barrier test data or partial-width data, proportioned up to full-width data (because they will be proportioned back down for a partial-width impact), are known for the vehicle in question. Piecewise fits may be useful, as shown by Strother et al. (14). It is not necessary to maintain $A = \sqrt{2BG}$] if another assumption fits the data better. For instance, setting B to zero leaves A as the constant force when the kinetic energy (instead of the test velocity) is seen to vary linearly with the crush depth from an intercept G.

In the CRASH3 data tables, anomalous values may be noted: large, opposed variations of B and G from the norm, most notably for the rear of Classes 4 and 5/6, less so for the rear of Classes 7 and 8 (B = 13, 70, 55, 25; G = 4986, 628, 818, 2373). Such opposed variations suggest chance rotation about a clump of data at a single test speed, with oppositely varying slope and intercept. For Classes 4 and 5 this can also be seen by inspection of Table 1; the intercept is high and the slope low for Class 4, conversely for Class 5.

This suggests that the data are based largely on 30-mph barrier crashes with a scattering of other data insufficient to well define the intercept. If Campbell's intercept data for 1974 GM large cars are a better guide for Classes 4–8, those intercepts might be changed to V' = 7.5, 7.5, 8.0, and 8.5 mph, giving G' = 1243, 1374, 1405, and 1586; $B' = B(30 - V'_o)^2/(30 - V_o)^2$ = 28.9, 57.1, 46.6, and 30.1; and $A' = \sqrt{2BG}$] = 268, 396, 362, and 291. This rotates the data fit about the data at 30 mph, altering the data only for the much less and much more severe cases. A similar adjustment for the front of vans is left to the reader.

In the process the anomalies have been reduced in the zero-velocity intercept of the same test data, $A/B = V_o/(dV/dC)$. A/B is also (as shown elsewhere by McHenry) the prestress distance required to establish a threshold level of A pounds for a nonreturning spring of rate B lb/in., with prestress energy expenditure of G. (This is a limited analogy, because the energy is actually lost to hysteresis at the time of the impact, not at the time of manufacture.) A/B is generally well behaved at 2.1 to 3.6 in. for the sides and 6.0 to 10.5 in. for the front of passenger vehicles, and 9.4 to 9.6 in. for the rear in Classes 1, 2, and 3. The foregoing adjustment has modified the anomalous values of A/B = 27.9(!), 4.2, 5.5, and 13.7 in. for the rear of Classes 4–8 to 9.27, 6.93, 7.76, and 9.66 in., which are credible horizontal intercepts or prestress distances.

Monk and Guenther (15) updated the 1983 CRASH damage tables, but although they list the sources and the data analysis program, there are no source data. The final data are both tabulated and plotted, but the two presentations fail to agree by substantial amounts (Table 2). Obviously the user should consider both versions until the uncertainty can be resolved; the amounts listed in Table 2 are possible upward or downward corrections applicable respectively to the speed changes found for light and for heavy damage.

All this is of little help to the CRASH user with no access to modify the present tables programmed into CRASH; the most he can do is to choose another vehicle classification with table entries closer to his known data.

Following from Equation 3.1, the further assumption that energy per unit width is a constant allows evaluation of the

TABLE 1 CRASH3 DATA EXPRESSED AS BARRIER IMPACT DATA

Class	1*	2	3	4	5	1*	2	3	4	5
	Velocity Intercept (mph)					Slope (mph/inch)				
Front	7.7	6.5	6.8	9.2	7.6	1.20	1.09	1.19	0.87	0.87
Side	2.2	2.8	3.7	3.0	3.7	1.06	1.35	1.21	1.07	0.98
Rear	10.4	10.1	9.9	15.0	5.1	1.08	1.06	1.06	0.54	1.20

* W=2173+296=2469 (was 2469+296) so M=6.39 (was 5.70)

TABLE 2 MONK APPENDIX E EXCESS OVER CRASH3 DATA

Class	1*	2	3	4	5	1*	2	3	4	5
	At zero crush (mph)					At crush for 30 mph (mph)				
Front	<1	<1	<1	-1.5	<1	1	<1	<1	<1	<1
Side	<1	<1	<1	<1	1.5	8	-4	<1	-2.5	<1
Rear	<1	-5	<1	9	1	-1.5	-6	-3	-3	-2.5

* W=2173+296=2469 (was 2469+296) so M=6.39 (was 5.70)

energy for any crush contour by using the integral across the width,

$$E_c = \int (G + AC + 0.5BC^2)dw \quad (3.2a)$$

or by using a piecewise linear fit ($C = mx + b$) as suggested by Wooley et al. *(7)*:

$$E_c = \Sigma w_j\{G + (A/2)(C_{j-1} + C_j) + (B/6)(C_{j-1}^2 + C_{j-1}C_j + C_j^2)\}] \quad (3.2b)$$

(j = 1 to n) for n trapezoidal segments of independent widths w_j.

This permits any number of arbitrarily sized segments to fit any profile; this seems preferable to the specific trapezoidal integrals for two, four, and six equispaced points used in CRASH. As in the CRASH3 User's Manual (*5*, Section 11.3), further equations can be written for the depth and offset from the midpoint of the geometric center of the crush area to refine the crush centroid definition.

Equation 3.2b defines the work done in crush only for axial impact. If the direction of deformation is oblique and if the structure is isotropic (homogeneous), then it appears obvious that the measurements to be used above must be taken along and normal to the direction of deformation or corrected to those axes. But because this is not the treatment provided in CRASH, and there are related considerations to discuss, oblique deformation will be considered later.

DAMAGE DYNAMICS

Consider two vehicles in central impact along the X-axis separated by a distance $z = x_1 - x_2$ and acted on mutually by the impact force F_r, which causes crush of each vehicle. For now, neglect possible slippage between the vehicles and assume that the kinetic energy expended in impact equals the total crush energy

$$\Sigma E_k = \Sigma E_c \quad (4.1)$$

Now the relative acceleration due to the force will be

$$\ddot{z} = \ddot{x}_2 - \ddot{x}_1 + F_r/M_2 + F_r/M_1 = F_r\Sigma M^{-1} \quad (4.2a)$$

where $\Sigma M^{-1} = M_1^{-1} + M_2^{-1} = (M_1 + M_2)/M_1M_2 = (1 + R^{-1})/M_1$. Solving for the force and integrating over the distance gives

$$\Sigma E_k = \int F_r dz = \int \ddot{z}dz/\Sigma M^{-1} = \int (d\dot{z}/dt)dz/\Sigma M^{-1}$$
$$= \int \dot{z}d\dot{z}/\Sigma M^{-1} = 0.5(\dot{z}_o^2 - \dot{z}_f^2)/\Sigma M^{-1}$$
$$= 0.5\dot{z}_0(1 - \varepsilon^2)/\Sigma M^{-1} \quad (4.2b)$$

which is the loss in kinetic energy during the impact, the kinetic energy of approach (E_k) less the kinetic energy of separation. The latter was assumed to be zero in CRASH; the more general treatment here will always reduce to the CRASH treatment by the assumption that $\varepsilon = 0$. It is so readily included that it will be done at this point rather than later.

The quantity $\dot{z}_o$ is the closing speed, and $\dot{z}_f = \varepsilon\dot{z}_o$ is the speed of separation (in the opposite direction except in the case of perforation—the total penetration of a target by a bullet—when ε is negative). Their sum is the change in differential speed along the X-axis during the entire impact,

$$\Delta\dot{z} = \dot{z}_o + \dot{z}_f = (1 + \varepsilon)\dot{z}_o$$

or, by substituting Equation 4.2b solved for $\dot{z}_o$ and then Equation 4.1,

$$\Delta\dot{z} = (1 + \varepsilon)\sqrt{2\Sigma M^{-1}\Sigma E_c/(1 - \varepsilon^2)]} = \sqrt{2E^*\Sigma M^{-1}]} \quad (4.3a)$$

For $\varepsilon = 0$ this reduces to the CRASH equation (*5*, Section 9.1b, Equation 12) obtained by using the further assumption of simple linear stiffness of the structure, which restriction evidently is superfluous. In Equation 4.3a, two corrections for rebound have been consolidated into the definition of an equivalent energy of deformation,

$$E^* = \Sigma E_c(1 + \varepsilon)^2/(1 - \varepsilon^2) \quad (4.3b)$$

For each vehicle the divisor $(1 - \varepsilon^2)$ gives, from the actual crush energy E_c, the kinetic energy of approach along the crush axis, the quantity evaluated as E_k in the usual barrier test, so that part of the correction is built into the available data. The multiplier $(1 + \varepsilon)^2$ enables the determination of the speed change of the mass from approach to departure, as if more damage but no rebound had been found, or else (in the case of perforation) less damage but no exit velocity.

In Equation 4.3a, the time integral of Equation 4.2a, the individual speed changes for Vehicles 1 and 2 occur in inverse proportion to the affected mass (which is the principle of conservation of momentum), giving

$$U'_{\Delta 1} = \Delta\dot{x}_1 = (M_2/M_t)\sqrt{2E^*\Sigma M^{-1}}]$$
$$= (1 + R^{-1})^{-1} \sqrt{2E^*(1 + R^{-1})/M_1}]$$
$$= \sqrt{2E^*/M_1(1 + R^{-1})}] \quad (4.4a)$$

$$U'_{\Delta 2} = \Delta\dot{x}_2 = (M_1/M_t)\sqrt{2E^*\Sigma M^{-1}}]$$
$$= U'_{\Delta 1}/R \quad (4.4b)$$

where $M_t = M_1 + M_2$ is the total mass and $R = M_2/M_1$ as before.

Now, although the angular energy involved in any accompanying rotation is neither considered nor restricted, the effect of offset of the mass center from the impact force F_r studied earlier may be found. Assume a fixed offset,

$$h_i = y_i \cos(\beta_i + \xi_{\Delta i}) - x_i \sin(\beta_i + \xi_{\Delta i}) \quad (4.5a)$$

(i = 1, 2) where x_i and y_i are the body coordinates of the crush centroid and $\beta_i + \xi_{\Delta i}$ (with β_i usually zero) is the direction of the impact force in body axes. Then the acceleration at the point of impact is

$$\ddot{x}_{pi} = \ddot{x}_i - h_i\ddot{\psi}_i = (-F_r/M_i) - h_i(F_r h_i/M_i k_i^2)$$
$$= -F_r/\gamma_i M_i = \ddot{x}_i/\gamma_i \quad (4.5b)$$

(i = 1, 2) where

$$\gamma = k^2/(h^2 + k^2) \quad (4.5c)$$

So at the crush centroid the ratio of force applied to the vehicle to resulting linear acceleration $(F_r/\ddot{x}_{pi})$, which may be called the effective mass of each vehicle, is $\gamma_i M_i$. The force is still found from the deformation by conservation of energy; when that force is offset from the mass center, the acceleration at the crush centroid is $\ddot{x}_{pi}$ and at the mass center is $\dot{U}_{\Delta i} = \ddot{x}_i \gamma_i \ddot{x}_{pi}$.

Substituting effective masses in Equations 4.4 and letting M' denote their total, the speed change at each mass center in offset impact is obtained:

$$U'_{\Delta 1} = \gamma_1(\gamma_2 M_2/M')\sqrt{2E^*\Sigma(\gamma_i M_i)}]$$
$$= \sqrt{2E^*/M_1(\gamma_1^{-1} + R^{-1}\gamma_2^{-1})}] \quad (4.6a)$$

$$U'_{\Delta 2} = \gamma_2(\gamma_1 M_1/M')\sqrt{2E^*\Sigma(\gamma_i M_i)}] = U'_{\Delta 1}/R \quad (4.6b)$$

Evidently linear momentum continues to be conserved. Exclusive of oblique impact, the CRASH derivation (*5*, Section 9.1) covered the same physics as Equations 3.1 through 4.6 for the special case of linear crush stiffnesses and zero rebound, with no errors due to the omission of angular momentum or energy terms for that case, contrary to the reservations expressed by Wooley et al. (*7*).

CRASH correctly incorporates no moment coefficient of restitution as proposed by Brach (*10, 13, 16*), who remarks (*16*) that "the point of application of the (collision force) is never known precisely, (so) the resultant must consist of both a force (at some arbitrary point) and a moment." Means of obtaining that moment are then postulated, but this postulates knowledge of the unknown. If the error is known, the correction can be made; if it is not known, the moment coefficient of restitution is not known. All experimental values of that coefficient are no more and no less than discoveries of the investigational error in centroid location. Incorrect results can issue if the moment coefficient of restitution is assigned a generalized value other than unity, which applies zero correction to the centroid location.

APPROACH VELOCITIES BASED ON DAMAGE ANALYSIS

Whether for angular or for axial impact, to find the approach velocity of each vehicle by using the damage data, the components along the approach course are taken and the speed change is subtracted from speed at separation, giving

$$U'_{oi} = U_{si} \cos(\nu_{si} - \nu_{oi}) - U'_{\Delta i} \cos\xi_{\Delta i} \quad (5.1)$$

(i = 1, 2). For an exactly in-lane, centered, northbound rear impact, for the overtaken vehicle the angles are zero, the cosines are +1, and the separation velocity exceeds the approach velocity; for the overtaking vehicle $\xi_{\Delta i}$ is 180 degrees, its cosine is −1, and the approach velocity exceeds the separation velocity.

The cosines remain close to unity for impacts close to axial, so zero and 180 degree inputs could used for impacts within about 10 degrees of axial. For large angles, the exact form should be used. At about 20 degrees the angular solution based on location data becomes credible. In comparing the result from Equation 5.1 with the result from Equation 2.1, because the separation velocities are the same for both, these results will differ only to the extent that the speed changes during impact differ.

The inferred closing velocity will be the result of substitution of the values from Equation 5.1 in Equation 2.2, which in axial impact reduces to a simple subtraction.

OBLIQUE IMPACT

As previously mentioned, by using the Campbell structural model the impact force and energy can be inferred from the depth of crush, based on their observed interdependence in barrier crashes. If the force (F_r in Figure 1a) is along the normal to the surface, the crush of the jth segment (C_j in Figure 2a) is normal to the surface and the width w_j is along the surface. If the force is oblique at the angle ξ_r, logically the crush (C_j' in Figure 2b) is also oblique at the angle ξ_c in general and can be so measured, with the associated width w_j' measured normal thereto. There is no further correction for oblique impact for that structural assumption and those directions of measurement; the treatment is complete.

But field measurements along a field-selected direction of crush are not only inconvenient but presumptuous, because a new direction might be assigned later. It is more prudent to take measurements along and normal to a major axis of the vehicle and subsequently convert to the oblique measurements. The oblique depth and width are then given by

$$C_j' = C_j/\cos\xi_c; \; w_j' = w_j \cos\xi_c + (C_{j-1} - C_j) \sin\xi_c \quad (3.2c,d)$$

In programming these can be used directly in Equation 3.2b. Neglecting the component due to change of depth, from Equation 3.2a the crush energy becomes

$$E_c = \int [G + A(C/\cos\xi_c) + 0.5B(C/\cos\xi_c)^2] (\cos\xi_c)dw$$

$$= \int (G\cos\xi_c + AC + 0.5BC^2/\cos\xi_c)dw$$

showing multiplication of G by the cosine of the angle and division of B by the same quantity. If G were large and C were small (light oblique end impact), this could reduce, rather than increase, the energy for a given contour. (Less measured width, despite more measured crush, can mean less energy.)

Monk and Guenther (*15*, p. 48) began a similar departure from CRASH3 but omitted the cosine effect in depth, the sine effect in width, and G; the result was zero sensitivity to ξ_c. The experimental evidence discussed was so subject to uncertainty with regard to the structural resistance as to be inconclusive with regard to angularity effects, except that it did appear to confirm the absence of a $1 + \tan^2\alpha$ effect.

The nonisotropic assumption ($\xi_c \neq \xi_r$; $\alpha \neq 0$; crush not in the direction of the applied force) is discussed at length by Fonda (*8*) and appears to account for the different treatment of oblique impact in CRASH (with inconsistencies). However, even though the latter have been eliminated in the present treatment, it is preferred to assume the structure to be isotropic, as in SMAC. Then both skins (not just one) must buckle in corner impact, and, much as in SMAC, the structural characteristics in the deformed corner are conveniently divided along the corner trace, shown in Figure 2c. Each part is evaluated independently, by using the respective values of G, A, and B, and the results are summed. This weights the structural properties according to the involved width, varying smoothly from fully frontal to fully side deformation (not possible in CRASH).

Proceeding to the question of damage dynamics in oblique impact, the possibility of intervehicle sliding (scraping), which often occurs in oblique impact, is included. Using the force and motion vectors shown in Figure 2d, allowing for a nonisotropic structure ($\alpha \neq 0$), given each of the crush forces F_{ci}, the intervehicle force is

$$F_r = (F_c/\cos\alpha)_1 = (F_c/\cos\alpha)_2$$

which acts through a distance that is the cosine component of the crush distance plus the sine component of the scrape length,

$$dz = dz_1 + dz_2 = \Sigma(dC \ \cos\alpha + ds_v \ \sin\xi_r)_i$$

So the total work done, the product of force and distance, is

$$\Sigma E_p = \int F_r dz = \int (F_c/\cos\alpha)\ (dC \ \cos\alpha)_i + \int F_r(ds_v \ \sin\xi_r)]_i$$

$$= \Sigma E_c + \int F_r(\sin\xi_r)ds_v]_i \qquad (4.1a)$$

This expresses all of the potential energy expended as work done, scraping included. The scraping component will be discussed later.

Absent scraping, the total work done is no more and no less than the total crush energy; there is no correction for oblique impact. This is reasonable, because no other work is done; and this should have been the result reached in CRASH2 and 3. Instead, by using an inconsistently nonisotropic assumption, the force F_r was assumed to act through an excessive distance $dC/\cos\alpha$ shown in Figure 2d, increasing by the factor $\tan^2\alpha$ the distance traversed and hence the work done. There was no such extra work done; the derivation was and is incorrect. For the case of oblique impact with no rebound and no scrape, CRASH3 will overvalue the impact energy. (A method of adjusting the CRASH3 PDOF and damage midpoint was found, but was rather cumbersome and is not offered.)

THE GENERALITY AND VALIDITY OF CRASH

This completes the rederivation of the CRASH equations; it is hoped that the treatment has been clarified while the algebraic length has been reduced and its generality has been illuminated. The following points have been shown:

1. In CRASH there are no violations of the principles of conservation of momentum and conservation of energy. Mutual forces acting between two bodies inherently result in a conservation of linear momentum, in offset as in central impact, whatever the effect of offset on the intervehicle forces. When those forces are found from the damage assessment, the resulting moment due to offset determines the change in angular momentum. Lacking two flywheels interacting on a common shaft, angular momentum is not conserved. The principle of conservation of energy is applied. Criticisms of CRASH for failure to consider angular momentum, conservation of angular momentum, and conservation of energy are ill founded.

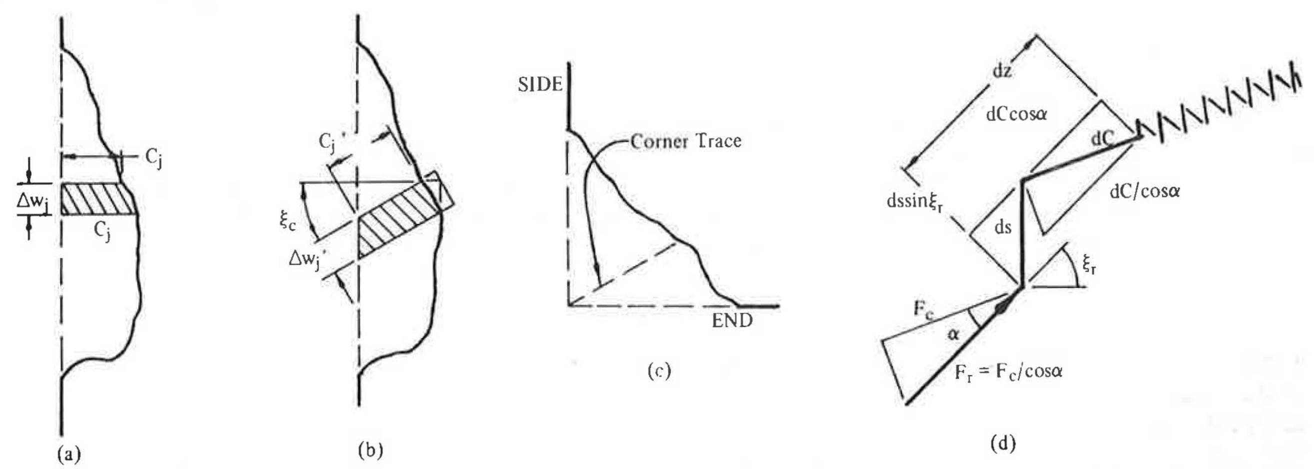


FIGURE 2 Crush motion relationship: *(a)* normal deformation, *(b)* oblique deformation, *(c)* corner deformation, *(d)* slide and crush.

2. The assumption in the damage dynamics equations of CRASH of linear crush characteristics of the vehicle structures can be seen as merely illustrative, a convenience in derivation. The rederivation here avoids any assumption as to the distance or time pattern of the intervehicle force, so that only the damage assessment so much as implies any particular force pattern for crush of the structure. Further research could alter that model whereas the remainder of CRASH would remain fully applicable.

3. CRASH3 overvalues damage and speed change for oblique impact without scraping and rebound, which are never considered.

The following results are obtained in this paper from the site data alone in all equations through 2.4 and from the damage data alone in the remaining equations, except that Equation 5.1 uses both:

Variable	*Equation*
Speed at separation	1.1 or 1.6
Speed at impact	2.1, 5.1
Closing velocity	2.2, 5.1
Velocity change in impact for Vehicle 1	2.3, 4.6a (i=1)
Velocity change in impact for Vehicle 2	2.3, 4.6b (i=2)
Direction of force from normal	2.4, examination
Damage energy	3.2
Effective damage energy	4.3b

For axial impacts, Equations 2.1 through 2.4 do not apply, but for angular impacts the separately obtained magnitudes and directions of the speed changes should be compared and reasonable input data revisions should be adopted when they result in greater compatibility of the independent results.

CRASH likewise furnishes two pairs of speeds for angular impacts, but only gives components inferring the computed PDOF and outlines no technique of refinement of the reconstruction. If the CRASH equations are as represented, results identical to those of CRASH could be obtained if the CRASH assumptions were reinstated. Although site data, cartesian-to-polar data reduction, crush centroid determination, trajectory simulation, and SMAC setup have not been attempted, all of the accident reconstruction results of CRASH have been duplicated or refined.

EXTENDED CRASH

In the process of showing the generality of the established CRASH equations, CRASH has already been extended by refining the common velocity check, by allowing irregular crush contour segmentation, by replacing the $1 + \tan^2\alpha$ correction with a proper consideration of oblique crush (including subdivision of corner crush along the corner trace), and by introducing the effects of rebound.

In the spin analysis, the yaw rate and spin time in Equations 1.6b and 1.6c (which were found internally in CRASH) and the post-spin speed (U_f) and time ($U_f/\theta\mu g$) can be furnished as program outputs.

The reconstruction can be further extended by means of certain further computations. These are informative in themselves and help to refine the reconstruction by providing additional pairs of values for the same quantity evaluated independently from different data. This proceeds according to logical equation number.

By extending the angular impact analysis, the peak force of impact may be approximated by assuming the $A = G = 0$, $B_1 = B_2 = B$ case of the structure and $M_1 = M_2 = M$—in effect, the barrier impact case with no velocity intercept. Then the affected structure will undergo harmonic motion of frequency $\sqrt{K/M]} = \sqrt{Bw/M]} = 17.6(dV/dC)$ rad/sec. The peak acceleration is then the frequency times the initial velocity, giving the peak intervehicle force as

$$F = 17.6(dV/dC)MU_{\Delta i} = 0.80(dV/dC)WU \quad (2.5)$$

where $17.6/g = 17.6/22 = 0.80$. As dV/dC is typically somewhat more than 1 mph/in., with good reason the peak impact acceleration can be approximated as a little under 1 g/mph of speed change. This is consistent with the 12.5 g/ft and 0.9 in./mph cited by Mason and Whitcomb *(17)*; $0.9\ (12.5/12) = 0.9375$ g/mph. In the metric system this is $(17.6/35.3)dV/dC = 0.50dV/dC$, or a little over $^1/_2$ g/kph. A less approximate treatment (revoking the simplifying assumptions) could be developed from these principles.

In either axial or angular impact, with location data, by combining the mass center speed at separation in the (confirmed) direction of the principal force with the velocity at the crush centroid location induced by the yaw rate, the speed of separation at the crash centroid in the direction of the force may be found:

$$U_{\Delta s} = \Sigma U_{si} \cos\{\nu_{si} - (\nu_{oi} + \xi_{\Delta i})\} - h_i\dot{\psi}_{si}] \quad (2.6)$$

(i = 1, 2), with the centroid offset h found from Equation 4.5a. As the vehicles separate in the direction of the forces, this is inherently positive and there is inherent subtraction of respective velocity components.

This is use of the damage location data without regard to damage severity in impulse analysis. It neglects the speed loss of each vehicle due to tire forces during impact up to the instant of actual separation; for side impacts this loss can be 1 or 2 mph, but is likely to be in much the same direction for both vehicles and will not significantly affect their speed of separation.

The coefficient of restitution is evaluated by dividing by the corresponding closing speed (inherently negative):

$$\varepsilon = -U_{\Delta s}/\Sigma U_{oi} \cos\xi_{\Delta i} - h_i\dot{\psi}_{oi}] \quad (2.7)$$

where the initial yaw rate $\dot{\xi}_{oi}$ normally is zero.

The same result could be obtained from CARR1 *(10)* if those equations (74ff) were used strictly to solve for the coefficient of friction and the coefficient of linear restitution, with a 1.0 value assumed for the coefficient of moment restitution.

As part of the damage analysis, it will be useful to evaluate for each vehicle the mean final crush depth in the direction of crush:

$$\begin{aligned} C_i^* &= \Sigma w_j'(C_{j-1} + C_j)/2]_i/\Sigma w_j']_i \quad (i = 1,2) \\ &\cong \Sigma[(w_j \cos\xi_c)\,(C_{j-1} + C_j)/2\ \cos\xi_c]/\Sigma w_j \cos\xi_c \\ &= \Sigma[w_j(C_{j-1} + C_j)]/2w_t \cos\xi_c \end{aligned} \quad (3.3)$$

and thence the average of the mean final crush depths, $C^{**} = 0.5(C_1^* + C_2^*)$. This will allow the independent estimation of the coefficient of restitution from the damage data by the following method, from SMAC *(3)*.

SMAC finds all velocities as the results of structural (and tire) forces applied to inertias, and stops the elastic recovery according to the coefficient of recovery $c = (\delta_{max} - \delta_f)/\delta_{max}$, as distinct from the coefficient of restitution $\varepsilon = \dot{\delta}_f/\dot{\delta}_0$. The values of ε that were the basis of the published values of c will be reconstituted.

Writing the ratio of the net work done on the structure to the gross work done before rebound (with the spring rate a constant, strictly a SMAC assumption) equated to the ratio of the respective kinetic energy losses by the impacting mass results in

$$\begin{aligned}\text{Net } E/\text{gross } E &= 0.5K\delta_f^2/0.5K\delta^2_{max}\\ &= 0.5M(\dot{\delta}_0^2 - \dot{\delta}_f^2)/0.5M\dot{\delta}_0\\ &= (1 - c)^2 = 1 - 2c + c^2\\ &= 1 - \varepsilon^2;\ \text{whence } \varepsilon = \sqrt{2c - c^2}\end{aligned} \tag{3.4a}$$

It is desired to find ε from the data available for the SMAC equation,

$$c = C_0 - C_1\delta + C_2\delta^2 = C_0\,(1 - \delta/\delta_l)^2 \quad (\delta \le \delta_l)$$

giving c as a function of crush depth with $C_1 = 2\sqrt{C_0C_2}$, as is imposed for the SMAC data. Because typically $C_0 = 0.064$ in SMAC, c^2 is negligible in Equation 3.4a and $\varepsilon = \varepsilon_0(1 - \delta/\delta_l)$ (for $\delta \le \delta_l$), or in the present notation,

$$\varepsilon' = \varepsilon'_0 = (1 - C^{**}/C_f) \quad (C^{**} \le C_f) \tag{3.4b}$$

which is a straight line between the intercepts $\varepsilon'_0 = \sqrt{2C_0}$ and $C_f = \sqrt{C_2/C_0}$. The currently standard SMAC inputs give $\varepsilon'_0 = 0.358$ and $C_f = 36.8$ in., or essentially $\varepsilon = 0.36 - 0.01C^{**}$. Of course, other expressions might be used [Smith and Tsongas (*18*, p. 47)].

For either vehicle the total crush force in the direction of crush, assuming that structural damping forces have subsided during the impact, is

$$\begin{aligned}F_c' &= \Sigma[(w_j \cos\xi_c)\,(A + B(C_{j-1} + C_j)/2\,\cos\xi_c)]\\ &= w_t\,(A\,\cos\xi_c + BC^*)\end{aligned}$$

Incorporating the possible angularity α due to nonisotropic structure, the total intervehicle force is

$$F'_{ri} = F'_{ci}/\cos\alpha_i = w_{ti}(A\,\cos\xi_{ci} + BC_i^*)/\cos\alpha_i \tag{3.5}$$

($i = 1, 2$). The damage-based values for the two vehicles should be in reasonable agreement with each other and with the location-based result from Equation 2.2c. In unusually light or heavy impacts, structural property adjustment by rotation about the 30-mph case as previously noted might substantially improve agreement between the forces. If crush data exist for only one of the vehicles, it is reasonable to reconstruct crush data for the missing vehicle by assuming a matching contour and peak force. The present approximations are not expected to closely evaluate the actual force of impact but rather to assist in refinement of the reconstruction.

The work done in intervehicular scraping will now be considered. Neither CRASH nor (to the author's knowledge) any other reconstruction treatment has considered the work done in scraping (intervehicular sliding), but it is entirely practicable.

Referring to the target vehicle, if the shear force has reached its friction limit ($\xi_r = \arctan\mu$) and scraping has occurred, additional work has been done on that vehicle by the shear component of the intervehicle force ($F_r \sin\xi_r$) moving along the shear surface through a distance s_v. Equivalently, it may be stated that the intervehicle force acts through the sine component of the shear motion plus the cosine component of its crush depth, as already expressed in Equation 4.1a. This expresses all of the potential energy as work done, scraping included.

The kinetic energy already found in Equation 4.2 does all of this work. If the same treatment is applied to the potential energy and if the direction of the force and the ratio of sliding to crushing are constant during the impact, substitution of the force from Equation 4.2a into 4.1a gives

$$\begin{aligned}\Sigma E_p &= \Sigma[E_c + (\sin\xi_r)\textstyle\int \ddot{z}ds_v/\Sigma M^{-1}]\\ &= \Sigma[E_c + (\sin\xi_r)\textstyle\int (d\dot{z}/dt)\,(ds_v/dz)\,dz/\Sigma M^{-1}]\\ &= \Sigma E_c] + (z_s/z_t)\Sigma E_k] = \Sigma E_k]\end{aligned}$$

so that the work done only in crush is

$$\Sigma E_c = (1 - z_s/z_t)\Sigma E_k = (z_c/z_t)\Sigma E_k = \Sigma E_k/R_s \tag{4.1b}$$

with

$$R_s = 1/(1 - z_s/z_t) = z_t/z_c = 1 + z_s/z_c \tag{4.1c}$$

where $z_t = z_s + z_c$; $z_c = \Sigma C' \cos\alpha = \Sigma C'$ if $\alpha = 0$; $z_s = \Sigma s_v \sin|\xi_r| = \Sigma s_v \sin|\xi_c + \alpha| = \Sigma s_v \sin(\arctan\mu) = \mu\Sigma s_v/(1 + \mu^2) \cong \mu\Sigma s_v$ so that R_s is the ratio of the total intervehicle motion to the component due to crush, or 1 plus approximately μ times the ratio of slide distance to crush depth. Equations 4.3 become

$$\begin{aligned}\Delta\dot{z} &= (1 + \varepsilon)\sqrt{2\Sigma R_s M^{-1}\Sigma E_c/(1 - \varepsilon^2)\cos^2\alpha}\\ &= \sqrt{2E^*\Sigma M^{-1}}\end{aligned} \tag{4.3a}$$

where

$$E^* = \Sigma E_c/\cos^2\alpha]R_s(1 + \varepsilon)^2/(1 - \varepsilon^2) \tag{4.3b}$$

is the equivalent energy of deformation, incorporating the multiplying factor for scraping; the obsolescent CRASH2-CRASH3 correction for oblique impact, if desired; and the two rebound correction factors previously discussed.

There must be appreciable vehicle crush to provide a measure of the normal and hence the shear forces. Data collection will now include observation, identification, and measurement of the scrape marks on the surfaces of the vehicles. For each set of simultaneous scrape marks, lest the same distance be counted twice, it is necessary to consider one vehicle as the target vehicle, which provides a relatively flat surface traced by a limited area of the bullet vehicle. The marks on the bullet vehicle, made at the same time, are not informational. However, there could be another set of marks that occurred before or after the first, in which the vehicle roles are reversed; the scrape process must be visualized carefully.

With damage data from the yaw acceleration found in Equation 4.5b, the yaw rate for each vehicle is inferred:

$$\dot{\psi}_{si} = \int (Fh_i/M_i k_i)dt = -U'_{\Delta i}(h_i/k_i^2) + \dot{\psi}_{oi} \qquad (4.6c)$$

($i = 1, 2$), with $\dot{\psi}_{oi}$ normally zero. For each vehicle this provides a second separation yaw rate to compare with the first.

The speed of separation at the crush centroid along the DOPF based on the deformation data is

$$U'_{\Delta s} = \varepsilon U'_{\Delta 1}(\gamma_1^{-1} + R^{-1}\gamma_2^{-1})/(1 + \varepsilon) \qquad (4.6d)$$

This gives values that can exceed 10 mph, much in excess of the hard limit of under 5 mph for an "acceptable" trajectory solution in CRASH3; yet it can also give small values that $U_{\Delta s}$ from Equation 2.6 should not greatly exceed.

Because $U'_{\Delta 1}$ will increase but ε will decrease as the crush increases, $U'_{\Delta s}$ will not vary rapidly with assumed crush depth. It is therefore a fairly reliable value to use in correcting the trajectory data, which obtain the separation velocity only from the difference in the trajectories to rest and could be considerably in error in individual cases.

The following equations provide six additional pairs of quantities of interest and of value in refining the reconstruction:

Variable	*Equation*
Separation yaw rate for Vehicle 1	1.6d, 4.6c (i=1)
Separation yaw rate for Vehicle 2	1.6d, 4.6c (i=2)
Force of impact for Vehicle 1	2.5, 3.5 (i=1)
Force of impact for Vehicle 2	2.5, 3.5 (i=2)
Separation speed at crush centroid	2.6, 4.6d
Coefficient of restitution	2.7, 3.3b

The degree of correlation to be expected between these sets of values will have to be found by experience. But inherently these all serve as validity checks whereby the reconstruction is checked for internal consistency and the bracketing is narrowed.

For purposes of statistical accident data collection, coefficients of restitution for angular (intersection) impacts between two vehicles are available for the first time by using Equation 2.8. The method requires no instrumentation; the result is reliable within some range according to uncertainties in the location data. This suggests application of effort in the statistical collection of empirical coefficients of restitution from real accidents, subject to avoidance of systematic errors in the site exam and the intervehicle coefficient of friction. DOT's continuing interest in occupant injury exposure data requires only a good damage exam, which may be all that is possible by the time DOT investigators arrive, but full reconstructions of intersection impacts, when possible, will give empirical coefficients of restitution and also improve the injury exposure data.

SUMMARY

The validity of CRASH in general has been confirmed, although some details have been revised and some limitations avoided. CRASH has never contained any errors due to omission of angular motion considerations and has never had a true limitation to the assumption of linearity of crash force with deformation.

The original treatment of CRASH gave damage-basis speed change overvaluation in the case of diagonal deformation and undervaluation by omission of rebound velocity and scraping. Overall, CRASH generally undervalues impact speed changes.

The new equations in this paper extend CRASH to give new results: peak impact force, individual speed changes including rebound, individual directions of speed change, individual yaw rates, joint speed of separation at the crush centroid, and joint coefficient of restitution in impact, all (at least in the case of intersection impacts) in pairs of values independently derived from different input data. These are of interest in themselves and allow input data refinement and increased accuracy of reconstruction.

As with the original CRASH programs, these solutions become practicable only when programmed for automated solution, as shown elsewhere *(8)*. Programming in BASIC allows full user review of the programming as well as the physics and algebra of the treatment. Whereas CRASH incomprehensibly treated or invisibly programmed is precarious for forensic or other critical purposes, the present paper provides CRASH techniques in a form acceptable for demanding applications.

REFERENCES

1. E. Marquard. Progress in the Calculations of Vehicle Collisions. *Automobiltechnische Zeitschrift*, Vol. 68, No. 3, 1966, pp. 74–80.
2. K. L. Campbell. Energy Basis of Collision Severity (SAE 740565). In *Proc., Third International Conference on Occupant Protection*, Troy, Michigan, July 10–12, 1974, Society of Automotive Engineers, Warrendale, Pa.
3. R. R. McHenry, I. S. Jones, and J. P. Lynch. *Mathematical Reconstruction of Highway Accidents*. Calspan ZQ-5341-V-2. U.S. Department of Transportation, Dec. 1974.
4. R. R. McHenry. *A Comparison of Results Obtained with Different Analytical Techniques for Reconstruction of Automobile Accidents*. SAE 750893. Society of Automotive Engineers, Warrendale, Pa., Oct. 1975.
5. *CRASH3 User's Guide and Technical Manual*. Report DOT HS 805-372. National Center for Statistics and Analysis, NHTSA, U.S. Department of Transportation, Feb. 1981; revised April 1982.
6. R. A. Smith and N. G. Tsongas. *Crash Phase Accident Reconstruction*. SAE 860209. Vol. P-174, Society of Automotive Engineers, Warrendale, Pa., Feb. 14–28, 1986.
7. R. L. Wooley, C. Y. Warner, and M. D. Tagg. *Inaccuracies in the CRASH3 Program*. SAE 850255. Vol. P-159, Society of Automotive Engineers, Warrendale, Pa., Feb. 25–March 1, 1985.
8. A. G. Fonda. *CRASH Extended for Desk and Handheld Computers*. SAE 870044. Society of Automotive Engineers, Warrendale, Pa., Feb. 23, 1987.

9. W. F. Milliken, D. W. Whitcomb, W. Close, C. L. Muzzey, and A. G. Fonda. Research in Automobile Stability and Control and in Tyre Performance. *Proceedings of the Automobile Division, Institute of Mechanical Engineers*, No. 7, 1956–1957.
10. J. E. Martinez. *Computer Aided Automobile Accident Reconstruction—Theory, Graphics, Animation.* SAE 850250. Vol. P-159, Society of Automotive Engineers, Warrendale, Pa., Feb. 1985.
11. R. M. Brach. *Least Squares Collision Reconstruction.* SAE 870429. Society of Automotive Engineers, Warrendale, Pa., Feb. 1987.
12. R. L. Wooley. *Accident Investigation Techniques.* Workshop Notes. Accident Investigation Techniques Task Group, Society of Automotive Engineers, Warrendale, Pa., Sept. 26, 1985.
13. R. M. Brach. Vehicle Collision Analysis. *Automotive Engineering and Litigation,* 1986 supplement.
14. C. E. Strother, R. L. Wooley, M. B. James, and C. Y. Warner. *Crush Energy in Accident Reconstruction.* SAE 860371. Vol. P-174, Society of Automotive Engineers, Warrendale, Pa., Feb. 24–28, 1986.
15. M. W. Monk and D. A. Guenther. *Update of CRASH2 Computer Model Damage Tables.* Report DOT HS 806-446. U.S. Department of Transportation, March 1983.
16. R. M. Brach. *An Impact Moment Coefficient for Vehicle Collision Analysis.* SAE 770014. Society of Automotive Engineers, Warrendale, Pa., 1977.
17. R. P. Mason and D. W. Whitcomb. *The Estimation of Accident Impact Speed.* Report YB-3109-V1. Calspan, Buffalo, N.Y., Aug. 1972.
18. R. A. Smith and J. T. Tsongas. *Accuracy and Sensitivity of CRASH.* Report DOT HS 806-152. U.S. Department of Transportation, March 1982.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.