TRANSPORTATION RESEARCH

RECORD

No. 1485

Safety and Human Performance

Human Performance and Safety in Highway, Traffic, and ITS Systems

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

TRANSPORTATION RESEARCH RECORDS, which are published on an irregular basis throughout the year, consist of collections of papers on specific transportation modes and subject areas. The series primarily contains the more than 600 papers prepared for presentation at Transportation Research Board Annual Meetings; occasionally the proceedings of other TRB conferences or workshops are also published. Each Record is classified according to the subscriber category covered in the papers published in that volume. The views expressed in the papers are those of the authors and do not necessarily reflect the views of the sponsoring committee(s), the Transportation Research Board, the National Research Council, or the sponsors of TRB activities. The Transportation Research Board does not endorse products or manufacturers; trade and manufacturers' names may appear in a Record paper only because they are considered essential to its object.

PEER REVIEW OF PAPERS: All papers (Annual Meeting papers and those presented at other TRB conferences or submitted solely for publication) in the Transportation Research Records have been reviewed and accepted for publication by the Transportation Research Board's peer review process established according to procedures approved by the Governing Board of the National Research Council. Papers are refereed by TRB technical committees as identified in each Record. Reviewers are selected among committee members and other outside experts. The Transportation Research Board requires a minimum of three reviews; a decision is based on reviewer comments and resultant author revision.

TRANSPORTATION RESEARCH BOARD PUBLICATIONS are available by ordering individual publications directly from the TRB Business Office or by annual subscription through organizational or individual affiliation with TRB. Affiliates and library subscribers are eligible for substantial discounts. For further information or to obtain a catalog of TRB publications in print, write to Transportation Research Board, Business Office, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418 (telephone 202-334-3214).

July 1995

TRANSPORTATION RESEARCH BOARD
1995 EXECUTIVE COMMITTEE

Chairman: Lillian C. Borrone, Director, Port Department, The Port Authority of New York and New Jersey, New York City
Vice Chairman: James W. van Loben Sels, Director, California Department of Transportation, Sacramento
Executive Director: Robert E. Skinner, Jr., Transportation Research Board

Edward H. Arnold, Chairman and President, Arnold Industries, Lebanon, Pennsylvania
Sharon D. Banks, General Manager, AC Transit, Oakland, California
Brian J. L. Berry, Lloyd Viel Berkner Regental Professor and Chair, Bruton Center for Development Studies, University of Texas at Dallas
Dwight M. Bower, Director, Idaho Transportation Department, Boise
John E. Breen, The Nasser I. Al-Rashid Chair in Civil Engineering, Department of Civil Engineering, University of Texas at Austin
William F. Bundy, Director, Rhode Island Department of Transportation, Providence
David Burwell, President, Rails-to-Trails Conservancy, Washington, D.C.
A. Ray Chamberlain, Vice President, Freight Policy, American Trucking Associations, Alexandria, Virginia (Past Chairman, 1993)
Ray W. Clough (Nishkian Professor of Structural Engineering, Emeritus, University of California, Berkeley), Structures Consultant, Sunriver, Oregon
James C. DeLong, Director of Aviation, Denver International Airport, Colorado
James N. Denn, Commissioner, Minnesota Department of Transportation, St. Paul
Dennis J. Fitzgerald, Executive Director, Capital District Transportation Authority, Albany, New York
James A. Hagen, Chairman of the Board, Conrail Inc., Philadelphia, Pennsylvania
Delon Hampton, Chairman and CEO, Delon Hampton & Associates, Chartered, Washington, D.C.
Lester A. Hoel, Hamilton Professor, Department of Civil Engineering, University of Virginia, Charlottesville
Don C. Kelly, Secretary, Kentucky Transportation Cabinet, Frankfort
Robert Kochanowski, Executive Director, Southwestern Pennsylvania Regional Planning Commission, Pittsburgh
James L. Lammie, President and CEO, Parsons Brinckerhoff, Inc., New York City
Charles P. O'Leary, Jr., Commissioner, New Hampshire Department of Transportation, Concord
Jude W. P. Patin (Brig. Gen., U.S. Army, retired), Secretary, Louisiana Department of Transportation and Development, Baton Rouge
Craig E. Philip, President, Ingram Barge Company, Nashville, Tennessee
Darrel Rensink, Director, Iowa Department of Transportation, Ames
Joseph M. Sussman, JR East Professor and Professor of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge (Past Chairman, 1994)
Martin Wachs, Director, Institute of Transportation Studies, School of Public Policy and Social Research, University of California, Los Angeles
David N. Wormley, Dean of Engineering, Pennsylvania State University, University Park
Howard Yerusalim, Vice President, KCI Technologies, Inc., Hunt Valley, Maryland

Mike Acott, President, National Asphalt Pavement Association, Lanham, Maryland (ex officio)
Roy A. Allen, Vice President, Research and Test Department, Association of American Railroads, Washington, D.C. (ex officio)
Andrew H. Card, Jr., President and CEO, American Automobile Manufacturers Association, Washington, D.C. (ex officio)
Thomas J. Donohue, President and CEO, American Trucking Associations, Inc., Alexandria, Virginia (ex officio)
Francis B. Francois, Executive Director, American Association of State Highway and Transportation Officials, Washington, D.C. (ex officio)
Jack R. Gilstrap, Executive Vice President, American Public Transit Association, Washington, D.C. (ex officio)
Albert J. Herberger (Vice Adm., U.S. Navy, retired), Administrator, Maritime Administration, U.S. Department of Transportation (ex officio)
David R. Hinson, Administrator, Federal Aviation Administration, U.S. Department of Transportation (ex officio)
T. R. Lakshmanan, Director, Bureau of Transportation Statistics, U.S. Department of Transportation (ex officio)
Gordon J. Linton, Administrator, Federal Transit Administration, U.S. Department of Transportation (ex officio)
Ricardo Martinez, Administrator, National Highway Traffic Safety Administration, U.S. Department of Transportation (ex officio)
Jolene M. Molitoris, Administrator, Federal Railroad Administration, U.S. Department of Transportation (ex officio)
Dharmendra K. Sharma, Administrator, Research and Special Programs Administration, U.S. Department of Transportation (ex officio)
Rodney E. Slater, Administrator, Federal Highway Administration, U.S. Department of Transportation (ex officio)
Arthur E. Williams (Lt. Gen., U.S. Army), Chief of Engineers and Commander, U.S. Army Corps of Engineers, Washington, D.C. (ex officio)

TRANSPORTATION RESEARCH RECORD

No. 1485

Safety and Human Performance

Human Performance and Safety in Highway, Traffic, and ITS Systems

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1995

Transportation Research Record 1485
ISSN 0361-1981
ISBN 0-309-06122-9
Price: $37.00

Subscriber Category
IVB safety and human performance

Printed in the United States of America

Sponsorship of Transportation Research Record 1485

GROUP 3–OPERATION, SAFETY, AND MAINTENANCE OF TRANSPORTATION FACILITIES

Chairman: Jerome W. Hall, University of New Mexico

Users and Vehicles Section
Chairman: Conrad L. Dudek, Texas A&M University

Committee on Transportation Safety Management
Chairman: Howard Dugoff, ICF, Inc.
Secretary: Michael M. Finkelstein, Michael Finkelstein & Associates

James A. Arena, Michael B. Brownlee, Noel C. Bufe, Terry M. Burtch, Leanna Depue, Adele Derby, Georges Dobias, Robert D. Ervin, Eugene Farber, John D. Graham, Philip W. Haseltine, Paul B. Jones, Dennis C. Judycki, Eugene R. Russell, Sr., Albert J. Slechter, Judith Lee Stone, Eileen Eckhart Strauch, David C. Viano

Committee on Vehicle User Characteristics
Chairman: Alison Smiley, Human Factors North, Inc.
Secretary: Thomas A. Ranney, Liberty Mutual Research Center
Gerson J. Alexander, Vivek D. Bhise, Linda Clifford, Robert E. Dewar, Eugene Farber, Brian A. Grant, Wolf-Dieter Kappler, Alexander C. Landsburg, Herbert Moskowitz, Michael Perel, Val Pezoldt, Thomas H. Rockwell, Kare Rumar, Michael Sivak, Anthony C. Stein, E. Donald Sussman, Richard van der Horst, Robert C. Vanstrum, Jerry Wachtel, Kenneth Ziedman, Helmut T. Zwahlen

Committee on Travelers Services
Chairman: Robert B. Helland, Federal Highway Administration
Brian C. Bonnett, Thomas J. Browne, Frank J. Cihak, J. Edwin Clark, Robert F. Dale, Michael S. Della Rocca, Cheryl Janne Gribskov, Robert C. McClure, Joseph B. Nolan, Michael A. Perfater, Neil D. Schuster, George T. Snyder, Jr., Samuel C. Tignor

Committee on Simulation and Measurement of Vehicle and Operator Performance
Chairman: R. Wade Allen, Systems Technology, Inc.
Elizabeth Alicandri, Paul S. Fancher, Jr., Deborah M. Freund, Slade Hulbert, Wolf-Dieter Kappler, P. Robert Knaff, Rodger J. Koppa, William H. Levison, Ian Noy, Charles E. Radgowski, Thomas A. Ranney, Arthur W. Roberts III, Gary L. Rupp, Alison Smiley, Jerry Wachtel, David H. Weir, Reginald T. Welles, Helmut T. Zwahlen

Committee on User Information Systems
Chairman: Arthur W. Roberts III, New Jersey Department of Transportation
Kevin N. Balke, David L. Helman, David B. Knies, Harold Lunenfeld, Richard Arnold Olsen, Jeffrey F. Paniati, Martin T. Pietrucha, James W. Stoner, James I. Taylor, Gary B. Thomas, Leslie Ann Whitaker, David M. Zaidel

Committee on Traffic Law Enforcement
Chairman: Don D. Hinton, Hinton and Associates
William E. Anderson, John W. Billheimer, George W. Black, Jr., R. Quinn Brackett, Susan N. Bryant, John J. Compston, Terry W. Conner, Olin K. Dart, Jr., David B. Daubert, Lawrence E. Decina, Douglas B. Gurin, Roy E. Lucke, Jerry McCoy, James W. McMahon, Jerry G. Pigman, Paul A. Pisano, Ted Purtle, John H. Strandquist, Fredrick M Streff

Committee on Traffic Records and Accident Analysis
Chairman: Paul P. Jovanis, University of California-Davis
Secretary: Jeffrey F. Paniati, Federal Highway Administration
Sherry S. Borener, David J. Bozak, Thomas E. Bryer, Myung-Soon Chang, Benjamin V. Chatfield, Ted Chira-Chavala, Barbara Hilger DeLucia, Mark Lee Edwards, C. Arthur Geurts, Alan F. Hoskin, Patricia S. Hu, John J. Lawson, Joseph C. Marsh IV, Judson S. Matthias, Ronald C. Pfefer, Carol Lederhaus Popkin, Howard S. Stein, John G. Viner, Phyllis E. Young

Committee on Motor Vehicle Technology
Chairman: Ben C. Parr, State Farm Insurance Companies
Secretary: Robert W. Mathews
R. Wade Allen, Michael R. Appleby, William T. Birge, Ovi Colavincenzo, Robert F. Fleischmann, A. Keith Gilbert, Joseph J. Innes, Lee L. Lowery, Jr., Charles A. Preuss, Stephen H. Richards, R. Lewis Sabounghi, Dean L. Sicking, Robert L. Ullrich

Transportation Research Board Staff
Robert E. Spicher, Director, Technical Activities
Richard F. Pain, Transportation Safety Coordinator
Richard A. Cunard, Engineer of Traffic and Operations
Nancy A. Ackerman, Director, Reports and Editorial Services

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1994.

Transportation Research Record 1485

Contents

Foreword

The papers in this volume are evidence of the breadth of safety interests and research on highway and traffic systems. Advanced technology, particularly the role of humans in new ITS systems, roadway design and features, and more effective use of traffic research are paving the way to safer highway transportation. To continually improve the safety record of a system that marked its 3 millionth fatality in 1995 appears to be an unending task. As vehicles, amount of traffic and travel, roadway complexity, and changes in the driving population influence safety, research such as presented in this Record is conducted to understand system changes, create improved systems and countermeasures, and evaluate performance. The reviewed papers presented are intended to provide information needed in the quest for a safer highway transportation system.

Key Human Factors Research Needs in Intelligent Vehicle-Highway System Crash Avoidance

LOUIS TIJERINA

The Intelligent Vehicle-Highway System (IVHS) program attempts to enhance surface transportation safety, efficiency, and comfort by applying advanced technologies, including information processing, communications, control, and electronics. One goal of IVHS is to reduce the number of crashes as well as the severities of crashes that do occur. Crash avoidance systems (CASs) are therefore a focus of research in the IVHS program and related programs abroad (e.g., DRIVE and PROMETHEUS). An important step in CAS research and development is to understand the sizes of various crash problems and their etiologies. This provides the necessary background for the development of CAS concepts. It is equally important to understand the time and distance budgets required and available for crash avoidance to assess the role that the driver might play. An IVHS program sponsored by the Volpe National Transportation Systems Center (VNTSC) and the Office of Crash Avoidance Research of NHTSA pursues this important step. A brief synopsis of key findings from the VNTSC crash problem studies is provided, and key research needs in the area of human factors are addressed. The issues are organized in terms of driver–CAS interface issues, driver response to CAS activation, the secondary effects of CASs on safety, and comprehensive crash avoidance. The topics discussed include the need to understand driver behavior, the impact of novel displays on driver acceptance, the use of graded alarms, driver reliability and reaction time, the effects (both positive and negative) of CAS false alarms on drivers, the feasibility of drivers taking evasive maneuvers, decreased driver attention to the driving task, increased hazard exposure on the roadway, change in driver behavior with the presence of CASs, expectancy violations, and the design of an integrated CAS.

The Intelligent Vehicle-Highway System (IVHS) program attempts to enhance surface safety, efficiency, and comfort by applying advanced technologies, including information processing, communications, control, and electronics (*1*). One goal of IVHS is to reduce the number of crashes as well as the severities of crashes that do occur. Crash avoidance systems (CASs) are therefore a focus of research in the IVHS program and related programs abroad (e.g., DRIVE and PROMETHEUS).

An important step in CAS research and development is to understand the sizes of various crash problems and their etiologies. This provides the necessary background for the development of CAS concepts. It is equally important to understand the time and distance budgets required and available for crash avoidance to assess the role that the driver might play. An IVHS program sponsored by the Volpe National Transportation Systems Center (VNTSC) and the Office of Crash Avoidance Research of NHTSA pursues this important step. The present paper provides a brief synopsis of key findings from the VNTSC crash problem studies and addresses key research needs in the area of human factors.

Battelle, 505 King Avenue, Columbus, Ohio 43201.

SYNOPSIS OF CRASH ANALYSES AND CAS CONCEPTS

Figure 1 delineates the crash types analyzed for crash problem studies and the percentages of U.S. police-reported crashes in 1 year accounted for by each crash type. These percentages are derived from a series of reports prepared as part of IVHS program crash problem studies (*2,3–10*) and other studies (*11–14*). Since the sample sizes of detailed crash cases examined in the various crash problem studies were limited, the samples do not necessarily represent the nationwide population of crashes. However, the figure provides some indication of the relative incidences of different crash types.

As a point of reference, approximately 6,110,000 police-reported crashes occurred in 1991 according to the statistics of the General Estimates System. The eight crash types analyzed account for roughly 68 percent of all crashes. As indicated in Figure 1, rear-end crashes account for almost one of every four police-reported crashes that occur in the United States each year. Single-vehicle roadway departures account for approximately 20 percent of all such crashes, and intersection-related crashes (signalized and unsignalized straight-crossing-path crashes and crashes involving a left turn across the path at an intersection) account for roughly 17 percent of all such crashes. The category labeled "other" in Figure 1 includes all other crash types such as animal strikes, untripped rollovers, and at-grade railroad crossing crashes.

Table 1 lists each crash type and its problem size estimate along with crash subtypes and causal factors (associated factors thought to be key contributors to crash occurrence). The data in Table 1 were also derived from the series of reports generated for IVHS program crash problem studies (*2–10*). These were determined by a clinical analysis of crash cases carried out by expert crash investigators at Calspan Corporation. The materials to support such analyses were detailed crash reports from the Crashworthiness Data System of the National Accident Sampling System. These reports included driver and witness statements, police comments, coded variables on reporting sheets, scaled schematics, measurements taken at the crash scene, and photographs of the involved vehicles.

Rear-end crashes are the most common type of crash. The analysis indicated that for about three of every four such crashes, the lead vehicle was stationary for some period of time before the impact. In the remaining crashes the lead vehicle was moving. Driver inattention and use of an inappropriate following distance or headway accounted for almost 93 percent of such crashes according to the cases analyzed. This suggests that there is potential for forward obstacle detection (for the case of a stationary lead vehicle), headway detection (for the case of a moving lead vehicle), or intelligent cruise control to alleviate at least some of these crashes.

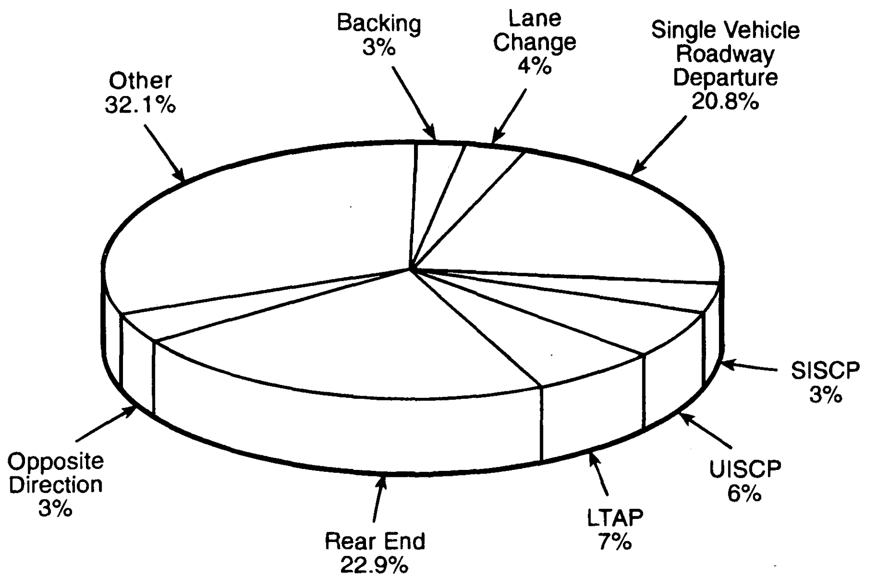


FIGURE 1 Distribution of crashes analyzed in VNTSC crash problem studies.

Single-vehicle roadway departures (SVRDs) are almost as prevalent as rear-end crashes. Roughly 6 of 10 roadway departures occur at curves; this suggests that a driver speed warning system cued to the presence and design speed of the curve (adjusted for slippery road conditions) would be promising. The remaining 4 of 10 roadway departures occur on straight road segments and may benefit from a lane-keeping aid. Unlike rear-end crashes, the etiology of SVRD crashes is much less uniform. According to the analysis, driver incapacitation is associated with roughly one of four such crashes, with slippery roads (20 percent), excessive speed (14 percent), and inattention and/or evasive maneuvering (29 percent) accounting for smaller percentages. Pavement monitoring (combined with speed warning) and driver impairment detection and warning may prove to be useful for countermanding some of these causal factors.

Backing crashes are of two main subtypes. One subtype involves the subject vehicle (SV) backing into a stationary or slowly moving object such as parked vehicles and pedestrians or pedal-cyclists. The SV is usually traveling at a slow speed. For this problem rear-zone detection and warning systems should be of benefit. The second subtype involves the SV backing out and moving into the path of a principal other vehicle (POV) that is traveling in a crossing path, usually at a substantially higher speed. Avoidance of such crashes will likely require a system that takes into account the POV time to arrival or gap to warn the SV driver (or perhaps the reverse, to warn the POV driver).

Lane change crashes occur predominantly because the SV driver was not aware that there was another vehicle in the adjacent travel lane. In the proximity subtype the POV was found to be not only in the blind spot but also occasionally beside or even slightly forward of the SV (although still overlapping laterally). This finding suggests that side-zone detection and warning may be beneficial but that detection only in the blind spot is not likely to be sufficient. The fast approach subtype, in which the one vehicle changes lanes unaware that another vehicle is fast approaching, was infrequent among the cases analyzed. The reasons for this were not found, but it may be due to increased alertness on the part of all drivers in situations in which fast-approach crash hazards are most likely (e.g., in merging traffic).

The intersection crashes analyzed involved both signalized (traffic light present) and unsignalized (stop or yield sign present) intersections and included both straight-crossing-path and left-turn-across-path crashes. Although the associated casual factors vary by intersection crash type, driver inattention and misperception of the POV are dominant. CAS concepts that are cued to signal state, sign presence, and gap or time to collision are likely to be of greatest benefit. In particular, a comparison of the time gap for a POV to arrive in a collision zone with the estimated time needed to complete an intersection maneuver may be especially useful for crash avoidance.

Based on the sample data, opposite-direction crashes are most often associated with an intoxicated driver; this is followed by evasive maneuvers (perhaps caused by a sudden intrusion into the SV's travel lane), slippery roads, and driver inattention (*10*). Interestingly, improper passing is seldom the cause of the opposite-direction crash, perhaps because drivers are usually alert when engaging in such a maneuver. Lane drift and vehicle control loss, along with the causal factor of driver intoxication, might be alleviated by driver impairment detection, lane detection and warning, and pavement monitoring and speed warning systems. Given the multiplicity of different crash circumstances, a single CAS concept will likely not be sufficient.

With this synopsis as background, the following sections will introduce key human factors issues in IVHS crash avoidance. These issues suggest research needs and challenges to bring IVHS crash avoidance to fruition.

DRIVER–CAS INTERFACE ISSUES

The user interface is the driver's gateway to CAS functions. Internal logic and features are only apparent through the interface, and the usability of that interface determines the extent to which the person gains access to the capabilities that the product or system has to

TABLE 1 Crashes, Crash Subtypes, Causal Factors, and CAS Concepts Analyzed

Crash Type	Problem Size	Crash Subtypes	Key Causal Factors	Key CAS Concepts
Rear-end Crashes	23%	• Lead Vehicle Stationary (75%) • Lead Vehicle Moving (25%)	• Driver Inattention and/or Following Too Closely (93%)	• Forward Obstacle/Headway Detection
Single Vehicle Roadway Departure (SVRD)	20%	• Straight Road Departure (39%) • Curve Road Departure (61%)	• Driver Incapacitation (DUI, Drowsy, Seizure) (25%) • Slippery Roads (20%) • Excessive Speed (14%) • Inattention and/or Evasive Maneuver (29%)	• Driver Impairment Detection • Pavement Monitor and Speed Warning • Curve Detection/Speed Warning • Lane Detection and Warning • Forward Obstacle/Headway Detection
Backing Crashes	1%	• Encroachment (43%) • Crossing Paths (57%)	• Did not see (61%) • Improper backing (Did not look?) (27%)	• Rear-zone Monitoring • Gap Monitor (Cross Traffic) and Warning
Lane Change Crashes	4%	• Proximity (93%) • Fast Approach (7%)	• Driver unaware of POV	• Side-zone Detection
Signalized Intersection Straight Crossing Paths Crashes (SISCP)	3%	• Straight Crossing Paths (100%)	• Deliberately Ran Signal - Ran Red Light (23%) - Tried to Beat Signal Change (16%) • Inattention (36%) • Intoxicated Driver (13%) • Vision Obstructed (4%)	• Signal-cued Warning/Control • Driver Impairment Detection
Unsignalized Intersection Straight Crossing Paths Crashes (UISCP)	6%	• Did Not Stop (42%) • Stopped, Proceeded Against Traffic (58%)	• Misperception of POV (49%) • Inattention (23%) • Vision Obstructed/Impaired (15%) • Intoxicated Driver (3%) • Deliberate Sign Violation (3%)	• Sign-cued Warning/Control • Gap Monitor (Cross Traffic) and Warning • Driver Impairment Detection
Intersection Left Turn Across Path Crashes (LTAP)	7%	• Did Not Stop Before Turn (72%) • Stopped, Then Turned (28%)	• Misperception of POV (28%) • Looked Did Not See (24%) • Vision Obstructed (24%) • Deliberate Traffic Control Device Violation (16%)	• Gap Monitor (Opposite Traffic) and Warning
Opposite Direction Crashes (ODC)	3%	• Lane Drift (47%) • Vehicle Control Loss (48%) • Improper Passing (5%)	• Intoxicated Driver (37%) • Evasive Maneuver (18%) • Slippery Roads (15%) • Inattention (7%)	• Driver Impairment Detection • Forward Obstacle/Headway Detection • Lane Detection and Warning • Pavement Monitoring/Speed Warning

offer. Although many reports discuss the ergonomics of displays and controls (*15*), a selected set of issues emerged from the analyses conducted in the crash problem studies.

Need To Understand Driver Behavior

A basic research need is to better understand typical driver behavior, the modifiability of that behavior, and the implications of these to CAS design, implementation, and assessment. Farber and Paley (*16*) presented data collected at New Mexico's I-40 near Albuquerque that showed that almost 30 percent of drivers were driving with time headways of 1.0 sec or less, which greatly increases the risk of a rear-end crash. How modifiable is such behavior if one were to introduce a headway monitoring system that alerts the driver that he or she is following too closely? In the backing case what are the typical eye, head, and body movements used in backing? This bears on the success of a driver warning system that is visual in nature. Although an auditory warning may alleviate the problem, what happens if multiple CASs are present in the vehicle and the driver must decide to what the warning refers? In the lane change crash what are useful indicators of the driver's intent to change lanes? This would be very helpful to tailor the CAS to present warnings only when they are appropriate. Even if some drivers use turn signal indicators, what is the sequence of turn signal use to lane change (before, during, or after)? How modifiable might turn signal use be if a CAS were provided with instructions that benefits will only accrue when the turn signals are used? In the case of left turns across path at intersections, what will a driver do when a CAS warns against turning yet the driver cannot visually verify a crash hazard (as may occur in cases of obstructed vision or driver misperception)? Answers to these questions are likely to require collection of data over a relatively long period of time to see both the time course and the steady-state behaviors of drivers.

A need to gain more information about driver behavior in crash circumstances also exists. For example, in the lane change crash analysis no information was available on the distributions of lane change times and the lateral accelerations associated with common lane changes (not evasive steering). In the case of straight crossing path crashes at intersections there is a need for information on the distributions of the velocities and accelerations that drivers exhibit while approaching and crossing intersections. Data on normal lane keeping would be helpful for the design of a lane detection and warning system. Finally, human factors engineering has learned the "fallacy of the average man," that is, trying to design a system to fit the 50th percentile person. Although distributional information may be useful for modeling and simulation in support of system development, a CAS may have to be tailored to the individual driver.

Novel Displays and Driver Acceptance

The nature of the displays for CASs will require careful research to arrive at acceptable solutions. Two examples will clarify this point.

Ward and colleagues (*17*) reported on a field study of a contact analogue head-up display (HUD) that provided infrared camera output directly on the windshield superimposed on the actual objects in the road scene. Compared with drivers who had no HUD, drivers using the prototype HUD drove more slowly and reported a higher subjective workload. Tijerina et al. (*18*) point out that speed reduction is a common technique that drivers use to manage high workloads, so these results are consistent with other human factors data. In an oral presentation of the paper and a videotape of the contact analogue HUD of Ward et al. (*17*), it was indicated that it was quite difficult to drive with the display because of the time delay in superimposing the infrared image with the real object and in the ghostly appearance of the infrared images. What is clear is that drivers will have difficulty in getting accustomed to the HUD imagery that is likely to be feasible (at least with infrared sensors) in the near term. Systems that cause greater workloads and travel times will likely find a lack of acceptance.

Consider next the use of kinesthetic-tactile displays in CASs. This approach uses torque shifts in the steering wheel and counterforce on the accelerator pedal to signal to the driver both a warning and indications of what to do. Theory and laboratory investigation suggest that such displays provide a stimulus and feedback that afford fast and accurate responses with little driver attention. However, drivers have sometimes described such displays as distracting and disturbing (*19*), and drivers might be unwilling to have such systems in their vehicles (*20*). Furthermore, when coupled with automatic vehicle control, kinesthetic-tactile displays were the least preferred from among several alternative CAS interfaces (*21*). Thus, the use of a novel display system with demonstrated performance-enhancing benefits may nonetheless fail to be accepted by drivers.

Another dimension to the acceptance of CAS interfaces may be social in nature. COMSIS (*22*) alluded to a social factor in interface design in which a warning presentation to a driver while one or more passengers are in the vehicle could be a cause of embarrassment. Research is needed to assess (a) how important such a factor might be, (b) what driver behaviors change in light of negative social consequences, and (c) how interface design might ameliorate or aggravate such social effects. For example, the kinesthetic-tactile display may come to be preferred to a buzzer or other more obvious warning delivery system by virtue of its subtlety.

Use of Graded Alarms and Transitions Between Levels of Intervention

Human factors literature suggests that graded alarms (e.g., crash possible, probable, or imminent) enhance performance over an all-or-nothing alarm system (*23*). It is therefore reasonable to consider the use of graded alarms for CAS implementation.

Several components are needed to make graded alarms feasible. First, the hazard or potential hazard must be detected by the CAS early enough to provide for alarm gradation. The time budgets may be inadequate to support the graded alarm concept in such circumstances as sudden intrusion by a POV, pedestrian, or pedal-cyclist into the SV travel lane. Second, there must not be an adverse consequence to the traffic system in general. For instance, graded alarms for rear-end crash avoidance would have to be carefully selected so that traffic flow is not degraded. Third, there must be sufficient structure in the driving environment to support graded alerts without introducing a multiplicity of false alarms. For example, one analysis conducted to provide a warning to the POV driver of an oncoming SV driver in the scenario of a straight crossing path crash at a signalized intersection determined that, at least for the conditions assessed, it was not feasible to provide graded warnings if the warning is cued to the SV's location from the intersection and anticipated decelerations given an assumed travel speed (*6*). On the other hand, graded warning or intervention for the SV driver at a signalized intersection (*6*) or stop sign (*7*) seemed more feasible, assuming that signal status and stop sign location could be known in advance. Finally, Horowitz (oral presentation, IVHS Human Factors Workshop, 73rd Annual Meeting, TRB, Washington, D.C., 1994) has raised concern about the psychological refractory period. This is a phenomenon found in the human performance laboratory in which presentation of a signal delays the response to a subsequent signal. Although such effects diminish with the interstimulus interval and disappear after an interval of 0.5 sec, this has not yet been assessed in the crash avoidance arena.

DRIVER RESPONSE TO CAS ACTIVATION

Driver Reliability and Reaction Time

Modeling CAS effectiveness involves some assessment of (a) the probability that the driver will respond appropriately and (b) the latency of that response. In general, models of CAS effectiveness assume full and accurate compliance with the CAS's warning. Farber (*24*) presented an example of human reliability that assumed that drivers were operating in parallel with the CAS. Given the nature of reliability of parallel systems, Farber (*24*) demonstrated that such a system will be highly reliable. On the other hand, Tijerina et al. (*6*) note that a series system may be the more appropriate reliability model for a CAS that is warning an inattentive or unaware driver. In a series system the system fails if any component fails. Tijerina et al. (*6*) demonstrated that if the CAS really works in series with the driver who is unaware of the crash hazard, then the total system is less reliable than either component alone. Research into the nature of how drivers respond to CAS activation from a reliability standpoint would be informative and might suggest a means for improving overall reliability.

The issue of latency of response is the next issue worthy of research. Taoka (*25*) favors modeling surprisal brake reaction time with a log-normal distribution. In particular, the data of Sivak et al. (*26*) were used in several analyses to assess, given a maximum time budget available for delay compatible with crash avoidance, what proportion of the driver population could respond as fast or faster than the maximum time available. An interesting point is to consider the validity of this approach for determining driver latency to respond to a CAS warning.

Figure 2, taken from Forbes (*27*), presents the cumulative distribution functions for brake reaction time collected in a variety of field settings by different methods and with different subjects. The distributions varied. What is of concern is that in most of the studies the drivers knew that they were involved in an experiment. Although they may not have anticipated the surprise stimulus that they experienced during the trial(s) represented in the data, they were nonetheless likely to be highly alert. What is needed are data, either obtained through field observations or extrapolated from theory, that will provide reaction time distributions associated with CAS warnings. In particular, it is interesting to speculate that the reaction time distribution will shift given (a) uncertainty about the threat (its location or existence) and (b) the time to select the appropriate response. Since systems in general and CASs in particular will not be 100 percent reliable or accurate, the impact of this may be manifested in a search-and-verification step on the part of the driver, a step that will take time and that should be factored into driver reaction time estimates to support both CAS effectiveness modeling and CAS design. Finally, given that older drivers exhibit slower responses, inclusion of data on older drivers in IVHS CAS effectiveness prediction and CAS design efforts is warranted.

CAS False Alarms and Their Effects on Drivers

Signal detection theory states that for a fixed level of sensitivity moving the decision threshold to afford a greater probability of correct detections will of necessity involve a greater incidence of false alarms (*28*). The problem of false alarms is difficult to assess in a limited experimental framework. Even though it may be possible to assess false alarms in a laboratory over the course of an hour, there may be little validity of such results to real-world conditions in which false alarms may occur over days or weeks rather than minutes. The costs and benefits of correct detections and false alarms and the probability of a hazard will also affect driver tolerance of false alarms.

Farber (*29*) recently presented quasi-Monte Carlo simulation results assessing the effectiveness of a rear-end collision warning for the lead vehicle moving. He compared the stopping distance

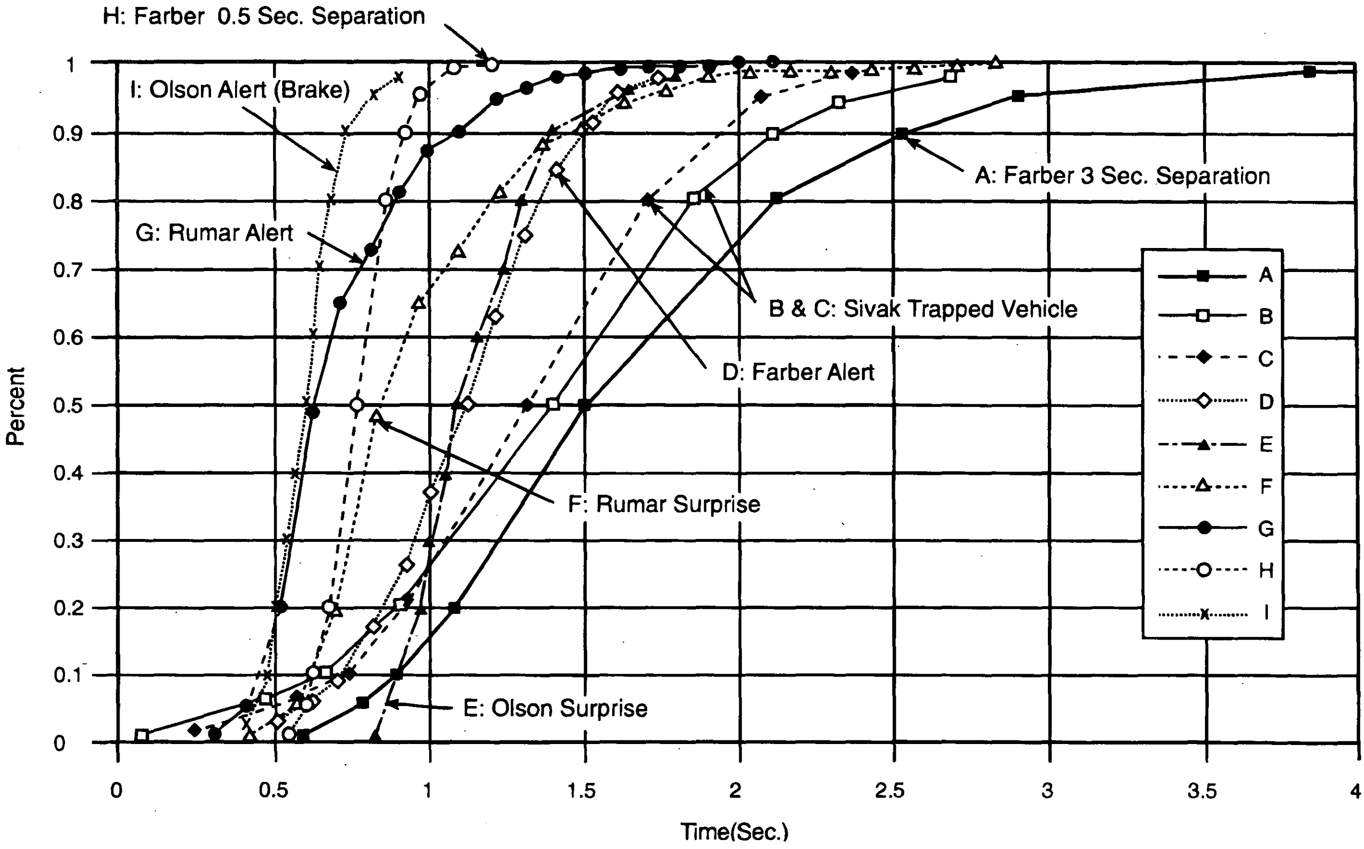


FIGURE 2 Brake reaction time distributions from various studies (*27*).

algorithm reported by Knipling et al. (*3*) with a closing rate algorithm using time headway data collected from more than 31,000 vehicle pairs on I-40 in New Mexico. Farber found that although the stopping distance algorithm had effectiveness rates approaching 100 percent, it gave approximately 1,100 false alarms for every crash, whereas the closing rate algorithm gave 6 to 11 false alarms per crash. Farber correctly points out that drivers may not be willing to respond to such frequent warnings, the great majority of which will be false alarms.

Close consideration of Farber's analysis suggests research issues associated with false alarms. First, the high incidence of false alarms results primarily from the fact that the New Mexico data included data for substantial proportions of drivers traveling with headways of less than 1.0 sec, which would lead to a preponderance of stopping distance algorithm alarms. It is plausible that, and perhaps worthy of a longitudinal study to determine if, drivers would modify their car-following behavior in the presence of a rear-end CAS. A study by Janssen and Nilsson (*20*) suggests that this will happen. In their simulator research they found that in most instances the distribution of headways with a CAS was shifted to reduce the proportion of close headways. Second, Farber (*24*) reports that drivers in the United States average about one reportable rear-end crash every 50 years. Amortizing 1,100 false alarms (recall that this is the number of warnings per crash) over 50 years, this amounts to less than 2 false alarms a month, on average. If the CAS provides a warning under alarming circumstances and such near-miss situations arise more frequently than actual crashes, perhaps the false alarm problem is really a blessing in disguise. One might hypothesize that such warnings provide familiarization to the driver and intermittent reinforcement to honor the CAS. Intermittent reinforcement has proven to be an excellent means of building resistance to extinction of a conditioned response (*30*) and may serve a similar function in IVHS crash avoidance. Thus, the issue of false alarms is indeed fundamental to CAS development, but it may involve both good and bad properties that should be assessed in a variety of ways, including longitudinal assessments of CASs perhaps in IVHS demonstrations.

Feasibility of Driver Response

In most kinematic analyses of crash avoidance, drivers are assumed to make braking responses alone or steering responses alone. Seldom are combined steering and braking responses considered. Allen (*31*) explicitly addressed this point. Figure 3 depicts stability limits on combined steering and braking maneuver accelerations on the basis of data for rear-wheel-drive and front-wheel-drive vehicles. Figure 3 shows that single responses may be more aggressive, but if both steering and braking inputs are applied, the vehicle may become directionally unstable, thus leading to secondary crash consequences. The abilities of drivers to provide the expected response is suspect and may require automatic control intervention to assist the driver and maintain safe control of the vehicle at all times during the emergency response. Alternatively, if the driver is provided with an early alert or warning, it is less likely that the driver will perform extreme and potentially unstable maneuvers.

SECONDARY EFFECTS OF CAS

A CAS could inadvertently undermine safety in a variety of ways. Some key concerns that have arisen over the course of the various crash analyses are presented.

Decreased Driver Attention to Driving Task

A major concern is that CASs will result in decreased driver attention to driving conditions. This might be manifested in decreased time spent looking at the road scene and less frequent mirror sam-

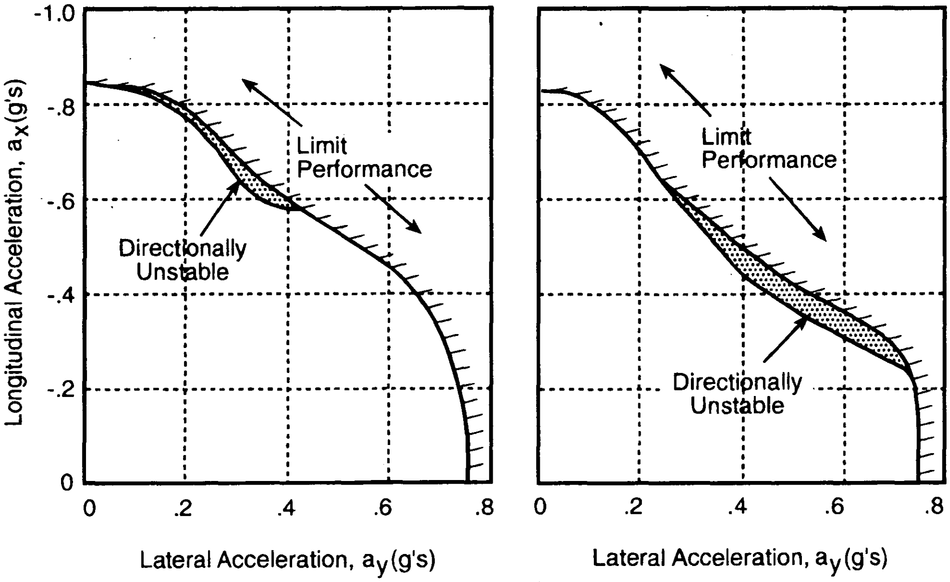


FIGURE 3 Limits on combined steering and braking maneuvering accelerations: *left*, rear-wheel drive; *right*, front-wheel drive (*31*).

pling. Drivers might develop a false sense of security that the CAS will protect them even when this is not true because of miscalibration or changes in settings, environmental degradation, or situations beyond the performance envelope of the CAS. Assessment of this concern should be a high-priority human factors research need. If it is found to be true further research will be needed to find ways to countermand this adverse consequence.

Increased Exposure on Roadway

Another secondary effect that CASs may have is increased hazard exposure. If, for example, CASs allow for greater traffic volume or invite more drivers onto the roadway under conditions that those drivers might previously have avoided (e.g., vision enhancement at night or during bad weather), then this puts more vehicles on the road and may increase crash probabilities given that CASs, vehicles, and infrastructures are less than perfectly reliable. Assessment of such effects will likely require longitudinal studies or large-scale demonstration projects.

Changes in Driver Behavior

Modifications in driver behavior have been mentioned previously. That some behavioral changes might adversely affect safety can be illustrated by a sample of studies. Nilsson (*32*) described work by Nilsson and Alm (*33*) who investigated vision enhancement in a driving simulator by simulating driving under clear conditions, foggy conditions, and foggy conditions with a simulated vision enhancement system that consisted of a monitor positioned on the hood near the windshield (i.e., simulating a HUD). On the monitor a clear picture of the road and its environment was presented to the driver. Drivers in the simulator with the HUD chose higher travel speeds than those chosen by drivers without the vision aid. Enhanced visibility benefits could be negated by higher speeds, especially if reduced visibility due to weather is accompanied by poorer traction or if higher speeds are not expected by other drivers sharing the roadway.

HUDs are supposed to provide the driver with an image on the windshield so that the driver does not have to take his or her eyes off the road. However, because of packaging constraints, only a portion of the road scene ahead will be subject to enhancement; this is called the HUD "eye box." The scene outside the HUD eye box will remain without enhancement. It is possible that the benefits of HUD vision enhancement will be offset by a reduced rate of detection of events in the periphery. Bossi et al. (*34*) conducted a simulator study of this and found significant impairment of peripheral target detection and identification performance under conditions intended to simulate night. These results need to be replicated by other methods since it was a simulator study rather than real-world driving, the targets were symbols presented in various locations rather than actual objects, and the driver's response time was to activate the high-beam stalk. However, it appears that the HUD for vision enhancement may capture a driver's visual attention and decrease a driver's attention to objects outside the eye box.

Consider a rear-end crash avoidance example to introduce the notion of modified safety margins. Janssen (*35*) presented a hypothetical effect of the presence of a CAS on the distribution of time to collision under normal car-following conditions (Figure 4). If there is a fixed criterion it is possible that drivers, over time, will compress the distribution of time headways or time to collision from the right-hand side because they are confident that the CAS will warn them. Although the warning thresholds presumably will be judiciously chosen, it is possible that there will be an increased crash potential if the CAS fails, if there is environmental degradation, or if there is a change in the warning threshold setting that the driver does not fully comprehend.

Expectancy Violations

Drivers depend a great deal on expectancies while they are on the roadway (*36*). Expectancies refer to a driver's readiness to respond to conditions, situations, events, and information in predictable and successful ways. Expectancies are of two different types: a priori and ad hoc. A priori expectancies are those that drivers bring to the

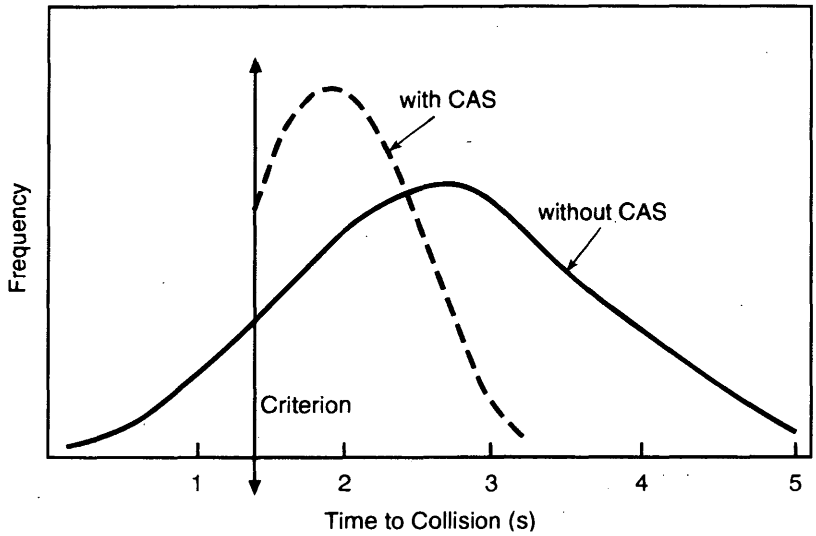


FIGURE 4 Hypothetical effect of CAS on distribution time to collision (*35*).

driving task on the basis of long experience, learning, culture, or upbringing. Ad hoc expectancies are those that develop from situation-specific factors encountered while driving. Violation of driver expectancies can have adverse consequences for safety.

As mentioned earlier adverse safety consequences might arise as a result of violated expectancies. Ervin (oral presentation, IVHS America Safety and Human Factors Committee and NHTSA Workshop on Collision Avoidance Systems, Reston, Va., 1994) has presented an interesting example in which the presence of a CAS in one vehicle leads to a crash because of violated expectations on the part of another driver. On a foggy night one driver of a vehicle without vision enhancement is waiting at a stop sign on a side street to make a left-hand turn onto a major road. The driver notices in the distance the dim headlights of an oncoming car. Expecting that no sane driver would be driving at high speeds with such poor visibility, the driver on the side street begins to pull out. Too late, the driver realizes the approaching vehicle is indeed traveling fast and a crash ensues. The hapless driver on the side street made use of expectancies developed through past experience that did not, and that perhaps could not, take the new vision enhancement technology into account. Assessments of such problems will be an important, although difficult, part of CAS evaluations.

Another potential complication in need of research is the extent to which performing maneuvers to avoid one crash leads to another crash as a by-product. Examples might include emergency braking to avoid rear-ending a lead vehicle. This could nonetheless precipitate a rear-end crash, albeit with the emergency braking vehicle now being the struck rather than the striking vehicle. Work analyzing crash avoidance opportunities to date has focused on single vehicles or pairs of vehicles for preliminary assessment of crash avoidance potential. Examination of the impacts of CASs on the traffic system in a more global perspective is warranted. It is likely that there will indeed be some instances in which there is no simple means of crash avoidance. Therefore, the design of a CAS that can gather evidence and quickly suggest an optimum solution or response is unlikely. This will probably be true if for no other reason than that which is optimum depends on the variables and the weights assigned to those variables in a cost function; these may vary by driver and circumstance.

COMPREHENSIVE CRASH AVOIDANCE AND HUMAN FACTORS IMPLICATIONS

The presence of more than one CAS in the vehicle (or an integrated CAS that provides warnings of more than one type of hazard) poses the potential for driver overload and confusion. The COMSIS Corporation (*22*) discusses means of prioritizing warnings and the need to convey different crash hazards to the driver in an effective manner. In a recent review of that document several human factors researchers in transportation were doubtful that multiple crash hazards are all that likely. Thus, the problem of multiple hazards and multiple warnings may be moot. On the other hand, a recent analysis of Crashworthiness Data System crash cases in support of the development of intersection crash avoidance specifications suggests that the issue bears further investigation. In particular, it was noted that a large proportion of crashes that occur at intersections are actually rear-end crashes. Presumably, a CAS that sensed a rear-end crash threat as the foremost crash threat would warn the driver of that first. Additional in-depth analyses of crash circumstances might uncover such scenarios of potential hazards so that further study can be undertaken to assess the likelihood that two crash hazards would arise simultaneously or in close succession.

CONCLUSIONS

CASs are a unique part of the IVHS mission because they are intended to have a direct link to safety. As such it is perhaps appropriate to verify that above all such systems do no harm. Thus, research needs should address both the positive effects and the negative effects that such systems might have. This work will benefit by taking a variety of approaches to answer specific questions. Laboratory work and studies in part-task simulators with rapid prototypes of CAS–driver interfaces may be useful for understanding and enhancing their human factors properties. Vehicular control studies to assess stability are well suited to the test track. High-fidelity simulation on a system such as the National Advanced Driving Simulator will be useful for testing driver performance under simulated crash circumstances in realistic scenarios that cannot be replicated on the roadway. Systems such as the portable Data Acquisition System for Crash Avoidance Research may provide a means of capturing real-world driving behavior and vehicle performance to augment simulator research. Many secondary effects on the driver and the traffic system as a whole will likely be understood only in the context of large-scale demonstration programs that take place over a long period of time.

CAS design, implementation, and evaluation involve many other issues besides human factors issues. These other factors include fail-safe operation, hardware reliability, and interoperability for cooperative systems, maintainability, and cost-effectiveness. It is hoped that a comprehensive, systems-oriented approach to CAS development will lead to systems that truly enhance traffic safety for many years to come.

REFERENCES

1. Najm, W. G. A Review of IVHS Crash Avoidance Technologies. Paper presented at IVHS America Safety and Human Factors Committee and NHTSA Workshop on Collision Avoidance Systems, Reston, Va., March 1994.
2. Hendricks, D., J. Allen, L. Tijerina, J. Everson, R. Knipling, and C. Wilson. *VNTSC IVHS Program Topical Report 1: Rear End crashes*, Vol. I and II. Battelle, Columbus, Ohio, 1992.
3. Knipling, R. R., M. Mironer, D. L. Hendricks, L. Tijerina, J. Everson, J. C. Allen, and C. Wilson. *Assessment of IVHS Countermeasures for Collision Avoidance: Rear-End Crashes.* Report DOT HS 807 995. NHTSA, U.S. Department of Transportation, 1993.
4. Hendricks, D., J. Allen, L. Tijerina, J. Everson, R. Knipling, and C. Wilson. *VNTSC IVHS Program Topical Report 2: Single Vehicle Roadway Departures*, Vol. I and II. Battelle, Columbus, Ohio, 1992.
5. Tijerina, L., D. Hendricks, J. Pierowicz, J. Everson, and S. Kiger. *Examination of Backing Crashes and Potential IVHS Countermeasures.* Reports DOT HS 808 016 and DOT-VNTSC-NHTSA-93-1. John A. Volpe National Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1993.
6. Tijerina, L., J. D. Chovan, J. A. Pierowicz, and D. L. Hendricks. *Examination of Signalized Intersection, Straight Crossing Path Crashes and Potential IVHS Countermeasures.* Reports DT HS 808 143 and DOT-VNTSC-NHTSA-94-1. John A. Volpe National Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1994.
7. Chovan, J. D., L. Tijerina, J. A. Pierowicz, and D. L. Hendricks. *Examination of Unsignalized Intersection, Straight Crossing Path Crashes and Potential IVHS Countermeasures.* Reports DOT HS 808 152 and DOT-VNTSC-NHTSA-94-2. John A. Volpe National Transportation

Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1994.

8. Chovan, J. D., L. Tijerina, G. Alexander, and D. L. Hendricks. *Examination of Lane Change Crashes and Potential IVHS Countermeasures.* Reports DOT HS 808 071 and DOT-VNTSC-NHTSA-93-2. John A. Volpe National Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1994.
9. Chovan, J. D., L. Tijerina, J. H. Everson, J. A. Pierowicz, and D. L. Hendricks. *Examination of Intersection, Left Turn Across Path Crashes and Potential IVHS Countermeasures.* Reports DOT HS 808 154 and DOT-VNTSC-NHTSA-94-4. John A. Volpe National Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1994.
10. Chovan, J. D., J. J. Everson, D. L. Hendricks, and J. Pierowicz. *Analysis of Opposite-Direction Crashes.* Battelle, Columbus, Ohio, 1994.
11. Knipling, R. R., D. L. Hendricks, J. S. Koziol, Jr., J. C. Allen, L. Tijerina, and C. Wilson. A Front-End Analysis of Rear-End Crashes. *Proc., IVHS America 1992 Annual Meeting*, 1992, pp. 17–20.
12. Knipling, R. R. IVHS Technologies Applied to Collision Avoidance: Perspectives on Six Target Crash Types and Countermeasures. *Proc., IVHS America 1993 Annual Meeting*, 1993, pp. 249–259.
13. Najm, W. G., J. S. Koziol, Jr., L. Tijerina, J. A. Pierowicz, and D. L. Hendricks. Comparative Assessment of Crash Causal Factors and IVHS Countermeasures. *Proc., IVHS America 1994 Annual Meeting*, 1994, pp. 412–421.
14. Mironer, M., and D. L. Hendricks. *Examination of Single Vehicle Roadway Departure Crashes and Potential IVHS Countermeasures.* Reports DOT HS 808 144 and DOT-VNTSC-NHTSA-94-3. John A. Volpe National Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1994.
15. Peacock, B., and W. Karwowski. *Automotive Ergonomics.* Taylor and Francis, London, 1993.
16. Farber, E. I., and M. Paley. Using Freeway Traffic Data to Estimate the Effectiveness of Rear-End Collision Countermeasures. *Proc. IVHS America 1993 Meeting*, 1993, pp. 260–268.
17. Ward, N. J., L. Stapleton, and A. M. Parkes. Behavioral and Cognitive Impact of Night Time Driving with HUD Contact Analogue Infra-Red Imaging. Paper presented at XIVth International Technical Conference on the Enhanced Safety of Vehicles, May 1994.
18. Tijerina, L., B. H. Kantowitz, S. Kiger, and T. H. Rockwell. Driver Workload Assessment of In-Cab High Technology Devices. Paper presented at XIVth International Technical Conference on Enhanced Safety of Vehicles, Munich, Germany, May 1994.
19. Farber, B., K. Naab, and J. Schumann. *Evaluation of Prototype Implementation in Terms of Handling Aspects of Driving Tasks.* DRIVE Project V1041. Traffic Research Center, University of Groningen, Groningen, The Netherlands, 1991.
20. Janssen, W., and L. Nilsson. *An Experimental Evaluation of In-Vehicle Collision Avoidance Systems.* DRIVE Project V1041. TNO Institute for Perception, Soesterberg, The Netherlands, 1990.
21. Nilsson, L., H. Alm, and W. Janssen. *Collision Avoidance Systems—Effects of Different Levels of Task Allocation on Driver Behavior.* DRIVE Project V1041. TNO Institute for Perception, Soesterberg, The Netherlands, 1991.
22. *Preliminary Human Factors Guidelines for Crash Avoidance Warning Devices.* COMSIS Corporation, Silver Spring, Md., 1993.
23. Sorkin, R. D., B. H. Kantowitz, and S. C. Kantowitz. Likelihood Alarm Displays. *Human Factors,* Vol. 30, 1988, pp. 445–459.
24. Farber, E. I. Intelligent Vehicle Highway System Benefits and Public Policy. Paper presented at 13th International Technical Conference on Experimental Safety Vehicles, Paris, 1991.
25. Taoka, G. T. Brake Reaction Times of Unalerted Drivers. *ITE Journal,* March, 1989, pp. 19–21.
26. Sivak, M., P. L. Olson, and K. M. Farmer. Radar Measured Reaction Times of Unalerted Drivers to Brake Signals. *Perceptual and Motor Skills,* Vol. 55, 1982, p. 494.
27. Forbes, L. M. Lyman M. Forbes discussion of Wade Allen's Paper: The Driver's Role in Collision Avoidance Systems. Paper presented at IVHS America Safety and Human Factors Committee and NHTSA Workshop on Collision Avoidance Systems, Reston, Va., March 1994.
28. Kantowitz, B. H., and R. D. Sorkin. *Human Factors: Understanding People–System Relationships.* John Wiley and Sons, New York, 1983.
29. Farber, E. I. Using the Reamacs Model to Compare the Effectiveness of Alternative Rear End Collision Warning Algorithms. Paper presented at XIXth International Technical Conference on the Enhanced Safety of Vehicles, Munich, Germany, 1994.
30. Houston, J. P. *Fundamentals of Learning and Memory*, 2nd ed. Academic Press, New York, 1981.
31. Allen, R. W. The Driver's Role in Collision Avoidance Systems. Paper presented at IVHS America Safety and Human Factors Committee and NHTSA Workshop on Collision Avoidance Systems, Reston, Va., March 1994.
32. Nilsson, L. Simulator Studies and Driver Behavior Research. In *Driving Future Vehicles* (A. Parkes and S. Franze, eds.), Taylor and Francis, London, 1993, pp. 401–407.
33. Nilsson, L., and H. Alm. *Effects of a Simulated Vision Enhancement System on Driver Behavior and Driver Attitudes.* PROMETHEUS Report VT1. Swedish Road and Traffic Research Institute, Linkoping, Sweden, 1991.
34. Bossi, L. L. M., N. J. Ward, and A. M. Parkes. The Effect of Simulated Vision Enhancement Systems on Driver Peripheral Target Detection and Identification. Paper presented at 12th Triennial Congress of International Ergonomics Association, Toronto, Canada, Aug. 15–19, 1994.
35. Janssen, W. H. *The Impact of Collision Avoidance Systems on Driver Behavior and Traffic Safety: Preliminaries to Studies Within the GIDS Project.* DRIVE Project V1041. TNO Institute for Perception, Soesterberg, The Netherlands, 1989.
36. Alexander, G., and H. Lunenfeld. Positive Guidance and Driver Expectancy. In *Automotive Engineering and Litigation*, Vol. 3 (G. A. Peters and B. A. Peters, eds.), Garland Law Publishing, New York, 1990, pp. 617–679.

Publication of this paper sponsored by Committee on Vehicle User Characteristics.

Analysis of Stated Route Diversion Intentions Under Advanced Traveler Information Systems Using Latent Variable Modeling

SAMER M. MADANAT, C. Y. DAVID YANG, AND YING-MING YEN

One of the benefits of advanced traveler information systems (ATISs) is their ability to divert travelers to alternative routes during traffic incidents to alleviate congestion. ATISs may effectively convince travelers to divert to alternative routes by providing information that is considered useful. Therefore, it is important to identify the factors that explain drivers' route diversion behaviors to properly assist in the design and implementation of ATISs. An application of latent variable models to determine the factors that affect drivers' stated intentions to divert from their usual routes when faced with traffic congestion is described. Two latent variables were identified: drivers' attitudes toward route diversion and their perceptions of the reliability of information provided by radio traffic reports (RTRs) or changeable message signs (CMSs). These two latent variables were determined to be significant explanatory variables of route diversion intentions. Some drivers' travel and socioeconomic characteristics and the type of information provided by RTRs and CMSs were also found to be important explanatory variables.

Advanced traveler information systems (ATISs) are being developed to provide information that affects travel choices such as en route diversion, route selection, and departure time decisions. However, the benefits of ATISs are achieved only if the commuters respond to the information conveyed by ATISs in a positive manner. The study presented here discusses and identifies some of the relevant factors that influence drivers' stated diversion propensities.

A literature review of previous work in the area of modeling drivers' route diversion behaviors under conditions of real-time information presentation is presented first. Then, a description of the stated preference data used for the study is provided. The overall response patterns are then discussed. The core of the paper is the methodology used to estimate latent variables of travelers' attitudes toward route diversion and their perceptions of the reliabilities of radio traffic reports (RTRs) and changeable message signs (CMSs) and the discrete choice analysis used to identify the contributing factors that influence route diversion propensity. Finally, conclusions are presented.

LITERATURE REVIEW

Earlier studies found that prescriptive and descriptive traffic information encourages route diversion (*1–7*). Those studies also indicated that drivers expressed a higher propensity to divert their routes when they were experiencing increasing delays and congestion, when travel times and travel distances on drivers' preferred routes were longer, when congestion was caused by an unexpected incident rather than a recurring event, when trip direction was from home to work, and as their familiarity with the alternative routes increased. Finally, it was reported that young, male, and unmarried drivers are more likely to divert to alternative routes.

Khattak et al. (*8*) used a questionnaire survey of downtown Chicago commuters to study the factors that affect diversion from and return to regular routes. They used joint multinomial logit models to model drivers' choices among three alternatives: no diversion, diversion and no return, and diversion and return. They determined that drivers who were risk seekers, who stated a higher diversion preference, who were familiar with several routes, or who were making longer trips have higher diversion rates. Drivers who were making longer trips were also found to be more likely to return to the original route after a temporary diversion. On the other hand, commuters who were classified as risk seekers were less likely to return to their regular routes after diversion.

Khattak and colleagues (*9*) conducted a study in 1993 in the San Francisco Bay Area to identify variables that affect the diversion propensities of commuters. That study used a linear regression model to explore the effects of different types of information on route diversion. The study found that the diversion rate increases as the amount of travel information increases and that prescriptive information might be sufficient to achieve a high diversion rate. They also found that route diversion propensity depends on the presence of opportunities to divert, personality characteristics, and weather conditions.

Researchers at the University of California at Davis used bivariate probit models to determine the factors that influence the use of traffic information by commuters and their propensity to use alternative routes (*10*). They found that long-distance commuters, females, college graduates, or respondents who reported uncertainty in travel time as a major problem are more likely to use traffic information. Furthermore, they found that drivers with higher incomes and levels of education and who often use traffic information have a higher probability of using alternative routes. Those researchers also estimated negative binomial models of route-changing frequency and identified several influential factors: perceptions of the accuracy of traffic information and variations in traffic condition, travel time, and travel distance.

Jou and Mahmassani (*11*) used Poisson regression models to relate route, departure time, and joint switching frequencies to three

---

School of Civil Engineering, Purdue University, 1284 Civil Engineering Building, West Lafayette, Ind. 47907-1284.

factors: characteristics of the commuters, work environment, and traffic network. They found that commuters tend to change their departure times, routes, or both more frequently in the morning than in the evening. Furthermore, they showed that route and departure time decisions are interdependent on each other. Finally, they found that all three factors mentioned earlier are important determinants of the departure time and route-switching behavior.

Adler et al. (*12*) used an in-laboratory interactive microcomputer simulation to collect data for the study of en route behavior under ATISs. They estimated a structural equation model for modeling the en route behaviors of drivers. In addition, they identified four latent factors that affect diversion and information acquisition by drivers and investigated the interrelationships between these decisions. They viewed these latent factors as arousal and motivation concepts that would lead drivers to divert their routes or acquire information.

To pursue a deeper understanding of travelers' route diversion behaviors, the study presented in this paper further explores the determinants that affect drivers' decisions regarding route diversion. Like other researchers, the richness of stated preference data, specifically, rating data, has been exploited. Moreover, this study incorporated stated preference data into a structural equation model of route diversion behavior and allowed the latent factors representing drivers' perceptions of the reliability of traffic information and attitudes toward compliance with traffic advisories to be captured.

DESCRIPTION OF SURVEY AND DATA SET

Factors that may influence drivers' willingness to divert from a regular route were summarized by Khattak et al. (*7,8*):

1. Characteristics of congestion such as length and cause of the delay;
2. Source of delay information such as radio traffic reports, CMSs, or personal observation of congestion;
3. Attributes of delay information received such as accuracy and reliability;
4. Attributes of regular and alternative routes such as travel time and safety;
5. Attributes of commuters such as socioeconomic characteristics and personality;
6. Trip characteristics such as trip origin and destination; and
7. Situational factors such as time pressure, time of the day (i.e., daylight hours or nighttime hours), and weather conditions.

Based on these factors, a questionnaire survey was designed to collect the necessary stated preference (SP) data. The relevant parts of the questionnaire were organized as follows. First, the respondents were asked about the characteristics of their commute trip. Then, the respondents were asked a series of questions concerning their attitudes toward route diversion and perceptions of the reliability of traffic information provided by RTRs and CMSs. A five-point Likert scale ranging from "strongly agree" to "strongly disagree" was used as the response format for these questions on attitudes and perceptions. Next, hypothetical situations were presented to the respondents to explore drivers' stated diversion propensities. The responses for these hypothetical questions were simple binary choices, that is, either "yes" or "no." Finally, the socioeconomic characteristics of the respondents were recorded.

The data for the study were collected through a phone survey of households in the northwestern part of Indiana that included Lake County, Porter County, and La Porte County. The respondents who participated in the phone survey were randomly selected from telephone directories. These respondents were first asked if they were frequent users of the Borman Expressway. A total of 491 valid observations were collected through the telephone survey.

The Borman Expressway is a section of I-94 that stretches from the county line of Indiana's Lake County and Porter County to the state line of Illinois and Indiana. Besides the Borman Expressway, three other east–west routes are located in the study area. These east–west routes include the I-90 Toll Road, US-12, and US-20. The Borman Expressway is one of the most heavily congested freeways in the Midwest and is characterized by very high truck traffic: about 30 percent of daytime traffic consists of commercial trucks.

Since SP data were used in the study to evaluate drivers' stated diversion propensities, a general understanding of the strengths and weaknesses of SP data is important. According to Ben-Akiva et al. (*13*), the advantages of SP data are that

1. They can educe preferences on new (nonexisting) alternatives;
2. Attributes are prespecified and error free;
3. Multicollinearity among attributes can be avoided; and
4. Attributes that are not easily quantified, such as safety and comfort, can be incorporated.

However, one major drawback of SP data causes this type of data not to be widely used in model estimation: the reliability of the elicited information under hypothetical scenarios and its consistency with actual market behavior (*13*). The reliability of SP data has two different aspects: validity and stability. Discrepancies between stated and actual behavior may exist because of policy or justification biases, and this is referred to as a lack of validity. Lack of stability relates to the magnitude of random errors in SP data (*14*). In one case reported in the literature, Wardman (*15*) found that the residual standard deviation of an SP choice model differed from that associated with a reveal preference (RP) model by 20 percent. However, Wardman also indicated that the 20 percent difference would not lead to critical differences between RP and SP choice probabilities. Nevertheless, caution should be taken whenever SP data are used.

SP data were used in the present study because the focus of the research is to investigate the latent variables that influence drivers' route diversion propensities. This type of information can only be obtained by asking some hypothetical questions by the SP approach.

OVERVIEW OF RESULTS

This section summarizes the results obtained from the telephone survey. It should be emphasized that the results presented in this section only describe the relative importance of various factors in determining a tendency toward diversion but do not give any absolute patterns of diversion.

Most respondents (87.6 percent) stated that they would divert to an alternative route when the Borman Expressway is congested. Almost 68 percent of respondents who participated in the study indicated that they cannot tolerate more than 15 min of traffic delay. More respondents stated that they would divert to an alternative route to avoid traffic delay and congestion during daylight hours instead of nighttime hours and during good weather conditions instead of bad weather conditions. It was also found that approximately the same number of respondents stated that they would divert to an alternative route to avoid traffic delay and congestion whether they are driving from home to work or from work to home. Such a result might indicate that the time pressure factor is not very significant to these respondents.

Drivers' attitudes toward route diversion and perceptions of the reliability of the information provided by RTRs and CMSs were identified through a series of questions that required the responses to be given on a five-point Likert scale. Table 1 reports the distributions of responses along with the statements used in the survey.

Most respondents stated that they have positive attitudes toward route diversion by answering agree or strongly agree to the first three statements (Table 1). Respondents were also asked about their perceptions of RTRs and CMSs in terms of information attributes (relevance, reliability, and accuracy) as shown in Statements 4 to 7 of Table 1. More than half of the respondents rated RTRs and CMSs better than average on these three information attributes by stating that they agree or strongly agree on Statements 4 to 6 and disagree or strongly disagree on Statement 7. This indicates that traffic information disseminated through current information sources is perceived positively by regular users of the Borman Expressway. It is important to point out that a high percentage of participants have positive attitudes toward diversion and good perception of RTRs and CMSs, as indicated in Table 1. This is because the respondents sampled in the study are frequent users of the Borman Expressway and are familiar with the highway network of the area. Finally, it is noteworthy that Statement 7 is designed as an opposite of Statement 5 to check the validities of the responses given by survey participants.

Respondents were then presented with five hypothetical scenarios and were asked whether they would divert to an alternative route from the Borman Expressway. The scenarios were characterized by different types of information conveyed. The results and the hypothetical scenarios are presented in Table 2.

The diversion rate is lowest when the information provided is qualitative (Table 2); such a result is expected since this type of information does not provide additional details compared with what was available to the drivers when they first found out about the congestion. Table 2 also indicates that the diversion rate increased as the amount of information provided increased. This suggests that some commuters might be restrained from diverting their routes because of not having enough information about their alternative route at present. Finally, the largest stated propensity to divert the route was obtained when RTRs and CMSs provided prescriptive information. Such a finding is also expected, because prescriptive information implies that the alternative route is the best option. These high diversion rates provide a good indication that drivers are responsive to the information given by RTRs and CMSs under incident conditions. Once again, the percentages of respondents who indicated a preference for route diversion during traffic congestion reported in Table 2 are relatively high because these respondents are frequent users of the Borman Expressway.

Finally, some socioeconomic characteristics of the 491 respondents who participated in the survey are summarized in Table 3. According to Table 3 the majority of people who participated in the study range in age from 20 to 65 years (91.8 percent). Such results are reasonable because most people in these age groups are working people.

TABLE 1 Distributions (Percent) of Responses to Questions on Attitudes and Perceptions

Statement	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1. For my typical trip using the Borman Expressway, I am familiar with at least one other alternative route besides the Borman Expressway.	4.3	4.1	0.4	36.8	54.4
2. I often change my planned route while driving.	12.2	26.9	3.7	29.9	27.3
3. I am willing to divert to alternative routes to avoid traffic delays/congestion.	1.8	3.5	2.0	37.9	54.8
4. I frequently listen to radio traffic reports (RTR) or take notice of changeable message signs (CMS).	7.5	19.6	2.9	40.3	29.7
5. RTR or CMS usually provide information useful to me.	6.9	11.0	5.9	49.3	26.9
6. I often change my route in response to RTR or CMS.	9.4	28.5	4.9	36.4	20.8
7. RTR or CMS do not provide any relevant information.	30.7	51.1	6.9	8.2	3.1

TABLE 2 Route Diversion Behavior in Response to Five Hypothetical Scenario Questions

Scenario	Respondent Who Indicated a Preference for Route Diversion (%)
Qualitative Information	
1. Congestion is reported by RTR or CMS but no information is conveyed concerning expected delay time or possible alternative routes.	59.7
Quantitative Information	
2. Congestion is reported by RTR or CMS and information is given regarding the expected delay time, but no information regarding alternative routes is conveyed.	71.3
3. Congestion is reported by RTR or CMS and the information conveyed includes the expected delay time and specific instruction on the direction of your best alternative route.	86.8
4. Congestion is reported by RTR or CMS and the information conveyed includes the expected delay time, specific instruction on the direction of your best alternative route, and travel time on your best alternative route.	90.8
Prescriptive Information	
5. Congestion is reported, and RTR or CMS urges you to take your best alternative route.	94.1

LATENT VARIABLE MODELING

Latent variable models have been used in the social and behavioral sciences for many years. Recently, researchers from other disciplines such as economics and transportation have also used the concepts of latent variables. Ben-Akiva et al. proposed an analytical framework for incorporating psychometric data in the modeling of travel decisions (*13*). Figure 1 presents the framework used in the present study. In Figure 1 latent variables are those terms inside the ellipses, whereas the measurable (manifest) variables are inside the rectangles. According to Figure 1 drivers' preferences are influenced by two major components: (a) manifest variables that include socioeconomic characteristics and traffic information and (b) latent variables that include attitudes and perceptions. Since attitudes and perceptions cannot be measured directly, attitudinal indicators and perceptual indicators (i.e., Statements 1 to 7 from Table 1) were used to measure drivers' attitudes and perceptions. The present study used the framework in Figure 1 to identify the latent variables that influence drivers' route diversion intentions. First, the methodology used for the analysis of latent variables will be presented. Then, the results of the analysis are described.

TABLE 3 Socioeconomic Characteristic Distributions of Respondents

Classification of Socioeconomic Characteristics	Distribution (%)
Age Groups	
Less than 20 years old	2.1
20 ~ 29 years old	24.6
30 ~ 39 years old	28.5
40 ~ 49 years old	21.8
50 ~ 65 years old	16.9
Greater than 65 years old	6.1
Marital Status	
Single	38.3
Married	61.7
Gender	
Male	56.2
Female	43.8

Methodology

The analysis of latent variables for the present study was accomplished by using the LISREL model. The LISREL model consists of two parts: structural equations and measurement equations.

Structural equations specify the relationships between the latent variables. For the purpose of the present study the structural equation specifies the relationship between the unobservable factors that influence route diversion propensity. According to Everitt (*16*) a structural equation that relates two types of latent variables, dependent and explanatory, can be expressed as

$$\eta = B\eta + \Gamma\xi + \zeta \qquad (1)$$

where

$\eta' = [\eta_1, \ldots, \eta_m]$, a vector of dependent latent variables,
$\xi' = [\xi_1, \ldots, \xi_n]$, a vector of explanatory latent variables,
$\zeta' = [\zeta_1, \ldots, \zeta_m]$, a vector of residuals representing both errors in equations and random disturbance terms,
B = the matrix that contains regression weights for predicting η's from other η's, and
Γ = the matrix that contains regression weights for predicting η's from ξ's.

The direct causal effects of η variables on other η variables are represented by the elements of B; hence, the diagonal elements of B are

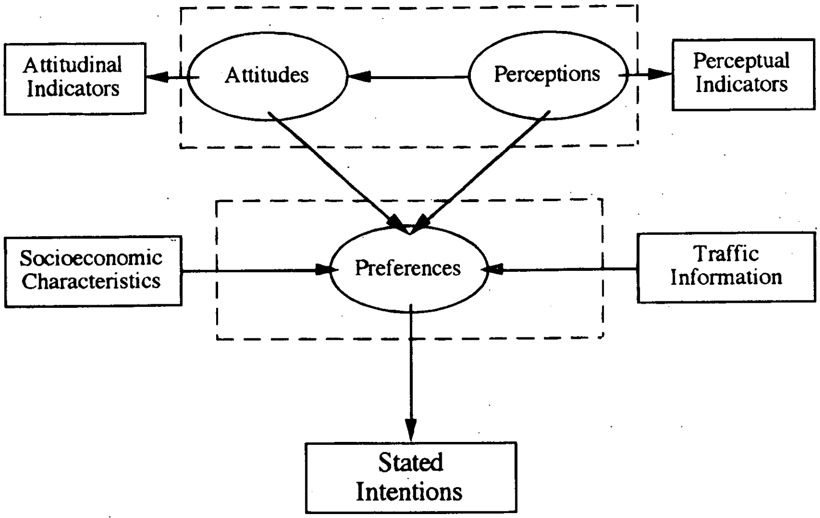


FIGURE 1 Framework for analysis of drivers' stated route diversion intentions.

zero. Similarly, the elements of Γ represent the direct causal effects of ξ variables on η variables. ζ and ξ are assumed to be uncorrelated.

Measurement equations specify how the latent variables relate to the observed (manifest) variables. Two sets of observed variables correspond to the two types of latent variables mentioned earlier: $y' = [y_1, \ldots, y_q]$ and $x' = [x_1, \ldots, x_p]$. The y variables are regarded as the indicators for η, the dependent latent variables. The indicators of the explanatory latent variables, ξ, are the x variables. Hence, the measurement part of the LISREL model that relates the manifest and latent variables can be written as

$$y = \Lambda_y \eta + \epsilon \tag{2}$$

and

$$x = \Lambda_x \xi + \delta \tag{3}$$

where

Λ_y = the matrix that contains regression weights of y on η, $(q \times m)$,
Λ_x = the matrix that contains regression weights of x on ξ, $(p \times n)$, and
ϵ and δ = vectors of error terms corresponding to y and x, respectively.

The methodology discussed in this section is used to analyze the data and identify the relevant latent variables. It is important to keep in mind that the LISREL model assumes that the manifest variables are independent of one another given the values of the latent variables. This assumption is termed *conditional independence*, and it implies that the observed relationships between the manifest variables are produced by the latent variables.

Discussion of Results and Results of LISREL Model

Survey participants' responses from the seven questions on attitudes and perceptions presented in Table 1 were used to identify the latent variables. It was hypothesized that the latent aspects of drivers' route diversion propensities can be represented by two latent factors: (a) drivers' attitudes toward route diversion (denoted η_1) and (b) drivers' perceptions of the reliability of traffic information (denoted η_2). Hence, the structural equation can be expressed as

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \tag{4}$$

where

$\eta' = [\eta_1, \eta_2]$, the vector of dependent latent variables,
$\zeta' = [\zeta_1, \zeta_2]$, the vector of residuals, and
β_{12} = the direct effect of η_2 on η_1.

The measurement equation for the present study has seven indicators for the two latent variables that were identified. These seven indicators correspond exactly to the seven statements described in Table 1; thus, the first indicator (denoted y_1) corresponds to Statement 1, the second indicator (denoted y_2) corresponds to Statement 2, and so forth. The measurement equation is expressed as

$$\begin{vmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{vmatrix} = \begin{vmatrix} \lambda_{11} & 0 \\ 1 & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & \lambda_{42} \\ 0 & 1 \\ \lambda_{61} & \lambda_{62} \\ 0 & \lambda_{72} \end{vmatrix} \begin{vmatrix} \eta_1 \\ \eta_2 \end{vmatrix} + \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \end{vmatrix} \tag{5}$$

where

$y_1, \ldots, y_7$ = indicators for η_1 and η_2,
$\lambda_{11}, \ldots, \lambda_{72}$ = regression weights of the indicators on η_1 and η_2, and
$\epsilon_1, \ldots, \epsilon_7$ = error terms.

In the Λ_y matrix from Equation 5, λ_{21} and λ_{52} have been set equal to 1 to normalize the scale of η_1 and η_2. Also noted from Equation 5, several elements of the Λ_y matrix were set equal to zero to denote that some indicators do not load onto a particular latent variable. For example, λ_{12} equal to zero means that the first indicator (i.e., Statement 1, For my typical trip using the Borman Expressway, I am familiar with at least one other alternative route besides the Borman Expressway) is not an indicator of a driver's perceptions of information reliability (i.e., η_2). It is also evident from Equation 5 that the fourth indicator (i.e., Statement 4) and the sixth indicator (i.e., Statement 6) extend over both latent variables, and thus, these two indicators were loaded on both η_1 and η_2. This overlap implies that these indicators are related to both latent variables, which is established by the results presented in Table 4.

Examining the results in Table 4, Indicators 1, 2, and 3 (i.e., λ_{11}, λ_{21}, and λ_{31}, respectively) can be classified as attitudinal indicators since these three indicators show a significant relationship to η_1, drivers' attitudes toward route diversion: they have high coefficients and *t*-statistic values. Similarly, indicators 4, 5, 6, and 7 (i.e., λ_{42}, λ_{52}, λ_{62}, and λ_{72}, respectively) were classified as perceptual indicators because they have high coefficients related to η_2, drivers' perceptions of information reliability. Furthermore, Indicators 4 and 6 also have a noticeable relationship to η_1, as discussed earlier.

The element of matrix B, β_{12}, represents the direct causal effects of η_2 on η_1. Therefore, a positive β_{12} value indicates that a driver who has a good perception of traffic information provided by RTRs and CMSs is likely to have a positive attitude toward route diversion. Such a finding is reasonable and expected. Also shown in Table 4 is the chi-square value for the null hypothesis that the predicted covariance matrix of the *y* variables is equal to the observed covariance matrix of the *y* variables. This statistic equals 15.93, which is less than the 95th percentile of the chi-square distribution with 11 degrees of freedom, 19.68. This suggests that the LISREL model used here does provide an adequate fit for the data.

Finally, the estimated values of the two latent variables, known as factor scores, were obtained by using the LISREL model. Every observation collected for the study has its corresponding factor scores, or $\eta_{1,i}$ and $\eta_{2,i}$, where *i* is equal to 1, . . . , 491. These factor scores are used in the following section to determine the factors that influence drivers' stated route diversion propensities by using discrete choice models.

TABLE 4 Estimation Results for LISREL Model

Parameter	Estimated Coefficient	t-Statistic
β_{12}	0.3143	4.6004
λ_{11}	0.6570	7.3781
λ_{31}	0.4604	6.4173
λ_{41}	0.1123	1.7397
λ_{42}	0.8832	14.7491
λ_{61}	0.3521	5.1483
λ_{62}	0.6950	12.0282
λ_{72}	-0.6512	-14.3891
Summary Statistics:		
No. of Observations = 491		
Degree of Freedom = 11		
Chi-Square Value = 15.93		

BINARY CHOICE MODEL

Methodology

The five hypothetical scenario questions (Table 2) in the survey that were presented to the respondents to explore drivers' stated diversion propensities required simple binary responses of yes or no. Since the choice set in this situation contains exactly two choices, binary choice modeling is the appropriate analysis tool to be used. For the purpose of the present study a binary logit model is used.

The model can be represented by the following equation:

$$U = \alpha p + \beta q + \gamma\eta + \nu$$

$$U^* = \begin{cases} 1 \text{ if } U \geq 0 \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

where

U = unobserved variable representing a respondent's propensity to divert,
U^* = observed choice (1 if the respondent diverts to an alternative route and 0 otherwise);
p = vector of dummy variables that represents various types of information provided to drivers, in this case, the five hypothetical scenarios described in Table 2;
α = coefficient vector corresponding to p;
q = vector of manifest variables that influence choices, which includes the travel and socioeconomic characteristics of respondents;
β = coefficient vector corresponding to q;
η = vector of latent variables that influence route diversion decision, in this case, values of $\eta_{1,i}$ and $\eta_{2,i}$ obtained from the LISREL model;
γ = coefficient vector corresponding to η; and
ν = error term.

The αp portion of Equation 6 can be rewritten as

$$\alpha p = \alpha_0 + \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \alpha_4 p_4 \tag{7}$$

where

α_0 = alternative specific constant,
p_1 to p_4 = dummy variables corresponding to Scenarios 2 to 5 in Table 2, respectively, and
α_1 to α_4 = coefficient vector corresponding to p_1 to p_4, respectively.

To better explore the effects of different types of information and other factors on commuters' stated route diversion intentions, respondents' stated preferences to the five hypothetical scenarios described in Table 2 were pooled to produce a data set of 2,455 observations.

Estimation Results for Binary Logit Model

The results of the estimation obtained by using the binary logit model are presented in Table 5. The estimated coefficient for the constant term is negative according to the results presented in Table

TABLE 5 Estimation Results for Binary Logit Model

Variable Name	Estimated Coefficient	t-Statistic
constant	-3.5599	-7.5442
dummy 1	0.7454	4.5482
dummy 2	2.1209	10.6283
dummy 3	2.6841	11.9990
dummy 4	3.3055	12.7843
time	-0.5640	-7.7865
age	0.1925	1.2133
marital status	-0.1684	-1.1779
attitude	0.5426	8.8990
perception	0.8142	14.0592
Summary Statistics:		
No. of Observations = 2455		
L(O) = -1701.7		
L(B) = -789.39		
Rho-Square = 0.536		

5. This indicates that drivers will not likely divert to alternative routes when only qualitative information (i.e., Scenario 1 in Table 2) is provided by RTRs and CMSs. Route diversion rates of drivers will increase as the information provided by RTRs and CMSs changes to quantitative information (i.e., Scenarios 2, 3, and 4 in Table 2) and the amount of information provided contains increasingly more detail, especially relating to the alternative routes. Such a conclusion is inferred on the basis of estimated coefficients of the variables Dummy 1, Dummy 2, and Dummy 3 since they correspond to Scenarios 2, 3, and 4, in Table 2, respectively. The values of the estimated coefficients for these three variables increase from Dummy 1 through Dummy 3. The variable Dummy 4 indicates prescriptive information (i.e., Scenario 5 in Table 2); it has the highest value of all dummy variables. Thus, the route diversion rate increases even further when prescriptive information is provided. All of these variables are statistically significant; thus, these variables play an important role in explaining drivers' route diversion behaviors.

The time variable indicates the total delay time a driver can tolerate before considering using an alternative route. The estimated coefficients for the time variable in Table 4 have negative values, which means that the longer the delay a driver can tolerate, the more likely he or she is not to divert to alternative routes. Such a finding is expected.

The variable age represents the ages of the respondents who participated in the phone survey. According to the results given in Table 5, it can be concluded that young drivers are more likely to divert their routes. Even though the value of the *t*-statistic for this variable is relatively low, this variable was kept in the final model because it has been recognized as an important determinant of route diversion behavior, and the finding here is consistent with the result reported by other researchers. The estimated coefficient of the variable marital status is negative; therefore, it can be inferred that single drivers are more likely to divert to alternative routes to avoid traffic congestion. Other researchers reported a similar finding on the effects of marital status on route diversion. The *t*-statistic for this variable is also relatively low.

The variables attitude and perception are the latent variables identified earlier by using the LISREL model. The variable attitude characterizes drivers' attitudes toward route diversion. According to the estimated coefficient listed in Table 5, commuters are more likely to divert their routes to avoid traffic delays and congestion when they have positive attitudes toward route diversion. The variable perception captures drivers' perceptions of the reliability of traffic information provided by RTRs and CMSs. The estimated coefficient for this variable is positive as well; thus, drivers who have good perceptions of RTRs and CMSs would more likely follow the recommendations provided. It should be noted that the *t*-statistic values reported in Table 5 for these two variables are overstated because the predicted values of the latent variables were used as explanatory variables (*17*).

Finally, it is recognized that the estimation results presented in Table 5 may be inconsistent because a sequential estimation approach instead of a joint estimation of the LISREL model and the discrete choice model was used for this model system. The sequential estimation approach results in inconsistent parameter estimates, given that the discrete choice model is a nonlinear model (*10*). Despite this potential inconsistency, the major contribution of the present study should be recognized. It identified two latent variables that are very important in determining drivers' route diversion behaviors. As shown in the estimation results of the binary logit model, the addition of latent variables provided a better understanding of drivers' decision-making process in regard to route diversion.

CONCLUSION

The effects of drivers' travel and socioeconomic characteristics, attitudes toward route diversion, perceptions of the reliability of traffic information provided, and the types of information provided by RTRs and CMSs on their willingness to divert their routes were investigated.

The results indicated that drivers who have low tolerances toward traffic delays, have positive attitudes toward route diversion, and perceive RTRs and CMSs to be reliable sources of information are more likely to divert from their usual routes in cases of traffic incidents. In addition, various types of information conveyed through RTRs and CMSs greatly influence drivers' route diversion intentions. Drivers' willingness to divert their routes increased as the amount of information provided by RTRs and CMSs became increasingly more elaborate. It was found that drivers are most likely to divert their routes when elaborate quantitative information (i.e., Scenario 4 in Table 2) or prescriptive information (i.e., Scenario 5 in Table 2) is conveyed through RTRs and CMSs. These results are consistent with findings reported by other researchers.

The findings from the present study could be used to form useful policy guidelines in developing ATISs. For example, ATISs can effectively influence drivers' route diversion decisions by providing detailed descriptions of alternative routes on the network or by conveying quantitative rather than qualitative information.

ACKNOWLEDGMENTS

The research was partly supported by Hughes Aircraft Co. through a research grant to Purdue University. The authors benefited from the comments of three anonymous reviewers.

REFERENCES

1. Heathington, K., R. Worrall, and G. Hoff. Attitudes and Behavior of Drivers Regarding Route Diversion. In *Highway Research Record 363*, HRB, National Research Council, Washington, D.C., 1971, pp. 18–23.
2. Huchingson, R. D., and C. Dudek. Delay, Time Saved, and Travel Time Information for Freeway Traffic Management. In *Transportation Research Record 722*, TRB, National Research Council, Washington, D.C., 1979, pp. 36–39.
3. Dudek, C., R. Huchingson, and R. Brackett. Studies of Highway Advisory Radio Messages for Route Diversion. In *Transportation Research Record 904*, TRB, National Research Council, Washington, D.C., 1983, pp. 4–9.
4. Mannering, F. L. Poisson Analysis of Commuter Flexibility in Changing Routes and Departure Times. *Transportation Research*, Part B, Vol. 23B, 1989, pp. 53–60.
5. Mahmassani, H. S., C. Caplice, and C. Walton. Characteristics of Urban Commuter Behavior: Switching Propensity and Use of Information. In *Transportation Research Record 1285*, TRB, National Research Council, Washington, D.C., 1990, pp. 57–69.
6. Allen, R. W., D. Ziedman, T. J. Rosenthal, A. C. Stein, J. F. Torres, and A. Halati. Laboratory Assessment of Driver Route Diversion in Response to In-Vehicle Navigation and Motorist Information Systems. In *Transportation Research Record 1306*, TRB, National Research Council, Washington, D.C., 1991, pp. 82–91.
7. Khattak, A. J., F. S. Koppelman, and J. L. Schofer. Stated Preferences for Investigating Commuters' Diversion Propensity. *Transportation*, Vol. 20, No. 2, 1993, pp. 107–127.
8. Khattak, A. J., J. L. Schofer, and F. S. Koppelman. Commuters' Enroute Diversion and Return Decisions: Analysis and Implications for Advanced Traveler Information Systems. *Transportation Research*, Part A, Vol. 27A, No. 2, 1993, pp. 101–111.
9. Khattak, A. J., A. Kanafani, and E. Le Colletter. Stated and Reported Route Diversion Behavior: Implications on the Benefits of ATIS. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.
10. Abdel-Aty, M. A., K. M. Vaughn, R. Kitamura, P. P. Jovanis, and F. L. Mannering. Models of Commuters' Information Use and Route Choice: Initial Results Based on a Southern California Commuter Route Choice Survey. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.
11. Jou, R. C., and H. S. Mahmassani. Comparability and Transferability of Commuter Behavior Characteristics Between Cities: Departure Time and Switching Decision. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.
12. Adler, J. L., T. F. Golob, and M. G. McNally. A Structural Model with Discrete Choice Variables for Predicting Enroute Behavior Under ATIS. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.
13. Ben-Akiva, M., T. Morikawa, and F. Shiroishi. Reliability of Stated Preference Techniques. *Proc., 5th World Conference on Transportation Research*, Yokohama, Japan, 1989.
14. Ben-Akiva, M., and T. Morikawa. Estimation of Travel Demand Models from Multiple Data Sources. *Proc., 11th International Symposium on Transportation and Traffic Theory*, Yokohama, Japan, 1990.
15. Wardman, M. Stated Preference Methods and Travel Demand Forecasting: An Examination of the Scale Factor Problem. *Transportation Research,* Part A, Vol. 25A, Nos. 2/3, 1991, pp. 79–89.
16. Everitt, B. S. *An Introduction to Latent Variable Models.* Monographs on Statistics and Applied Probability. Chapman & Hall, New York, 1984.
17. Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, Mass., 1985.

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

Experimental Analysis and Modeling of Advice Compliance: Results from Advanced Traveler Information System Simulation Experiments

KENNETH M. VAUGHN, RYUICHI KITAMURA, AND PAUL P. JOVANIS

Computer-based microsimulation is evolving as a useful tool for the collection of travel behavior data. Analysis of the route choice problem in particular demands sequential data to capture the behavioral dynamics involved. The use of microsimulations to collect data of this type is in its infancy, because microcomputers powerful enough for this type of simulation have only recently become available. One such simulation recently completed at the University of California at Davis resulted in a data set that will support dynamic modeling. The simulation collected 32 sequential binary route choice decisions made by 343 subjects under various experimental conditions. The experimental factors included information accuracy, feedback, provision of descriptive rationale for route advice, indication of one route alternative as a freeway, and control for stops on the side road route. An analysis of the experimental treatments used in the simulation is presented, and a dynamic probabilistic model of subjects' advice compliance is developed. A regression approach was used to estimate the factor effects of an analysis of variance model of the experimental treatments. Dynamics were introduced into the model by the development of a perception variable that, when it is incorporated, leads to the adaptive expectations model. A linearized model of relative frequencies incorporating lagged dependent variables to account for behavioral dynamics is formulated and estimated. Econometric methods of pooled cross-sectional, time-series analysis are used to estimate models that account for heteroskedasticity and autocorrelation.

The application of advanced technologies to the traffic environment is an area that will have significant impacts on individual driver behavior. If real-time, accurate information on the characteristics of the travel environment can be provided to travelers before their departure and while they are en route, will behavior be altered in such a way as to improve the individual accessibility of drivers or improve the overall characteristics of the travel environment resulting in accessibility gains for all drivers, or will the individual benefits of such systems conflict with systemwide improvement goals? To accurately model the macrolevel effects of advanced traveler information systems (ATISs) the microlevel effects of these systems on individual driver behavior must be analyzed and understood.

A thorough understanding of behavioral dynamics is critically important when behavior cannot be represented properly on the basis of cross-sectional observation. This arises when time lags exist between a change in the travel environment and a behavioral change in response. Such time lags can be caused by the lack of information; experimentation and learning; the psychological, time, and monetary costs of searching; organized behavior based on planning; perception thresholds; constraints; and apparently irrational preferences for habitual behavior. These would lead to behavioral inertia, resistance to change, and differential speeds of adjustment, which in turn may lead to asymmetric responses to change (*1*).

Dynamic decisions are decisions in context and in time. This means that the decision maker must consider the consequences of each decision for future decisions, that he or she is constrained by earlier decisions, and that he or she may sometimes be able to correct problems caused by earlier decisions in later decisions (*2*). Prediction of the magnitude and timing of behavioral responses to changes in contributing factors demands the use of sequential data. The recognition of behavioral dynamics in the decision process points to the need to observe behavioral units repeatedly over time to accurately predict and forecast behavioral response.

Brehmer (*2*) states that experiments on dynamic decision making simply cannot be made by ordinary experimental methods because of the dynamic and interactive nature of the experimental tasks and claims that experimentation in this area is impossible without interactive computer simulation, which has only recently become available in the laboratory. Computers make it possible to create dynamic simulations and to study how subjects interact with such simulations. Brehmer labels such simulations "microworlds." These microworlds simulate some of the essential features of the dynamic system being investigated and are designed to reflect three main characteristics of real-world decision problems: complexity, dynamics, and opaqueness. They are complex in that they require the subjects to consider many different elements, dynamic in some or all aspects, and opaque in that all characteristics are not automatically revealed to the subject, requiring the formation and testing of hypotheses about their state and characteristics. The use of computer simulation in the framework of a controlled laboratory experiment is the approach that was taken in the present study to analyze the dynamic nature of the route choice process.

DESCRIPTION OF SIMULATION EXPERIMENTS

An experiment to investigate drivers' learning and pretrip route choice behavior under ATISs was performed by using an interactive route choice simulation experiment carried out on a personal computer (PC) (*3*). All of the experiments subjected drivers to 32 simulated days in which they were to choose one of two possible routes. For each travel day an amount of delay was randomly assigned to each of the two routes. The units of delay assigned to a

---

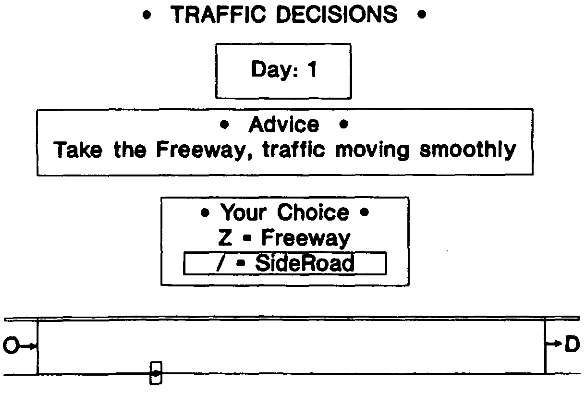
Institute of Transportation Studies, 2028 Academic Surge, University of California, Davis, Calif. 95616.

particular route are proportional to the travel time experienced on the route. The delay was distributed over the 32 trials such that the mean delay for each route was equal but the variance differed. In this manner routes with potentially faster travel times but with a greater amount of uncertainty (as one might expect on a freeway) can be compared with routes with slower travel times but with a greater amount of certainty (similar to surface street routes). The screen display of the simulator is shown in Figure 1. In Figure 1 the double-line link represents the freeway and the single-line link represents the side road. When the subject selects a route, a red blinking cursor (depicted by the small box on the side road link) moves across the screen from the origin (O) to the destination (D). The speed at which the cursor moves represents the average travel speed on that link for that travel day.

The simulation begins by presenting a set of instructions to the subject describing how the program operates. The subjects are told that they have purchased a new traffic watch device that will provide them with traffic information before they select a route. The subjects are also told that the device will not always be accurate, but they are not given any indication of its overall accuracy. Before beginning the simulation the subjects are shown examples of the fastest and slowest possible times on each of the routes, and they may repeat the examples as often as necessary to get familiar with the system. Subjects are instructed that their main task is to minimize their overall travel time by deciding when and when not to follow the advice provided by the traffic information system. Subjects are also told that their decision and response times are being measured and that they should try to respond as quickly as they can make a good decision.

The simulation was developed such that various experimental treatments could be applied and then data could be collected under these different conditions. The treatments that could be applied to the simulation included the following:

1. Accuracy. The accuracy level of the advice provided to subjects could take on values of 60, 75, or 90 percent. Accuracy as defined within this experiment means that for any given trial day *i*, the probability that the information on day *i* is correct (P_i) is equal to the accuracy level of the experiment. For example, condition 1 of experiment 1 used an accuracy level of 60 percent; thus, on any given trial day *i* P_i is equal to 0.6, or on average, over the 32 trial days, subjects experienced 19 trials in which correct information was provided and 13 days in which incorrect information was provided (subjects were not aware of the level of accuracy assigned in the experiment).


• TRAFFIC DECISIONS •

Day: 1

• Advice •
Take the Freeway, traffic moving smoothly

• Your Choice •
Z • Freeway
/ • SideRoad

FIGURE 1 Screen display of simulator.

2. Stops. A simulated stop on the side road route could be applied.
3. Rationale. A justification statement as to why the subject should follow the advice could be provided.
4. Feedback. Feedback could be provided at the end of each trial in the form of actual simulated travel times on the two routes for that trial.
5. Freeway. An identification of the routes as freeway and side road as opposed to simply Routes A and B.
6. Road. The display could provide the simulated origin and destination with the two route links as shown in Figure 1 or with no network display provided, and the travel time could be simulated by a blinking cursor located in the center of the screen.

Three separate experiments were carried out to collect data under various conditions. The three experiments and the conditions under which the simulation has been run to date are discussed elsewhere (*3*). The first experiment was used to investigate the accuracy requirements of ATISs. The experiment was structured as described earlier, but with three levels of information accuracy provided. Three separate groups of 23, 25, and 29 subjects were run through the simulation at three levels of accuracy: 60, 75, and 90 percent, respectively. The analysis and results of this first experiment can be found elsewhere (*3*). In the second and third experiments the information accuracy was held constant at 75 percent and other experimental conditions were varied. This paper provides an analysis of the data collected in the second and third experiments. The second and third experiments resulted in data from 266 subjects giving 8,512 individual choices. The computer program automatically recorded and stored data from each subject for 32 sequential trials. Test subjects were all undergraduate students in the Psychology Department at the University of California at Davis. Although the representativeness of students may be questionable, if the goal is to investigate how humans perceive information, learn, and adapt over time, then this pool of subjects may be as representative as any other within their homogeneous age group. Although some may question the "humanness" of undergraduate students with respect to basic human behavioral characteristics, the sample should be representative; however, it will be biased by age. With regard to other characteristics, 42 percent of the sample was male, which is underrepresentative of the general population. To measure the level of driving experience in the sample, all subjects were asked to rate their driving frequency. One hundred eleven subjects indicated that they either currently commute or do not now but formally commuted, 98 subjects indicated that they do not commute but drive frequently, and 57 subjects indicated that they either drive infrequently or do not drive.

ANALYSIS OF EXPERIMENTAL TREATMENTS BY ANOVA

In this section the effects of the various treatment combinations are analyzed by analysis of variance (ANOVA) techniques. ANOVA models are used for studying the relation between a dependent variable and one or more independent variables for experimental and observational data. The strength of the ANOVA model, and the main reason that it is applied here, is that it does not require the investigator to make assumptions about the nature of the statistical relation (except for covariate terms), nor does it require the independent variables to be quantitative (*4*). The first treatment, accuracy, has three levels, and the effects of this treatment have been documented previously (*3*). Generally, increasing information accuracy results in increased advice compliance, and subjects tend to quickly identify the level of accuracy of the system and accept advice at a rate equivalent to the level of accuracy.

The significant findings from the first data set (*3*) are summarized as follows:

1. Acceptance of advice increases with increasing information accuracy.
2. Males are more willing to accept advice than females and also make their decisions faster than females.
3. Experienced drivers are not as willing as less experienced drivers to accept advice, and they also make their decisions faster.
4. A "freeway bias" exists with subjects more willing to accept freeway advice.
5. Although males are more willing to accept advice, they are also less likely to purchase an information system.
6. Although less experienced drivers are more likely to follow advice, they are also less likely to purchase an information system.

In a parallel project that is investigating commuters' acquisition and use of traffic information and its effect on route choice decisions using survey methods (*5,6*), it was found that females were more willing to accept pretrip traffic information than males, which contradicts findings provided elsewhere (*3*). A study by Mannering et al. (*7*) of the influence of traffic information on route, mode, and departure time choice, based on travel surveys, also found that males were less likely to be affected by pretrip traffic reports for all three decisions. These contradictory findings spurred further investigation into the effects of gender. In the data presented elsewhere (*3*) it was found that at information accuracy levels of 65 and 90 percent male acceptance of information was significantly higher than female acceptance, whereas at a 75 percent level of accuracy no significant difference in information acceptance was found between males and females. The net effect that was reported was that males had higher acceptance rates than females, but these results were not independent of accuracy level. In the data from the second and third experiments, which have much larger sample sizes and all observations are at a 75 percent accuracy level, it is found that females do accept pretrip information more than males, which agrees with the similar findings from the surveys.

The analysis by Vaughn et al. (*3*) focused on the effects of accuracy on subjects' agreement with advice, the decision time required to make a route selection, and the subjects' willingness to purchase a system on the basis of their experiences with the simulation. The analysis performed here focuses on these same dependent variables but addresses the effects of four of the remaining experimental treatments as defined earlier: feedback, rationale, freeway, and stops. ANOVA models were estimated for each of the three dependent variables, and the final models that retained the significant main effects and interactions are presented in Table 1. The model of subjects' agreement with advice indicates that the effects of feedback and rationale are individually significant and that the interaction effects of stops and rationale are strongly significant, whereas the individual effects of stops and freeway are only moderately significant. In the model of subjects' decision time, feedback, rationale, and stops are all individually significant, and again, the interaction effects of stops and rationale are strongly significant. In the model of subjects' willingness to purchase an information system, all of the main and interaction effects in the model are strongly significant.

ANOVA Regression Models

The constant vectors of the ANOVA factor effects model give an indication of the effects of the within-factor levels of the grouping variables on the dependent variable. These factor-level constants can be estimated by a regression approach equivalent to the ANOVA model.

The factor effects model can be represented in the matrix form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathrm{x}\tau + \epsilon \qquad (1)$$

where

$\mathbf{X}$ = a matrix of dummy variables that represent the levels of the treatments,
x = a matrix of covariates,
$\boldsymbol{\beta}$ and τ = coefficient vectors, and
ϵ = error vector.

When a background variable (covariate) is strongly related to the dependent variable, an analysis of covariance may increase the precision of comparisons between treatments by reducing the within-group variability in the dependent variable due to the influence of the covariate. Since all treatments in this analysis have only two levels, the ANOVA regression model has the same number of parameters as the parent ANOVA model has effects. For a detailed explanation of the ANOVA regression procedure see Neter et al. (*4*).

The objective is therefore to test appropriate hypotheses about the treatment effects and to estimate these effects. For hypothesis testing the model errors are assumed to be normally and independently distributed with zero mean and constant variance σ^2. Additional assumptions of the covariate model are that within each group the dependent variable has a linear relationship with the covariate and that the slope of the regression for each covariate is the same in each group.

Because of the sequential nature of these data the method of ordinary least squares (OLS) for estimating the regression is not appropriate. In this data set $\mathbf{Y}$ is a cross-sectional time-dependent vector of observed behavior with elements $Y_i(t)$, which is the observation for individual i at time period t, and i is equal to 1 to N and t is equal to 1 to T, where N is the number of subjects (266) and T is the number of sequential, equally spaced observations on each subject (n = 32). ϵ is a cross-sectional time-dependent vector of errors with elements $\epsilon_i(t)$. x is an $NT \times k$ matrix of k, $NT \times 1$ vectors of explanatory variables whose elements may be either cross sectional (gender) or time dependent (traffic information on day t), or both (perception variables). $\boldsymbol{\beta}$ is a $k \times 1$ parameter vector that is assumed to be the same for all i. Kmenta (*8*) presents a framework for estimating models with cross-sectional, time-series data. The assumptions of the Kmenta model allow for cross-sectional heteroskedasticity and assume cross-sectional independence and first-order autocorrelation AR(1). Estimation of this model can be performed by using feasible generalized least squares (FGLS) and maximum likelihood methods (*8,9*).

TABLE 1 ANOVA Table

Dependent →	Agreement[a]	Decision Time[b]	Purchase[c]
Independent ↓	**F-stat[d]**	**F-stat[d]**	**F-stat[d]**
Feedback	31.01	17.63	596.09
Rationale	13.19	22.53	147.89
Freeway	1.07	0.22	12.62
Stops	1.76	41.00	87.90
Stops/Rationale interaction	13.11	47.10	448.22

[a] Agree (1=yes,0=no)
[b] Decision time is time in seconds
[c] Purchase (1=extremely likely to buy, to 7=would not buy)
[d] F critical = 3.84 (α=0.05,1,∞)

The factor effects model with ANOVA regression was estimated with a pooled, cross-sectional, time-series model by the FGLS method with an econometrics computer package (*9*) for the dependent variables agreement and decision time. Tests for heteroskedasticity were performed on the pooled model for these dependent variables, and significant violation of the homoskedastic assumption was indicated. The pooled, cross-sectional, time-series model corrects for cross-sectional heteroskedasticity, and log likelihood ratio tests indicated significant improvement in log likelihood values for the pooled, cross-sectional, time-series models. The dependent variable purchase is purely cross-sectional, and this model was estimated by OLS procedures. In addition to the main and interaction effects of the ANOVA models, the covariates included in the models were the trial number, a dummy variable indicating side road advice, a male dummy variable, a driving experience variable, and a perception variable. The results of the model estimation are presented in Table 2.

The first two models presented in Table 2 are static models of behavior; observed behavior at time t is only dependent on contributing factors also observed at time t. The assumption underlying this model is that behavior (**Y**) changes immediately in response to a change in contributing factors (**X**) and that behavior does not depend on past history. The second two models presented an attempt to capture the behavioral dynamics by including a time-

TABLE 2 ANOVA Regression Models

Dependent →	Agreement[a]	Decision time[a]	Agreement[b]	Decision time[b]	Purchase[c]
Independent ↓	**Coeff.(t)**	**Coeff.(t)**	**Coeff.(t)**	**Coeff.(t)**	**Coeff.(t)**
Intercept	.7446(50.23)	3.8124(55.94)	-.0991(-1.04)	4.3959(62.08)	4.0524(100.1)
Trial number	.0031(6.39)	-.0741(-34.01)	-	-	-
Side road advice	-.0607(-10.07)	.3119(19.51)	-.0624(-10.27)	.3172(19.64)	-
Sex(M=1,F=0)	-.0394(-4.10)	-.2907(-6.53)	-.0399(-4.18)	-.2805(-6.61)	-.1131(-3.61)
Driving experience	.0043(0.72)	.0464(1.75)	.0069(1.18)	.0484(1.90)	.0198(1.00)
Feedback	.0217(4.45)	.0420(1.82)	.0223(4.58)	.0488(2.22)	-.3922(-24.67)
Rationale	.0243(3.53)	.1536(5.01)	.0274(3.98)	.1229(4.17)	-.2785(-12.14)
Freeway	-.0060(-0.87)	.0074(0.25)	-.0091(-1.32)	.0373(1.33)	.0838(3.67)
Stops	-.0108(-2.20)	-.1305(-5.63)	-.0087(-1.78)	-.1337(-5.95)	-.1514(-9.47)
Stops/Rationale interaction	.0188(3.58)	-.1691(-6.90)	.0213(4.08)	-.1838(-7.72)	-.3660(-21.41)
Perceived accuracy of information	-	-	1.3034(9.52) λ=.97,ic=.65	-3.4078(-41.27) λ=.89,ic=0	-
Log likelihood	-4100.98	-13436.40	-4081.31	-13303.80	-14839.0
[d]Buse R^2	.777	.674	.780	.698	-
R^2-adj	-	-	-	-	.144

[a] Static, pooled cross-sectional, time-series model (FGLS)
[b] Adaptive expectations, pooled cross-sectional, time-series model (instrumented FGLS)
[c] Static, cross-sectional model (OLS)
[d] See reference (*10*)

varying covariate that is formulated as the subjects' perceived (or expected) accuracy of information. This formulation leads to a model known as the *adaptive expectations model* (*8*).

The first model of subjects' agreement with advice indicates that having feedback and providing a rationale for the advice significantly increases the level of agreement with advice. The effects of having one route labeled freeway and having stops on the side road result in a decreased level of agreement. This result is due to subjects being less willing to accept side road advice when one route is labeled a freeway or when the side road route has stops. The interaction of stops on the side road and rationale significantly increases the level of agreement. This result is due to subjects' overcoming the aversion to the side road route with stops when given some rationale as to why they should take that route (e.g., there is a problem on the freeway route today). The covariate terms indicate that less experienced drivers are more likely to follow advice, whereas male subjects are less likely to follow advice, and side road advice results in a reduction in the level of agreement. These results are supported by similar findings from a previous analysis (*3*), previous survey efforts (*5,6*), and other sources (*7*).

The inclusion of a time train variable (the trial number) is an attempt to capture the dynamics of learning in a static framework. The coefficient on this variable gives an indication of the effect of time on the behavior. The coefficient indicates that subjects are increasing their level of advice compliance over time, presumably because they are learning that the information provided is accurate enough to improve their performance in route selection.

To introduce a dynamic relationship into the model it is assumed that the behavior is history dependent; that is, behavior at time t is assumed to be a function of contributing variables measured at time $t - s$ through time t. This model may be formulated as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^1\tau + \ldots + \mathbf{Z}^m\mu + \epsilon \qquad (2)$$

where

$\mathbf{Y}$ = a cross-sectional, time-dependent, $NT \times 1$ vector of observed behavior with elements $Y_i(t)$, which is the observation for individual i at time period t, i equals 1 to N and t equals 1 to T, where N is the number of subjects and T is the number of sequential, equally spaced observations on each subject;
ϵ = an cross-sectional, time-dependent vector of errors, with elements $\epsilon_i(t)$;
$\mathbf{X}$ = an $NT \times k$ matrix of k, $NT \times 1$ vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ of explanatory variables with elements $x_{ji}(t)$, where j = 1 to k_j;
$\boldsymbol{\beta}$ = $k \times 1$ parameter vector that is assumed to be the same for all i_j;
$\mathbf{Z}^m$ = NT $\times$ l matrices of l, $NT \times 1$ vectors $(z_l, \mathbf{z}_2, \ldots, \mathbf{z}_l)$ of explanatory variables with elements $z_{ji}(t - m)$, where $j = 1$ to l, $m = 1$ to s_j, and
τ and μ = $l \times 1$ parameter vectors that are assumed to be the same for all i.

In this formulation, each $\mathbf{Z}^m$ matrix is a subset of the exogenous variables in X, lagged one time period from $t - 1$ to $t - s$, where $t = s + 1, \ldots, T$. $\mathbf{Z}^m$ could also include all exogenous variables in X, in which case, l is equal to k.

Frequently, some restrictions are placed on the regression coefficients so that the number of the regression parameters is reduced (*8*). The most common is the restriction that the parameters be declining in a geometric progression resulting in a geometric lag distribution and what is known as the *adaptive expectations model.* Suppose that the dependent variable is modeled not as a linear function of some explanatory variable in the same time period but as a function of the expected or perceived value of the variable. Most travel decisions exemplify this relation because the decisions are dependent on the expected characteristics of the travel environment whose true value is not revealed until after the decision has been made. In the model estimated here it is assumed that a subject has some expectation or perception as to whether the information that they receive is accurate and is a function of past experiences.

This model can be formulated as

$$Y_i(t) = \alpha + \beta X_i^*(t + 1) + \epsilon_i(t) \qquad (3)$$

where

$$X_i^*(t + 1) = (1 - \lambda)X_i(t) + \lambda X_i^*(t) \qquad (4)$$

and $0 \leq \lambda < 1$. This formulation assumes that the expected or perceived value of X at time $t + 1$ is a weighted average of the current value of X and the expected value of X in the current time period. This formulation is based on the idea that the current expectations are derived by modifying previous expectations in light of the new information from the current experience (i.e., dynamic updating) (*8*). This weighted average updating function was also used to update perception variables in earlier work (*3,11*).

Equation 4 can also be expressed as

$$\begin{aligned} X_i^*(t + 1) = (1 - \lambda)[&X_i(t) + \lambda X_i(t - 1) \\ &+ \lambda^2 X_i(t - 1) + \ldots.] \end{aligned} \qquad (5)$$

which results in the following geometric lag distribution when substituted into Equation 3:

$$\begin{aligned} Y_i(t) = \alpha + \beta(1 - \lambda)[&X_i(t) + \lambda X_i(t - 1) \\ &+ \lambda^2 X_i(t - 1) + \ldots.] + \epsilon_i(t) \end{aligned} \qquad (6)$$

In this form the effect of X on Y extends indefinitely into the past, but the coefficients are declining geometrically such that the distant values of X become negligible. The magnitude of the parameter λ then becomes a relative measure of an individual's learning. A large value of λ indicates strong memory effects, with experiences in the past significantly contributing to the current expectation, whereas small values indicate that the effects of past experiences have very little or no effect on current expectations.

This model can be estimated by assuming values for λ and the initial conditions for $X_i^*(t)$ and then using Equation 4 as an instrumented variable in Equation 3. This instrumented variable method was used in a pooled cross-sectional, time-series model to estimate iteratively across values for λ and the initial conditions. The convergence criteria used for the iteration was to maximize the log likelihood value of the estimate. The third and fourth models presented in Table 2 are adaptive expectations models where the perceived accuracy of the information variable has been substituted into the previous models for the time train variable.

In the second model of subject agreement in Table 2, all of the static variables maintain their previous interpretations, which is desired. The perceived accuracy of information is shown to strongly influence the level of agreement with advice and indicates that as the perception of the accuracy increases, so does the level of agreement. The log likelihood value in this model shows significant improvement over the static model. The value of λ was found to be 0.97, indicating very strong memory effects in updating the perception of accuracy. Iida et al. (*12*), who conducted similar simulation experiments, have also found strong memory effects in the updating of subjects' expected travel time on a route.

The static model of subjects' decision times indicates that feedback and rationale significantly increase the amount of time that subjects spend making a decision, whereas stops on the side road and the stops-rationale interactions decrease the time spent making a decision. The covariates in this model indicate that more experienced drivers and male subjects make their decisions faster, whereas advice to take the side road increases the decision time. The time train covariate indicates the improvement in decision making that occurs over time, with subjects spending less time making a decision over time. In the dynamic model of subjects' decision time, there is a significant improvement in the log likelihood value over that in the static model. The perceived accuracy of information is shown to be strongly significant and indicates that an increase in the perception of accuracy leads to reduced decision time. The value of λ again suggests the significance of past experiences in current choice behavior.

The last model in Table 2 is a static cross-sectional model of subjects' stated willingness to purchase an information system. In this model all of the treatment main effects and the stops-rationale interaction are shown to be significant. Feedback, rationale, stops, and the interaction term all increase the likelihood of purchasing a system, whereas the freeway treatment leads to lowering the likelihood of a purchase. This indicates that increasing the level of information available in the system increases subjects' willingness to purchase a system. The negative effect of the freeway treatment may be due to the strong freeway bias that is exhibited. If subjects have a strong bias for freeway use and it is in their choice set of route alternatives, then the attractiveness of a system that will be advising them away from their preferred route may be lessened. The covariates in this model indicate that more experienced drivers and male subjects were less likely to purchase an information system, which supports previous results (*3*).

DYNAMIC MODEL OF ADVICE COMPLIANCE

Modeling Relative Frequencies

In this section the estimation of a binary choice model that uses choice frequency data that contain a nonnegligible number of limit cases is presented. The model presented here has a binary choice probability formulation that uses a logistic function. Relative frequencies for advice compliance are ratios of the number of agreements with advice over the total number of trips made. Modeling of relative frequencies has been performed by using linear or linearized models (models that by simple mathematical manipulations can be converted into a linear form), nonlinear methods, and limited dependent variable methods (*13*).

When an individual decision maker faces two or more alternatives and a single observation is available for each instance of a set of explanatory variables, a probabilistic model of the choice of alternative can be formulated on the basis of random utility theory. This formulation leads to the classical utility maximization form of the logit or multinomial logit model when the disturbances are assumed to be independent and identically distributed random variables with Weibull distributions. When an individual faces two alternatives and repeated observations are available for each instance of a set of explanatory variables, then sample proportions can be used as estimates of the probabilities of choosing either alternative. A linear probability model of relative choice frequencies, transformed to approximate a logit model, allows for efficient parameter estimation by generalized least-squares methods (*14*). The first use of this method is generally attributed to Berkson (*15*).

Linear and Linearized Models

Assume that an individual choice maker i makes n_i choices between two alternatives and let y_i be the number of times that alternative 1 is chosen and let $y_i - n_i$ be the number of times that alternative 2 is chosen. A linear probability model of the sample proportions, $p_i = y_i/n_i$, can be formulated as a linear combination of explanatory variables as

$$\mathbf{p} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{7}$$

Unbiased, consistent estimates of the model parameters can be obtained by generalized least-squares procedures. A significant problem with the linear model is that the predicted proportions do not necessarily lie between zero and 1.

The true proportions P_i are related to the sample proportions p_i by $p_i = P_i + e_i$. The logit transformation is obtained by assuming that the true proportions are related to the set of explanatory variables by a logistic function, giving

$$P_i = 1/[1 + \exp(-X_i'\beta)] \tag{8}$$

It can be shown that the odds ratio $P_i/(1 - P_i)$ is simply $\exp(X_i'B)$, and Equation 8 can be written as the logarithm of the odds ratio (log-odds):

$$\ln[P_i/(1 - P_i)] = X_i'\beta \tag{9}$$

This equation can be rewritten in terms of the empirical probability (*14*) as

$$\ln[p_i/(1 - p_i)] = X_i'\beta + u_i \tag{10}$$

where u_i is equal to $e_i/[P_i(1 - P_i)]$. Note that the empirical log-odds, $\ln[p_i/(1 - p_i)]$, can be expressed as $\ln[y_i/m_i]$, where $m_i = n_i - y_i$ (*13*). This is called the *logit*.

The error term u_i in Equation 10 can be shown to have $E[u_i]$ equal to 0, and its variance can be approximated as (*14*).

$$\mathrm{Var}(u_i) = 1/[n_i p_i(1 - p_i)] \tag{11}$$

These equations can be used to estimate the parameter vector in the logit. Note that the variance of u_i is heteroskedastic and may be serially correlated as well, depending on the nature of probability P_i.

One problem arises when the observed frequencies, y_i or m_i, are equal to zero. In this case the log-odds, and thus the empirical probabilities, are undefined (*13*). Haldane (*16*) devised a method to circumvent this problem in which the value ½ is added to all the frequencies. The dependent variable for this type of methods can be written as

$$Y_i = \ln[(y_i + \delta)/(m_i + \delta)] = X_i'\beta + u_i \qquad (12)$$

where δ is a positive constant ($\delta = {}^1/_2$ by Haldane's method). The logit, Y_i, is now defined for all cases and can be used as the dependent variable in the estimation.

Results

A model of advice compliance was estimated by using the linearized modeling framework and Haldane's method discussed in the previous section and is presented in Table 3. For each individual, the 32 trials were broken into four blocks of 8 trials each. The empirical probability for individual i in trial block j therefore is $P_{ij} = k_{ij}/t_{ij}$, where i is equal to 1 to 266, j is equal to 1 to 4, and t_{ij} is equal to 8 because all subjects made the same number of choices in all blocks. The dependent variable in the model is the logit of advice compliance and is defined as $\ln[(k_{ij} + {}^1/_2)/(m_{ij} + {}^1/_2)]$, where k_{ij} is the number of agreements with advice in trial block j, and m_{ij} is the number of disagreements with advice. By grouping the trails into blocks, the dependent variable becomes a cross-sectional, time-dependent vector of observations, and Kmenta's method (*8*) of estimating models with cross-sectional, time-series data accounting for first-order autocorrelation and heteroskedasticity is the method used to estimate this model.

The cross-sectional independent variables in the model included a male gender dummy, and the significant main and interaction effects of the experimental treatments that were identified in the ANOVA model, including feedback, rationale, freeway, stops, and stops-rationale interactions. The time-dependent independent variables include three trial block dummies, the number of times that correct information was provided in the current block, the number of times that incorrect information was provided in the previous block, and the number of times that a correct choice was made in the previous block.

The effects of the experimental treatments in this model are very similar to those discussed earlier for the ANOVA models. Providing feedback and an indication that one of the routes was a freeway significantly increases the probability that subjects will follow the system's advice, whereas having stops on the side road significantly reduces the probability that the subjects will comply. The effect of providing a decision rationale to the subjects has the opposite sign in this model compared with that in the ANOVA model,

TABLE 3 Linearized Probability Model of Advice Compliance[a]

Dependent →	**In (Ratio$_t$)**[b]
Independent ↓	**Coeff.(t)**
Feedback provided[c]	.2238(16.71)
Decision Rationale provided[c]	-.0222(-1.04)
Freeway route indicated[c]	.1817(8.55)
Stops required on side road[c]	-.2376(-16.92)
Stops & Rationale interaction[c]	.1031(7.67)
Trial block number two[d]	.4228(5.94)
Trial block number three[d]	.5015(6.63)
Trial block number four[d]	.6222(7.92)
Number of correct advice in current block	.0436(5.02)
Number of bad advice in previous block	-.0901(-8.52)
Male gender[d]	-.0733(-3.34)
Number of correct choices in the previous block	-.0367(-3.49)
Constant	.7218(11.60)
Log likelihood	-1009.9
Buse R^2	.967
Prediction Rate	75.1%

Agree with Advice	0	1
Actual	2110	6402
Correctly Predicted	9	6381
Market Share	2217.8	6294.2

[a] Dynamic, pooled cross-sectional, time-series model (FGLS)
[b] Ratio$_t$ = (number of agreements in block t)+½ / (number of disagreements)+½
[c] 1/-1 dummy variable
[d] 1/0 dummy variable

but interpretation of this coefficient is not reliable here because of the low significance level. The interaction effects of stops and rationale significantly increase the probability of compliance, which also agrees with the ANOVA results. The coefficient on the gender variable again indicates that males are less likely to comply with advice.

The dynamic variables in the model give some indication of how subjects are learning and modifying their behaviors over time. The trial block dummy variables are all strongly significant and indicate that as subjects progress through the trials their compliance behaviors change. The greatest change in probability of compliance occurs between the first and second trial blocks, with the probability of compliance continuing to increase from the second through the fourth trial block. This finding supports earlier results indicating that subjects could quickly identify the level of accuracy of the information. Here a significant increase in advice compliance from the first to second trial block can be seen, indicating that subjects have learned that the accuracy of the information provided is at such a level that compliance with advice improves their travel choices. The coefficient on the number of times that correct advice was provided in the current block indicates that even within a block subjects could recognize improved accuracy, and as accuracy improves compliance improves.

As was shown in the ANOVA regression model, the agreement behavior is dependent on the accumulation of past experiences. These effects are captured in this model through the use of lagged variables. The effect of increased levels of bad advice in the previous trial block negatively affects compliance in the current block. The coefficient on the correct number of choices in the previous block is also significant and has a negative effect, which is not intuitive. One would think that as the number of correct choices in the past increases compliance would increase. One must also realize that a correct choice as defined here includes choosing the correct route regardless of advice. If a subject chooses the correct path on a day that it is not the advised path, this could have a negative impact on advice compliance.

CONCLUSION

Analysis of the experimental treatments used in the simulation found that providing subjects with feedback significantly increased the level of agreement with system advice and that subjects were much more willing to purchase systems that provided feedback. Providing subjects with descriptive information in the form of rationale with the route advice also significantly increased agreement level and willingness to purchase a system. A negative bias toward routes with stops is indicated, with subjects less likely to comply with advice to take a route with stops. This finding is interesting in that even with the stops on the side road, over all 32 trials there is no advantage of the freeway over the side road because the mean travel times for both routes across the trials were equal. This indicates the different values that subjects place on congestion delay versus stopped delay. The significance of the interaction between stops and rationale indicates that the provision of the descriptive information had the effect of reducing the negative bias toward routes with stops and increasing the probability that subjects would accept advice to take routes with stops.

The findings of the effects of gender clarify previous results and agree with real-world results from travel surveys (*3,5–7*). It was found that males were less likely to comply with pretrip route advice. These types of comparisons with real-world studies will help to validate the results of simulations.

The dynamic nature of the decision process was modeled by using an adaptive expectations model and a linearized probability model of discrete choice. The adaptive expectations model was used to model the effects of subjects' perception of the accuracy of information. This framework assumes that for each trial the subject has some perception or expectation of accuracy of the information provided and that this perception is updated over time on the basis of experiences. The results indicate the significance of this perception variable and also indicate that subjects have a strong memory effect, such that current perceptions rely more heavily on the accumulation of past experience as opposed to just recent past experience. A dynamic probabilistic model of advice compliance accounting for heteroskedasticity and autocorrelation was estimated by using a linearized relative frequencies model. The model provided similar results for effects of the experimental treatments and also indicates the significance of dynamic effects in complying with the system's advice.

The use of computer simulation has been found to provide a useful tool for the collection of sequential route choice data in the presence of ATISs. The simulation is simple and structured around a binary choice set (freeway versus side road), yet the results obtained indicate the significance of the behavioral dynamics involved in the decision process. Recognition of the simplicity involved in the simulation as well as the limited choice set and information available has lead to the development of a more complex PC-based simulator. This new simulator operates with a more complex traffic network and incorporates interconnected freeway and arterial and surface street link types to create a large choice set of alternatives from the origin to the destination. The type and level of information that can be provided to subjects have also been expanded to include pretrip route guidance, en route guidance, incident information, and congestion levels. A new set of simulation experiments that uses this expanded simulation software has recently been completed. In those experiments 100 regular commuters from the Sacramento, California, area were recruited, and each commuter completed 20 sequential trials with the simulator. This new set of experiments has provided a rich data source that is under analysis (*17*).

ACKNOWLEDGMENTS

This research was funded by the California Department of Transportation and the Partners for Advanced Transit and Highways.

REFERENCES

1. Kitamura, R. Panel Analysis in Transportation Planning: An Overview. *Transportation Research*, Vol. 24A, 1990, pp. 399–415.
2. Brehmer, B. Dynamic Decision Making: Human Control of Complex System. *Acta Psychologica, International Journal of Psychonomics*, Vol. 81, 1992, pp. 211–241.
3. Vaughn, K. M., et al. Experimental Analysis and Modeling of Sequential Route Choice Behavior under ATIS in a Simplistic Traffic Network. In *Transportation Research Record 1408*, Transportation Research Board, National Research Council, Washington, D.C., 1993.
4. Neter, J., W. Wasserman, and M. H. Kutner. *Applied Linear Statistical Models*. Irwin, Homewood, Ill., 1990.
5. Abdel-Aty, M. A., et al. *Survey of Route Choice Behavior: Empirical Results from Southern California and Their Implications for ATIS.*

Research Report UCD-ITS-RR-93-12. Institute of Transportation Studies, University of California at Davis, 1993.

6. Abdel-Aty, M. A., et al. *Understanding Commuters' Attitudes, Uncertainties, and Decision-Making and Their Implications for Route Choice*. Research Report UCD-ITS-RR-94-5. Institute of Transportation Studies, University of California at Davis, 1994.
7. Mannering, F., et al. Statistical Analysis of Commuters' Route, Mode and Departure Time Flexibility and the Influence of Traffic Information. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.
8. Kmenta, J. *Elements of Econometrics*, 2nd ed. Macmillian Publishing Company, New York, 1986.
9. *SHAZAM Econometrics Computer Program, User's Reference Manual v7.0.* McGraw-Hill Book Company, New York, 1993.
10. Buse, A. Goodness of Fit in Generalized Least Squares Estimation. *The American Statistician*, Vol. 27, No. 3, 1973, pp. 106–108.
11. Yang H., R. Kitamura, P. P. Jovanis, K. M. Vaughn, and M. A. Abdel-Aty. Exploration of Route Choice Behavior with Advanced Traveler Information using Neural Network Concepts. *Transportation*, 1993.
12. Iida, Y., et al. Experimental Analysis of Dynamic Route Choice Behavior. *Transportation Research*, Vol. 26B, 1992, pp. 17–32.
13. Goulias, K. G., and R. Kitamura. Analysis of Binary Choice Frequencies with Limit Cases: Comparison of Alternative Estimation Methods and Application to Weekly Household Mode Choice. *Transportation Research*, Vol. 27A, No. 1, 1993.
14. Judge, G. G., et al. *The Theory and Practice of Econometrics*. John Wiley and Sons, New York, 1980.
15. Berkson, J. Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, Vol. 39, 1944, pp. 357–365.
16. Haldane, J. B. S. The Estimation and Significance of the Logarithm of a Ratio of Frequencies. In *Annals of Human Genetics*, Vol. 20, Part 4, 1956, pp. 309–311.
17. Vaughn, K. M., et al. *Information Use and Learning with ATIS: Analysis of Network Simulation Experiments*. Research Report UCD-ITS-RR-94-10. Institute of Transportation Studies, University of California at Davis, 1994.

The views expressed in this paper are those of the authors and do not necessarily represent the views of the funding agencies.

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

Framework for Assessing Benefits of Highway Traveler Information Services

LAZAR N. SPASOVIC, MARIA P. BOILE, AND ATHANASSIOS K. BLADIKAS

The problem of estimating traveler benefits from an information system that is capable of forecasting traffic conditions on the roads of a network with a variable degree of accuracy is considered. Unlike the cases that have already appeared in the literature, in which furnishing actual times eliminates all uncertainty out of drivers' behavior, here the uncertainty about the occurrence of a particular traffic condition may simply be reduced instead of being completely eliminated. The primary purpose is to propose a methodological framework that can evaluate the benefits from the introduction of information services in highway networks. Given certain behavioral aspects of the traveler decision-making process, the focus is on quantifying what travelers gain by having perfect or partial information about highway traffic conditions. The secondary purpose is to use the proposed methodology to determine the optimal number of travelers to whom information on travel conditions should be provided.

Intelligent transportation systems (ITSs), which encompass advanced surveillance, communication, control, and computing systems and engineering management methods, are envisioned to be able to increase safety, reduce congestion, and improve the productivity of transportation systems. Within ITS advanced traveler information systems (ATISs) are envisioned to provide travelers with information on the status of highways either before they depart from their homes and workplaces or enroute so that they can make informed route choices and minimize their travel times. The information given to a traveler may involve transmission of observed travel times or forecasted future traffic conditions in the network so that congested areas can be avoided. Embryonic systems of traffic information dissemination and traveler guidance, such as Metro Traffic and Shadow Traffic, are already operational in metropolitan areas. The impact of traffic information on travelers' behavior, and subsequently on transportation network performance, is not obvious and must be researched.

Most research on the effectiveness of information systems has been undertaken in the area of evaluating the benefits from correcting drivers' perception (or misperception) of actual road link travel times (*1*). It assumes that drivers choose their routes on the basis of perceived travel times. As drivers assign themselves over the network an equilibrium point is reached when drivers cannot further decrease their perceived travel times by unilaterally changing routes. This assignment is deemed inefficient because drivers may be choosing inefficient (high-travel-time) routes since they are not aware of the conditions on all available alternatives. Most of the papers reviewed previously (*1*) conclude that substantial savings might be achieved if information on actual travel times on links is given to drivers so that an equilibrium traffic assignment based on actual rather than perceived travel times can be reached.

THE PROBLEM

The problem considered in this paper is that of estimating traveler benefits from an information system that is capable of forecasting traffic conditions on the roads of a network with a variable degree of accuracy. Unlike the cases that have already appeared in the literature, in which furnishing actual times eliminates all uncertainty out of drivers' behavior (perception of travel times), here the uncertainty about the occurrence of a particular traffic condition may simply be reduced instead of completely eliminated.

The primary purpose of this paper is to propose a methodological framework that can evaluate benefits from the introduction of information services in highway networks. Given certain behavioral and attitudinal aspects of the traveler decision-making process, the paper focuses on quantifying what travelers gain by having perfect or partial information on highway traffic conditions.

The secondary purpose of the paper is to use the proposed methodology to evaluate the value of information to address the impacts of market penetration of information services on travelers' behavior and network performance. Primarily, the methodology is used to determine the optimal number of travelers to whom advanced information on travel conditions should be provided.

It is expected that the cost of ITS technologies will substantially decrease with the mass production of devices such as transponders that can receive and send traffic information. This low cost could make ITS technologies widely available to travelers. In the beginning travelers who are using ITSs and services will benefit greatly because they will be able to take advantage of the real-time (or near-real-time) information on traffic conditions obtained via these systems. However, as the number of ITS users increases, alternate routes may also become congested, diminishing the benefits of information that may eventually disappear.

It is incorrect to view ITS as a panacea for transportation ills and to assume that when all travelers are given access to the same network information they all will be better off than they were when they had only historical information or no information at all. In a congested network a change in the traveler assignment pattern caused by providing information may substantially change the total network travel times. In turn this can increase an individual traveler's average travel time compared with that in the situation when he or she chooses routes on the basis of historical information or no information. There is a threshold point at which giving information to an additional traveler will make him or her worse off. This thresh-

---

L. N. Spasovic, School of Industrial Management and National Center for Transportation and Industrial Productivity, New Jersey Institute of Technology, Newark, N.J. 07102. M. P. Boile, Department of Civil and Environmental Engineering, Lafayette College, Easton, Pa. 18042–1775. A. K. Bladikas, Department of Industrial and Manufacturing Engineering and National Center for Transportation and Industrial Productivity, New Jersey Institute of Technology, Newark, N.J. 07102.

old point may designate the optimal market penetration, defined as the number of users who will benefit when given information. Beyond this point having the information is disadvantageous for travelers. Moreover, travelers with no information may do better than those with information.

The information dissemination impacts are analyzed first for a simple network and then for a more complex network in which two traveler services provide information to subscribers in either a cooperative or a noncooperative manner. The impact of introducing travel information on network performance is not well understood and is far from obvious. The understanding of the value of information concept and the ways to estimate the gains from using an information service are critical for the successful development of an ATIS.

DECISION-MAKING UNDER UNCERTAINTY: AN EXAMPLE

Two routes, designated Routes A and B, shown in Figure 1 are available to drivers commuting from Origin O to Destination D. The travel time experienced by drivers depends on the traffic conditions that are encountered on the routes. From past experience a traveler characterizes traffic conditions as either normal or congested. In traffic engineering parlance these conditions can be thought of as levels of service A, B, or C for normal traffic and D, E, or F for congestion. The travel times experienced by the average driver are 22 min under normal conditions and 58 min under congested conditions for Route A and 31 min under normal conditions and 39 min under congested conditions for Route B.

From past experience a driver estimates that 60 percent of the time he or she encounters normal traffic. Thus, he or she predicts that normal traffic conditions will continue to appear with a probability of 0.6 and that congestion will be encountered with a probability of 0.4. These probabilities are called *prior probabilities.* The expected travel time on the routes is then calculated as the weighted sum of experienced travel times for each traffic condition, where the weights are the probabilities of occurrence of the traffic conditions. The expected travel times on routes A and B, $E_A(T)$ and $E_B(T)$, respectively, are

$$E_A(T) = (0.6 \cdot 22) + (0.4 \cdot 58) = 36.4 \text{ min}$$

$$E_B(T) = (0.6 \cdot 31) + (0.4 \cdot 39) = 34.2 \text{ min}$$

Based on the criteria of minimization of expected travel times the driver chooses Route B.

It would be advantageous for a driver to obtain advanced information on traffic conditions before he or she chooses routes. For example, a driver who ordinarily travels on Route A would use the information about congestion to avoid Route A and would use Route B instead. This information that eliminates all uncertainty about the occurrence of a traffic condition in the decision-making process is called *perfect* information. (Note that prior probabilities, the percentage of time a traffic condition will occur, cannot be changed; the driver can only receive information about which traffic condition will occur before he or she chooses a route.) The best the driver can do to minimize his or her travel time is to use Route A 60 percent of the time and Route B the remaining 40 percent of the time. The minimum expected travel time that the driver can achieve with perfect information, *E(PI),* is

$$E(PI) = (0.60 \cdot 22) + (0.4 \cdot 39) = 28.8 \text{ min}$$

Assuming that it is possible to obtain perfect information, the amount a driver should be willing to pay for it is determined as the time savings between the expected travel time with prior information and the travel time with perfect information multiplied by the value of time. In the earlier example the difference is 5.4 min (i.e., 34.2 − 28.8). Assuming that the value of time is $15/hr ($0.25/min), the most that a driver should pay per trip is $1.35 ($0.25/min · 5.4 min/trip).

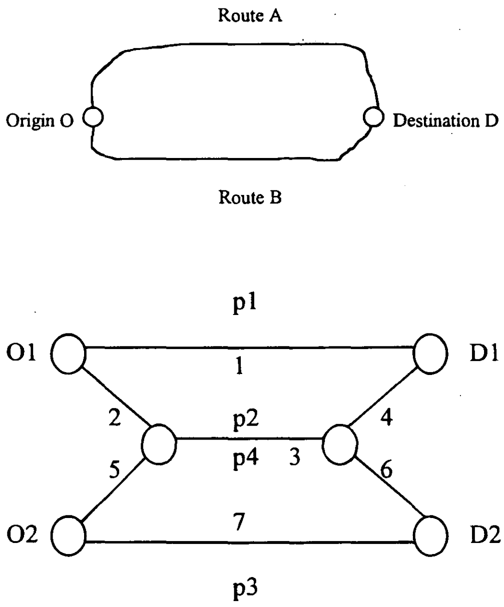


FIGURE 1. Example highway network: (*top*) single O-D pair, (*bottom*) two O-D pairs and two information services.

Formalized Approach

The example given earlier is typical of decision making under uncertainty. Clearly, more than two traffic conditions may exist (e.g., there may also be an incident). In general, a decision maker may undertake action a_k from the set of all possible actions $A = (a_1, a_2, a_3, \ldots, a_n)$. Several states of nature x_i included in the set $X = (x_1, x_2, x_3, \ldots, x_m)$ can occur, each with probability $p_i(x_i)$. For each action a_k that is undertaken when the state of nature x_i occurs, the decision maker receives a payoff (a reward or a loss), $V_{ki}(a_k, x_i)$.

In this example there are two states of nature (normal traffic and congestion), each occurring with a probability of 0.6 and 0.4, respectively, and two actions (choose Route A or B). The payoff $V_{ki}(a_k, x_i)$ is the travel time that a traveler experiences for selecting route k when condition i occurs.

The function $V(a,x)$ represents the set of all payoffs and is called the *gain function.* In the example this function was deterministic, but in general it could be a random variable and most likely a function of traffic volume. When the traffic conditions are broadly defined, as in the earlier example, the assumption that the gain function is deterministic is plausible.

In general, the decision maker–traveler will try to optimize the expected value of his or her gain function. In this particular example he or she minimizes the expected travel time of choosing action k when state of nature i occurs over the set of actions:

$$E[V(a,x)] = \min_k \sum_i p(x_i) \cdot V(a_k, x_i)$$

The expected gain with perfect information $E(PI)$, in which the traveler selects k to minimize $V(a_k, x_i)$ for each x_i, is given as

$$E(PI) = \sum_i p(x_i) \cdot \min_k[V(a_k, x_i)]$$

The value of perfect information to the traveler is based on the difference between the expected travel time with perfect information and the expected travel time with prior information.

More Complex Example

When traffic is forecasted traffic conditions are usually given for each road (or route) rather than as a general statement about network congestion as in the previous example. In the following, more complex example the same two-link network is considered, but now there are four traffic conditions:

1. Both routes have normal traffic (designated N_A and N_B),
2. Route A has normal traffic but Route B is congested (N_A, C_B),
3. Route A is congested and Route B has normal traffic (C_A, N_B), and finally
4. Both routes are congested (C_A, C_B).

Assume that these four conditions occur with the following probabilities:

$p(N_A, N_B) = 0.24$, $p(N_A, C_B) = 0.36$, $p(C_A, N_B) = 0.16$, $p(C_A, C_B) = 0.24$.

The outcome when a traveler chooses Route A or B depends on the conditions as follows:

	Traffic Condition			
Decision to Take	*(N_A, N_B)*	*(N_A, C_B)*	*(C_A, N_B)*	*(C_A, C_B)*
Route A	22 min	22 min	58 min	58 min
Route B	31 min	39 min	31 min	39 min

The expected travel times on route A and B, $E_A(T)$ and $E_B(T)$, respectively, are

$$E_A(T) = (0.24 \cdot 22) + (0.36 \cdot 22) + (0.16 \cdot 58) + (0.24 \cdot 58) = 36.4 \text{ min}$$

$$E_B(T) = (0.24 \cdot 31) + (0.36 \cdot 39) + (0.16 \cdot 31) + (0.24 \cdot 39) = 35.8 \text{ min}$$

A driver is expected to choose Route B, which has the least expected travel time.

The expected travel time with perfect information, $E(PI)$, is:

$$E(PI) = (0.24 \cdot 22) + (0.36 \cdot 22) + (0.16 \cdot 31) + (0.24 \cdot 39) = 27.52 \text{ min}$$

The difference between the expected travel time with perfect and prior information is 8.28 min (35.8 − 27.52). Assuming that the value of time is \$15/hr, the most the traveler should pay for perfect information is \$2.07 per trip (\$0.25/min · 8.28 min/trip).

Formalized Procedure

In the case describes in the previous section the state of each route can be described by random variables X for Route A and Y for Route B. The random variable can assume the state-of-nature normal traffic (N) or congested traffic (C). Then the traffic conditions on the network can be described by using a joint distribution of X and Y, $f_{X,Y}(x, y) = P(X = x, Y = y)$. The payoff $V_{kij}(a_k, x_i, y_j)$ is the time that a traveler experiences when selecting route k under state of nature - (i, j). The gain function $V(a, x, y)$ represents the set of all payoffs.

The minimum expected travel time that a traveler can achieve with prior information is

$$E[V(a, x, y)] = \min_k \sum_{i,j} p(x_i, y_j) \cdot V(a_k, x_i, y_j)$$

The expected benefit of having perfect information when a traveler selects route k to minimize V for each (x_i, y_j) state is

$$E(PI) = \sum_{i,j} p(x_i, y_j) \cdot \min_k[V(a_k, x_i, y_j)]$$

The difference between $E[V(a, x, y)]$ and $E(PI)$ multiplied by the monetary value of time determines an upper bound on what one should pay for perfect information.

INTRODUCTION OF INFORMATION SERVICE

In the second example normal traffic prevailed on either route 76 percent of the time $[p(N_A, N_B) + p(N_A, C_B) + p(C_A, N_B)]$. With perfect information a traveler would be able to predict this condition 100 percent of the time and use the facility that is not congested. The traveler could also benefit from imperfect information, however, as long as route conditions can be predicted better than prior information allows. Information, for example, that correctly predicts traffic conditions 90 percent of the time allows the traveler to take advantage of normal traffic conditions 68.4 percent (0.9 · 0.76) of the time.

Assume that in addition to the prior probabilities on the traffic condition a traveler is given conditional probabilities that describe the past performance of the information system. These probabilities indicate the frequency with which the information service forecasted a particular traffic condition, given that this condition indeed occurred. These probabilities are given in Table 1. The entry of 0.6 (in the upper left of Table 1) indicates that the service forecasted normal traffic on Routes A and B 60 percent of the time when normal traffic acually occurred on the routes. In addition, 15 percent of the time the service forecasted normal traffic on Routes A and B

TABLE 1 Conditional Probabilities

Forecasted	Actual Condition			
Condition	N_A,N_B	N_A,C_B	C_A,N_B	C_A,C_B
N_A,N_B	0.6	0.15	0.15	0.1
N_A,C_B	0.15	0.6	0.1	0.15
C_A,N_B	0.15	0.1	0.6	0.15
C_A,C_B	0.1	0.15	0.15	0.6

when actually a normal traffic condition occurred on Route A but congested traffic occurred on Route B.

These conditional probabilities can be used to update the prior probabilities and obtain *posterior probabilities*. The posterior probabilities represent the percentage of time that a particular traffic condition would occur given that it had been forecasted. The posterior probabilities can in turn be used to calculate expected travel times on the routes.

The following notation is needed before proceeding to solve the problem:

- x, y = traffic condition that occurred.
- m, n = traffic condition that was forecasted.
- $P_{X,Y}(x, y)$ = joint probability of condition x,y occurring.
- $Q_{M,N|X=x,Y=y}(m, n)$ = conditional probability of information service forecasting condition $(M = m, N = n)$, given that $(X = x, Y = y)$ occurred.
- $Q_{M,N|X=x,Y=y}(m, n) \cdot P_{X,Y}(x, y)$ = the joint probability of the forecast of condition m, n and the actual condition x, y.
- $Q_{M,N}(m, n)$ = marginal probability of traffic condition $(M = m, N = n)$ being forecasted for all possible occurrences of X, Y.
- $h_{X,Y|M=m,N=n}(x, y)$ = conditional probability of condition X, Y occurring given that the forecast indicated $(M = m, N = n)$.

Given this notation the procedure for computing posterior probabilities and expected travel times is as follows:

1. Gather the prior probabilities $P_{X,Y}(x, y)$ and the conditional probabilities $Q_{M,N|X=x,Y=y}(m, n)$.
2. Calculate the joint probabilities for each condition and its forecast, $Q_{M,N|X=x,Y=y}(m, n) \cdot P_{X,Y}(x, y)$.
3. Calculate marginal probabilities, $Q_{M,N}(m, n)$.
4. Calculate posterior probabilities $h_{X,Y|M=m,N=n}(x, y)$ as

$$h_{X,Y|M=m,N=n}(x, y) = Q_{M,N|X=x,Y=y}(m, n) \cdot P_{X,Y}(x, y)/Q_{M,N}(m, n)$$

5. Calculate expected travel times for each x,y using the posterior distributions as

$$E[V(a, x, y)] = \Sigma_k [V(a_k, m, n)] \cdot h_{X,Y|M=m,N=n}(x, y)$$

Numerical Example

Table 2 was developed by using the prior probabilities of traffic conditions given earlier and the conditional probabilities of traffic condition forecasts from Table 1. Table 2 (*a*) shows the joint probabilities for each traffic condition and its forecasts that are obtained by multiplying the prior probabilities by the conditional probabilities. The marginal probability of a forecasted traffic condition, $Q_{M,N}(m, n)$, is obtained by summing the joint probabilities over all actual conditions. For example, the marginal probability that N_A, N_B is forecasted equals 0.246(0.144 + 0.054 + 0.024 + 0.024).

Table 2 (*b*) shows the posterior probabilities that are obtained by dividing the joint probabilities by the marginal probability. For example, the posterior probability of N_A, N_B occurring given that it is forecasted equals 0.585 (0.144/0.246).

TABLE 2 Probabilities of Numerical Example

a. joint and marginal probabilities of actual traffic condition and its forecast

	Actual				Marginal
Forecast	N_A,N_B	N_A,C_B	C_A,N_B	C_A,C_B	Probability
N_A,N_B	0.144	0.054	0.024	0.024	0.246
N_A,C_B	0.036	0.216	0.016	0.036	0.304
C_A,N_B	0.036	0.036	0.096	0.036	0.204
C_A,C_B	0.024	0.054	0.024	0.144	0.246

b. posterior probabilities

	Actual			
Forecast	N_A,N_B	N_A,C_B	C_A,N_B	C_A,C_B
N_A,N_B	0.585	0.22	0.098	0.098
N_A,C_B	0.12	0.71	0.053	0.12
C_A,N_B	0.176	0.176	0.47	0.176
C_A,C_B	0.098	0.22	0.098	0.585

c. minimum time routes, expected travel times, and marginal probabilities

Forecast	Choose Route	Travel Time (min/veh)	Marginal Probability
N_A,N_B	A	22*(0.585 + 0.22) + 58*(0.098 + 0.098) = 29.078	0.246
N_A,C_B	A	22*(0.12 + 0.71) + 58*(0.053 + 0.12) = 31.368	0.304
C_A,N_B	B	31*(0.47 + 0.176) + 39*(0.176 + 0.176) = 33.754	0.204
C_A,C_B	B	31*(0.098 + 0.098) + 39*(0.22 + 0.585) = 32.6	0.246

The route choices resulting in the minimum expected value of travel time for each forecast and the marginal probabilities of the forecasts are given in Table 2 (*c*).

The minimum expected travel time for the average driver is

$$E[V(a, x, y)] = 0.246 \cdot (29.078) + 0.304 \cdot (31.368) + 0.204 \cdot (33.754) + 0.246 \cdot (32.6) = 32.79 \text{ min}$$

This time is obtained by the drivers using Route A when the forecast calls for either normal traffic on both routes or normal traffic on Route A and congestion on Route B (55 percent of the time) and using Route B when the forecast calls for either congested traffic on both routes or congested traffic on Route A and normal traffic on Route B (45 percent of the time).

The use of an information system resulted in the reduction of travel time by 3 min (35.8 − 32.79) compared with the case when only historical (or prior) information is used. The most that a traveler should be willing to pay for the service is his or her value of 3 min.

OPTIMAL NUMBER OF INFORMED TRAVELERS

It would be incorrect to assume that all travelers should be given information even if they are willing to pay for it. A highway network has limited capacity and experiences congestion. As congestion increases so does travel time. If travelers are given information indiscriminately it is possible that the change in commuting patterns will result in higher average travel times on routes, and thus higher individual travel times for all travelers, including those with information.

It is assumed that travelers are minimizing their individual travel times or behave according to Wardrop's First Principle (*2*). An inherent assumption of Wardrop's First Principle is that travelers have full information (actual travel times) about the routes. In the case in which travelers base their route choice on perceived travel times instead of actual ones, the principle states that at equilibrium the traveler cannot improve his or her perceived travel time by unilaterally switching routes. This principle leads to Stochastic User Equilibrium (*3*).

The minimization of individual traveler times must not be confused with the minimization of networkwide travel times, also known as Wardrop's Second Principle (*2*). It is common that in minimizing their individual travel times travelers can indeed increase the total networkwide travel time. In the absence of congestion Wardrop's First and Second Principles yield the same traffic assignment flows and costs.

In this paper it is assumed that travelers have an incentive to purchase information of a given accuracy as long as it can make them better off. This implies that travelers will be able to further minimize their individual travel times in comparison with the case of having only historical (prior) information. Therefore, the optimal (or desirable) number of travelers can be determined by an equilibrium pattern which the travel times for travelers who use the information service are at least as good as those for travelers with historical information. This equilibrium point determines the maximum market share for the information service.

Network Equilibrium

To analyze this problem the networkwide interactions among travelers must be investigated. The network highway links are congested, and the travel time on a link is estimated according to the Bureau of Public Roads (BPR) congestion curve (*4*) of the form

$$t_l = t_0 \cdot [1 + a(\text{link flow/link capacity})^x]$$

where

t_l = average travel time on link l,
t_0 = free-flow travel time on link l (the distance divided by the free-flow speed),
a = marginal increase in link time when accommodating an additional vehicle, and
x = coefficient.

The link capacities used in the BPR curve were calculated by using the methodologies set forth in the 1985 *Highway Capacity Manual* (*5*) for each road type (arterial and freeway).

Assume that the congestion curves for Routes A and B are

$$t_A = 17 \cdot [1 + 3 \cdot (\text{flow}/2{,}000)^2], \text{ and}$$

$$t_B = 30 \cdot [1 + 3 \cdot (\text{flow}/6{,}000)^2]$$

where Route A and B capacities of 2,000 and 6,000 vehicles per hour, respectively, are calculated for levels of service that represent stable flow. Depending on traffic conditions, normal and congested, each of these curves are divided into two areas. For normal traffic conditions on Routes A and B they are

$$t_A = 17 \cdot \{1 + 3 \cdot [(100 + f_A)/2{,}000]^2\}, \text{ and}$$

$$t_B = 30 \cdot \{1 + 0.3 \cdot [(1{,}500 + f_B)/6{,}000]^2\}$$

For congested traffic conditions on Routes A and B they are

$$t_A = 17 \cdot \{1 + 3 \cdot [(1{,}300 + f_A)/2{,}000]^2\}, \text{ and}$$

$$t_B = 30 \cdot \{1 + 0.3 \cdot [(5{,}100 + f_B)/6{,}000]^2\}$$

The graphical representation of the volume-travel time functions and areas representing normal and congested conditions are given in Figure 2.

If 500 travelers are assigned over the routes the resulting travel times are 21.59 min for normal conditions and 58.31 min for congested conditions on Route A and 30.99 min for normal conditions and 37.84 min for congested conditions on Route B. If normal conditions on Route A prevail 60 percent (24 percent + 36 percent) of the time and congestion prevails 40 percent (16 percent + 24 percent) of the time, the expected travel time is 36.28 min. If normal conditions on Route B prevail 40 percent (24 percent + 16 percent) of the time and congestion prevails 60 percent (36 percent + 24 percent) of the time, the expected travel time is 35.1 min.

When the model is solved with prior information only, all 500 travelers are assigned to the minimum time Route B, yielding an average travel time of 35.1 min (the minimum of 35.1 and 36.28 min for Routes B and A, respectively). When travelers are assigned by using the posterior information provided by the information service, all 500 travelers use Route A 55 percent of the time [the sum of the marginal probabilities of those choosing route A in Table 2(*c*)] and Route B 45 percent of the time, yielding an average travel time of 31.39 min. This lower travel time is achieved by using the information service to forecast normal traffic on Route A and assigning the travelers over that route during those periods and over Route B when traffic is congested on Route A.

If all 500 travelers were subscribers to the service they would have been better off than if they were using prior information only.

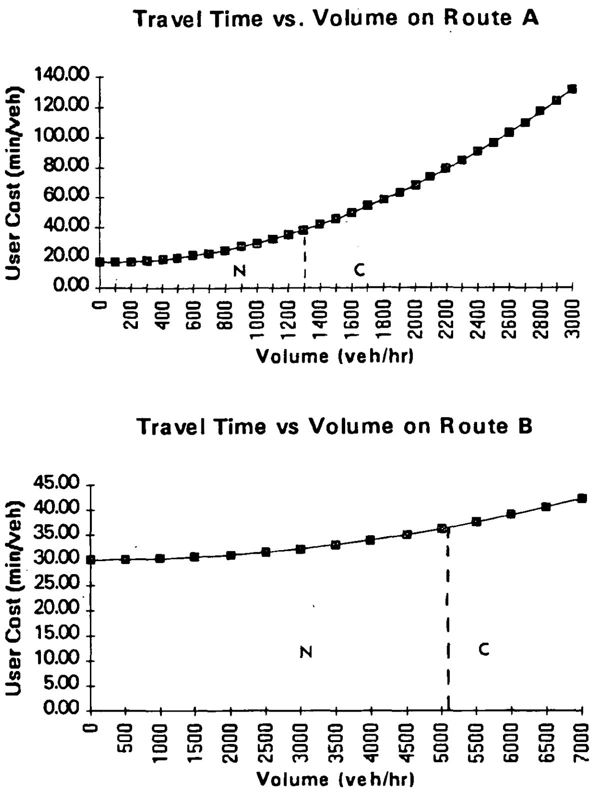


FIGURE 2. BPR congestion curves and traffic conditions.

By using the information their individual time decreased by 3.71 min (from 35.1 to 31.39 min). Finally, when the travelers are assigned with perfect information, all 500 are assigned to Route A 60 percent of the time and Route B 40 percent of the time. Thus, the travel times are further minimized to an average of 28.09 min.

The question that arises is determination of the optimal number of users to be given information. Common sense would indicate that travelers would have an incentive to use the information service only if it did not make them worse off than they were with no service at all. Consequently, the optimal market penetration is determined by a traffic assignment for which the average travel time is no less than that of an assignment obtained without information.

Returning to the previous example, it can be verified that when 800 travelers (instead of 500) are assigned according to the minimum expected travel time, the resulting travel times are 34.818 min when only prior information is available and 35.770 min with the information service (posterior). It is obvious that the switching of commuting patterns caused by the vehicles that obtained information has caused the average travel times on links to become higher than they were in the case in which they had prior information. One may conclude that it is highly unlikely that people will be using the service that makes them worse off (by 0.952 min in this case) than they were before they started using it. Therefore, the number of users to whom information services should be provided (i.e., an equilibrium among users with and without an information service) needs to be determined.

To determine the optimal number of users to be given information, assume that travelers are divided into two classes: those without and those with information, designated *fwo* and *fw*, respectively. The *fwo* travelers will be using only Route B, whereas *fw* travelers will be using Routes A and B in a mixed strategy. It should be recognized that the travel times on routes depend on both types of flows *fwo* and *fw*, because these flows use the routes simultaneously. Since travelers with information *fw* will be using Route A during certain times (i.e., 55 percent of the time), they will be removed from Route B. Thus, the following equation can be set up. The equation determines an equilibrium pattern in which the travel times for travelers who use the information service are at least as good as those for travelers with historical or prior information. The left side of the equation represents the travel times of travelers with historical information, whereas the right side represents the travel times of travelers who use the information service.

$$\begin{aligned}
&0.4\cdot\{30\cdot[1 + 0.3\cdot[(1{,}500 + fwo + fw - 0.55fw)/6{,}000]^2\}\\
&+ 0.6\cdot\{30\cdot[1 + 0.3\cdot[(5{,}100 + fwo + fw - 0.55fw)/6{,}000]^2\}\\
&= 0.246\cdot\{17\cdot[1 + 3\cdot[(100 + fw)/2{,}000]^2]\cdot 0.805 + 17\cdot[1\\
&+ 3\cdot[(1{,}300 + fw)/2{,}000]^2]\cdot 0.196\} + 0.304\cdot\{17\cdot[1 + 3\cdot[(100\\
&+ fw/2{,}000]^2]\cdot 0.83 + 17\cdot[1 + 3\cdot[(1{,}300 + fw)/2{,}000]^2]\cdot 0.173\}\\
&+ 0.204\cdot\{30\cdot[1 + 0.3\cdot[(1{,}500 + fwo + fw)/6{,}000]^2]\cdot 0.646\\
&+ 30\cdot[1 + 0.3\cdot[(5{,}100 + fwo + fw)/6{,}000]^2]\cdot 0.352\}\\
&+ 0.246\cdot\{30\cdot[1 + 0.3\cdot[(1{,}500 + fwo + fw)/6{,}000]^2]\cdot 0.196\\
&+ 30\cdot[1 + 0.3\cdot[(5{,}100 + fwo + fw)/6{,}000]^2]\cdot 0.805\}
\end{aligned}$$

Given that $fwo + fw = 800$, solving this equation yields flows for *fwo* and *fw* of 57.957 and 742.043 vehicles, respectively. The travel assignment times are equalized at 34.882 min. This result implies that approximately 742 vehicles will have an incentive to subscribe to the service, but only if it were free.

The resulting average travel time as a function of traffic volume for each information distribution strategy is given in Figure 3. There are three information distribution strategies: (a) all travelers have only prior information, (b) an information service gives information on traffic conditions to all travelers indiscriminately, and (c) the information service is provided to a selected group of travelers, whereas the rest of the travelers use historical information. The number of users with information indicates the optimum market penetration because the average travel time between an origin and a destination cannot be further improved by giving information to an additional traveler.

Figure 3 shows that up to a volume of 700 vehicles/hr, if all travelers are given the information an average traveler will experience a shorter travel time compared with that in the case in which he or she were to use prior information. After that volume the information should be given to only a portion of the total travelers. For example, for a volume of 900 vehicles/hr, the information should be given to 748 vehicles. This results in an equilibrium time of 35.080 min/vehicle. (Note that if the information is given to all 900 vehicles the average user cost would increase to 37.515 min.) For a volume of 1,000 vehicles/hr the number of travelers with information will be 755 vehicles/hr. It is apparent that the 11.1 percent inverse in volume (i.e., from 800 to 900 vehicles) will increase the market penetration of the service by 0.9 percent (i.e., from 748 to 755 vehicles). Thus, the majority of additional travelers will be given no information.

Further Extensions

The framework is also applied to a more complex network consisting of six nodes (two origin–destination pairs and two through nodes), seven links, and four paths. The network is shown in Fig-

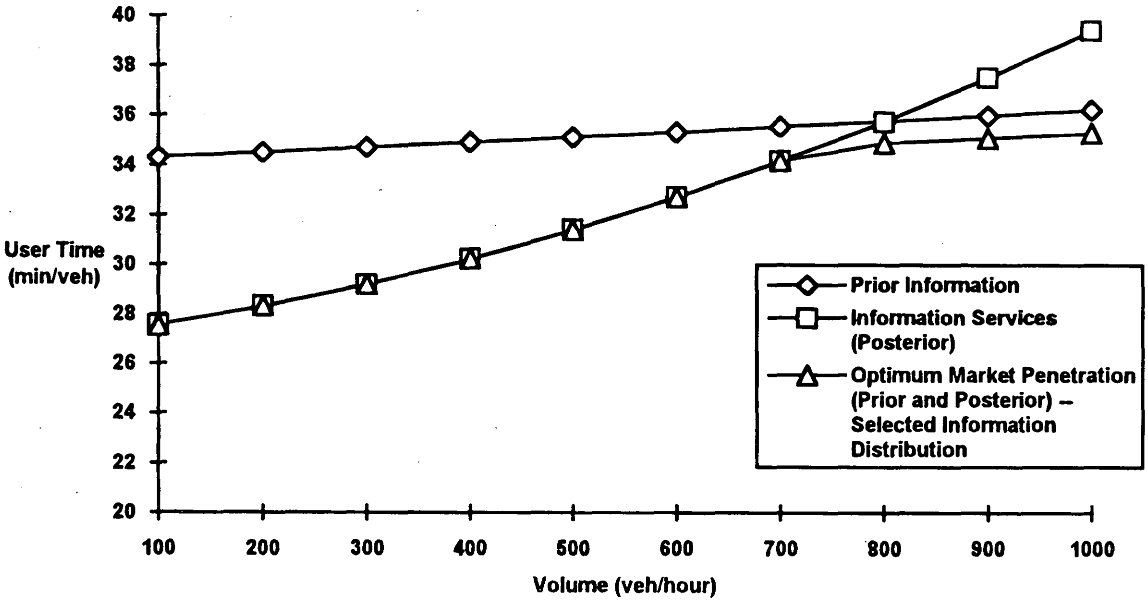


FIGURE 3. Travel times under various information distribution strategies.

ure 1(*b*). For the purpose of simplifying the discussion of the results it is assumed that the network is symmetric (i.e., Links 1 and 7 are identical and Links 2 and 4 are identical to Links 5 and 6). Two paths are available for travel from each origin to a destination. Paths P1 and P3 consist of Links 1 and 7, respectively, whereas Paths P2 and P4 consist of Links 2, 3, and 4 and Links 5, 3, and 6, respectively. The prior and posterior probabilities of normal and congested traffic occurring on Paths P1 and P2 and Paths P3 and P4 are assumed to be identical to those given in Tables 1 and 2 for Routes A and B. It is also assumed that there are two information services available to travelers: Poly-Traffic serving the O1–D1 pair and Shade-Traffic serving the O2–D2 pair. There are 500 vehicles/hr between each origin and destination.

The results presented in Table 3 indicate that when the services made their decision as to whom to give information and what commuting strategies to suggest independently from each other, they gave the information to all 500 vehicles from each origin. The vehicles ended up with the strategy of choosing Paths P2 and P4 55 percent of the time and Paths P1 and P3 45 percent of the time. Since the services made their decisions in a vacuum, they did not consider the possible strategies of their opponents. Thus, they perceived that the resulting vehicle travel times would be 31.39 min.

TABLE 3 Optimal Market Penetration and Travel Times Under Various Information Dissemination Strategies

Information Type	Select Routes	Users with Information Service (veh/hour)	Expected Travel Time (min/veh)	
			Perceived	Actual
Prior	P1 and P3 100% of the time	0	35.1	35.1
Posterior -- No Cooperation between Services	P2 and P4 55% of the time, and P1 and P3 45% of the time	1,000	31.39	38.91
Posterior -- Cooperation between Services	P2 and P4 55% of the time, and P1 and P3 45% of the time	748	34.69	34.69

However, since Link 3 is shared by Paths P2 and P4, the congestion on Link 3 caused by an additional 1,000 vehicles going over it 55 percent of the time resulted in an actual travel time of 38.91 min. This decision to suggest to all vehicles to choose the same commuting strategy made all vehicles worse off than they would have been if they had stayed on Paths P1 and P3 100 percent of the time. Had they stayed on Paths P1 and P3, as they would have under prior information, their travel times would have been 35.1 min.

When the services either cooperated with each other or attempted to predict each other's responses in terms of a likely strategy, the optimal assignment indicated that they each gave information to use Paths P2 and P4 55 percent of the time to only 374 vehicles. This resulted in a travel time of 34.69 min. The remaining 126 vehicles from origin O1 were not given information and thus stayed on Path P1 100 percent of the time. The remaining 126 vehicles from origin O2 stayed on Path P3 100 percent of the time.

Future Framework Extensions

The framework presented in this paper can be expanded to take into account traveler utilities rather than travel times. The utility, in addition to travel time, may include other impedances such as out-of-pocket cost and other qualitative measures of a commute such as the scenery along the route, perceived safety of the surrounding area, and so forth.

The paper assumed that drivers have linear utilities (i.e., they place the same value on 1 min saved on a 22-min trip as well as on a 58-min trip). This is a rather strong assumption, because people value time savings higher on a shorter trip than on a longer trip (*6*). A candidate nonlinear utility function that can be used in the framework has been provided elsewhere (*6*). The process of deriving nonlinear utility functions has been given previously (*7*).

The framework can also be expanded to take into account the fact that more than two traffic conditions can arise (and be perceived by a driver) on the route. In addition, instead of using discrete distributions a continuous probability distribution can be used to describe traffic conditions. Consequently, the payoffs can be expressed as expected values of a random variable, the traffic flow. Various methods for estimating congestion functions need to be incorporated into the model as well.

The framework presented here is rather aggregate. It does not recognize the time dimension of a decision-making process. The methodology presented needs to be expanded to take into account the dynamic aspects of decision making. There needs to be a feedback loop between the traffic conditions arising at various links in the network at various moments in time and travelers' decisions.

CONCLUSIONS

A framework was presented for assessing the benefit of information from an ATIS. It provides transportation professionals with a tool to evaluate the value of an information service and to compare it with the case in which only historical information was available. In addition, the framework can be used to evaluate the value of the information service in comparison with the case in which travelers have perfect information. The methodology can also be used to estimate the characteristics and accuracy of the information service. For example, given a certain target market penetration, the methodology can be used to determine the system accuracy (i.e., probabilities of detecting traffic conditions) so that an average potential user receives a certain level of benefits. The method can be further improved by considering additional traffic conditions that occur on the routes. More appropriate congestion functions can be derived and used in the method. The procedures presented in this paper can be used within a more comprehensive framework that uses state-of-the-art traffic assignment techniques of mathematical programming and simulation to better ascertain both the potential value of information to customers and the optimal number of people to whom the information should be given. Finally, the framework may complement behavioral studies that determine travelers' attitudes toward various types of information delivery technologies.

ACKNOWLEDGMENTS

This research was partially supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program, through the National Center for Transportation and Industrial Productivity at the New Jersey Institute of Technology. This support is gratefully acknowledged.

REFERENCES

1. Hamerslag, R., and E. C. van Berkum. Effectiveness of Information Systems in Networks With and Without Congestion. Presented at 70th Annual Meeting of the Transportation Research Board, Washington, D.C., 1991.
2. Wardrop, J. G. Some Theoretical Aspects of Road Traffic Research. *Proc., Institute of Civil Engineers*, Part II, 1952.
3. Sheffi, Y. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1985.
4. *Traffic Assignment Manual.* U.S. Bureau of Public Roads, U.S. Department of Commerce, 1964.
5. Transportation Research Board. *Special Report 209: Highway Capacity Manual,* TRB, National Research Council, Washington, D.C., 1985.
6. Chen, K., and T. Reed. Nonlinear Disutility of Travel Delay: Implications for Traffic Diversions Benefit and Strategy. *IVHS Journal*, Vol. 1, No. 2, 1993, pp. 115–134.
7. Winston, W. *Operations Research: Applications and Algorithms.* ITP, 1994.

Publication of this paper sponsored by Committee on Travelers Services.

Automated Information Systems at Interstate Welcome Centers as Element of Intelligent Vehicle-Highway Systems

LISA H. DEAN AND ROY C. LOUTZENHEISER

No standard for the distribution of information at information centers located in Interstate rest areas and welcome centers exists. Some states use basic automated information systems (AISs). A second-generation AIS would be a method of distributing information through Interstate rest areas and welcome centers that is compatible with the future plans of the Intelligent Vehicle-Highway System. A questionnaire was sent to each state to identify its method of distribution of information at information centers. At least 17 states use an AIS to provide information to travelers. A second questionnaire was sent to the operators of an AIS. Most AISs can be classified into two categories: those containing paid advertisements and those not containing paid advertisements. A second-generation AIS (AIS-2) can be developed by modifying the AIS. The physical system should be a network of terminals linked to a master system. Expert system programming would help manage the data in a "yellow pages" method of advertisements. Computer-integrated manufacturing could be used to tie together all of the different aspects of the AIS-2 and allow for change as technology changes over time.

Thousands of rest areas and welcome centers across the United States operate to assist motorists as they travel on the Interstate system. The primary functions of these rest areas and welcome centers are to provide the traveler, either the single driver or several persons in a vehicle, with a place to relax and to use rest room facilities. A secondary function is to provide a source of information about the surrounding area and the state.

No uniform method of distributing information at Interstate rest areas and welcome centers exists. Most rest areas and welcome centers provide staff and brochures to inform the public. This combination allows travelers (a) to interact with the staff to find specific information and to carry this information away with them in the form of a brochure or (b) to review brochures only and take the ones that appear to be of interest. However, two major problems occur when this combination of information distribution methods is used. First, it is not financially feasible to operate a staffed rest area continuously. During the hours that the main building is closed it is not possible for the traveler to receive data through the information center. The second problem is that much waste is generated by some travelers. They may grab numerous brochures and, at a later time, read them, keeping those with the information desired and then discarding those pamphlets that they do not need.

An expanded form of an information center has recently been implemented in several states. An automated information system (AIS) allows a traveler to automatically obtain information about a specific subject in a timely manner without receiving excess information. However, AISs have various limitations, such as the availability of information.

Through further development of the AIS, a much greater range of information could be made available to the traveler. Information on (a) local restaurants and hotels, (b) local attractions, and even (c) road construction in the area as well as a host of other things could be programmed into the AIS. With a wider variety of information about the local area available to the traveler, it is likely that more tourism dollars would be spent in an area. On a broader scope of improving travel on the U.S. roadway system, the Intelligent Vehicle-Highway System (IVHS) program has been initiated and includes plans to develop AISs to assist travelers (*1*).

The research described here specifically addressed the concept of AISs at information centers, their current use, and their potential for further development. A survey of the various state agencies responsible for rest areas and welcome centers in the United States was made to determine the current methods of providing travelers with relevant information. The existing AISs, referred to as the first-generation systems, were analyzed in detail. Successful elements and apparent problems were noted. Finally, a conceptual design of a second-generation AIS was proposed on the basis of the analysis. This proposed system would serve as a partial link between the AIS in use today and the future systems in the IVHS program. The research concluded with recommendations on how a second-generation AIS could be implemented.

BACKGROUND

Roadways, present and past, are built to provide travelers with a way to move from one place to another, usually in a relatively safe and easy manner. However, the roadway by itself has not always been sufficient. Throughout most of the ages resting places adjacent to the roadside have also been established. These places allow travelers to rest so that they are more alert when they continue their journeys. These rest areas, however, are only as useful as the facilities provided. As travelers' needs and journey purposes have changed over time, so have the purposes and features of the rest areas changed.

The original travelways were footpaths used for hunting, primarily created by movement of many animals traveling the path (*2*). Early humans usually did not travel far from home, traveling only far enough so that they could return home for shelter at night. Rest areas were only a place for brief relaxation or for protection from bad weather.

L. H. Dean, 910 Willow Avenue, Apt. Je-1, Cookeville, Tenn. 38501. R. C. Loutzenheiser, Department of Civil and Environmental Engineering, Tennessee Technological University, P.O. Box 5015, Cookeville, Tenn. 38501.

As the mode of travel improved and humans began to travel farther from home and to trade with other villages, travel back home each night was not practical. The purpose of the rest area began to expand into providing overnight facilities. Travelers turned to existing structures for shelters, usually a religious sanctuary, rather than just an area adjacent to the road. These first organized rest areas provided safety from bandits, as well as a place to rest and obtain food. Local people began to realize the financial benefits of these travelers, who bought local goods and services, thus increasing activity in the local economy. The local people would tell travelers about the happenings in the community and possibly about what to expect as their journey continued (*3*). Thus, the exchange of information became a part of the purpose of the rest area: the local people gain in economic development and travelers know more about local activities and the roadway ahead.

As Americans were moving west in the colonial period, the rest areas, usually called taverns, were an important element in expansion. They were privately operated establishments that provided food and overnight accommodations for regular travelers in early American history. These taverns also provided a place for the exchange of local information by the people who visited them. For example, fur traders on their regular trading routes were frequent customers. They could eat and rest while they learned of areas where game was abundant, and in exchange they would share their traveling knowledge with others.

Motor Vehicle Age

With the rapid development and use of the motor vehicle in the early 20th century, the emphasis of travelers' needs at rest areas changed again. Trips were longer both in time and in distance. It became important that travelers have a chance to relax, a need met by most rest areas. Knowing that rested drivers meant fewer accidents on the travelways, the early highway designers began to realize the need to add additional space adjacent to the travelways as turnouts. When the initial turnouts were few and far between, drivers would create their own turnouts by stopping on the side or shoulder of the roadway or sometimes even on adjacent private property. With the early roads being narrow, the created turnouts were frequently hazardous to moving vehicles, particularly if the stopped vehicles reduced sight distance. Many landowners also became increasingly disturbed by the constant trespassing onto their lands. To reduce the problem the state of Michigan began constructing turnouts in 1919. These are now considered the first rest areas on public roads (*4*).

Highway engineers began to look at the ideal frequency and spacing of turnouts with respect to motorists' needs. Turnouts were becoming mandatory to provide a safer roadway, as well as to meet the needs of the drivers. During the 1920s and 1930s turnouts became more frequent and evenly spaced along the major highways in the United States. In 1938 the federal government recognized their importance by granting federal money for their construction and maintenance. As a result many turnouts were upgraded to wayside areas (or roadside parks), with the addition of picnic tables, trash cans, shade, and possibly well or spring water. Many new wayside areas were also built from 1940 to 1960, and some included a small building with limited rest room accommodations. These wayside areas met the needs of the typical motorist during this time period. With the advent of the Interstate system in the late 1950s, motorists' needs changed again.

The Interstate system allowed motorists to travel direct routes from city to city on a limited-access highway. More people owned cars, and vehicle miles of travel increased more than ever before in history. Much of this travel was for recreational purposes rather than strictly for business. People were traveling into new areas in the United States and now needed information about the sights to see, motels and restaurants, and special events. The old turnouts or wayside areas did not provide the details about a particular state. They just provided basic food and a place to rest.

The new wayside areas were called roadside rest areas and provided a place to rest, relax, and obtain food, as well as a place to obtain tourist information. Many locations began to distribute information to the traveling public so that they would know more about the state or region. Some of the roadside rest areas evolved into information distribution centers, with food and rest being a secondary purpose. These rest areas, known as welcome centers, were usually located near the state line to provide information to motorists as they entered the state. Although other needs of motorists were met, the main purpose of the facility was to provide motorists with information on the state or region.

The Future: IVHS Program

Modern-day roads are continually becoming more and more congested. Transportation engineers have begun to realize a growing need for new technologies to increase the efficiency of the roadway system rather than for continual building of more roadways. In 1991 the Intermodal Surface Transportation Efficiency Act included provisions for IVHSs. To accomplish IVHS goals, the IVHS program was divided into six areas. Although each area is unique, all are interlocking. These areas are

- Advanced traffic management systems,
- Advanced traveler information systems,
- Advanced vehicle control systems,
- Advanced public transportation systems,
- Commercial vehicle operations, and
- Advanced rural transportation systems.

The area critical to the present research was advanced traveler information system (ATIS) technologies, which attempt to assist the traveler with planning, perception, analysis, and decision making to improve the convenience and efficiency of travel. ATISs acquire, analyze, communicate, and present information to assist travelers from their origins to their destinations. "A major component of the ATIS is providing information to the driver of a vehicle" (*5*).

METHODOLOGY

To various degrees information to travelers who use the Interstate highway system is now provided at rest areas and welcome centers. The overall object of the present research was to determine if there is a more effective process of providing travelers with this information. The research was accomplished in three general steps. The first step (through an initial questionnaire) was to determine details about the existing systems that were being used and the degree of automation. A second, more specific questionnaire (step two) was developed to determine additional details about the existing first-generation AISs. The third step was to combine the good parts of operating systems, particularly, the first-generation AISs, into a second-generation AIS that would be compatible with the IVHS program.

Step 1: Existing Information Centers

Since published information about specific practices at Interstate rest areas is limited, a survey of government agencies was conducted. The format of the survey was a written questionnaire that was mailed to the operating agency of information centers in each of the 50 states. The questions for the initial questionnaire (Figure 1) were developed to determine how operators of the information centers acquire and distribute traveler information. The main areas that were addressed in the questionnaire were the (a) sources of the information, (b) method of distribution of the information, and (c) type of information available. To encourage a quick response the length of the questionnaire was kept short.

The questionnaire was mailed to the state agency in each of the 50 states that deals with tourism. The agencies and their addresses were obtained from the *National Directory of State Agencies*. The cover letter requested that the questionnaire be forwarded if there was someone more knowledgeable on the subject. A self-addressed, stamped envelope was enclosed to encourage a quick and easy response. If there was no response within 1 month a copy of the questionnaire was sent to an official at the state department of transportation.

An evaluation of the returned questionnaire included the following relationships: information sources, distribution methods, number of automated systems, method of automation, and agency responsible for operation. No attempt was made to quantify the quality of service, since the users are the primary source for such an evaluation.

Step 2: First-Generation AISs

A second questionnaire was developed to acquire more detailed data on existing AISs. In particular, this questionnaire (Figure 2) inquired about specific operating details of the AISs. This questionnaire was sent to a state agency or private business, or both, that operated an AIS at an Interstate rest area or welcome center. When necessary a follow-up telephone call was made to clarify or amplify the answer.

With the data received from this second questionnaire an analysis of each AIS was performed by classifying the options offered on every system. The most common attributes of every system and the unique attributes of every system were identified.

Step 3: Second-Generation AISs

The components for a second-generation AIS were designed by using the common and the unique elements of the existing systems. The object was to put together the good elements and, where possible, minimize the undesirable elements. Where possible the second-generation AIS would use as much of the existing system as practical, particularly the hardware, to encourage a cost-effective system.

Also taken into consideration in the development was how this second-generation AIS could fit into the future expectations of the IVHS program. The major association would be with the in-vehicle communication system. Prototype vehicles have now been equipped with monitors and personal computers to allow access to an electronic telephone directory.

EXISTING INFORMATION CENTERS

The first questionnaire was distributed to all 50 states. Seven of the questionnaires were returned to Tennessee Technological University because of an incorrect mailing address. A second copy of the questionnaire was mailed to the correct address, determined by using the telephone directory of the city where the agency is located.

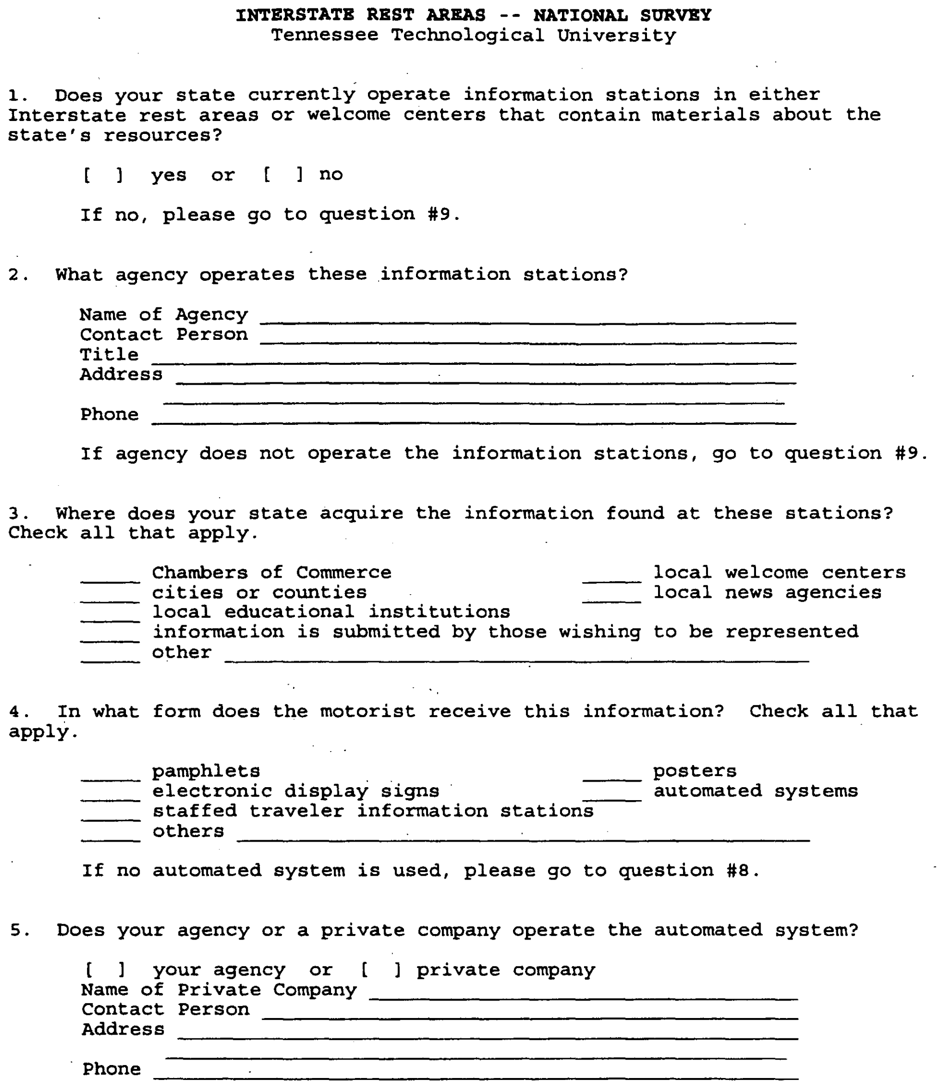
Thirty-nine responses were received. One of the remaining agencies referred the authors to another state agency to obtain the information requested; however, a questionnaire mailed to that agency was not returned. For the 10 states not responding, a second attempt was made to receive the information by mailing the questionnaire to each state department of transportation. This resulted in the return of four more responses, for a total of 43 usable responses.

Information centers are operated in the welcome centers and rest areas of at least 40 states. One of the remaining 10 states (Arizona) planned to open its first visitor information center in 1994. For the analysis in the present research, only the 40 states that responded to the survey and that had information centers were considered.

The distribution of information is the responsibility of many different state agencies, as shown in Table 1. The state agency that deals with tourism or travel manages the information centers in most states (75 percent). Often these agencies are part of or are within the state department that deals with economic development or commerce (45 percent). Like the days of the Roman Empire, travelers still bring economic growth to the areas to which they journey. Other states have a separate agency that handles only tourism (28 percent). In five states (Delaware, Michigan, New Hampshire, Ohio, and Texas) the information centers are run by the state department of transportation, and in two states (South Carolina and Utah) the information centers are run jointly by the state department of transportation and the state department of parks, recreation, and tourism.

Another factor that makes each information center distinct is the source of the information that is passed along to travelers. The majority of states receive part of their information from chambers of commerce (90 percent of the states), from submitted information (85 percent), or from local governmental units (83 percent). Other sources of information include educational institutions (35 percent), local welcome centers (13 percent), or local news agencies (8 percent). Most agencies use a combination of sources to fulfill their needs. The most common combination is the chamber of commerce, local governmental units, and submitted information. This allows a wide range of information to be gathered from just a few sources.

Of specific interest are the methods of distribution of the information. All information centers provide brochures to motorists. Virginia charges a provider fee to distribute the brochure to the traveler. All centers except one (New Hampshire) have at least part-time staff available to assist motorists. Several states use posters (50 percent) or electronic display signs (10 percent). Seventeen states have some sort of automated system available for motorists. In addition, three states (Colorado, South Carolina, and Texas) have automated systems in the developmental stage. Other forms of information distribution include lighted transparencies, taped information for the visually impaired, and videotapes.



INTERSTATE REST AREAS -- NATIONAL SURVEY
Tennessee Technological University

1. Does your state currently operate information stations in either Interstate rest areas or welcome centers that contain materials about the state's resources?

[ ] yes or [ ] no

If no, please go to question #9.

2. What agency operates these information stations?

Name of Agency ______________________________
Contact Person ______________________________
Title ______________________________
Address ______________________________
____
Phone ______________________________

If agency does not operate the information stations, go to question #9.

3. Where does your state acquire the information found at these stations? Check all that apply.

_____ Chambers of Commerce
_____ cities or counties
_____ local educational institutions
_____ information is submitted by those wishing to be represented
_____ other ______________________________
_____ local welcome centers
_____ local news agencies

4. In what form does the motorist receive this information? Check all that apply.

_____ pamphlets
_____ electronic display signs
_____ staffed traveler information stations
_____ others ______________________________
_____ posters
_____ automated systems

If no automated system is used, please go to question #8.

5. Does your agency or a private company operate the automated system?

[ ] your agency or [ ] private company
Name of Private Company ______________________________
Contact Person ______________________________
Address ______________________________
____
Phone ______________________________

FIGURE 1 Interstate rest area questionnaire.

Most agencies (98 percent) use the combination of brochures and staff to inform travelers. This allows travelers to talk with someone who can give specific information about a place, whereas the brochures allow them to carry something along with them as they continue their journeys.

Some of the states use an automated system to assist the traveler in finding the proper information. Of the 17 states with automated systems, 5 states operate the systems themselves, 10 states use private agencies, and 2 states operate the systems jointly with a private agency or university. Most offer only touch-screen capabilities and printout on demand. Two states (Idaho and Minnesota) have their automated systems encased in a building similar to those used for automated teller machines (ATMs) for 24-hr availability.

Of the 17 states with automated systems, most offer a variety of information including information on motels (94 percent), statewide attractions (88 percent), and state and national parks (88 percent). Other information available on some systems includes local attractions (76 percent), local restaurants (65 percent), and local stores or malls (29 percent). Four states (Montana, Minnesota, Idaho, and Missouri) offer information on cost, and one state (Wyoming) offers information on availability.

Overall, a good response was received from the first questionnaire, with 40 of the 50 states responding with a usable reply (Table 2). The most frequent source for the information located in the information center was the chamber of commerce: this was followed by information submitted from various sources and local governmental units such as cities or counties. By far the most common methods of distribution were brochures and staff, and a combination of these methods was used in every state that responded except one. Seventeen states operated AISs in their information centers, and 3 more states had AISs in development. The most common types of information found on these AISs were information on motels, state and national parks, and statewide attractions.

AUTOMATED INFORMATION SYSTEMS

The second step of the research was to acquire specific information about the AISs currently in use. The data on each system needed included the number of advertisers, the cost of advertisements, and how advertisers are initially located.

6. If your state uses an automated system, what are the features of this system? Check all that apply.

_____ touch screen
_____ personal computers
_____ print out automatic to motorist
_____ print out with graphics
_____ automated systems enclosed in own building for 24-hour/day use (similar to ATM's used at banks)
_____ others ______________________________
_____ audio capabilities
_____ print out available to motorist

If printouts are available, please enclose one along with the survey if possible.

7. If your state uses an automated system, what types of information are offered to the motorist? Check all that apply.

_____ statewide attractions
_____ local restaurants
_____ local stores or malls
_____ availability
_____ local attractions (such as plays, festivals, & sporting events)
_____ others ______________________________
_____ state and national parks
_____ motels
_____ cost information

If printouts are available, please enclose one along with the survey if possible.

8. Has your state performed any research since 1985 on the use of automated systems as information centers at Interstate rest areas and welcome centers.
[ ] yes or [ ] no

If yes, would you please send me a copy of the report. If there is a cost for the report, please let me know in advance.

9. Person filling out form (or send a business card):

Name ______________________________
Title ______________________________
Address ______________________________
____
Phone ______________________________

May I contact you if more information is required? [ ] yes or [ ] no

Would you like a copy of the results of this survey? [ ] yes or [ ] no

Please return this survey by August 31, 1993 to

Lisa Dean
Tennessee Technological University
Box 5015, ext. 85
Cookeville, TN 38505

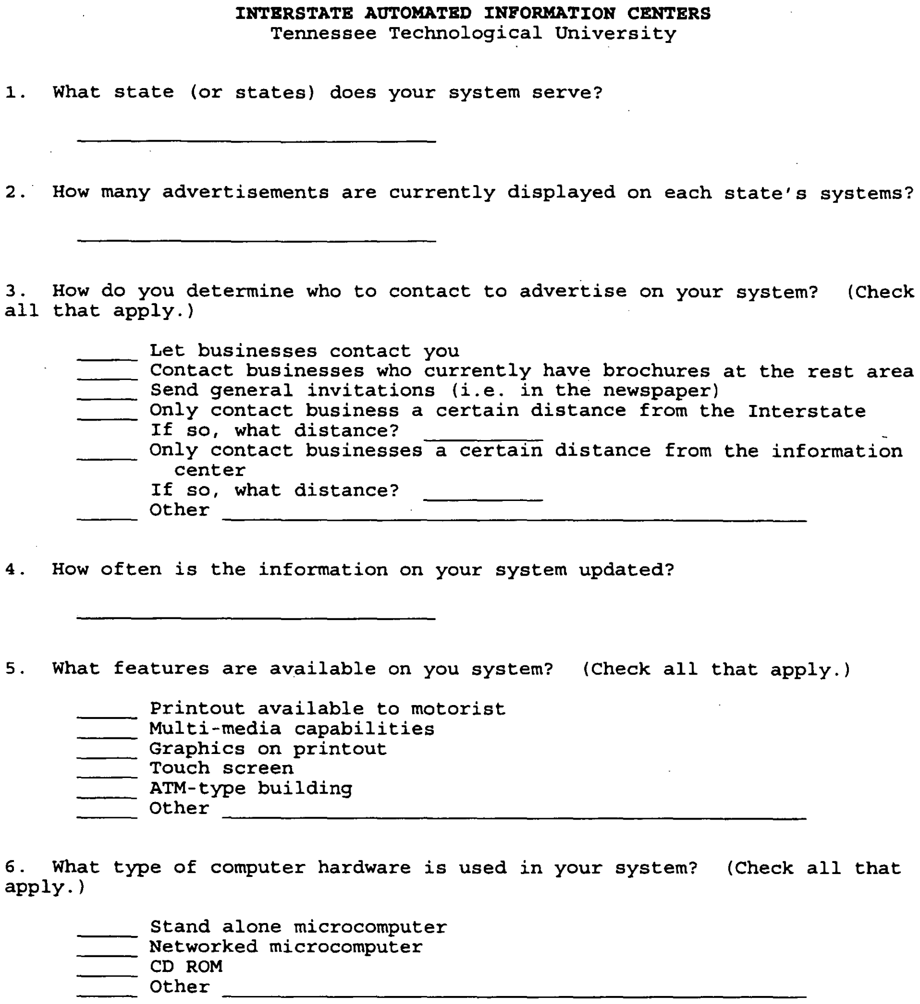
FIGURE 1 *(continued)*

Based on the first questionnaire, 17 states had AISs at Interstate rest areas or welcome centers. The operators of these systems were diverse in nature. In some cases the state agency that operated the information center also operated the AIS. In other cases private businesses were contracted to run the systems. Sometimes a combination of sources worked together to operate the system. Educational institutions and tourism associations also helped to run these systems.

The second questionnaire was developed to acquire more information about the operators of the AISs in the 17 states that had them. After 4 weeks information on systems in 10 of the 17 states with AISs was received. A phone call was made to each operator not responding to obtain information about the AISs in the additional states. As a result of these phone calls three more states returned the questionnaire. For the analysis presented here only the 13 states responding to the survey were considered, with the results summarized in Table 2.

After evaluating the data from the second questionnaire, two distinct types of systems were detected. The first type of system relied mostly on paid advertisements to generate the bulk of information found on the AIS (54 percent). The second type of system relied on free advertisements to relay information (46 percent). Although all of the systems had common characteristics, no two systems were exactly alike. A system's unique elements usually strengthened its effectiveness.

The first group of systems that had many common characteristics comprised systems based on paid advertisements. The principal income of the system was through businesses choosing to pay to advertise on the AIS. However, some advertisements were allowed on the systems free of charge. For example, information about public attractions (i.e., state and national parks) was allowed on the system for free. This allowed for more information to be relayed to the traveler. The money received from private businesses (i.e., motels and restaurants) to advertise was usually the major source of income, whereas information about attractions provides the users with a full range of knowledge. It is possible for these systems to generate a profit. Seven of these cases were studied, and in all cases this type of system was operated by a private business.



INTERSTATE AUTOMATED INFORMATION CENTERS
Tennessee Technological University

1. What state (or states) does your system serve?

____

2. How many advertisements are currently displayed on each state's systems?

____

3. How do you determine who to contact to advertise on your system? (Check all that apply.)

_____ Let businesses contact you
_____ Contact businesses who currently have brochures at the rest area
_____ Send general invitations (i.e. in the newspaper)
_____ Only contact business a certain distance from the Interstate
If so, what distance? __________
_____ Only contact businesses a certain distance from the information center
If so, what distance? __________
_____ Other ____

4. How often is the information on your system updated?

____

5. What features are available on you system? (Check all that apply.)

_____ Printout available to motorist
_____ Multi-media capabilities
_____ Graphics on printout
_____ Touch screen
_____ ATM-type building
_____ Other ____

6. What type of computer hardware is used in your system? (Check all that apply.)

_____ Stand alone microcomputer
_____ Networked microcomputer
_____ CD ROM
_____ Other ____

FIGURE 2 AIS questionnaire.

The second group of systems that emerged was a system based on no paid advertisements. In many cases the respondents of the questionnaire believed that they offered no advertisements, only information, to the public. The cost of the system was primarily funded through the state. No revenue was generated by the system; its only purpose was to inform the traveling public. These systems were generally run by departments of the state, but in one case (Kentucky) it was operated by a private company.

Some systems also have unique features that appear to allow the system to better inform the traveling public. In three states (Iowa, Kansas, and Michigan) users can watch videos of different locations around the state. In Michigan the advertiser can update material more frequently, if the advertiser is willing to pay for it. In three states (Idaho, Minnesota, and South Dakota) the AIS was enclosed in an ATM-type building for continuous use by the traveler, even when the rest area or welcome center is closed. One company, Interactive Technologies, had a network of systems in the western states of Idaho, Montana, Utah, and Wyoming. These four states had telephones nearby to allow travelers to call for reservations.

DEVELOPING AIS-2

The levels of service of the AISs varied. In developing an updated second-generation system (AIS-2), it would be important to consider further use of the AIS with the IVHS program. AIS-2 should be able to integrate all of the information needs of travelers today and to be flexible to allow for expansion to meet the needs of travelers tomorrow. AIS-2 should not merely be a two-dimensional system that matches the needs of the traveler with technology. It should be a system of three dimensions that also includes the ability to change as new technologies are developed over time.

Several factors that should be considered with AIS-2 include the hardware and software of the system, the handling of data, and the compatibility of the system with the IVHS program. These factors not only need to be considered at present, but how they may possibly change in the future should also be considered. A method of handling complex automation has been implemented in the business world. Computer-integrated manufacturing (CIM) is the most powerful concept to date for bringing people, technology, products, and processes together into a single integrated system (*6*). The concept

7. What complications or difficulties have you experienced with your system? Check all that apply.

_____ Insufficient amount of advertisers to encourage use by travelers
_____ Recurring system failure
_____ Children playing on the system, not allowing access to adults
_____ Print out with graphics
_____ Others ____

8. What is the current fee(s) for an advertisement on your system?

____

9. Are any advertisements free of charge? If so, what advertisements?

____

10. Would you allow your company's name to be used in a published report about automated systems as information centers in rest areas?

[ ] yes or [ ] no

11. May I contact (or possibly interview) you if more information is required?

[ ] yes or [ ] no

Person filling out form (or send a business card):

Name ____
Title ____
Address ______________________________________
____
Phone ____

Please return this survey by February 17, 1994 to

Lisa Dean
Tennessee Technological University
Box 5015, ext. 85
Cookeville, TN 38505

FIGURE 2 ***(continued)***

of CIM can be used with AIS-2, tying together all of the different factors and allowing them to be modified as technology changes.

Physical System

Most of the AISs operated on a stand-alone microcomputer and had to be individually updated at each location. Therefore, it was not financially feasible to update the AISs often. From the information gathered from the second questionnaire, on average, the AIS at each location was updated twice a year. If the AISs at all locations were linked to a master terminal, updates could be sent by telecommunication waves and the information on the AIS would be current. With a networked system the availability of space (i.e., motel rooms and concert tickets) would be current information that could be distributed accurately. Also, information such as road construction, weather conditions, and traffic congestion could be transmitted to the users.

The question quickly arose as to what information should be distributed on this system. Too much information simply confuses the user, whereas too little information does not adequately inform the user. Most systems currently in use charge private businesses (i.e., hotels and restaurants) to display advertisements. The prototype information system of the IVHS system used in Orlando simply had a yellow pages in electronic form (*7*). A combination of these methods can be used for maximum benefit. All tourist-related activities should be available in the system. This allows users to know that their questions will be answered thoroughly and will not be limited by only the specific business that chose to advertise on the system.

In the 1960s Iowa studied the information needs of drivers and developed the Infosite idea. Infosites were buildings containing informational posters and displays to help travelers (*8*). Advertisers had to pay to display these signs. Only 20 percent of the space allotted for displays was used. Travelers did not rely on the information found at Infosites because it was incomplete. The program failed because of a lack of information offered to the public. With a yellow pages method of advertising, most businesses would be represented; therefore, travelers could depend on the AIS to receive complete information. As a source of revenue, advertisers could pay to have their basic advertisements enlarged.

One problem that may occur with allowing excessive advertisements on the AIS is an overabundance of information, which could confuse the traveler. Expert system software is a type of software that uses a degree of expertise in problem solving that is comparable to that of a human expert. By using expert system software, the user could narrow the field of choices by answering simple questions. For example, instead of listing every motel in a certain geographic area, the AIS might query travelers on their preferences on cost and specific features (i.e., swimming pool or hot tub). This would allow the AIS to inform the traveler of only a limited number of motels, permitting the traveler to choose from a reasonable number of sources.

TABLE 1 Agencies That Operate Information Centers

State	Specific Agency	Governing Agency
AL		Bureau of Tourism and Travel
AK		Jointly by various agencies
AZ	Office of Tourism	Office of the Governor
AR		Department of Parks & Tourism
CO	Tourism Board	Office of the Governor
DE		Department of Transportation
GA		Department of Industry, Trade & Tourism
FL	Division of Tourism	Department of Commerce
ID		Department of Parks & Recreation
IN	Tourism Department	Department of Commerce
IL	Bureau of Tourism	Department of Commerce & Community Affairs
IA		Department of Economic Development
KS	Tourism	Department of Commerce & Housing
KY	Tourism Cabinet	Department of Travel Development
LA	Office of Tourism	Department of Culture, Recreation & Tourism
ME		Publicity Bureau
MD		Dept. of Economic & Employment Development
MA	Office of Travel & Tourism	Department of Commerce & Development
MI		Department of Transportation
MN	Office of Tourism	Department of Energy & Economic Development
MS		Department of Economic Development
MO	Division of Tourism	Department of Economic Development
MT	Travel Montana	Department of Commerce
NH		Department of Transportation
NM		Department of Tourism
NJ	Division of Travel & Tourism	Department of Commerce & Economic Development
NC	Division of Travel & Tourism	Department of Commerce
ND		Tourism Department
OH		Department of Transportation
OK	Division of Travel & Tourism	Tourism & Recreation Department
OR	Tourism Division	Economic Development
RI	Division of Tourism	Department of Economic Development
SC		Department of Parks, Recreation & Tourism and Department of Transportation
SD	Tourism	Department of State Development
TN		Department of Tourist Development
TX		Department of Transportation
UT		Travel Council, Travel Regions, Department of Transportation
VA	Division of Tourism	Department of Economic Development
WV	Division of Tourism & Parks	Department of Commerce
WI	Division of Tourism	Department of Development
WY	Division of Tourism	Department of Commerce

TABLE 2 Results of Questionnaires

Interstate Rest Areas Questionnaire	
Number of questionnaires sent	50
Number of responses	43
Number of information centers	40
Number of states with AIS	17
Most common sources of information	1) Chamber of Commerce 2) submitted 3) local governmental unit
Most common methods of distribution	1) brochures 2) staff 3) posters
Most common type of information found on AIS	1) motels 2) state and national parks 3) statewide attractions
Automated Information Systems Questionnaire	
Number of questionnaires sent	17
Number of responses	13
Number of AIS with paid advertisements	7
Number of AIS without paid advertisements	6
Unique elements on some systems	1) telephones 2) videotapes 3) ATM-type building

Data Management

CIM could be used to manage all of the data in AIS-2. All of the data should be input into a master system and then transferred to the other systems in the network. Not only will the main agency operating the system need to have access to data entry, but each private business represented on the system will also need to enter their specific time-dependant data (such as reservations) to keep the system updated.

The data output will be the information received by the traveler. Expert system software can be a means of controlling the amount of data received by the traveler. Real-time output will be in the form of screen display, printed general information, and printed detailed information.

Once the traveler has received the information from AIS-2, in many cases a response to the business is desired. Currently, telephones are placed near some of the systems so that the user can call to make reservations. This step could be improved by installing a modem in the microcomputer. Then, just by pressing a button (or by touching a screen) the user could request AIS-2 to call and make reservations, providing that the businesses (i.e., motels and restaurants) have the computer facilities to answer this type of call.

AIS-2 will be a complex system involving state-of-the-art components, large quantities of data, and real-time information needs. With the IVHS program being the major program for the U.S. surface transportation system, the development of AIS-2 must complement the development of IVHSs. CIM is the umbrella that can manage the physical system and data management needs of AIS-2 while considering the future needs of the IVHS program.

CONCLUSIONS

Based on the results of the present study the following conclusions were drawn.

1. Seventy-five percent of the Interstate information centers were operated by the state agency that deals with tourism. In five states the information centers are operated by the state's department of transportation.
2. The most common sources of information for the centers were chambers of commerce, private submittals, and local governmental units.
3. All of the states surveyed used brochures to distribute information, and all of the states except one used at least part-time staff.
4. Seventeen states had an AIS to assist motorists in obtaining information about the state. Three more states had AISs in development.
5. The type of information available from some AISs included information on motels, statewide attractions, parks, local attractions, restaurants, and stores. Four states offered information on cost, whereas one state offered information on availability.
6. The first-generation AISs fall into two categories: (a) those with paid advertisements and (b) those without paid advertisements.
7. Development of AIS-2 should be compatible with the IVHS program. The physical system should be a networked system of terminals linked to a master computer. Data input and updating, data output, and traveler response could be managed by CIM. Information for the traveler could be processed by expert system programming.

RECOMMENDATIONS

1. Various types of information can be included on an AIS. Additional research is needed to determine what type of information (i.e., weather conditions or road construction) is functional and practical for AIS-2.
2. The range or geographic extent of the information (i.e., surrounding states, regions, and counties) to be offered on an AIS is another area that needs further evaluation.
3. Pilot IVHS studies have included in-vehicle AISs. Airports, shopping centers, and other populated centers have provided information for the traveling public. An evaluation of the optimum type of location(s) of AISs for the traveling public is necessary.

ACKNOWLEDGMENTS

The authors express appreciation to the Center for Manufacturing Research and Technology Utilization, Tennessee Technological University, for sponsoring this research.

REFERENCES

1. *Transportation Research Circular 412: Primer on Intelligent Vehicle Highway Systems.* TRB, National Research Council, Washington, D.C., 1993, pp. 5–28.
2. Gregory, J. W. *The Story of the Road.* Macmillan Company, New York, 1938.
3. *Transportation Research Circular 275: Issues Relating to National Policy on Motorist Information and Roadside Amenities.* TRB, National Research Council, Washington, D.C., 1984, p. 2.
4. Cardone, S. M. Maintenance Cost of Rest Areas in Michigan. In *Highway Research Record 93,* HRB, National Research Council, Washington, D.C., 1965.
5. *Strategic Plan for Intelligent Vehicle-Highway Systems in the United States.* IVHS America, Washington, D.C., 1992.
6. Gerelle, E., and J. Stark. *Integrated Manufacturing Strategy, Planning, and Implementation.* McGraw-Hill Book Company, New York, 1988.
7. Collier, W. C., and R. J. Weiland. Smart Cars, Smart Highways. *IEEE Spectrum,* April 1994, pp. 27–33.
8. Griswold, W. E. Information Needs of Interstate Highway Motorist in Iowa. In *Highway Research Record 285,* HRB, National Research Council, Washington, D.C., 1969.

Publication of this paper sponsored by Committee on Travelers Services.

Should Emergency Call Boxes Be Placed in Freeway Medians?

JAMES H. BANKS

Current guidelines related to emergency call boxes call for them to be placed on the right-hand side of the unidirectional roadway. The advisability of also placing call boxes in freeway medians or adjacent to high-occupancy vehicle (HOV) lanes in medians is considered. The opinions of California Highway Patrol officers and various professionals involved in the provision of call boxes in California and elsewhere were surveyed; opinions of survey respondents were mixed, but they were generally negative concerning median or HOV lane call boxes except in the case of barrier-separated HOV lanes. A review of accident reports to determine the number of persons struck while attempting to cross freeway lanes to access call boxes revealed that such accidents are exceedingly rare. A cost study showed that the costs of call boxes are quite modest compared with alternative means of providing motorist assistance. Based on the accident study it is concluded that installation of call boxes in median areas is not warranted except in the case of barrier-separated HOV facilities, where there would otherwise be no access to motorist assistance services.

Emergency call boxes are increasingly being used in California to allow drivers of disabled vehicles to communicate conveniently with the California Highway Patrol (CHP) so that roadside assistance can be provided. They are especially valuable on limited-access facilities, where the distance to off-facility telephones is often excessive, and where the potential accident exposure of disabled vehicles and pedestrians walking along the roadway to summon help is greatest. As of July 1994, 14 Service Authorities for Freeway Emergencies (SAFEs) had been authorized to provide call box service in California, and 11 were already in operation. These cover most major metropolitan areas in California and several rural areas. Statewide, some 14,500 call boxes were in operation.

Several alternative means of providing motorist assistance also exist. In addition to call boxes several local agencies provide freeway service patrols (FSPs). The goal of the FSPs is to reduce congestion by removing disabled vehicles and debris from traffic lanes as soon as possible. FSP programs generally consist of special fleets of dedicated tow trucks or other emergency vehicles that patrol selected sections of roadway during specific times (*1*). Besides these publicly provided services, increasing numbers of private vehicles are equipped with cellular telephones, and these have begun to play a significant role in providing CHP with notification of traffic incidents.

To date, however, only the call box program provides universal access to emergency services. FSPs are severely limited as to their hours of operation, and many vehicles are not equipped with cellular phones. Consequently, although traffic-blocking incidents are apt to receive timely responses even in the absence of call boxes, routine breakdowns may not. This is especially true if they occur at off-peak hours or on less critical facilities not served by FSPs.

The spacing and locations of call boxes are governed by a set of guidelines adopted jointly by the California Department of Transportation (Caltrans) and CHP (*2*). The typical call box installation consists of solar-powered cellular telephones spaced at intervals of 0.4 to 3.2 km, depending on the cost and anticipated use. Normal practice in California has been to locate call boxes on the right-hand side of the unidirectional roadway, just outside the edge of the shoulder. In most cases, left-hand shoulders are also available; consequently, some disabled vehicles may stop on the left-hand side of the roadway.

In addition, several types of high-occupancy vehicle (HOV) lanes are provided in California freeway medians. In some cases these are physically separated from the rest of the freeway by traffic barriers, so that the vehicles using HOV lanes are not able to reach the right-hand shoulder of the main freeway. Although the usual practice is to place call boxes on the right-hand shoulder only, at least one such HOV facility, the I-15 reversible HOV roadway in San Diego, has call boxes that are located in the median area.

Concern exists that drivers who stop their vehicles on the left-hand shoulder or on the side of a median HOV lane may attempt to cross the main freeway lanes to reach a call box and risk being hit by a moving vehicle. This concern was expressed as a part of a Caltrans safety stand-down in 1992, and as a result a study was conducted to investigate the need for and value of locating call boxes in medians or adjacent to median HOV lanes (*3*).

The objective of the study was to investigate the advisability of installing call boxes in medians or adjacent to HOV lanes in medians. This included investigating the safety implications of installing or not installing such call boxes, the experience and concerns of existing call box providers and law enforcement personnel, and the potential cost of such installations.

The study included (a) a telephone survey of SAFE administrators, vendors, and consultants to SAFEs, Caltrans headquarters and district personnel who work directly with the call box program, and CHP administrators who are involved in the call box program; (b) a written survey of CHP officers from Southern California who are familiar with call boxes; (c) a telephone survey of professionals outside California who are involved in providing call boxes in freeway medians or alongside median HOV lanes; (d) a detailed study of dismounted pedestrian accidents in Southern California that was intended to determine the number of accidents involving people crossing freeway lanes to access call boxes; and (e) an analysis of the costs of installing and operating call boxes in medians or alongside median HOV lanes.

SURVEYS

Survey of California Call Box Professionals

A telephone survey of various types of professionals involved in the current California call box program was conducted during the sum-

---

Civil and Environmental Engineering Department, San Diego State University, San Diego, Calif. 92182-1324.

mer of 1993. A total of 27 interviews were conducted. Of the respondents, 13 were administrators of SAFEs, 4 represented call box vendors or consultants to SAFEs, 6 were Caltrans headquarters or district personnel involved in the call box program, and 4 were CHP administrators involved in the call box program.

Each respondent was asked (a) whether call boxes are needed in freeway medians and why or why not, (b) whether call boxes are needed for median HOV facilities and why or why not, and (c) if call boxes are to be provided in medians or for median HOV facilities, how far apart they should be and whether any particular locations should be included or avoided. In addition, respondents from San Diego County were asked about their experience with the call boxes on the I-15 reversible HOV facility.

A majority of the respondents in this survey were opposed to median call boxes, with about 40 percent favoring them and 60 percent opposing them. The majority of the respondents in all groups surveyed except vendors and consultants were opposed; vendors and consultants were unanimously in favor of call boxes in freeway medians.

In addition, some respondents gave qualified answers. Some of the more important qualifications were (a) that there be adequate room to stop [in one case, a median shoulder width of at least 3 m (10 ft)] and (b) that priority be given to establishing median call boxes on wider freeways. Some respondents who said that median call boxes are needed added that they may not be needed on freeways with only two lanes in one direction or stated that top priority should be given to freeways with four or more lanes in one direction. At the same time other respondents stressed that if median call boxes are not universally available motorists may become confused. Several also stated that under the present funding arrangements, some counties would lack the resources to install call boxes in medians without compromising spacing or coverage; hence, it is almost certain that median call boxes will not be universally installed.

Respondents were also asked their reasons for favoring or opposing median call boxes. Most of those favoring them stated that it is inevitable that some motorists with mechanical problems will stop on left-hand shoulders. If they remain on the left-hand shoulder for very long, they will be more vulnerable than if they were on the right-hand shoulder, primarily because of the higher speeds in the left lane. Also, motorists disabled on left-hand shoulders may feel stranded, and hence may be tempted to cross the freeway lanes.

Those opposing median call boxes tended to agree that left-hand shoulders are more dangerous than right-hand shoulders, but they argued that to install call boxes in the medians would encourage motorists to stop there. Furthermore, they argued that if people do stop on the left side of the roadway they should be encouraged to remain in their vehicles, because their exposure will be even greater if they are out of their vehicles walking to or using call boxes. Several commented that currently, the overwhelming majority of stops are on the right-hand side of the road and that to install call boxes in the median would be to risk changing this. In addition, a few respondents commented that people might actually be tempted to cross from the right shoulder to the left shoulder, especially if a call box on the right shoulder were not functioning. Finally, several respondents emphasized that placing call boxes in medians is sometimes impossible because of a lack of space. This is especially reported to be a problem in the Los Angeles area.

When asked whether call boxes should be installed alongside median HOV facilities, respondents tended to distinguish between situations in which such facilities are physically separated from the mixed-flow lanes by means of concrete traffic barriers and those in which they are not. In the latter case HOV lanes are typically provided on both sides of the freeway median barrier and are separated from mixed-traffic lanes in the same direction by pavement markings, traffic cones, or a buffer strip. Most respondents favored installing call boxes alongside barrier-separated HOV facilities, but the respondents were about evenly divided as to whether they should be used where such facilities are not separated.

In the case of barrier-separated HOV facilities, the most common reason given for favoring installation of call boxes was the impossibility of vehicles reaching the right-hand shoulder. In the case of nonseparated facilities, an important reason given in opposition to installing call boxes was the lack of adequate shoulders. It was stated that where such facilities are retrofit measures, median shoulders were often sacrificed to provide for the HOV lanes. Otherwise, the reasons for favoring or opposing installation of call boxes alongside median HOV facilities were similar to those for favoring or opposing installation of call boxes in freeway medians.

When asked about call box spacing, respondents most commonly replied that the spacing of median or HOV lane call boxes should be the same as that of call boxes on the right-hand shoulder or that it should be in accordance with the existing call box guidelines. Only a few respondents mentioned specific distances.

Respondents were also asked about specific locations where call boxes should or should not be placed. The most common response to the question of where they should be placed was that they should be at the same locations as call boxes on the right-hand shoulder. Respondents believed that this would minimize the likelihood of people attempting to cross the freeway to use a call box. Common answers to the question of where call boxes should not be placed included

- Locations where there is insufficient shoulder width.
- Locations where there is inadequate sight distance.
- Locations at entrances or exits to either the freeway or an HOV facility.
- Other locations of special hazard.

Call box professionals working in San Diego County were also asked about their experience with the call boxes on the I-15 reversible HOV lanes and were asked to provide any related statistical information they might have, such as usage rates. All reported that there had been no problems with them. Based on statistics supplied by the San Diego SAFE, from 50 to 100 calls are placed from the 15 call boxes concerned each month. According to one respondent most of these calls are from motorists in the mixed-flow lanes. This respondent also stated that breakdown and accident rates for the HOV facility are very low.

Written Survey of CHP Officers

A written survey was administered to 40 CHP officers representing the San Diego, Orange County, Los Angeles, and San Bernardino-Riverside areas. A total of 39 responses were received. Questions were similar to those in the survey of California call box professionals.

A majority of the respondents in this survey opposed the installation of median call boxes, but there were considerable differences in the responses from different geographical areas.

The reasons for favoring or opposing median call boxes were similar to those advanced by California call box professionals.

Those favoring installation of call boxes in medians usually stated that some motorists will stop on left-hand shoulders and will be tempted to cross the freeway lanes if call boxes are provided on the right side but not on the left. Those opposed to median call boxes tended to argue that providing call boxes in medians will encourage motorists to stop in medians, where they are more vulnerable than on the right shoulder because speeds are higher in the left lane than in other lanes. Also, as in the case of the California call box professionals, some officers pointed out that other means of communication are available.

The second question was whether call boxes are needed alongside median HOV facilities and why or why not. Unlike the California call box professionals, CHP officers tended not to differentiate between barrier-separated and non-barrier-separated facilities, although assumptions as to which type of facility was meant may have influenced their answers, which were almost evenly divided.

When asked about spacing for median call boxes, CHP officers tended to respond with specific distances, which were mentioned in 21 responses. Of these, 12 recommended spacings of 0.4 km or less, 3 recommended 0.8 km, and 3 recommended 1.6 km or greater. In addition, three respondents listed a range of spacings.

Responses to the question of where call boxes should be placed included CHP enforcement areas and places with wide shoulders. Responses to the question of where they should not be placed included locations with inadequate sight distances, locations with inadequate median widths, elevated transition roads or bridges, and undercrossings. This last location was said to involve noise problems.

Experiences of Jurisdictions Outside California

A third survey was conducted to determine the experiences of jurisdictions outside California that provide call boxes either in medians or alongside median HOV lanes. Selected members of the TRB Freeway Operations Committee and FHWA representatives were polled in an attempt to identify such installations. Telephone surveys were then conducted to determine the details of the installations, the experiences of their operators, and the respondent's recommendations as to whether California should provide similar facilities. Only two installations were identified, one of them a toll bridge and the other a tunnel. Since both involved somewhat unusual situations, experience with them was of limited relevance to the study.

ACCIDENT STUDY

An accident study was conducted to establish the numbers and severities of accidents that might be prevented by providing median call boxes. The Traffic Accident Analysis and Surveillance System (TASAS) data base was used to identify all dismounted pedestrian accidents occurring in Caltrans Districts 7, 8, 11, and 12 during the 3-year period from January 1, 1990, to December 31, 1992. These four Southern California districts were selected because of their geographical proximities to San Diego State University and because they contain the majority of call boxes in California. Collision reports on file in the headquarters of these Caltrans districts were reviewed to establish the exact circumstances of these accidents.

Dismounted pedestrian accidents result from a variety of circumstances, most of which have nothing to do with call box use. To identify accidents that might have been prevented had a median call box been available, the following criteria were adopted: (a) a vehicle was stopped on the left-hand shoulder or in or next to a median HOV lane, (b) some individual (driver or passenger) from this vehicle was struck while attempting to cross the freeway lanes, and (c) there was cause to believe that the individual had attempted to cross the freeway lanes to use a call box.

A total of 642 dismounted pedestrian accidents were identified through the TASAS data base. Collision reports were located and reviewed for 602 of these. Table 1 summarizes the results.

Of the 602 accidents reviewed, only 2 clearly involved the three elements set forth earlier. One of these was fatal and the other involved injuries to the pedestrian. In addition, four accidents involved unexplained attempts to cross the freeway from the left-hand shoulder. These might have been related to attempted call box use, but there was no positive evidence that this was so. Three of these accidents resulted in fatalities; the fourth involved collisions among vehicles attempting to avoid a pedestrian who was not struck. None of these six accidents involved an HOV facility.

In addition to accidents meeting the criteria stated above, there were a number of accidents in which call box use was involved but was clearly not the cause of the accident. These included one case in which a victim had successfully crossed from the left-hand shoulder to use a call box, had returned successfully, and then was struck while standing on the left-hand shoulder. In another case a person other than the one crossing to use the call box was struck while standing on the left-hand shoulder.

TABLE 1 Results of Study of Dismounted Pedestrian Accidents

	Accident Circumstances				
Caltrans District	Call Box Use Involved	Unexplained Freeway Crossing	Definitely No Call Box Use	Report Not Found	Total
7	1	2	344	22	369
8	0	0	87	9	96
11	0	2	107	4	113
12	1	0	58	5	64
Total	2	4	596	40	642

Based on this accident study it appears that the number of accidents involving people struck while crossing freeway lanes to access call boxes is on the order of one per year in Southern California. By way of contrast, numerous dismounted pedestrian accidents result from people exiting vehicles stopped in the lanes because of accidents or mechanical problems, people ejected from vehicles during accidents who are subsequently struck by a vehicle, and people exiting vehicles parked on the shoulder (most often the right-hand shoulder) and either standing on the shoulder or working on a vehicle. A rather large proportion of dismounted pedestrian accidents occur at night, and a very large percentage of them involve drivers who are intoxicated or otherwise impaired. In any case call box use is not a significant factor in such accidents.

COST ANALYSIS

A cost study was undertaken to determine the approximate cost of installing call boxes in medians or alongside median HOV facilities. The unit costs of installing and operating call boxes in medians were assumed to be similar to those for call boxes on right-hand shoulders. A recent estimate by the San Diego SAFE of the cost of extending call box service to Imperial County was used as a basis for calculating unit costs. In addition, the costs of extending FSPs as an alternative to providing median or HOV lane call boxes were calculated for purposes of comparison.

The average annual cost of installing and operating a call box, assuming a 10-year life and interest rates of between 5 and 10 percent, was estimated to be between $1,320 and $1,400. In areas for which guidelines specify 0.4-km spacing, installation of one additional call box in the median opposite the existing call boxes would require 2.5 additional boxes per kilometer, at an annual cost of approximately $3,500.

By way of comparison, average FSP costs in Los Angeles reported by Finnegan (*1*) are about $50/hr/truck, with an average of 0.25 trucks per directional kilometer served required to provide the current level of service. Provision of such service 24 hr/day for 365 days/year would cost about $110,000/km/year; provision of service for 8 hr/day on weekdays only (similar to the service currently provided) costs about $25,000/km/year.

POTENTIAL SAFETY IMPACTS

Installation of emergency call boxes in freeway medians and alongside median HOV facilities is warranted only if they promise to provide a distinct safety advantage at a reasonable cost. The results of the accident study and the cost analysis indicate that although the costs are quite reasonable (especially compared with alternative means of providing motorist assistance), there is no reason to believe that there would be a significant safety advantage.

The major argument in favor of installing median call boxes is that they can prevent accidents that result when persons are struck while attempting to cross freeway lanes to access a call box. The accident study showed that such accidents are exceedingly rare. Meanwhile, many of the respondents in the surveys of California call box professionals and CHP officers expressed the opinion that median call boxes would actually cause more accidents than they would prevent.

The argument that median call boxes might increase accident risk is based on the assumptions that (a) they encourage motorists to stop on the left-hand shoulder, and (b) accident risks are distinctly higher on left-hand shoulders than on right-hand ones. To show this conclusively it would be necessary to establish the fact that placing call boxes in medians increases the exposure of persons and vehicles on median shoulders and that the risk of accidents is greater on the left-hand shoulder than on the right-hand shoulder when the different levels of exposure are considered.

Establishing these facts is quite difficult, since establishing the levels of exposure requires measuring the times spent by both vehicles and dismounted pedestrians on left-hand and right-hand shoulders. Such data are not readily available, and their collection was well beyond the scope of the study.

The only information found that seems to bear directly on the exposure issue is data regarding the location of FSP contacts with disabled vehicles. These data are broken down according to whether the vehicle serviced was stopped on the left shoulder, in the lanes, on the right shoulder, on a ramp, or at some other location. Exact figures vary somewhat on the basis of the service area and the time period considered, but in general, about 5 or 6 percent of the contacts occur on the left-hand shoulder and about 80 to 85 percent occur on the right-hand shoulder (*4*). In other words disabled vehicles are 13 to 17 times as likely to be on the right-hand shoulder as on the left-hand shoulder. It should be understood that these figures may not be entirely representative, since FSPs are limited as to the areas and the times of day that they serve. Also, they reveal only the numbers of vehicles stopped in various locations and not the amounts of time that they would remain there in the absence of the FSP.

Of the dismounted pedestrian accidents considered in the accident study, about 8 to 10 percent occurred on the left-hand shoulder and 25 to 35 percent occurred on the right-hand shoulder; that is, the accidents were 3 or 4 times more likely to occur on the right-hand shoulder than on the left-hand shoulder. Comparing the ratio for exposure to that for the number of accidents, it appears that the accident risk for dismounted pedestrians might be 3 or 4 times greater on the left-hand shoulder than on the right-hand shoulder. Such a comparison is certainly not conclusive, but it does suggest that there is a basis in fact for the belief that accident risk is greater on left-hand shoulders.

The question of whether median call boxes encourage motorists to stop on the left side of the freeway is even more difficult to address. At this time the only median call boxes in California are those along the I-15 reversible HOV roadway in San Diego. Analysis of calls from boxes along this section of freeway over a 1-month period showed that of 213 calls made during times when the HOV lanes were closed, 24 were made from median call boxes. That is, calls from the right-hand shoulder are about 8 times as likely as calls from the left-hand shoulder. It should be noted that in this case call boxes are located on only one side of the HOV facility at any given point and are therefore accessible from only one direction of the main lanes. Consequently, the rate of calls from median call boxes in this case should be roughly half that expected when median call boxes are accessible to motorists traveling in both directions. This leads to the conclusion that for normal median call box installations, the ratio of right-hand shoulder to median shoulder calls might be no greater than 4 to 1. If this is compared with the statistics on FSP contacts cited previously (disabled vehicles are 13 to 17 times more likely to be on the right-hand shoulder), it is clear that the median call boxes may very well be encouraging motorists to stop on the left-hand shoulder. Once again the analysis is not conclusive, but only suggestive, since the area involved is small and not necessarily typical of the freeway system as a whole.

To the extent that the issue can be decided, then, it appears that installation of median call boxes will not result in a significant safety benefit and that it might even be detrimental.

MEDIAN HOV FACILITIES

Although it does not appear that median call boxes are warranted in most situations, a somewhat stronger case may be made for them when there are barrier-separated median HOV facilities with adequate shoulders. In this case there is virtually no access to call boxes otherwise. Experience with call boxes alongside the barrier-separated HOV lane on I-15 in San Diego has generally been good, although there appears to be little use of them by traffic from the HOV lane. In all cases in which call boxes are installed along barrier-separated HOV facilities, it is important that adequate shoulders and sight distances be available.

In cases in which HOV lanes are not separated, there appears to be little reason to install median call boxes. Such HOV facilities are of two varieties: those that are separated from the mixed-traffic lanes by intermittent striping only, which are referred to as contiguous, and those that are separated by a paved buffer of 1.2 to 2.1 m (4 to 7 ft) with double yellow stripes on each side. In neither case is the difficulty of accessing the right-hand shoulder significantly greater than that from accessing it from the leftmost lane of a mixed-traffic facility of comparable width. Consequently, the arguments that apply to the advisability of placing call boxes in the medians of regular freeways also apply to nonseparated HOV lanes.

One problem that may be unique to nonseparated HOV lanes is that stalled vehicles can easily block an HOV lane, and if it is to retain its effectiveness, these must be removed promptly. For this purpose FSPs would appear to be more useful than call boxes, despite their higher costs. At present, all buffer-separated HOV facilities and some 85 percent of contiguous HOV lanes in California are served by FSPs. Those nonseparated HOV facilities that do not yet have FSP service are likely candidates for extension of patrol service, since they typically service high volumes of traffic and experience congestion in the main lanes.

CONCLUSION

The present study has considered the advisability of installing emergency call boxes in freeway medians or alongside median HOV facilities. Based on the evidence presented here, it does not appear that median call boxes are warranted except in the case of barrier-separated median HOV facilities with adequate shoulders. The costs of installing median call boxes are expected to be modest (about \$3,500/km/year for 0.4-km spacing), but it appears unlikely that any significant safety advantage could result. The types of accidents likely to be prevented by providing call boxes in medians are extremely rare; meanwhile, a number of call box professionals argue that median call boxes could increase accident risks by increasing the exposure of persons and vehicles on median shoulders. Although certainly not conclusive, evidence suggests that the accident risk is higher on left-hand shoulders than on right-hand shoulders and that the provision of call boxes in medians would encourage stops on the left-hand shoulder.

In the case of barrier-separated HOV lanes, installation of call boxes may be warranted by the fact that there is virtually no access to motorist assistance services otherwise. In the case of other types of median HOV facilities, access to call boxes on the right-hand shoulder is similar to that from the mixed-traffic freeway lanes, so that median call boxes are not warranted.

ACKNOWLEDGMENTS

This research was funded by Caltrans. Special thanks are due to Don Howe of the Caltrans Headquarters Division of Traffic Operations, Vince Zambrana of CHP, and all of the respondents in the surveys of California call box professionals and CHP officers, who are too numerous to mention by name.

REFERENCES

1. Finnegan, S. A. *Estimating Freeway Service Patrol Assists: An Analysis of the Los Angeles County Metro Freeway Service Patrol.* Los Angeles County Transportation Commission, Los Angeles, 1992.
2. *Caltrans Call Box and Motorist Aid Guidelines,* Final Interagency Draft. CHP California Department of Transportation and California Highway Patrol, July 1993.
3. Banks, J. H. *Median/HOV Lane Call Box Study, Final Report.* Civil Engineering Report Series 9401. San Diego State University, San Diego, Calif., 1994.
4. Borden, J. B. *1993 FHWA Administrator's Biennial Safety Awards Competition Submittal, California Department of Transportation Entry #4 for Freeway Service Patrol.* California Department of Transportation, 1993.

Publication of this paper sponsored by Committee on Travelers Services.

Simulator and Field Measures of Driver Age Differences in Left-Turn Gap Judgments

LOREN STAPLIN

Research evaluated the effect of varying approach vehicle speed on judgments of the last safe moment to initiate a left turn at an intersection ahead of oncoming traffic. Young (mean age, 33.3 years), young-old (mean age, 65.1 years), and old-old (mean age, 79.4 years) drivers were tested in a controlled field study and in laboratory studies by using varying simulation techniques. A repeated-measures design acquired the same responses from the same subjects by using the same stimuli under all methodologies. Reliable age differences in both target recognition distance and judged minimum safe gap distance were demonstrated, as was an age × speed interaction for gap judgment. Principal findings indicate a relative insensitivity to vehicle approach speed in left-turn situations by older drivers. It is argued that this produces a reliance on instantaneous judgments of perceived distance alone, disproportionately increasing the risk for older drivers when there is an isolated speeder in the opposing traffic stream. A countermeasure need is thus identified, but countermeasure effectiveness was not investigated in the present research. Furthermore, the present findings suggest that image and scene attributes including high resolution and correct size and perspective cues may be prerequisites for valid and generalizable driving simulation measures of visual sensory and perceptual task performance.

This paper reports on age-related differences in driver performance when individuals must judge the last safe moment to proceed with a left turn ahead of oncoming traffic. The findings of this investigation may be applied directly to the development of engineering countermeasures to reduce intersection traffic maneuver problems of this type. To this end key task demands of the left-turning situation, visual information-processing aspects of gap judgment, and relevant driver performance differences are first reviewed. Laboratory and controlled field experiments are then reported. These document the effects on gap judgment of driver age, vehicle approach speed, and various stimulus image characteristics associated with alternative simulation techniques.

Left-turn traffic maneuvers appear to create special problems for older drivers, as evidenced both by their overinvolvement in this accident category and by self-reports from older road users; furthermore, the specific reason cited most frequently for the involvements of older drivers in accidents in left-turn situations is failure to yield (*1,2*). Assuming that an oncoming vehicle has been detected by a driver waiting to turn left, two broad areas of concern for safe performance in this situation are the driver's understanding of his or her right-of-way status as conveyed by traffic control devices at the intersection and motion perception outcomes that yield appropriate gap judgments for the initiation of turning movements. Parallel research describing a cognitive engineering approach to improve left-turn signal displays has recently been documented (*3*). The present study addresses the sensitivity of gap judgments to driver age and task variables and the importance of such relationships to accident risk reduction for older drivers.

Prior investigations have addressed motion perception abilities pertinent to driving, including time-to-collision (TTC) and gap-acceptance judgments, although only a subset has compared older and younger subjects. In TTC estimates drivers estimate how long it takes, moving at a constant speed, to reach specified points in their paths. They are hypothesized to be based either on an optic-flow process, in which the driver's analysis of the relative expansion rate of an image (such as an oncoming vehicle) over time provides the estimate of TTC directly (*4–6*), or on a cognitive process in which TTC is estimated by using speed and distance information. In the first case the driver relies on two-dimensional information, that is, angular separation cues (the image gets larger), to estimate TTC; in the second case the driver calculates TTC on the basis of three-dimensional information. As reported later, a decline (possibly exponential) in the ability of older subjects to detect angular movement compared with that of young subjects can be described. By using a simulated change in the separation of taillights, indicating the overtaking of a vehicle, threshold elevations of greater than 100 percent were shown for drivers 70 to 75 years old compared with those for drivers 20 to 29 years old for brief exposures at night (*7*). Older persons may in fact require twice the rate of movement to perceive that an object's motion-in-depth is approaching, given a brief (2.0-sec) duration of exposure. Also, research has indicated that relative to younger subjects, older subjects underestimate approaching vehicle speeds, with greater errors of underestimation at higher speeds (*8*). Furthermore, a prior analysis of judgments of the last possible safe moment to cross in front of an oncoming vehicle traveling at lower speeds versus that for one traveling at higher speeds has shown that older persons allowed the shortest time margins at a 96-km/hr (60-mph) approach speed; in fact, older persons accepted a gap to cross at an average constant distance of slightly less than 152.4 m (500 ft), whereas younger subjects allowed a constant time gap and, thus, increased distance at higher versus lower speeds (*9*).

The present investigation sought to replicate the finding of relative insensitivity to vehicle approach speed for older drivers in the left-turn situation, specifically, by using alternative stimulus display techniques in a laboratory simulator, and then to apply a repeated-measures design to validate the gap judgments of the same test sample under controlled field conditions.

SUMMARY OF EXPERIMENTS

A laboratory experiment presented a young to middle-aged and two older driver groups with a single oncoming (target) vehicle

---

Human Factors Division, The Scientix Corporation, Brookview Corporate Center, Suite 3000, 1250 South Broad Street, Lansdale, Pa. 19446.

approaching at low and high speeds by using alternative displays to show the opposing leg of the intersection where the subject was waiting to make a left turn across traffic. Separate blocks of trials used a 20-in. television monitor, a large-screen video projection image, and a large-screen cinematic (35-mm) image to display the filmed approach of an identical target vehicle at the same location. All subjects viewed all target approaches by using all display techniques. In each trial subjects first performed a target recognition response when they could identify the approaching vehicle in the distance across the intersection; then, later during the target vehicle's approach, they indicated their judgment of the last safe moment to proceed with a left turn, yielding distance measures for each dependent variable.

In a following controlled field experiment, the same test sample watched opposing (target) vehicle approaches at the same speeds while sitting behind the wheel in an instrumented vehicle positioned so as to provide the same view across the same intersection as shown previously in the laboratory driving simulator. The same target vehicle filmed earlier for the laboratory stimulus preparation was used in the field trials. Again, subjects made responses of the last safe moment to proceed at the instant during the target's approach when they judged that it had become unsafe to initiate a left turn in front of the oncoming vehicle.

METHODOLOGY

Laboratory Experiment

Subjects

A total of 79 paid test subjects in three age groups (25 young/middle-aged, 29 young-old, and 25 old-old drivers) were recruited for this research through face-to-face, one-on-one solicitations at Pennsylvania photo license centers, where a person's birth date (month of year) is the determining factor as to who appears on any given day. The quasi-random sample obtained in this manner has been shown to provide a more representative range of visual capabilities in older age cohorts relative to those in samples obtained through newspaper advertisements or appeals to large groups of older persons (*10*). The mean age of the subjects in the young/middle-aged group was 33.3 years (range, 20 to 53 years), the mean age in the young-old group was 65.1 years (range, 56 to 72 years), and the mean age in the old-old group was 79.4 years (range, 75 to 91 years). Each age group included approximately 60 percent males and 40 percent females.

A study sample falling within age norms for visual and cognitive performance was defined by using a preliminary test battery: static acuity (corrected), contrast sensitivity, stereo depth perception, forward and reverse digit span, and understanding of spatial relationships indicated by performance on the WAIS-R block design subtest. No evidence of visual pathology or cognitive dysfunction sufficient to excuse any of the sample recruits from participation in this research was found.

Stimulus Materials

The test stimuli for the television, video projection, and cinematic trial blocks were produced from the filmed (30 frames/sec) approach of a white Mercury Marquis sedan on a two-lane highway at a speed of 48 km/hr (30 mph) from the perspective of a driver waiting to turn left onto an intersecting roadway. The approach of the target vehicle began out of sight behind a curve approximately 1.6 km (1 mi) from the camera position; an on-board distance-monitoring computer recorded the target vehicle's position every 1⁄30 sec during its approach, generating a look-up file of its separation distance from the subject (camera position) for every frame. All stimuli were recorded on 35-mm film stock by using a Panavision camera equipped with a 30-mm anamorphic lens. This lens provided a distortion-free field of view of approximately 72 degrees at an effective focal length of 1.5 cm (0.59 in.). A film-to-tape transfer was performed to produce the video master, which subsequently was stored on laser disc for stimulus presentation.

By using the display apparatus as described later, various image characteristics were presented to subjects in each block of trials in the laboratory. The video images were National Television Standards Committee (NTSC) quality; although this format theoretically permits 525 horizontal lines of resolution, postproduction and transfer to laser disc resulted in an effective resolution of only 300 to 350 scan lines. The 35-mm cinematic format, by comparison, displayed an image resolution of more than 3,000 lines. The large-screen display formats preserved correct size and perspective cues, such that the angular change associated with the target's motion in depth provided the same cues available to a driver viewing the scene through the windshield. The 20-in. television monitor display compressed the target stimulus, however, and did not present absolute changes in the angular size of the target that were accurate for its motion in depth as viewed under real-world conditions. Thus, the television monitor trials presented relatively lower resolution images, without correct size and perspective information; the projection video trials presented correct size and perspective information, also at lower resolution; and the cinematic trials presented correct size and perspective information at extremely high resolution.

The visual background of the stimulus scene was an uncluttered rural environment, and the target vehicle was the only vehicle visible in the scene. The luminance of the stimulus scenes in all display types exceeded 100 cd/m^2, as measured with a Pritchard 1980A photometer.

Apparatus

A driving simulator consisting of a Fiat 128 body and frame (with engine and gas tank removed) was used for the large-screen display trial blocks. A single-seat driving buck consisting of a frame without external body panels was used for the television monitor trials. In both data collection systems, a steering wheel-mounted response button was used to obtain target recognition responses, and a switch activated by brake pedal depression was used to record last safe moment to proceed responses.

A Stewart Lumiflex 180 rear-projection screen was used for the video projection and cinematic trials. The screen horizontal dimension was 274 cm (108 in.), and the viewing distance from the subject's eyes to the screen was 188 cm (74 in.). The viewing distance for the 20-in. (Sony) television monitor trials was 61 cm (24 in.). A Pioneer LD-V8000 laser disc player was used for all video trials. For the large-screen video trials a Barcodata 1001 projector was used, with additional signal enhancement provided by an Ikegami DSC-1O5OS digital scan converter. A 35-mm projector with an anamorphic lens was used to display the stimuli for the cinematic trials. A Society of Motion Picture and Television Engineers time code reader was used with the 35-mm projector to identify individual frames of film corresponding to a subject's button-push and brake pedal depression responses in the laboratory.

A 386/50 personal computer (PC) was used to control the presentation order of test stimuli, to initiate the displays of video stimuli, to record the times of the button-push and brake pedal depression responses, and for the video trials, to control the apparent speed of the target vehicle's approach. Since the target approach was filmed at 48 km/hr (30 mph), a PC command to double the playback speed for the laser disc was used to produce a 96-km/hr (60-mph) approach. The high-speed approach with cinematic stimuli was achieved through studio production of a copy of the stimulus film with every other frame removed.

Procedure

Data collection was conducted for one subject at a time in two successive visits to the laboratory. During the initial visit the visual and cognitive screening measures were obtained, and the subject performed the dependent measures by using the large-screen projection video display. During the following visit the television monitor and cinematic trials were performed. Trial order by target approach speed (48 and 96 km/hr) was counterbalanced within blocks for each age group, but all subjects performed the video projection trials first and then the cinematic trials and the television monitor trials. Unfortunately, the expense and limited availability of the 35-mm projection equipment precluded the complete counterbalancing of trials, that is, by display methodology, in the laboratory. At least 1 month elapsed between visits to the laboratory, however, reducing any possibility that learning from the large-screen video trials could have contaminated the cinematic data collection protocol.

After a subject was seated in the simulator and seat adjustments for his or her comfort were completed, a simple reaction time (RT) task was administered by using a button-push response to a light-emitting diode mounted on the dashboard and presented with random delay over seven trials. The quickest and slowest responses were discarded, and the mean of the remaining five trials was recorded as the simple RT or movement time for the subject. This was done to permit a correction for individual differences in RTs when calculating the recognition and gap judgment distances for each test condition, since movement time differences per se (i.e., independent of the information-processing operations underlying gap judgments for approaching vehicles) were not of direct interest in the study.

At the beginning of each trial the experimenter paused the first frame of the stimulus scene to deliver instructions. The target vehicle was not yet visible in the distance in this scene. The experimenter pointed out relevant scene elements to reiterate the scenario of a driver waiting to turn left onto the intersecting roadway and then reminded the subject that two responses were required: (a) press the button on the steering wheel at the earliest moment that you can identify the target vehicle approaching in the distance, and (b) depress the brake pedal at the last possible safe moment to turn in front of the target vehicle. The brake pedal response, although not typically associated with the initiation of a turn, nevertheless yielded cleaner data during pilot testing than, for example, a steering wheel movement in which the precise degree of deflection required to register a response was more ambiguous. None of the subjects evidenced any confusion or difficulty in understanding or performing the brake pedal response.

Controlled Field Experiment

Subjects

The test sample for the field experiment was retained from individuals participating in the prior laboratory testing. Individuals in the same age groups were sampled. Actual sample sizes within each age group in the field study are indicated in the data summary presented later in this section in Table 1.

Stimulus Materials

The same white Mercury Marquis sedan used during filming of the test stimuli for the laboratory study served as the target stimulus for the controlled field experiment. It was driven by a confederate who was in radio contact with the experimenter in the subject vehicle.

Apparatus

The subject vehicle was instrumented with hardware and software systems to monitor the distance traversed from a known reference point by the target vehicle on each trial and to record the subject's gap judgment response when the target reached the last possible safe moment for the subject to turn in front of it. A hand-held response button was used by each subject to perform the gap judgment dependent measure.

A microprocessor linked to the transmission in the target vehicle monitored the distance traveled from its (constant) starting point in each trial. This measurement system was accurate to the nearest 0.3 m (1 ft). The subject vehicle was equipped with a transmitter, activated by the hand-held button-push mechanism, that notified the distance-monitoring computer in the target vehicle the instant that the subject pushed the button to indicate his or her last safe moment response. This signal froze the display of the traversed distance, permitting calculation of target separation distance at the time of response by subtraction from its known distance [slightly under 1.6 km (1 mi)] at the starting point.

Procedure

Subjects were brought by van to the field data collection site, on NJ-29 in Hunterdon County, at the identical location where the laboratory test stimuli had previously been filmed. Traffic on this highway, although light during the data collection period of 10:00 a.m. to 3:00 p.m., was not controlled; therefore, the subject sat in the passenger seat position in the instrumented vehicle and the experimenter occupied the driver's position, to move out of the way of traffic if necessary. This protocol allowed subjects to attend solely to the approach of the target vehicle. Any interrupted trials were repeated.

Once the experimenter had positioned the instrumented vehicle properly at the intersection, she radioed the confederate in the target vehicle when to begin the approach. However, since this site was located on an open roadway, extraneous vehicles periodically entered the opposing lane between the subject and the target position or overtook the target vehicle at high speed from the rear. Data collection was aborted whenever this occurred, but some confusion was possible at the beginning of any given trial as to whether a just-detectable vehicle in the distance was in fact the target. Therefore,

TABLE 1 Mean (M) and Standard Deviation (SD) Distances in Meters for Each Age Group for Each Dependent Measure Under Each Experimental Methodology

	Experimental Methodology																							
	Laboratory: Television Monitor Display						Laboratory: Projection Video Display						Laboratory: Cinematic Display						Field: Instrumented Vehicle					
Age Group	Target Speed: Low			Target Speed: High			Target Speed: Low			Target Speed: High			Target Speed: Low			Target Speed: High			Target Speed: Low			Target Speed: High		
	Target Recognition Distance (m)																							
	n	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD						
Young/middle-aged	22	263	42	22	229	47	25	296	46	25	248	55	23	338	17	23	341	28						
Young-old	26	271	51	26	232	50	28	297	57	28	261	68	25	320	37	24	336	20						
Old-old	21	266	24	21	217	57	24	304	43	23	259	56	20	310	32	21	319	28						
	"Least Safe Gap" Distance (m)																							
	n	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD
Young/middle-aged	22	95	34	22	100	29	25	146	55	25	148	46	23	163	52	23	201	50	12	100	49	11	132	52
Young-old	26	137	56	26	124	42	28	228	67	28	192	68	25	203	67	25	230	56	13	156	50	11	158	33
Old-old	21	137	55	21	119	40	24	253	48	24	222	58	21	216	72	21	229	62	14	166	62	15	161	53

1m=3.28ft

contrary to the laboratory methodology, the experimenter verbally cued the subject to the presence of the target vehicle as it became visible as a point source in the distance, and no target recognition distance data were obtained in the controlled field experiment. After the subject pressed the hand-held response button to perform the dependent measure, the experimenter waited for the target vehicle to pass by and then pulled off the road onto the shoulder while the confederate repositioned the target vehicle at its starting point if another trial was to be performed.

RESULTS

The data from the laboratory and field experiments are summarized in Table 1. Table 1 shows the means and standard deviation for the recognition distance and gap distance measures for each age group as a function of target approach speed and the number of subjects completing data collection under each test condition. In addition, target recognition and gap distance data are provided in Figures 1 and 2, respectively, for each methodology. As noted earlier no recognition distance data were obtained in the controlled field trials. Statistical tests of the main effects and interactions for the age group and target speed variables on each dependent measure were performed by using the General Linear Models Procedure (PROC GLM) in SAS, with program options for repeated-measures designs within each block of trials corresponding to a single experimental methodology.

A significant effect of age group on the judged minimum safe gap was demonstrated by all laboratory test methodologies such that increasing subject age resulted in larger gap requirements. The same trend was observed in the controlled field data, but it failed to reach significance. The magnitude of the age group effects was given by $F = 20.66$ [degrees of freedom (df) = 2; $P < .0001$] by the video projection methodology, $F = 4.48$ (df = 2; $P < .01$) by the television monitor methodology, and $F = 3.21$ (df = 2; $P < .05$) when the target stimulus was presented cinematically. In addition, the effects of age group on target recognition distance were demonstrated by using the 35-mm film stimulus display methodology, because younger subjects recognized the target vehicle at significantly greater distances than older subjects ($F = 6.04$; df = 2; $P < .004$). No reliable effect of age group on this dependent measure was found for the video projection or television monitor methodologies, and target recognition distance was not measured in the controlled field trials.

Differences in target approach speed resulted in significant differences in both target recognition distance and the judged minimum safe gap for all laboratory methodologies. Overall, increasing target speed led to significantly shorter minimum safe gap judgments ($F = 9.57$; df = 3; $P < .0001$) as well as shorter target recognition distances ($F = 51.51$; df = 3; $P < .0001$) by using video projection to present the test stimuli. For the television monitor data the same pattern was observed: decreasing minimum safe gap judgments ($F = 5.63$; df = 1; $P < .02$) and shorter target recognition distances ($F = 71.83$; df = 1; $P < .0001$) with increasing target approach speed. With cinematic stimulus presentation, however, just the opposite results were demonstrated, that is, significant increases in the judged minimum safe gap ($F = 31.55$; df = 1; $P < .0001$) and target recognition distance ($F = 5.25$; df = 1; $P < .03$) as target speed increased. In the field trials only the gap judgment measure was obtained; for these data an increase in the judged minimum safe gap with increasing target approach speed ($F = 14.28$;

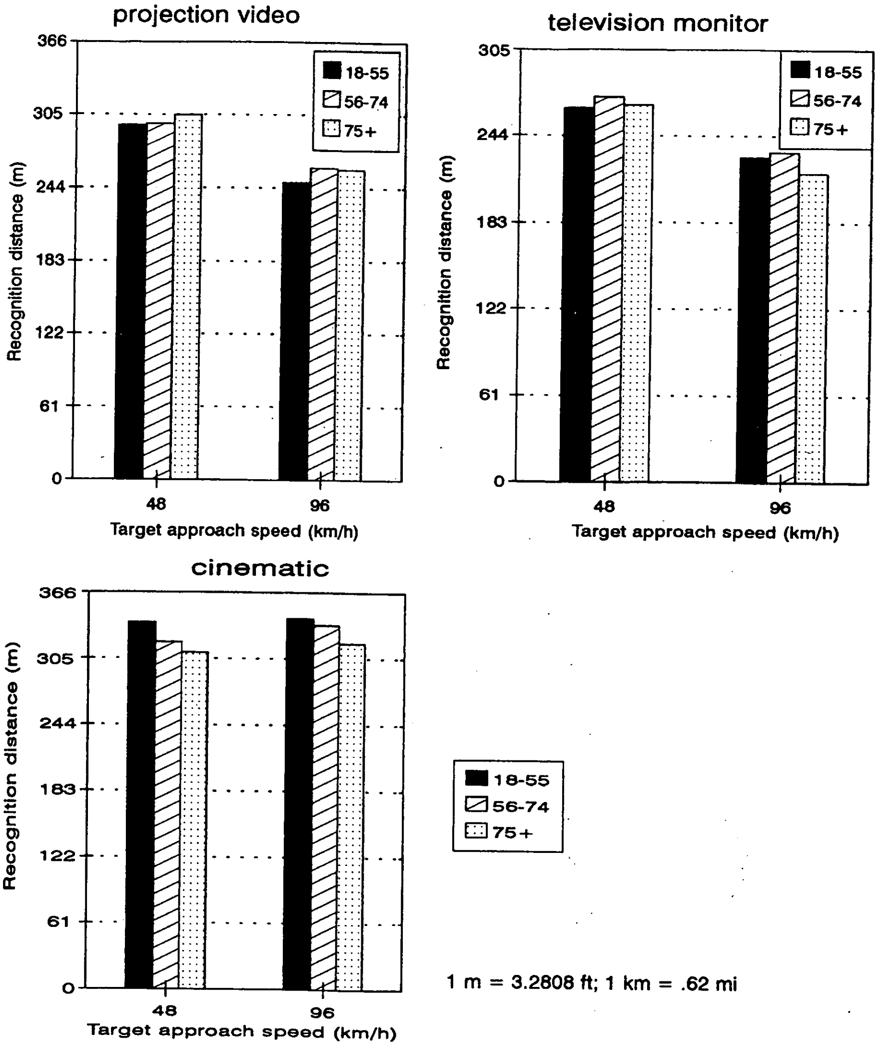


FIGURE 1 Target recognition distance as function of approach speed for three driver age groups for projection video, television monitor, and cinematic stimulus presentation methodologies.

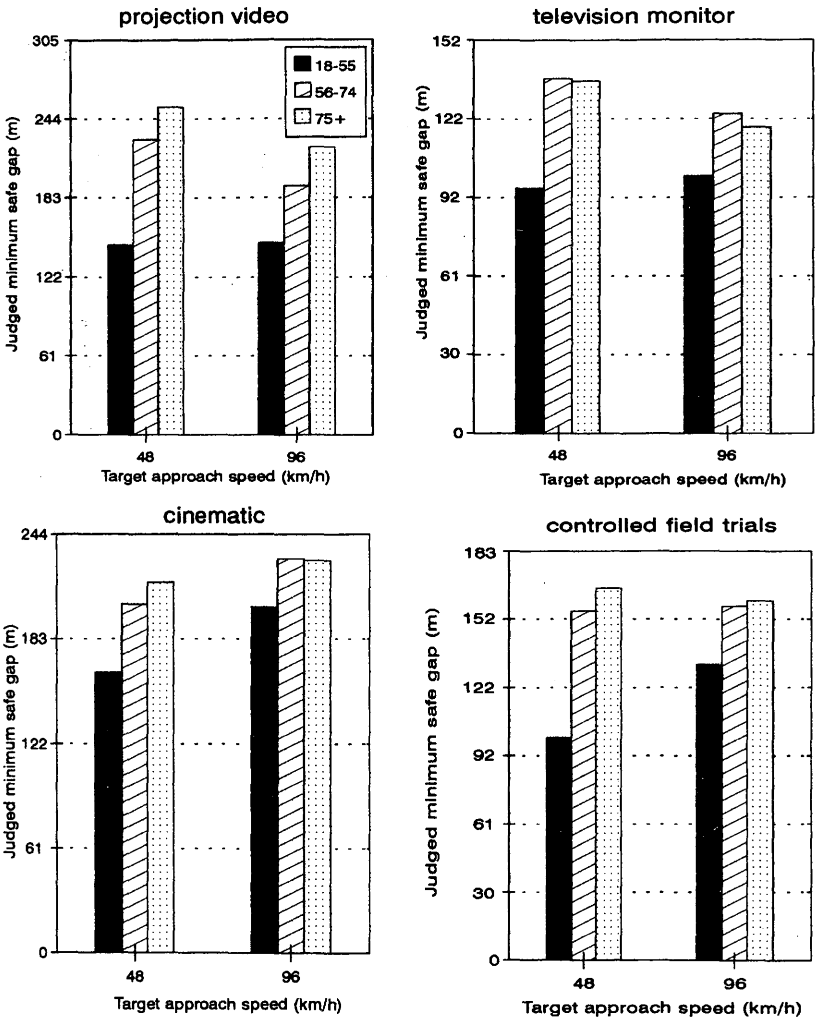
df = 1; $P < .0009$) mimicked the trend observed in the cinematic data in the laboratory.

Most interesting were significant interaction effects on gap judgment involving subject age group and target approach speed, which were demonstrated for the video projection and television monitor methodologies in the laboratory and for the controlled field trials. By using video projection to present test stimuli the interaction effect was demonstrated because the judged minimum safe gap for young/middle-aged drivers remained relatively constant across target speeds, whereas both the young-old and old-old drivers accepted smaller gaps as target speed increased ($F = 4.95$; df = 6; $P < .0001$). This identical pattern was also found in the television monitor data ($F = 3.14$; df = 2; $P < .05$). The results of the controlled field trials, however, differed markedly: the judged minimum safe gap of young/middle-aged drivers increased along with target speed, whereas the responses of young-old and old-old drivers were insensitive to this independent variable ($F = 4.49$; df = 2; $P < .02$). For the cinematic laboratory data, the pattern of differences paralleled that observed in the field, but at $P < .09$ it failed to reach significance.

GENERAL DISCUSSION OF RESULTS

The primary research hypothesis, that older drivers would experience a relative insensitivity to vehicle approach speed in left-turn situations, was reliably demonstrated in controlled field trials and was underscored by a consistent pattern of results by using cinematic stimuli in the laboratory. Under these test conditions, as one would hope, increasing conflict vehicle speed caused young drivers to increase their judgments of the minimum safe gap to initiate a turn while waiting at an intersection for oncoming traffic to clear. Older driver gap judgment distances, however, did not change significantly for a 48-km/hr (30-mph) versus a 96-km/hr (60-mph) target approach speed.

Not only did the proportional change in gap judgment from one target speed to another differ markedly as a function of driver age, but it also differed as a function of the method of stimulus presentation. Although significant interactions between age and target speed were demonstrated in the video projection and television monitor laboratory trials, these data were characterized by unchanging gap judgments across speed by younger subjects and by



1 m = 3.2808 ft; 1 km = .62 mi

FIGURE 2 Gap distance judgments as function of approach speed for three driver age groups for projection video, television monitor, and cinematic stimulus presentation methodologies.

substantial decreases in minimum safe gap size judgments by older subjects for faster target approaches.

The study's findings thus merit discussion in two important contexts: the design of engineering countermeasures to improve the safety of older drivers in performing left turns at intersections and the validity of simulator measurements of driver perceptual and cognitive responses by using various display techniques.

Two complementary countermeasure strategies have the highest potential to ameliorate older driver problems in turning situations: (a) cue the older (turning) driver to the presence of vehicles approaching at significantly higher-than-expected speeds (i.e., those that exceed the posted limit by a fixed amount) and (b) and slow down through traffic and make through drivers more aware of the potential for a conflict ahead with a turning vehicle. Potential means for achieving these behavioral goals include speed-actuated active warning devices for turning drivers, rumble strips for through traffic on intersection approaches, and special permissive phase signal treatments that may induce greater caution among turning drivers (e.g., flashing yellow or flashing red instead of the steady green ball now most widely used as the permissive phase treatment).

Somewhat less apparent is the extent to which the simulation techniques used in this research imposed limitations on the abilities of drivers of different ages to process motion cues to extract information about the speed-distance relationships of other vehicles. Since it must be assumed that display artifacts were absent from the controlled field data, the observed pattern of responses for younger and older drivers can be interpreted as reflecting a diminished capability of older subjects to process motion cues under conditions in which there is no deficiency of information in the image. When a deficiency of spatial information exists it might be expected to also result in distortions in the judgments of younger subjects. Specifically, the loss of high-frequency spatial cues might be expected to level performance across age groups, because (younger) subjects with the capability of using such information do not have it available to them.

The similarities of response patterns in the cinematic data to those in the field data suggest that the film methodology provided sufficient information for valid gap judgments, whereas differences associated with the video projection and television monitor data indicate one or more deficiencies in the information provided by these stimulus presentation methods. The key differences in the attributes of each display type are a loss of realistic target size and perspective cues with the NTSC signal of the television monitor, a loss of image resolution (high spatial frequency information) but preservation of correct perspective with video projection, and realistic size and perspective cues plus a high image resolution with the 35-mm cinematic stimuli.

The smaller minimum safe gap judgments of the older subjects by the video projection and television methodologies but not the cinematic methodology may reflect a floor effect in spatial information-processing capability for a degraded (low-resolution) image. Logically, more processing effort will be required to reach a confident judgment regarding target size when viewing a diffuse image versus a sharply defined image. The instantaneous processing of size cues for a diffuse target should therefore be interfered with to a greater extent by increasing target speed, and for observers with diminished motion perception capabilities, a minimum sampling interval may also be reached. This increased processing difficulty alleged to occur at higher speeds could produce a lower target recognition distance for degraded images, and the minimum sampling interval, the suggested information-processing floor effect, would correspondingly result in a lower gap judgment distance at 96 than at 48 km/hr. Although the target recognition distance also was reduced for the young/middle-aged subjects by using the video projection and television monitor displays, their hypothesized greater efficiencies in spatial information processing could have compensated to a degree for this limitation in the availability of relevant spatial cues.

With the cinematic display target recognition distance did not decrease for any age group for the lower versus the higher target approach speed. Thus, in the absence of significant image degradation, the spatial information that drivers seek for gap judgments was available at a distance sufficiently beyond the perceived minimum safe gap that afforded an adequate processing time to all subjects for the cognitive operations underlying speed or distance estimation; that is, no floor effect of the sort suggested earlier occurred. Under these circumstances it makes sense that the observed differences in the mean response magnitudes for gap judgments between groups can be accounted for strictly in terms of individual (group)-related diminished capabilities, as opposed to stimulus-bound factors.

An alternative interpretation of the present findings deserves comment. There is some appeal to couch the results of subjects' gap judgments in this research in terms of TTC values instead of target separation distances. Of course, it is individuals' perceived TTC that dictates their go–no go responses in this driving situation. Earlier research has demonstrated that there is a general underestimation of TTC such that as absolute TTC increases, error in judging TTC also increases (*11*). More critical is an understanding that perceived TTC is a derived construct: it depends on a prior perception of approaching target speed. Any interpretation of the present findings within a TTC framework must take this contingent relationship, as well as the perceptual distortion noted earlier, into account.

In conclusion, achieving valid measures of age differences in vehicle motion perception and associated maneuver decisions by using laboratory driving simulation techniques appears contingent on the highest possible realism in presenting the driving situation to which the results are to be generalized. At a minimum to understand the relationships between driver age and operational factors such as the driver's estimate of the speed of an oncoming vehicle, the preservation of a high image resolution plus the use of correct size and perspective cues appear to offer clear advantages in driving simulation research.

ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Transportation (FWWA) under the technical direction of Carole Simmons. The author, who served as principal investigator on this project, wishes to acknowledge Simmons' critical review of earlier drafts of the manuscript.

REFERENCES

1. Hauer, E. The Safety of Older Persons at Intersections. In *Special Report 218: Transportation in an Aging Society,* Vol. 1 and 2, TRB, Washington, D.C., 1988.
2. Staplin, L., and R. W. Lyles. Age Differences in Motion Perception and Specific Traffic Maneuver Problems. In *Transportation Research Record 1325,* TRB, National Research Council, Washington, D.C., 1992.
3. Staplin, L., and A. D. Fisk. A Cognitive Engineering Approach to Improving Signalized Left Turn Intersections. In *Human Factors,* Vol. 33, 1991, pp. 559–571.
4. Gibson, J. J. *The Senses Considered as Perceptual Systems.* Houghton Mifflin, Boston, Mass, 1966.
5. Lee, D. N. Visual Information During Locomotion. In *Perception: Essays in Honor of James J. Gibson* (R. B. McLeod and H. L. Pick, eds.), Cornell University Press, Ithaca, N.Y., 1974.
6. Cavallo, V., O. Laya, and M. Laurent. The Estimation of Time-to-Collision as a Function of Visual Stimulation. In *Vision in Vehicles* (A. G. Gale et al., eds.), Elsevier Science Publishers B. V., North-Holland, New York, 1986.
7. Lee, D. N. A Theory of Visual Control of Braking Based on Information About Time-to-Collision. In *Perception,* Vol. 5, 1976, pp. 437–459.
8. Scialfa, C., L. Guzy, H. Leibowitz, P. Garvey, and R. Tyrrell. Age Differences in Estimating Vehicle Velocity. In *Psychology and Aging,* Vol. 6, No. 1, 1991.
9. Hills, B. L., and L. Johnson. *Speed and Minimum Gap Acceptance Judgements at Two Rural Junctions.* Report SR515. Department of the Environment, Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1980.
10. Staplin, L., K. Lococo, and J. Sim. *Traffic Control Design Elements for Accommodating Drivers with Diminished Capability.* Report FHWA-RD-9-055. FHWA, U.S. Department of Transportation, 1990.
11. Schiff, W., and M. L. Detwiler. Information Used in Judging Impending Collision. *Perception,* Vol. 8, 1979, pp. 647–658.

The author assumes sole responsibility for the opinions and recommendations expressed herein.

Publication of this paper sponsored by Committee on Simulation and Measurement of Vehicle and Operator Performance.

Modeling of Motorway Operations

M. BRACKSTONE AND M. MCDONALD

Over the last 5 to 10 years rising levels of traffic flow on motorways have lead to the development of increasingly advanced intelligent vehicle-highway systems dedicated to improving capacity, stability, and safety. Many of these advances are to be achieved by modifying behavior or introducing a degree of vehicle control. To fully understand the effects of these mechanisms, however, it is necessary to develop and use appropriate microscopic simulation models. Accurate modeling of microscopic driver behavior presents several difficulties, ranging from the requirement for high-quality (dynamic) data on real driving behavior through the methodology used for cross-checking between simulated and observed behavior to the philosophical basis of modeling itself. These issues are discussed, the degree of shortfall present in current simulation and analysis techniques is examined, and potential guidelines for increasing the validity of microscopic simulation models are suggested.

With the rapid growth in motorway traffic driver behavior is perceived as becoming of increasing importance to safety and capacity. Measures designed to address these problems, formulated as part of the PROMETHEUS and DRIVE initiatives in Europe and VERTIS in Japan, face an increasingly difficult task because of the costliness of potential field trials, both financially and in possible increases in risk exposure to the driving public. Therefore, the use of simulation models as a cost-effective way of investigating intelligent vehicle-highway system (IVHS) solutions at a fundamental level has become an increasingly attractive alternative, allowing complex investigations to be conducted in a repeatable and controllable manner.

Although forming a highly diverse group, such models can broadly be divided by two criteria:

1. The scale of the traffic system modeled:
   –Urban regions, where vehicle interactions are primarily related to nodal points in a network (such as signalized intersections) and link behavior is relatively unimportant [e.g., CONTRAM, SATURN, NETSIM (*1–3*)], and
   –Interurban regions, where interactions typically occur through grade-separated junctions and link behavior becomes increasingly relevant because of the high proportion of travel time spent between the nodes, for example, SIMAUT (*4*).
2. The basis on which traffic dynamics are simulated:
   –Macroscopic formulation, in which variables such as average traffic density and speed over specific sections are used to define traffic behavior, for example, by expressing disturbances in traffic density as shock waves in a continuous medium through the use of continuity equations [e.g., META (*5*)], and
   –Microscopic formulation, in which the interactions of individual vehicles with each other is taken explicitly into account and calibrated or validated by using parameters such as speed, which can be observed as distributions at key points and reproduced through Monte Carlo-based simulations [e.g., Autobahn Simulator AS (*6*)].

As interest in the control of motorway networks has grown over the last 5 to 10 years, macroscopic interurban models have become comparatively commonplace and have been applied with some success. For example, motorway control algorithms have been tested through the simulation of motorway orbital routes surrounding major cities (*5*); these algorithms consist of a series of 10 to 20 junctions with spacings of about 1 km. However, as the complexity of IVHS technology rapidly increases, the degree of detail used in these models and available in their output is becoming more and more unsuitable. This is because consideration of the effect that advanced (individual) vehicle control systems (e.g., advanced intelligent cruise control) may have on the operation of the motorway as a whole is increasingly needed. This effect can be represented accurately only at a microscopic level.

Microscopic models are predominantly based on two fundamental behavioral mechanisms that are considered to describe the basic decision-making processes undertaken during motorway driving:

1. Car following, which describes the longitudinal speed–distance relationships adopted by a driver when following another vehicle. A range of such theories is in use. These are typically based on either the acceleration stimulus provided by the followed vehicle [according to relative speed and separation (*7*)] or maintenance of a set deceleration and an attempt to follow at a set stopping distance (*8*). For a more comprehensive review of these models see Brackstone and McDonald (*9*).
2. Lane changing, which governs the circumstances in which a vehicle moves into an adjacent lane. This may be considered in terms of a number of perception thresholds, the crossing of which activates the next stage in a maneuver or process. For example, the speed advantage that would be obtained if a vehicle were to change lanes to pass a slower vehicle in front may be required to exceed a set amount for the maneuver to be considered worthwhile (*10,11*).

These mechanisms are frequently enhanced by researchers as more scenarios are needed and more empirical evidence is accumulated. Typical modules attached to most models at a later stage of their development may include the following:

- Entry or exit behavior to cope with the short-headway lane change decision that is common near or at motorway merges and diverges, when drivers are forced to accept gaps that are substantially smaller than that which they would usually prefer to accept. This allows the simulation of technologies such as ramp metering or off-ramp control by vehicle type (*12*).
- Lane preference factors to allow the testing of alternate lane usage strategies, for example, goods vehicle lanes and speed banding. This typically involves continual refinement of the lane change

---

Transportation Research Group, Department of Civil Engineering, University of Southampton, Southampton, SO17 1BJ, United Kingdom.

algorithm to incorporate a wider range of factors contributing to lane choice, such as the effects of goods vehicles slowing on an incline (*13*).

- Intelligent vehicle applications, incorporating automatic speed or headway control algorithms. Modifications may also be required to model how driving behavior may change in the presence of such devices, that is, changes to the desired deceleration rates used by the driving population.

A brief list of the capabilities of the main models in use in European national and international research programs known to the authors is given in Table 1.

As this process of enhancement inevitably continues, the research community is forced to confront a growing need for data with which to validate these models. This is a far from straightforward task, however, with a range of almost philosophical traps awaiting the unwary:

- The use of inappropriate data.
- The use of inappropriate validation or calibration procedures caused by a lack of understanding of model sensitivity.
- An increase in discontinuities within the modeling.

These issues will be discussed in this paper, and ways in which microscopic modeling may need to develop in the remainder of this decade are suggested.

DEFINITIONS

In many references the distinction between calibration and validation is unclear, and the following definitions are given for clarification: *Validation* is undertaken at a macroscopic scale to ensure that the overall behavior of a model matches that readily observable. *Calibration* is undertaken at a microscopic level with regard to individual vehicle-to-vehicle interactions; it is the tuning of the many behavioral parameters comprising the decision-making processes.

USE OF INAPPROPRIATE DATA

By using the definitions given in the preceding section, two types of data are needed to check if a simulation model is correctly formulated. Validation data, which can be obtained in a comparatively straightforward manner and which typically consists of minute-by-minute records of flows, average speeds or headways, and traffic stream composition, can be easily extracted from loop data or video recordings of the road taken from a suitable viewpoint. Calibration data can also be obtained, but it is far more difficult to obtain these data because of the number and diversity of the parameters involved and because the parameters are often not directly related to easily observable quantities.

Typical records that would be required to examine the behavior or response of any one vehicle during the calibration process may include the following:

- Type, speed, relative speed, and distance to the vehicle being followed. In more critical situations this could be extended to the next but one vehicle to the front and even the vehicle to the rear.
- Types and speeds of adjacent vehicles in neighboring lanes. Data from these vehicles can be used to calculate the sizes and relative speeds of gaps into which the vehicle could move.
- Local flow or density and relative proportions of vehicle types, an approximate measure of how busy the driver believes the road to be.
- Geometry, that is, vertical and horizontal curvature of the road. This may affect visibility, desired speeds, and braking thresholds.

The majority of these parameters describe dynamic quantities, such as the relationship between the acceleration of a vehicle and changes in the relative speeds of and distances to surrounding vehicles. Thus, to fully understand such a process one must attempt to sample data at a rate similar to that at which changes may take place. This is an exceptionally difficult problem, however, because most of the variables are required relative to the probe vehicle, which is moving along the road and hence cannot be sampled sufficiently frequently by any method that relies on making observations at a set point.

TABLE 1 European Microscopic Interurban Simulation Models

	Details			
Model (Country of Origin, Reference)	Developer	Primary Current Research Use	Calibration Notes	Additional Modules
AS (D, 6)	Benz AG	PROMETHEUS		AVCS[a]
MISSION (D, 11)	IfV. Univ. Karlsruhe	National	Uses data from basic instrumented vehicle	AVCS[a]
SISTM (UK, 12, 13)	Wootton-Jeffreys-Atkins Ltd.	National (TRL/DoT)		EE[b], ML[c]
SPEACS (I, 14)	CSST	PROMETHEUS	Uses SCAN driving simulator, for direct input to data base	AVCS[a], EE[b]?
PELOPS (D)	IfK. RWTH, Univ. Aachen	PROMETHEUS		AVCS[a]
MONET3 (D)	RWTH, Univ. Aachen	PROMETHEUS		RC[d]

[a] Advanced Vehicle Control System modules
[b] Entry-Exit behaviour
[c] Enhanced for multi-lane vehicle distribution
[d] Roadside IVHS Communication protocols

A method frequently cited as being a solution to this problem is the monitoring of the behavior of a vehicle as it passes along a stretch of road by the use of a series of video cameras. This was attempted in the United Kingdom in 1993 (*13*), in which a series of video cameras was set up on a high embankment overlooking the M27 motorway. Each camera was set up to observe a 50-m longitudinal section of the road, and from the synthesis of data collected from the group of cameras over a total of 500 m, it was possible to track vehicles and measure their speeds and separations (Figure 1).

Analysis focused on attempting to find relationships between the relative speed between a lane-changing vehicle and up to four neighboring vehicles and the associated headway time (allowing an examination of questions such as, Do vehicles pull out earlier to lane change if they are moving faster?). Despite the presence of apparent trends in plots of these variables (Figure 2), regression analysis of the data did not enable any underlying relationships to be identified. A wide scatter was present in the data, and typical R^2 values (describing the degree to which the empirical data can be explained by the optimum linear relationship possible) were of the order of ~2 percent (i.e., only 2 percent of the variation observed could be explained by assuming the presence of a linear relationship).

Although the multicamera approach provided useful and previously unavailable data on lane-changing rates, the method did not allow the development of any significant findings on the individual processes under investigation, that is, gap acceptance on lane changing. This was considered to be primarily through the inaccuracy of the video-based measurement technique (5 to 7 percent in speed) and the inability to monitor vehicles over a sufficiently long stretch.

The main data collection alternative to this method is to place a single driver and his or her vehicle in a controlled environment where its response can be regularly and consistently measured. The obvious candidate here is to use vehicle simulators. However, at present these are generally restricted to simple scenarios, and by their nature they are an artificial environment in which errors in perception are not a potential matter of life and death. A more realistic option is to use an instrumented vehicle in which a vehicle driven in the traffic stream is used as a platform from which to observe the behaviors of adjacent drivers. Past applications of this technique have been very limited, but recent work (*15,16*) with laser and radar sensors seem to have decreased the measurement inaccuracy caused by so hostile an environment and may well yield high-quality behavioral data in the next few years as increased exploitation is made of sensor technology made available through advanced vehicle control system programs.

CALIBRATION AND VALIDATION PROCEDURE

Simulation models are validated to a large extent according to the particular phenomena of traffic flow that they are designed to investigate. Thus, microscopic models are validated according to the model's ability to produce flow breakdown and speed–flow relationships, whereas network-based models such as CONTRAM are assessed according to average flows observed on key links or nodes or queue lengths at particular junctions. This type of approach produces a validation or calibration that is necessarily restricted to a

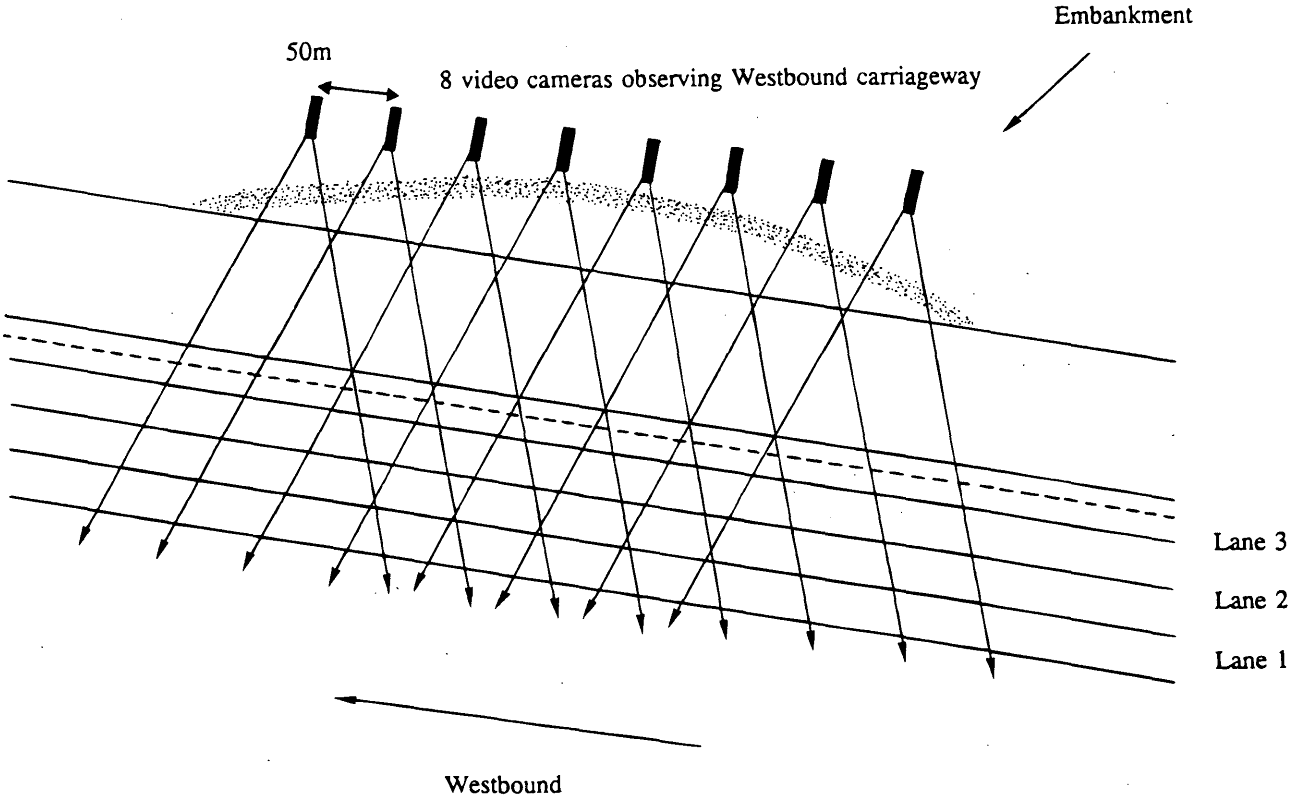


FIGURE 1 Schematic of video-based vehicle tracking method showing separation against closing speed between vehicle changing to a faster lane and an obstructing vehicle.

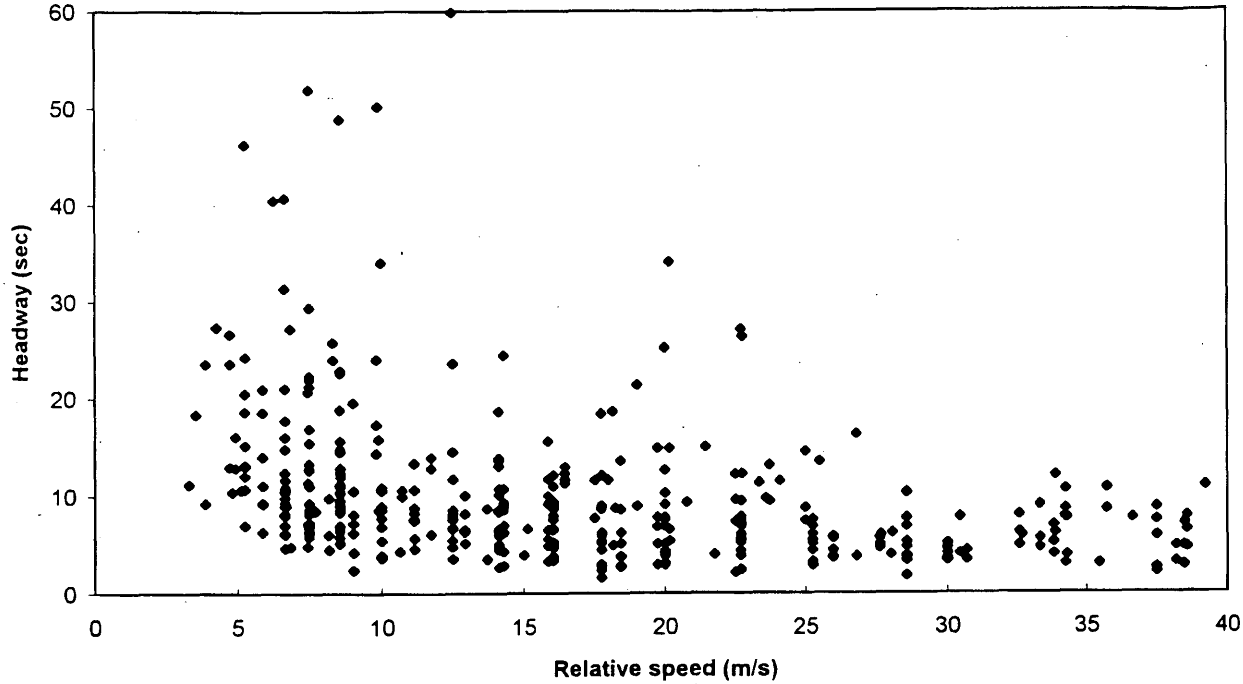


FIGURE 2 **Example of data collected from video-based tracking experiment.**

small number of readily observable phenomena that can be adequately replicated in most models, but that does not necessarily ensure that the model is an accurate descriptor over a wide variety of circumstances or in sufficient detail.

On attempting to validate a simulation model, the user is faced with the prospect of the variation of many parameters, all of which could potentially affect the particular measure under investigation. Although it may be possible to identify parameters that would have an easily detectable macroscopic impact through a common sense approach (e.g., increasing tolerance of speed differences would reduce lane-changing rates), this process is often hampered by an absence of information on model sensitivities. Although this may not initially seem to be a problem of any great significance, it poses potentially the most dangerous trap into which simulation-based traffic research can currently fall. However, with the current quality of the available data, no true microscopic calibration can be performed, and this has resulted in the use of a mesoscopic calibration in which an attempt is made to validate a number of more fundamental parameters that are not explicitly present in the model's formulation or that are not easily observable.

As an illustration of this point consider the calibration of a quantity such as the distribution of gap sizes accepted when moving into an offside lane. This quantity is not microscopic, because it is a by-product of many other factors contributing to the lane change decision, such as speeds and relative speeds of the vehicles involved. Nor is it macroscopic, however, because it cannot easily be measured without the use of specialized methods and does not itself comprise one of the fundamental measures of traffic flow.

Because of the use of this hybrid validation–calibration method, it is possible to arrive at a set of calibrated model parameters that produce a reasonable macroscopic fit and that may intuitively seem to be correct but that may nonetheless be an incorrect combination, producing inappropriate behavior in a measure not examined. For example, reproducing speed–flow relationships or lane splits to a high degree of accuracy does not guarantee that the simulated lane change rate (a potential measure of risk exposure) will be as accurate (*13*). The more detailed the behavior that one attempts to simulate, the greater the potential for a mismatch between the true microscopic calibration required and the mesoscopic calibration achieved.

Even if large amounts of new data were to become available, an improvement on this standard approach would still be required and would entail the testing of the decision-making processes in isolation from the simulation run as a whole. This would additionally involve the simulation of many individual events, varying the distributions of the behavioral parameters involved. Because many hundreds of runs of each scenario would thus be required, this would entail a program of investigation that would be of a magnitude similar to that of the data collection task itself.

MODELING DISCONTINUITIES

As this paper has progressed it has posed a series of questions and definitions as to what may constitute an accurate calibration. From the earlier sections it can be seen that such an answer requires a large amount of data to be matched under a wide range of circumstances by an equally large quantity of model output. Every time that a new question arises or a new set of behaviors is required, the model must be updated and revalidated.

By using traditional modeling philosophy an accurate calibration for universal application cannot be achieved, because the modeled behavior will be posed in terms of a growing set of if–then–do rules, which can easily become contradictory and discontinuous. The modeling task would therefore seem to require a forever-increasing amount of effort to produce each successive stage of detail or validity. The true driving process is obviously different, moving smoothly from one task to another, and is represented more accurately by the presence of a degree of error within each decision-making process. Until recently, such an approach to behavioral or control science would have been deemed conceptually impossible; however, with the introduction of so-called fuzzy reasoning such a treatment could now be undertaken. By this method a degree of confidence could be attached to each observable in a decision-making process, which itself would use linguistically defined variables by categorizing relative distance, for example, as close, very close, or dangerously close.

Although the use of fuzzy logic in traffic simulation is still in its infancy (*17*), it is clear that such an approach could provide a powerful tool in the simulation of driver error and overreaction and may enable investigators to bypass the ever-growing set of rules and parameters that characterizes microscopic models.

SUMMARY

A number of problems surrounding the use of microscopic motorway simulation models have been examined, and an attempt has been made to identify the steps required to produce a major upgrade in their validity and transferability. Two key points have been identified:

1. A more critical degree of calibration is necessary if models are to truly fulfill their description as microscopic, with individual driving processes being examined in isolation and receiving dedicated attention from specific data collection exercises, which themselves reflect the dynamic nature of the process.
2. A degree of uncertainty must be introduced into the microscopic decision-making processes. This may seem like a contradiction, but it is likely to produce a far more realistic description of real behavior, because drivers do not have perfect reactions, identical perceptions, or access to exact data from which to calculate their course of action.

In conclusion, for the validity of microscopic simulation models to advance for use as assessment tools in the development of IVHS, road design, and even driver education or legislation, it will probably prove to be necessary to shift emphasis away from traditional concepts, such as the quantity of data used in validation and calibration, to increasing the suitability of data and the relevance of the calibration processes and replicating the true nature of the behavior that is being modeled.

REFERENCES

1. Leonard, D. R., and P. Gower. *User Guide to CONTRAM Version 4.* Transportation Research Laboratory Supplementary Report 735. Transportation Research Laboratory, Crowthorne, United Kingdom, 1982.
2. Hall, M. D., D. Van-Vliet, and L. G. Willumsen. SATURN: A Simulation-Assignment Model for the Evaluation of Traffic Management Schemes. *Traffic Engineering and Control*, Vol. 21, No. 4, 1980, pp. 168–176.
3. Peat, Marwick, Mitchell and Co., General Applied Science Laboratories, Inc. *Network Flow Simulation for Urban Traffic Control System.* Final Report, Contract No. DOT-FH-11-7462. FHWA, U.S. Department of Transportation, 1971.
4. Morin, J. M. SIMAUT: Un Programme de Simulation du Trafic Autoroutier. Report 4. INRETS, France, 1985.
5. Papageorgiou, M., J. M. Bloseville, and H. Hadj-Salem. Macroscopic Modelling of Traffic Flow on the Boulevard Peripherique in Paris. *Transportation Research B*, Vol. 23, No. 1, 1989, pp. 29–47.
6. Benz, T. Traffic Flow Effects of Intelligent Cruise Control. *Proc., 26th ISATA Conference*, ATT/IVHS Symposium, Aachen, Germany, 1993, pp. 755–760.
7. Chandler, R. E., R. Herman, and E. W. Montroll. Traffic Dynamics: Studies in Car Following. *Operational Research*, Vol. 6, 1958, pp. 165–184.
8. Gipps, P. G. A Behavioural Car Following Model for Computer Simulation. *Transportation Research B*, Vol. 15B, 1981, pp. 105–111.
9. Brackstone, M, and M. McDonald. Validity of Microscopic Modelling of Motorway Traffic. *Proc., Intelligent Vehicles Symposium*, Paris, Oct. 24 to 26, 1994.
10. Wootton Jeffreys Consultants Ltd. SISTM, Batch Version User Guide. Transportation Research Laboratory, Crowthorne, United Kingdom, 1990.
11. CEC DRIVE Contract V1052 (ICARUS). Final Report. CEC DGXIII Brussels, Belgium 1992.
12. Harwood, N. *An Assessment of Ramp Metering Strategies using SISTM.* Project Report 36 (N012). Transportation Research Laboratory, Crowthorne, United Kingdom, 1993.
13. Brackstone, M., M. McDonald, and D. Jeffery. Simulation of Lane Usage Characteristics on 3 Lane Motorways. *Proc., 27th ISATA Conference*, ATT/IVHS Symposium, Aachen, Germany, 1994.
14. Broqua, F., G. Lerner, V. Mauro, and E. Morello. Co-Operative Driving: Basic Concepts and a First Assessment of "Intelligent Cruise Control" strategies. *Proc., DRIVE Conference on Advanced Transport Telematics in Road Transport.* Brussels, Belgium, Feb. 4 to 6, 1991, pp. 908–929.
15. Brackstone, M., and M. McDonald. An Instrumented Vehicle for Microscopic Monitoring of Driver Behaviour. *Proc., 1993 VNIS Conference*, Ottawa, Ontario, Canada, Oct. 1993.
16. Reiter, U. Empirical Studies as Basis for Traffic Flow Models. *Proc., 2nd International Symposium on Highway Capacity*, Sydney, Australia, Aug. 1994.
17. Yikai, K., J. Satoh, et al. A Fuzzy Model for the Behaviour of Vehicles to Analyse Traffic Congestion. *Proc., International Congress on Modelling and Simulation 1993*, University of Western Australia, Perth, 1993.

Publication of this paper sponsored by Committee on Simulation and Measurement of Vehicle and Operator Performance.

Review of Legibility Relationships Within the Context of Textual Information Presentation

HELMUT T. ZWAHLEN, MURALI SUNKARA, AND THOMAS SCHNELL

An extended review of the relevant legibility literature was conducted to provide normalized legibility performance data for a comparison and consolidation of past legibility research. The data were normalized by expressing the legibility performance in terms of visual angle subtended by the character height. The data revealed large variations in visibility performance among the reviewed studies, despite similar or even identical experimental treatments. The normalized data were grouped into sets, relating the visual angle to the width-to-height ratio W/H, the intercharacter spacing-to-height ratio S/H, and the stroke width-to-height ratio SW/H, for both negative and positive contrast. Second-order polynomial least-squares functions were established to obtain a proposed and tentative functional relationship between the visual angle and W/H, S/H, and SW/H. As expected the data indicated that positive-contrast characters generally require smaller stroke widths than negative-contrast characters and that more widely spaced characters show an increased legibility over closely spaced characters. The present investigation provides display designers with proposed analytical functional relationships between legibility performance (visual angle) and typographical properties.

Visual displays could be devices such as traffic signs, license plates, computer cathode ray tubes or flat panel displays, televisions, or even pages in a book. However simple or complex they are, visual displays are used to transmit visual information to a human receiver. For a display to be effective its message must be visible, distinguishable, and easily interpretable (*1*). For this reason it is important that the displayed material be maximally legible.

In one of the earliest studies Forbes (*2*) adopted the term *legibility* to indicate a subject's ability to read the characters on a traffic sign. In another early study Aldrich (*3*) indicated that the intrinsic legibility of license plates will depend on the combined effects of size and shape of the plate, height, width, style, stroke width, spacing, and grouping of characters.

Most studies reviewed in the present investigation agree that legibility is affected by factors such as but not limited to character height (H), character width (W), stroke width (SW), height-width ratio (H/W), height-stroke width ratio (H/SW), intercharacter spacing (S), interword spacing, and interline spacing, as well as possible interactions between those factors. Optimal legibility under given conditions may therefore be achieved by an arrangement of the stimulus material in which the typographical factors mentioned earlier are coordinated to produce optimal viewing conditions and easy and rapid reading with adequate comprehension. Designers generally increased the character size if better legibility was required in the past (not always possible because of aesthetics or limited space). There is still a large degree of uncertainty as to which fonts perform well in terms of legibility and which ones do not. Standard fonts have been established, especially for use on traffic signs on highways (*4*).

Many of the studies reviewed in the present investigation have focused their efforts on developing minimum or maximum permissible values for the various factors that affect legibility. However, most of those studies were generally concerned with the absolute size of the characters rather than expressing legibility performance as a function of the visual angle subtended by the characters.

STUDIES ON HEIGHT, WIDTH, STROKE WIDTH, AND SPACING OF CHARACTERS

Forbes et al. (*5*) established the legibility distances of highway destination signs in relation to H, W, and reflectorization using black-on-white standard series B (narrow) and series D (wide) characters for six different character heights: 15.24, 20.32, 25.4, 30.48, 45.72, and 60.96 cm (6, 8, 10, 12, 18, and 24 in.). The wider series D characters were more legible. Legibility distances of 15.24 m (50 ft) and 10.06 m (33 ft) per inch of character height were obtained for the wide and the narrow characters, respectively, under daylight and normal vision (6/6) conditions (nighttime values were 15 percent lower). Forbes investigated pure legibility without considering limited viewing time (in the driving context, typically from 0.2 to 0.8 sec), compromised visual acuity, or reduced contrast sensitivity (mostly in elderly individuals). The data from Forbes could be adjusted with correction factors to account for these constraints.

In a field experiment Uhlaner (*6*) studied the effect of SW on the legibility of a black-on-white 7.62-cm (3-in.) block (height = width) capital characters. An SW of 18 percent of H (SW = H:5.5) was recommended. It was suggested by Uhlaner that SW needs to be reduced as H/W is reduced. Uhlaner's study was mainly limited to daylight legibility with illumination levels between 2690 and 5918 lx (250 and 550 fc). The use of seven different H/SW ratios in the study gives the reader a fairly good idea about the parabolic nature of legibility performance as a function of the H/SW ratio (see Figure 6).

Berger (*7*) experimented with H, W, SW, form, and horizontal S of black numerals on a white background (negative contrast) and white numerals on a black background (positive contrast). Berger recommended SW/H ratios of 1:8 for a positive contrast and 1:13 for a negative contrast. Positive contrast was better recognized than

---

Human Factors and Ergonomics Laboratory, Department of Industrial and Systems Engineering, Ohio University, Athens, Ohio 45701-2979.

negative contrast. In another experiment conducted by Berger (*8*) five different numeral widths (1.5, 2.0, 2.75, 3.3, and 4.15 mm) for a numeral height of 6 mm were investigated by using the two black-on-white numerals 0 and 5. The legibility distance increased with increasing W. Berger (*9*) described the effects of varying both H and W on character legibility. Reportedly, legibility increased with increasing H and also with increasing W. Berger's experiments provide valuable quantitative information on the effects of the SW, H, and W of characters on legibility and were therefore used in the present investigation.

Kuntz and Sleight (*10*) established the H/SW ratio that was optimal for reading black-on-white and white-on-black numerals. The seven different H/SW ratios of 1:3.5, 1:4, 1:4.5, 1:5, 1:5.5, 1:6, and 1:6.5 were used. Kuntz and Sleight recommended an optimal H/SW ratio of 1:5 for reading both positive- and negative-contrast stimuli. It should be noted that the authors found no significant contrast polarity effect. This result is in conflict with the results of other researchers (*7,11*), who found an influence of contrast polarity on legibility performance. However, the lack of a contrast polarity effect may be attributed to a low display luminance.

TABLE 1 Studies Conducted on W/H Ratio of Characters

Author(s)/ Year	Ratios Investigated	Recommended Ratio	Polarity	Comments
*Forbes (1939)	0.43, 0.67	Not specified	Dark on Light	Legibility of highway destination signs
*Berger (1948)	0.25, 0.33, 0.46, 0.55, 0.69	Not specified	Dark on Light	Legibility of numbers
*Berger (1950)	0.25, 0.29, 0.36, 0.40, 0.45, 0.47, 0.57, 0.63, 0.73	Not specified	Dark on Light	Legibility of numbers and letters
*Forbes et al. (1951)	0.81	Not specified	Light on Dark	Legibility of highway signs
Soar (1955)	0.30, 0.45, 0.60, 0.75	0.60 or 0.75	Unavailable	Legibility of numbers
*Solomon (1956)	0.54, 0.73, 0.79	Not specified	Light on Dark	Legibility of highway signs
Brown et al. (1953)	0.55, 0.70, 0.85, 1	1	Light on Dark	Legibility of letters on aircraft control panels
Benson et al. (1988)	0.25, 0.41, 0.48, 0.56, 0.64, 0.72, 0.84, 1	Not specified	NA	Legibility of characters on visual display terminals
*Mace et al. (1993)	0.54, 0.67	Not specified	Dark on Light	Legibility of signs
*Mace et al. (1993)	0.79	Not specified	Light on Dark	Legibility of signs

The optimum width to height ratios were not specified in many studies because it was not the objective of these studies to determine the optimum ratio

In the Soar (1955) study no information regarding the polarity was found

The * indicates that the data from those studies were used in establishing the functional relationships shown in Figures 1 and 2

Forbes et al. (*12*) compared lowercase and uppercase characters displayed on highway signs. White-on-black series E capital characters and lowercase characters of approximately the same average W/H ratio were used. To approximate the effects of word patterns (as opposed to character legibility) and word familiarity, three sets of measurements were made by (a) using scrambled characters, (b) using California place names being viewed for the first time, and (c) using California place names being viewed for the second time.

In a field experiment Case et al. (*11*) analyzed the effects of intercharacter S and interline spacing on the legibility of 76.2-mm (3-in.)-high series E (SW/H ratio = 1:6) black-on-white and white-on-black characters. The two intercharacter spacings of 38.1 and 101.6 mm (1.5 and 4 in.) and two interline spacings of 50.8 and 101.6 mm (2 and 4 in.) were used. Widely spaced characters were much more legible than closely spaced characters. A significant interaction between contrast polarity and spacing was reported (the negative-contrast treatment was slightly more legible with close spacing; the positive-contrast treatment was considerably more legible with wide spacing). No conclusions were made regarding the interline spacing. The results are valid only for SW equal to 1:6.

Soar (*13*) studied the interaction between the W/H ratio and SW on numeral legibility. The four W/H ratios 3:10, 4.5:10, 6:10, and

TABLE 2 Studies Conducted on SW/H Ratio and S/H Ratio

A: Studies Conducted on the Stroke-Width to Height Ratio of Characters

Author(s)/ Year	Ratios Investigated	Recommended Ratio	Polarity	Comments
Aldrich (1939)	0.08, 0.125	0.125	Dark on Light	Legibility of license plates
*Uhlaner (1941)	0.08, 0.12, 0.16, 0.20, 0.24, 0.28, 0.32	0.18	Dark on Light	Legibility of letters
Berger (1944)	0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2	0.075	Light on Dark	Legibility of numbers
Berger (1950)	0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.3	0.125	Dark on Light	Legibility of numbers
Kuntz et al. (1950)	0.15, 0.17, 0.18, 0.20, 0.22, 0.25, 0.29	0.2	Light on Dark	Legibility of numbers
Kuntz et al. (1950)	0.15, 0.17, 0.18, 0.20, 0.22, 0.25, 0.29	0.2	Dark on Light	Legibility of numbers
Brown et al. (1949)	0.07, 0.1, 0.13, 0.17, 0.2	0.17	Light on Dark	Legibility of letters on aircraft control panels
Brown et al. (1951)	0.07, 0.12, 0.125, 0.14, 0.17	0.125	Light on Dark	Legibility of numbers on aircraft control panels
Soar (1955)	.0625, 0.1, 0.125, 0.2	0.1	Unavailable	Legibility of numbers
Baerwald et al. (1960)	0.08, 0.104, 0.125, 0.15	0.08	Dark on Light	Legibility of license plates
Baerwald et al. (1960)	0.08, 0.104, 0.125, 0.15	0.15	Light on Dark	Legibility of license plates
Hodge (1962)	0.08, 0.11, 0.14, 0.18, 0.22, 0.27	0.18	Dark on Light	Legibility of letters

The * indicates that the data from this study were used in establishing the functional relationships shown in Figure 6

B: Studies Conducted on the Spacing Between Characters to Height Ratio of Characters

Author(s)/ Year	Ratios Investigated	Recommended Ratio	Polarity	Comments
Case et al. (1952)	0.50, 1.33	NA	Light on Dark	Legibility of highway signs
Case et al. (1952)	0.50, 1.34	NA	Dark on Light	Legibility of highway signs
*Solomon (1956)	0.19, 0.228, 0.266, 0.286	0.28	Light on Dark	Legibility of highway signs
Baerwald et al. (1960)	0.08, 0.17, 0.28, 0.33, 0.42	NA	Dark on Light	Legibility of license plates
Baerwald et al. (1960)	0.08, 0.17, 0.28, 0.33, 0.43	NA	Light on Dark	Legibility of license plates

The optimum spacing between characters to height ratios were not specified studies because it was not the objective of these in many studies to determine the optimum ratio

The * indicates that the data from this study were used in establishing the functional relationships shown in Figures 1 and 3

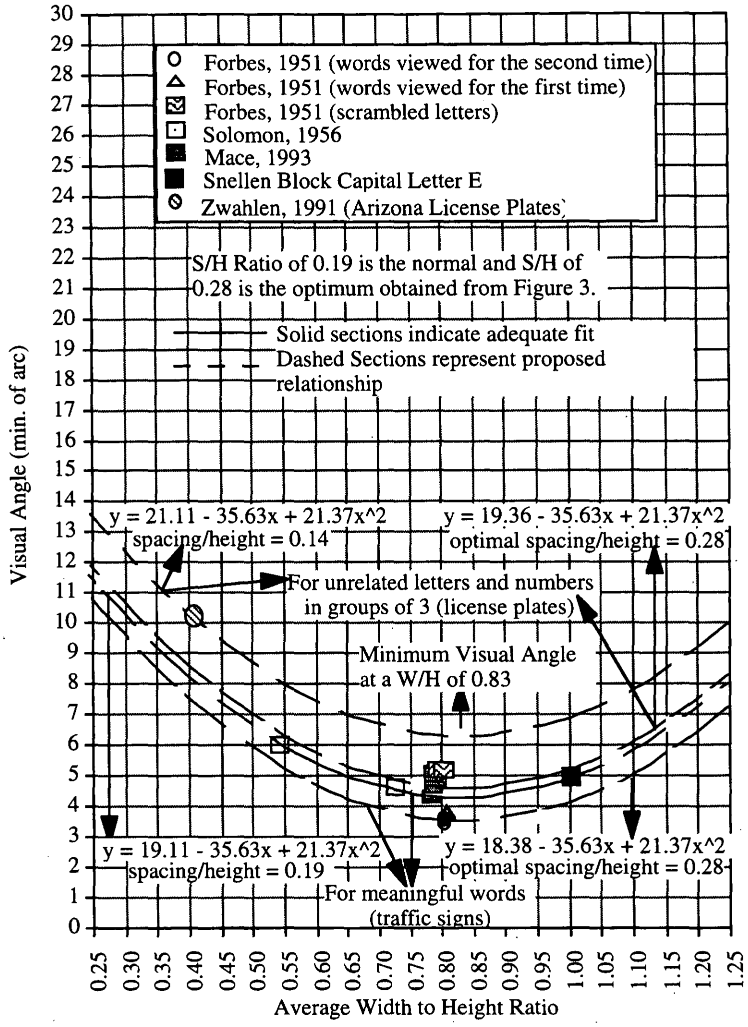


FIGURE 1 Visual angle as function of average W/H ratio for positive contrast.

7.5:10 at three different stroke widths were used. The SW/H ratios ranged from 1:5 to 1:8 for the widest SW and from 1:10 to 1:16 for the narrowest SW. A SW/H ratio of 1:10 was recommended to be optimal (no W/H and SW interaction). A W/H ratio of 10:7.5 was recommended for optimum legibility. However, Soar's experimental results suggest that most of the numerals showed consistent trends of increased legibility as W was increased.

Solomon (*14*) conducted a field experiment to determine the effects of W and intercharacter S on the nighttime legibility of highway signs with white-on-black characters. Characters with an H of 25.4 cm (10 in.), based on standard series C (narrow), series E (wide), and series ED (similar in form but slightly thinner than series E), were used in that study. S was increased from normal to 20, 40, and 60 percent above normal. Increasing S resulted in a legibility distance increase for all three font types up to the point where S was 40 percent above normal. The effects of increasing the H of a narrow series C character to the point where the area was equal to the area of a wide series E character were also studied. The two character types were equally legible for identical areas. This leads to the conclusion that a designer may use intercharacter S as an additional degree of freedom to compensate for constraints on H.

Baerwald et al. (*15*) investigated the factors affecting the legibility of automobile license plates using 76.2-mm (3-in.)-high standard series B numerals. A wide range of SWs was tested, and a significant interaction between SW and contrast polarity was found. An optimal SW of 11.113 mm (7/16 in.) or, alternatively, an optimal SW/H ratio of 1:6.8 was recommended for positive contrast. An optimal SW of 6.35 mm (1/4 in.) or, alternatively, an optimal SW/H ratio of 1:12 was recommended for negative contrast. Experiments conducted on S also indicated a significant interaction between S and the contrast polarity as well as between S and SW. Baerwald et al. concluded that the use of a thicker SW at a narrow internumeral S would have a greater effect on legibility than would a larger internumeral S. At a large internumeral S the positive-contrast numerals were more legible. However, when the internumeral S was decreased to a minimum, the negative-contrast numerals became more legible. These findings seem to confirm the previously described observations made by Case et al. (*11*).

Mace and Garvey (*16*) conducted a daytime study regarding the effects of increased H on sign legibility with a hypothesis that for a given color, contrast, and character series the legibility index will remain the same for different character heights. They further hypothesized that the legibility index would remain constant for different fonts, both contrast polarities, and different observer ages. They used two display groups with black series C and series D characters on a white background, respectively. Furthermore, they used one group with white characters on a green background and the modified series E font. Their results indicate that the legibility distance curves flatten between 30.48 and 40.64 cm (12 and 16 in.) of H for both young and older drivers and that extending H beyond 40.64 m (16 in.) might not be practical.

With the exception of the somewhat limited study conducted by Case et al. (*11*) (only two intercharacter S's and interline S's investigated) no other traffic sign-related studies dealing with intercharacter S and interline S were located. The visual display terminal legibility literature only provided the requirements of the Human Factors Society (*17*) (which state that "a minimum space of one character width shall be used between words and a minimum of two stroke widths or 15 percent of character height, whichever is greater shall be used for spacing between lines of text"). The *National Park Service Sign Manual* (*18*) provides some recommendations with regard to spacing, but no references were cited to support the recommended values.

Many other investigators, like Lauer (*19*), Aldrich (*3*), Mitchell and Forbes (*20*), Brown and Lowery (*21–23*), Schapiro (*24*), Crook et al. (*25,26*), Hughes (*27*), Hodge (*28*), and Benson and Farrell (*29*), have conducted studies dealing with either SW/H ratios or W/H ratios. These works were evaluated, and it was decided not to include the results in the present investigation because of the incompatibility or the incompleteness of the data presentation.

OBJECTIVES

The present investigation had a twofold objective: (a) to conduct an extensive literature review and to consolidate all of the relevant leg-

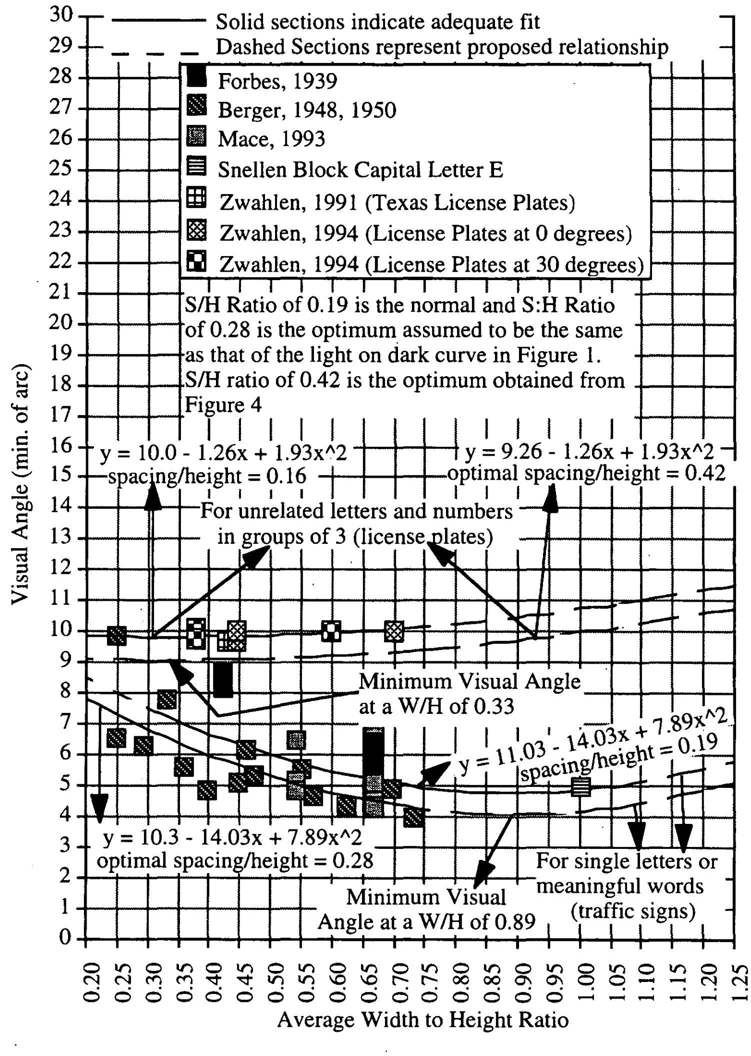


FIGURE 2 Visual angle as function of average W/H ratio for negative contrast.

ibility results from all the studies found in the literature and (b) to establish functional relationships based on least-squares curve-fitting techniques for legibility (in terms of visual angle subtended by H) as a function of the W/H ratio, intercharacter S/H ratio, and SW/H ratio for light on dark (positive contrast) and dark on light (negative contrast) characters.

METHOD

The studies discussed in the preceding literature review were first categorized into three major groups: (a) effects of W/H ratio on legibility, (b) effects of SW/H ratio on legibility, and (c) effects of intercharacter S/H ratio on legibility. Character dimensions and spacing dimensions were related to character height and were expressed in terms of dimensionless ratios (W/H, SW/H, S/H). Summary results of the previously reviewed studies are listed in Tables 1 and 2.

DEFINITIONS

Equation 1 shows the definition of contrast (*c*) as used in this investigation:

$$C = \frac{L_C - L_B}{L_B} \tag{1}$$

where L_C is the character luminance and L_B is the background luminance.

Equation 2 was used to calculate the visual angle (in minutes of arc) based on the character height and the legibility distance (in meters):

$$\text{Visual angle (minutes of arc)} = \frac{3438 \cdot \text{character height (m)}}{\text{Legibility distance (m)}} \tag{2}$$

The normalized visual angles expressed as a function of the W/H ratio, the intercharacter S/H ratio, and the SW/H ratio tended to be parabolically shaped (see Figures 1 to 6). Therefore, a second-order polynomial least-squares fit to the data is proposed as a method of obtaining a tentative functional relationship to express the legibility performance as a function of typographical properties for the given, limited data. The parabolas plotted in Figures 1 to 6 should therefore be considered proposed relationships. The dashed sections of these parabolas indicate extrapolated ranges (no data available) and are of a very tentative nature.

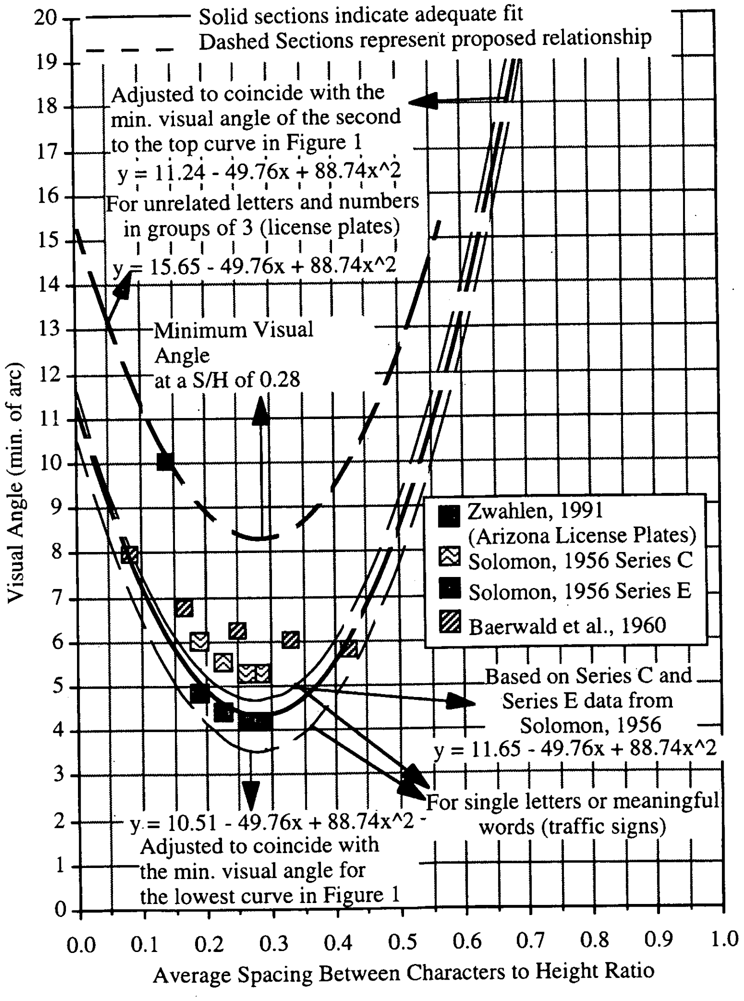


FIGURE 3 Visual angle as function of average intercharacter S/H ratio within a word for positive contrast.

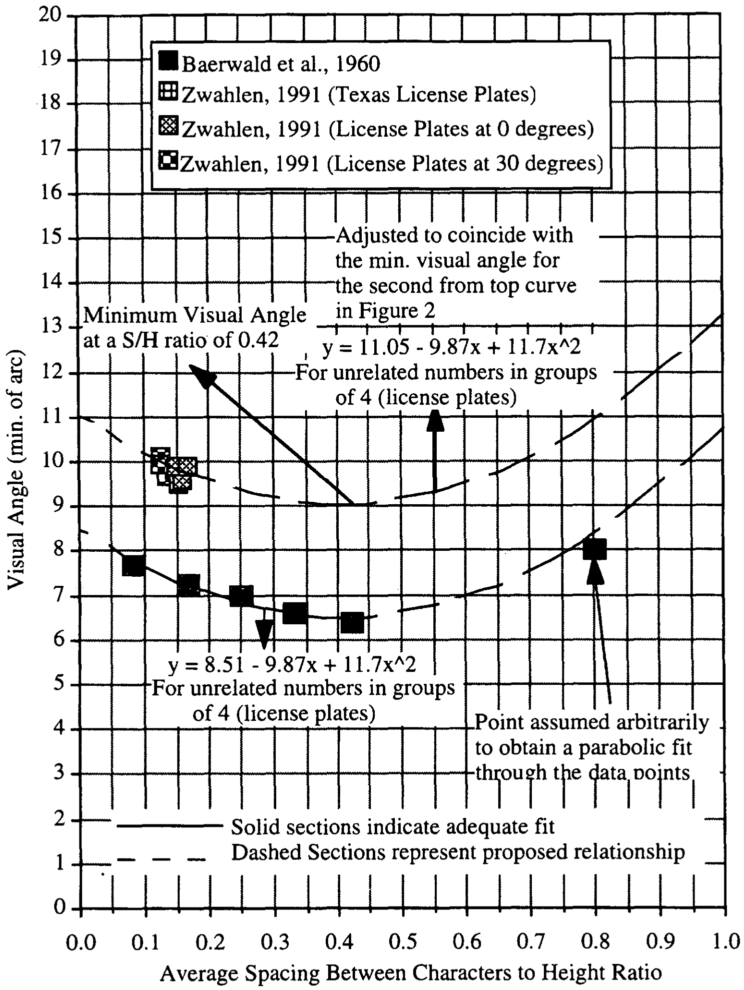


FIGURE 4 Visual angle as function of average intercharacter S/H ratio for negative contrast.

RESULTS AND GENERAL COMMENTS

It should be noted that the functional relationships and the optimal values obtained in this review are based on the information available in the literature, which was sometimes incomplete and most likely valid only for a limited set of conditions. Therefore, a number of assumptions and adjustments had to be made in almost all data sets to correct for experimental artifacts. In some cases it was not possible to establish the second-order polynomial fit because of incomplete experimental descriptions. Certain unknown and unspecified experimental conditions may be responsible for some of the rather large variations between the results of some of the studies. The use of a second-order polynomial least-squares fit to relate legibility to the W/H ratio, the S/H ratio, or the SW/H ratio is certainly not primarily justified by the presence of many and well-dispersed datum points. On the other hand the availability of relatively simple mathematically based functional relationships provide the opportunity to examine legibility trade-offs in an analytical manner.

Figure 1 shows the functional relationship between the visual angle (legibility measure) and the average W/H ratio for positive contrast. The Snellen E (block capital character in which the W/H ratio is 1) was added to have at least one average W/H ratio beyond 0.83. The bottom parabola and the second parabola from the top are based on an optimal S/H ratio of 0.28 (from Figure 3). The top parabola is based on an S/H ratio of 0.14. With the exception of the Snellen block capital E the second parabola from the bottom is based on an average S/H ratio of 0.19, which is fairly representative of series D characters. The remaining parabolas were vertically offset to fit the corresponding datum points. From Figure 1 it can be seen that meaningful words (traffic signs) appear to be more legible than unrelated characters (license plates) for positive contrast.

Figure 2 shows the functional relationship between the visual angle (legibility measure) and the average W/H ratio for negative contrast. For the bottom parabola an optimal S/H ratio of 0.28 was adopted from the positive-contrast data set (Figure 1). The second parabola from the bottom is based on an average S/H ratio of 0.19 (except for Snellen block letter). The top parabola in Figure 2 is based on an average S/H ratio of 0.16. For the second parabola from the top, an optimal S/H ratio of 0.42 was adopted from the data in Figure 4. From Figure 2 it can be seen that meaningful words (traffic signs) appear to be more legible than unrelated characters (license plates) for negative contrast.

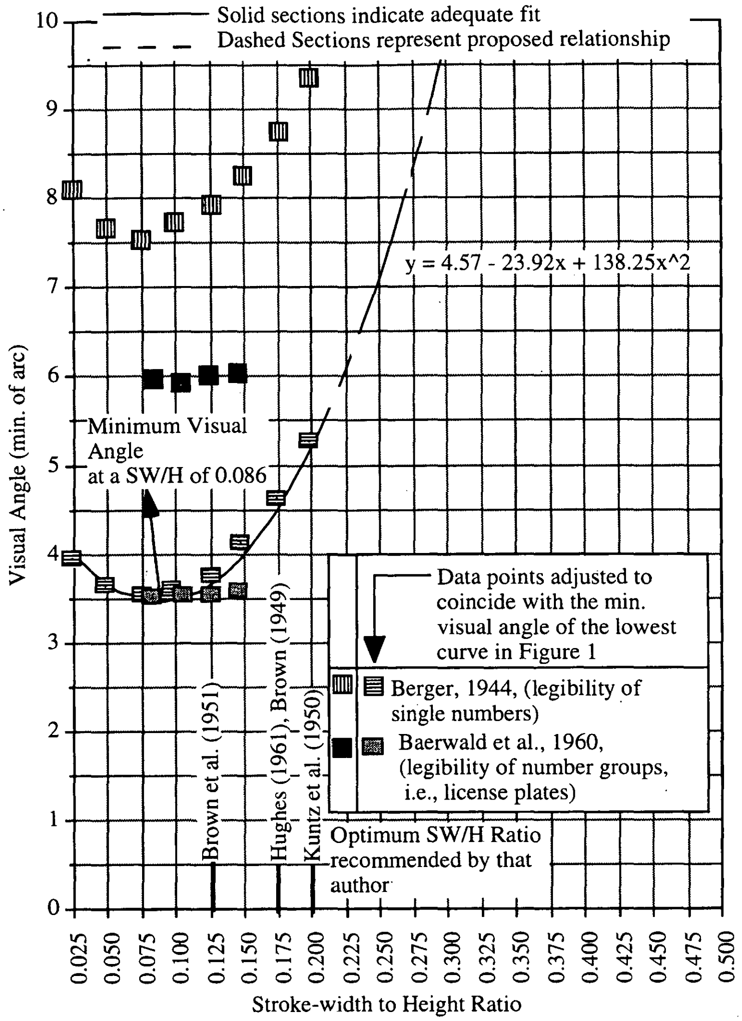


FIGURE 5 Visual angle and adjusted visual angle as function of SW/H ratio for positive contrast.

Figure 3 shows the functional relationship between the visual angle (legibility measure) and the average S/H ratio for positive contrast. The bottom parabola in Figure 3 was vertically adjusted to the minimum visual angle obtained from the bottom parabola in Figure 1. Similarly, the second parabola from the bottom was vertically adjusted to the minimum visual angle obtained from the second parabola from the top in Figure 1. The shape of the top parabola was assumed to be the same as that of the second parabola from the top (constant offset upward to 15.65 min of arc), because only one datum point was available.

Figure 4 shows the functional relationship between the visual angle (legibility measure) and the average S/H ratio for negative contrast based on data from Baerwald et al. (*15*) and Zwahlen (*30,31*). The top parabola was obtained by assuming the minimum visual angle to be equal to the minimum visual angle of the second parabola from the top in Figure 2.

Figure 5 shows the visual angle (legibility measure) and the derived functional relationships as a function of the SW/H ratio for positive contrast. As shown in Figure 5 the visual angle data obtained from Berger (*7*), which provides an optimum SW/H ratio of 0.075, is significantly different from the visual angle data obtained from Baerwald et al. (*15*), which provides an optimum SW/H value within a range of 0.075 to 0.15. Figure 5 also shows the recommended optimum SW/H ratios by

1. Brown and Lowery (*21*): SW/H ratio of 0.17 for the legibility of uniform stroke capital characters viewed in three character groups used on transilluminated control panels in military aircraft;
2. Hughes (*27*): SW/H ratio within a range of 0.125 to 0.17 for the legibility of single numbers;
3. Brown and Lowery (*22*): SW/H ratio of 0.125 for the legibility of uniform stroke capital characters viewed in three character groups used on transilluminated control panels in military aircraft; and
4. Kuntz and Sleight (*10*): SW/H ratio of 0.2 for the legibility of numbers.

The datum points presented in Figure 5 were then adjusted to coincide with the minimum visual angle of the bottom parabola in Figure 1.

Figure 6 shows the visual angle (legibility measure) and the derived functional relationships as a function of the SW/H ratios recommended by various investigators for negative contrast. The

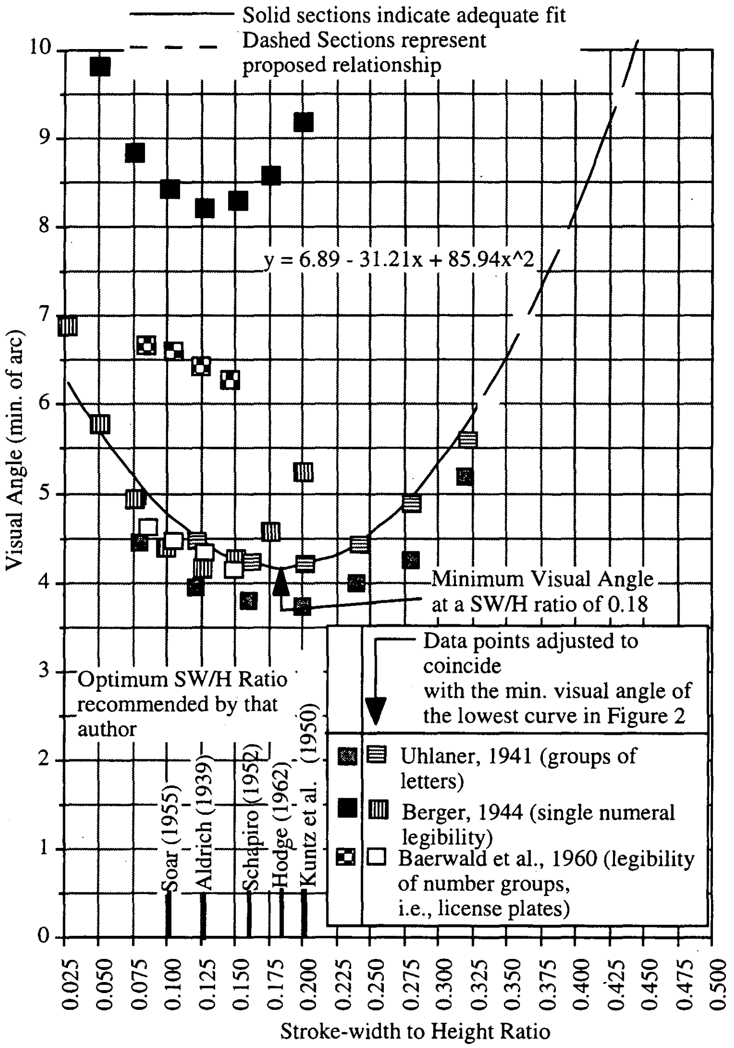


FIGURE 6 Visual angle and adjusted visual angle as function of SW/H ratio for negative contrast.

wide variation in these recommendations could be attributed to different experimental conditions. The datum points plotted in Figure 6 were adjusted to coincide with the minimum visual angle of the lowest parabola in Figure 2 to obtain the second-order polynomial functional relationship. Figure 6 also shows the optimum values recommended by other investigators, which are as follows:

1. Aldrich (*3*): SW/H ratio of 0.125 for the legibility of alphanumeric license plates with varying number of characters and numbers on each plate,
2. Berger (*7*): SW/H ratio of 0.125 for the legibility of single numbers,
3. Kuntz and Sleight (*10*): SW/H ratio of 0.2 for the legibility of numbers,
4. Soar (*13*): SW/H ratio of 0.1 for the legibility of single numbers,
5. Baerwald et al. (*15*): SW/H ratio of 0.15 for the legibility of license plates (four numbers on each plate),
6. Schapiro (*24*): SW/H ratio within a range of 0.125 to 0.2 for the legibility of single numbers, and
7. Hodge (*28*): SW/H ratio of 0.178 for the legibility of unrelated five- or six-character words.

SUMMARY AND CONCLUSIONS

Only a few studies that contained the necessary information to calculate the visual angles (legibility measure) as a function of either a character dimension or a spacing dimension were considered in the present investigation. Intercharacter spacing has been a subject of much controversy and has been studied by many researchers as a function of contrast polarity and a concept known as irradiation (for positive contrast). Irradiation may be highly dependent on the luminance.

It should be noted that the second-order polynomial functions Figures 1 to 6) extend in some cases into ranges of W/H, S/H, or SW/H, for which no data from the studies were available (tentative proposal by the authors). Until more appropriate values become available these proposed functional relationships could be helpful for display designs in which the overall available display space is limited or when displays are being designed for elderly individuals. For a fairly simple font a designer may want to optimize legibility by minimizing the visual angle on the appropriate parabola (Figures 1 to 6) and by selecting the corresponding optimal typographical factors (W/H, S/H, SW/H).

It was found that single characters and meaningful words are more legible than unrelated characters and numbers, both for positive and negative contrast. Although the recommended SW/H ratios varied considerably from study to study, it was generally found that characters displayed with a positive contrast require smaller SWs than characters displayed with a negative contrast. This may partially be attributed to irradiation, which may become a more and more serious problem as the display luminances increase. Furthermore, the reviewed data indicated that more widely spaced characters provide improved legibility over closely spaced characters. A possible explanation for this improved legibility was provided by Case et al. (*11*) in terms of eye fixation shifts.

REFERENCES

1. Kantowitz, B. H., and R. D. Sorkin. *Human Factors: Understanding People-System Relationships.* John Wiley and Sons, Inc., New York, 1983.
2. Forbes, T. W. *Human Factors in Highway Traffic Safety Research.* Wiley-Interscience, New York, 1972, pp. 95–109.
3. Aldrich, M. H. Perception and Visibility of Automobile License Plates. *HRB Proc.*, Vol. 17, 1939, pp. 393–412.
4. *Standard Alphabets for Highway Signs and Pavement Markings,* Metric Edition. Office of Traffic Operations, FHWA, U.S. Department of Transportation, 1977.
5. Forbes, T. W., and R. S. Holmes. Legibility Distances of Highway Destination Signs in Relation to Letter Height, Letter Width and Reflectorization. *HRB Proc.*, Vol. 19, 1939, pp. 321–355.
6. Uhlaner, J. E. The Effect of Thickness of Stroke on the Legibility of Letters. *Proc., Iowa Academy of Sciences,* Vol. 48, 1941, pp. 319–324.
7. Berger, C. Stroke-Width, Form and Horizontal Spacing of Numerals as Determinants of the Threshold of Recognition. *Journal of Applied Psychology,* Vol. 28, No. 3, 1994, pp. 208–231.
8. Berger, C. Some Experiments on the Width of Symbols as Determinant of Legibility. *Acta Ophthalmologica,* Vol. 26, No. 4, 1948, pp. 517–550.
9. Berger, C. Experiments on the Legibility of Symbols of Different Width and Height. *Acta Ophthalmologica,* Vol. 28, No. 4, 1950, pp. 423–434.
10. Kuntz, J. E., and R. B. Sleight. Legibility of Numerals: The Optimal Ratio of Height to Width Ratio of Stroke. *American Journal of Psychology,* Vol. 63, 1950, pp. 567–575.
11. Case, H. W., J. L. Michael, G. E. Mount, and R. Brenner. Analysis of Certain Variables Related to Sign Legibility. *Bulletin 60,* HRB, National Research Council, Washington, D.C., 1952, pp. 44–54.
12. Forbes, T. W., and K. Moscowitz. *A Comparison of Lower Case and Capital Letters for Highway Signs.* HRB, National Research Council, Washington, D.C., 1951, pp. 355–372.
13. Soar, R. S. Height-Width Proportion and Stroke-Width in Numeral Visibility. *Journal of Applied Psychology,* Vol. 39, No. 1, 1955, pp. 43–46.
14. Solomon, D. The Effect of Letter Width and Spacing on Night Legibility of Highway Signs. *HRB Proc.*, Vol. 35, 1956, pp. 600–617.
15. Baerwald, J. E., D. F. Karmier, and C. G. Herrington. The Functions and Design of Motor Vehicle License Plates. *University of Illinois Bulletin,* Vol. 58, No. 10, Sept. 1960.
16. Mace, D. J., and P. M. Garvey. The Effects of Increased Character Height on Sign Legibility, *Presented at 72nd Annual Meeting of the Transportation Research Board,* Washington, D.C., 1993.
17. *American National Standard for Human Factors Engineering of Visual Display Terminal Workstations.* Human Factors Society, 1988.
18. *National Park Service Sign Manual.* U. S. Department of Interior, 1988.
19. Lauer, A. R. Improvements in Highway Safety. *HRB Proc.*, Vol. 12, 1932, pp. 389–401.
20. Mitchell, A., and T. W. Forbes. Design of Sign Letter Sizes. *Proc., American Society of Civil Engineers,* Vol. 68, 1942, pp. 95–104.
21. Brown, F. R., and E. A. Lowery. *A Study of the Requirements for Letters, Numbers and Markings To Be Used on Trans-Illuminated Aircraft Control Panels. Part 1. The Effects of Stroke Width Upon the Legibility of Capital Letters.* Report TED NAMEL-609, ATI-205 706. Aeronautical Medical Equipment Laboratory, Naval Air Material Center, Philadelphia, 1949.
22. Brown, F. R., and E. A. Lowery. *A Study of the Requirements for Letters, Numbers and Markings To Be Used on Trans-Illuminated Aircraft Control Panels. Part 3. The Effect of Stroke-Width and Form Upon the Legibility of Numerals.* Report TED NAMEL-609, ATI-107 638. Aeronautical Medical Equipment Laboratory, Naval Air Material Center, Philadelphia, 1951.
23. Brown, F. R., and E. A. Lowery. *A Study of the Requirements for Letters, Numbers and Markings To Be Used on Trans-Illuminated Aircraft Control Panels. Part 4. Legibility of Uniform Stroke Capital Letters as Determined by Size and Height to Width Ratio and as Compared to Garmond Bold.* Report TED NAMEL-609, AD-10 211. Aeronautical Medical Equipment Laboratory, Naval Air Material Center, Philadelphia, 1953.
24. Schapiro, H. B. *Factors Affecting Legibility of Digits.* WADC Technical Report 127, Contract AF 33(038)18317, AD-1 117, 1952.
25. Crook, M. N., J. A. Hanson, and A. Weisz. *Legibility of Type as a Function of Stroke Width, Letter Width, and Letter Spacing Under Low Illumination.* WADC Technical Report 53–440, Contracts W33-038 ac-14559 and AF33(616)-2018, AD-43 309, 1954.
26. Crook, M. N., J. A. Hanson, and A. Weisz. *Legibility of Type as Determined by the Combined Effect of the Typographical Variables and Reflectance of Background.* WADC Technical Report 53–441, Contracts W33-038 ac-14559 and AF33(616)-2018, AD-43 309, 1954.
27. Hughes, C. L. Variability of Stroke-width Within Digits. *Journal of Applied Psychology,* Vol. 45, No. 6, 1961, pp. 364–368.
28. Hodge, D. C. Legibility of Uniform Stroke-Width Alphabet. I. Relative Legibility of Upper and Lower Case Letters. *Journal of Engineering Psychology,* Vol. 1–2, 1962–1963, pp. 35–46.
29. Benson, B. L., and J. E. Farrell. The Effect of Character Height to Width Ratio of CAT Display Legibility. *Society of Information Display: Digest of Technical Papers.* Palisades Institute of Research Services, Inc., New York, 1988, pp. 340–343.
30. Zwahlen, H. T. *Reflective License Plate Material—Evaluation of Conspicuity and Legibility Performance for a Standard License Plate Configuration Using Beads on Paint vs Reflective Sheeting.* Final Report FHWA-AZ91-358. Arizona Department of Transportation, Phoenix, 1991.
31. Zwahlen, H. T. Legibility and Conspicuity Performance of Dry and Wet Beads on Paint and Reflective Sheeting License Plates Under Low-Beam Illumination at Night. In *Transportation Research Record 1456,* TRB, National Research Council, Washington, D.C., 1994, pp. 112–124.

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

In-Traffic Driver Behavior: Development of Measures and Evaluation of Differences Between Finland and Michigan

JUHA LUOMA

A study was designed (a) to develop an initial set of measures to observe driver behavior in different countries and (b) to compare driver behavior in Finland with driver behavior in Michigan by using these measures. Development of the measures emphasized equating environmental factors and traffic rules. For such conditions the study produced a reliable research tool providing potentially valuable results. The measures were applied in one middle-sized city in both regions, Lahti in Finland and Ann Arbor in Michigan. The results indicated the following main differences in driver behavior. Drivers in Lahti (compared with those in Ann Arbor) signaled more frequently before the lane change or turning and came to a full stop at intersections with a stop sign more frequently. The following trends were found: drivers in Lahti exceeded the speed limit more frequently but decreased the speed earlier while approaching the intersection from the secondary road, accelerated more slowly after turning onto the secondary road, and accepted slightly longer gaps when entering the main road. The following aspects of driver behavior did not differ in the two cities: the variance of the speed in free-flow traffic, the proportion of short headways in car-following situations, the frequency of no stops compared with rolling and full stops at intersections with a stop sign, and yielding to pedestrians at intersections. The differences that were found differences are likely consequences of differences in values and norms.

The need for measures that can be used to collect descriptive and comparable data on actual driver behavior to improve traffic safety has been recognized for at least 20 years. Smeed (*1*) argued the following: "A knowledge of standards of driver behaviour would help to determine which aspects of such behaviour especially need to be improved. If these standards were assessed periodically, it would be possible to assess whether driver behaviour was or was not improving and assess the effects of policies directed to the improvement of driver behaviour. If the standards were assessed in different countries or in different parts of the same country, the results would be of value in explaining differences in accident rates in the countries or parts of the same country concerned."

Until recently, there has been no cross-national or longitudinal study on driver behavior that has collected data by (a) using a broad and unobtrusive set of measures and (b) matching or controlling environmental factors and traffic volumes properly. For example, Benjamin (*2*) reported drivers' close-following and speed behaviors on motorways in 11 countries. However, the sites varied substantially in terms of the posted speed limits and traffic volumes. Only two behaviors were compared in a study that investigated vehicle speeds while approaching an intersection from a secondary road and gap acceptance at one English and two German rural intersections (*3*). Traveling speeds on highways, drivers' visual scanning behaviors at intersections, and pedestrians' crossing behaviors at red traffic signals were compared in Canada, Japan, and Korea (*4*). However, that study did not attempt to match the road and traffic conditions.

Longitudinal studies have also collected data on many driver behaviors in different parts of the same city or country. Drivers' speed behavior, running red lights, use of turn signal when turning, and yielding to pedestrians were studied in Stockholm (*5*). In Finland several measures of driver behavior were obtained (*6*): speed behavior, headways of vehicles driving in car-following situations, use of safety belt, use of daytime running lights, use of turn signal while turning, proportion of drunk drivers, use of bicycle helmet, crossing on red lights by pedestrians, and use of retroreflectors by pedestrians. More specific definitions and control of traffic volumes for each measure would improve the comparability of the results of these studies with those of other studies.

The present study [based on a detailed technical report (*7*)] was designed (a) to develop an initial set of measures that can be used to observe driver behavior unobtrusively in different countries (or in different parts of a given country or in longitudinal studies) and (b) to investigate driver behavior in Finland and Michigan.

METHODS

Measures

The following measures were included in the study: Measure 1, speed behavior in a free-flow traffic situation; Measure 2, headways in car-following situations; Measure 3, use of turn signals before a lane change; Measure 4, speed while approaching an intersection from a secondary road; Measure 5, speed after turning onto a secondary road; Measure 6, use of turn signals before turning; Measure 7, stopping behavior at intersections with a stop sign; Measure 8, gap acceptance when entering a main road; and Measure 9, yielding to pedestrians at intersections.

The underlying logic in the selection of measures was that these measures would show different aspects of driver behavior that have potential safety consequences. Measures 1, 3, 6, 7, and 9 focused primarily on obeying a specific traffic rule. Measures 2, 4, 5, and 8 focused on behavior in situations with a general rule for safe driving. However, most of the measures have many aspects. Speed behavior includes two major aspects: speed deviation and exceeding the speed limit. Short headways mean that drivers have less time to react if the lead vehicle suddenly brakes. Speed while approaching an intersection from a secondary road and, especially,

---

The University of Michigan Transportation Research Institute, 2901 Baxter Road, Ann Arbor, Mich. 48109-2150.

gap acceptance when entering a main road were expected to reveal possible differences in safety margins. Speed after turning onto a secondary road shows how rapidly (aggressively) drivers accelerate. Use of a turn signal either before a lane change or before turning indicates that the driver obeys specific traffic rules and also how well drivers show their intentions to other road users. Stopping behavior at intersections with a stop sign primarily indicates whether the driver obeys traffic rules, but perhaps also provides information about safety margins. Yielding to pedestrians at intersections indicates how well drivers give a right-of-way to vulnerable road users, as well as how well they obey a particular rule.

It is acknowledged that in most cases the safety effects of a particular driver behavior are only potential. For example, one cannot quantify the safety effects of violating a specific traffic rule. In general, however, the violation of rules means a negative attitude toward traffic safety, an unwillingness to be concerned with other road users, or a willingness to accept a smaller safety margin.

Sites

The measures were applied in middle-sized cities in both countries, Lahti in Finland and Ann Arbor in Michigan. The population of each city is about 100,000. Measures 1, 3, 6, 8, and 9 were obtained from two sites in each city, whereas Measures 4 and 5 were measured at one site each, and Measure 7 was measured at three sites each (Table 1).

The study attempted to match the environment type and function, as well as sight distances, number and width of lanes, curvature, and gradient (*7*). The surface was dry asphalt that was in good or fairly good condition. Atmospheric conditions were good, with no rain, snow, or fog.

Equipment

All data except speed and headways were collected by video recording or note taking. Speed and headways were measured by a traffic counter connected to detector loops (Measures 1 and 2) or photocell pairs (Measures 4 and 5).

Drivers

The number and characteristics of the drivers for each measure are given in Table 2. The study focused on driver behavior, and thus the study excluded pedestrians and bicyclists, as well as police cars, ambulances, fire engines, taxicabs, cars driven by student drivers, motorcycles, mopeds, buses, and trucks. It is noteworthy that the minimum age for a driver's license is 18 years in Finland and 16 years in Michigan.

Vehicle Population

It is assumed that differences in vehicle populations are relatively minor and that they do not have a major impact on the behaviors studied. The most substantial difference is that cars in Finland are generally equipped with manual transmissions, whereas automatic

TABLE 1 Road and Traffic Conditions at Sites of Measures

		Speed Limit (km)		Number of Lanes for Observed Vehicles		Traffic Volume (Vehicles per Hour)	
Measure (Site)	Environment	Lahti	Ann Arbor[a]	Lahti	Ann Arbor	Lahti	Ann Arbor
Speed behavior (1a) and headways (2a)	Suburban streets	70	64	2	2	360	510
Speed behavior (1b) and headways (2b)	Suburban streets	70	72	2	2+1[b]	670	1,160
Use of turn signals before a lane change (3a, b)	Urban and suburban streets	50 or 60	48 or 56	2 or 3	2 or 2+1[b]	400-1,000	600-1,000
Speed while approaching an intersection (4), speed after turning (5), stopping behavior (7c), and gap acceptance (8a)	Suburban intersections with a stop sign	50 80[c]	56 72[c]	1	1	140 390[c]	140 650[c]
Use of turn signals before turning (6a) and stopping behavior (7a)	Urban intersections with a stop sign	50 50[c]	48 48[c]	1	1	60 80[c]	30, 40 90[c]
Use of turn signals before turning (6b), stopping behavior (7b), and gap acceptance (8b)	Suburban intersections with a stop sign	50 80[c]	40 72[c]	1	1	80 540[c]	80 550[c]
Yielding to pedestrians (9a, b)	Controlled downtown intersections	50 50[c]	48 48[c]	1	1	80, 140	60, 100

[a] converted from miles per hour.
[b] a center lane for left-turning vehicles because of a residential cross-street to the left.
[c] a crossing street (volumes in both directions).

TABLE 2 Number and Characteristics of Drivers for Each Measure

	Number of Drivers		Male Drivers (%)		Estimated Driver's Age (%)					
					<25	25-65	>65	<25	25-65	>65
Measure (Site)	Lahti	Ann Arbor	Lahti	Ann Arbor	Lahti			Ann Arbor		
Speed behavior (1a)	1,153	1,458	-	-	-	-	-	-	-	-
Speed behavior (1b)	380	106	-	-	-	-	-	-	-	-
Headways (2a, b)	1,485	1,665	-	-	-	-	-	-	-	-
Use of turn signals before a lane change (3a)	303	225	-	-	-	-	-	-	-	-
Use of turn signals before a lane change (3b)	281	285	-	-	-	-	-	-	-	-
Speed while approaching an intersection (4)	164	170	85.7	63.5	13.0	86.3	0.6	5.6	90.6	3.8
Speed after turning (5)	199	209	-	-	-	-	-	-	-	-
Use of turn signals before turning (6a)	110	115	72.8	71.6	21.9	70.2	7.9	5.9	93.1	1.0
Use of turn signals before turning (6b)	248	267	87.8	62.4	11.4	86.2	2.4	13.3	84.7	2.0
Stopping behavior (7a)	248	183	73.7	67.1	20.2	67.6	12.1	4.9	92.0	3.1
Stopping behavior (7b)	62	79	85.2	55.1	14.8	82.0	3.3	17.9	80.8	1.3
Stopping behavior (7c)	100	91	89.9	64.4	14.1	84.8	1.0	6.9	88.5	4.6
Gap acceptance (8a)	89	127	81.5	65.4	11.1	88.9	0.0	3.8	91.4	4.8
Gap acceptance (8a)	117	164	85.4	60.0	12.5	83.3	4.2	14.6	83.1	2.3
Yielding to pedestrians (9a, b)	154	165	70.5	69.5	17.9	76.9	5.1	25.0	73.5	1.5

\- data are unavailable

transmissions are the norm in Michigan. Also, cars are generally smaller in Finland than in Michigan.

Rules and Enforcement

Measures were taken in traffic situations and environments where similar rules applied (Table 1). The risk of being caught while violating the rules was assumed to be similar. In both countries drivers may believe that they are more likely to get a ticket because of exceeding the speed limit than because of an incomplete stop at an intersection with a stop sign. Furthermore, they may believe that they are even less likely to get a ticket because of failure to use a turn signal or because of not yielding to pedestrians. No objective data on the magnitude of the enforcement were collected, because these indicators were considered not to be directly comparable. For example, the same number of hours of enforcement in different regions is unlikely to mean the same strength of enforcement.

Procedure

Data were collected between 9 a.m. and 4 p.m. on Tuesdays through Thursdays in May 1993 (Lahti) and September and October 1993 (Ann Arbor). The observation car was a standard car or van parked in a normal manner. However, while investigating the use of turn signal before a lane change, the observation car was moving. For all measures that involved the use of the video recording, drivers were not able to see the camera before or while the behavior occurred.

To compare driver behavior it was appropriate to select only certain drivers for further analysis or to classify behaviors into categories (Table 3).

RESULTS

Speed Behavior in Free-Flow Traffic

Bartlett's test was performed to test the difference of variances (Table 4). The test was computed between the cities and between the first sites that included the majority of data. No significant differences were found between the two cities in either case.

The proportion of the drivers exceeding the speed limit was different at the two sites in Ann Arbor [$\chi^2(1) = 29.5, p < .00001$], but not in Lahti. At each site in Ann Arbor the proportion of the drivers exceeding the speed limit was smaller than the average in Lahti [for the first Ann Arbor site, $\chi^2(1) = 7.02$ and $p < .008$; for the second Ann Arbor site, $\chi^2(1) = 47.4$ and $p < .00001$]. There was no significant difference between the cities in the proportions of drivers exceeding the speed limit by more than 15 km/hr, but the difference between the sites was significant in each city [for Lahti, $\chi^2(1) = 25.5$ and $p < .0001$; for Ann Arbor, $\chi^2(1) = 19.7$ and $p < .001$].

Headways in Car-Following Situations

This analysis was performed for combined sites for each city by traffic volume, which varied between 200 and 899 vehicles per hour. The proportions of drivers with a short headway varied between 15 and 25 percent in Lahti and between 15 and 30 percent in Ann Arbor. There were no systematic differences in the proportion of short headways for a given traffic volume. However, the proportion of short headways was smaller in Lahti than in Ann Arbor when the traffic volume was 200 to 299 vehicles per hour [$\chi^2(1) = 5.26, p < .03$]. None of the other differences was significant.

TABLE 3 Selection of Drivers and Classification of Behavior

Measure	Selection of Drivers	Classification of Driver Behavior
Speed behavior in a free-flow traffic situation	Vehicles with a minimum headway of 10 s between the actual vehicle and the vehicle ahead	
Headways in car-following situations	Vehicles with a maximum headway of 5 s and the maximum speed difference of 10 s between the actual vehicle and the vehicle ahead	A short headway < 1 s
Use of turn signals before a lane change	(a) A vehicle was observed before a lateral movement or signalling (b) Other traffic travelling in the same direction as the observed car	A driver was categorized as using a turn signal if he/she signalled before crossing a lane marking.
Speed while approaching an intersection, and speed after turning onto a secondary road	(a) Vehicles with a minimum headway of 20 s from a vehicle ahead (b) Only turning drivers were included	
Use of turn signals before turning		A driver was categorized as using a turn signal if he/she signalled before the wheels began to turn or the vehicle stopped.
Stopping behavior at intersections with a stop sign	(a) Vehicles with a minimum headway of 20 s between an approaching vehicle and a vehicle ahead (b) No oncoming vehicles, pedestrians, or bicyclists that affected driver behavior (c) Drivers who accepted the first gap of the traffic flow on the main road	(a) Full stop = the wheels of the vehicle did not roll (b) Rolling stop = the vehicle speed was about the same as the walking speed (c) No stop = the vehicle speed was constant or might be reduced, but the speed was higher than in a rolling stop
Gap acceptance when entering a main road	(a) Vehicles with a minimum headway of 20 s in front of them (b) Vehicles that approached the stopping line in a situation when a gap (lag) on the main road was 10 s or less	
Yielding to pedestrians at intersections	Left-turning drivers (and pedestrians who had a green phase at the same time) when no oncoming vehicles or bicycles	For drivers: (a) Drove on (b) Reacted (braked, weaved, or stopped) For pedestrians: (a) Walked on (b) Reacted (slowed down, stopped, ran, or retreated)

Use of Turn Signals Before Lane Change

As shown in Figure 1, drivers in Lahti signaled more frequently than those in Ann Arbor [$\chi^2(1) = 19.7$, $p < .0001$]. The differences between the two routes were not statistically significant in either city.

Speed While Approaching an Intersection from a Secondary Road

The mean initial speed at a distance of 120 m before the intersection was 58.6 km/hr in Lahti and 64.2 km/hr in Ann Arbor, reflecting the differences in the posted speed limits. The proportions of drivers exceeding the speed limit initially (82.9 percent in Lahti and 81.4 percent in Ann Arbor), right-turning drivers, and drivers accepting the first gap when entering the main road were not significantly different.

Figure 2 shows for each city the mean approach speed at distances of 120, 90, 60, and 30 m before the intersection. On average the speed change (in comparison with the initial speed at 120 m) was greater in Lahti than in Ann Arbor at 90 m [4.1 versus 0.6 km/hr; $F(1,324) = 51.9$, $p < .0001$] and at 30 m [20.4 versus 18.0 km/hr; $F(1,309) = 7.38$, $p < .02$), but not at 60 m (9.1 versus 9.0 km/hr). In Ann Arbor (but not in Lahti) there was a

TABLE 4 Speed Behavior in Free-Flow Traffic by Site

	Lahti		Ann Arbor	
Aspect	a	b	a	b
Mean and median speed (km/h)	80	77	73	75
Standard deviation (km/h)	9	7	9	9
85-percentile of the speed, v_{85} (km/h)	89	84	82	85
Exceeding speed limit (%)	88.3	84.5	84.0	63.2
Exceeding speed limit by more than 15 km/h (%)	24.6	12.1	23.2	4.7

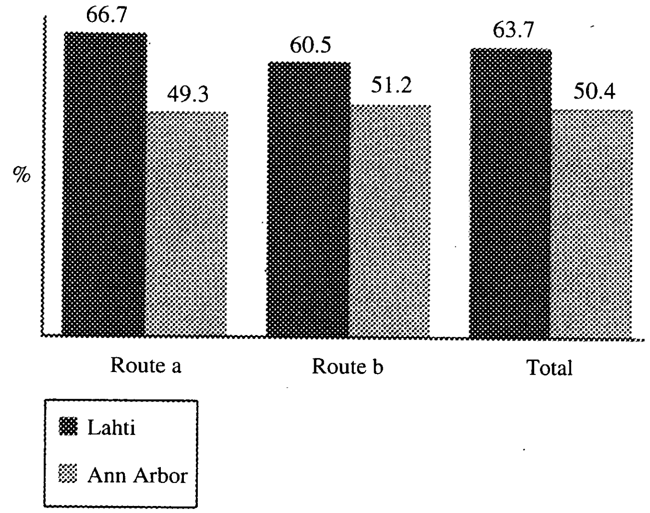


FIGURE 1 **Proportion of drivers using turn signal before lane change by route.**

slight incline uphill that helped to decrease speed. Consequently, the results suggest that the drivers in Lahti approached the intersection from the secondary road slightly more cautiously than the drivers in Ann Arbor. The standard deviations of the speed distributions at different distances from the intersection varied between 5.9 and 9.7 km/hr in Lahti and between 5.9 and 8.3 km/hr in Ann Arbor.

The speeds at the four locations were submitted to an analysis of variance by using the following three main factors: city, sex, and acceptance of the first gap while entering the main road. The effect of city was significant, mostly because of the higher initial speed. More interestingly, the effect of gap acceptance was significant at each distance: 120 m [$F(1,286) = 7.09, p < .01$], 90 m [$F(1,286) = 6.89, p < .01$], 60 m [$F(1,286) = 5.63, p < .02$], and 30 m [$F(1,286) = 12.9, p < .001$]. Drivers accepting the first gap drove faster at each location, with a tendency for this difference to be greater in Lahti than in Ann Arbor. The effect of sex was not significant (as in all of the following measures in which sex was recorded), as were all interactions.

Speed After Turning onto Secondary Road

The drivers in Lahti tended to accelerate more slowly than those in Ann Arbor (Figure 3). However, the result must be interpreted cautiously because of the higher speed limit and the slight incline downhill in Ann Arbor. The standard deviation of the speed was 5.6 to 6.4 km/hr in Ann Arbor and 4.7 to 6.1 km/hr in Lahti. The last speed measurement at a distance of 120 m from the intersection showed that 69.3 percent of the drivers in Lahti and 56.5 percent of the drivers in Ann Arbor exceeded the speed limit [$\chi^2(1) = 7.24, p < .01$].

Use of Turn Signals Before Turning

Drivers in Lahti signaled more frequently than those in Ann Arbor (Figure 4). This was the case at the urban intersections [for left turn, $\chi^2(2) = 22.9$ and $p < .00001$; for right turn, $\chi^2(2) = 16.3$ and $p < .00001$] and at the suburban intersections [for left turn, $\chi^2(2) = 2.05$ and p was not significant; for right turn, $\chi^2(2) = 16.3$ and $p < .0001$]. The effect of the turn direction was not significant at each intersection.

One could assume that there might be drivers who signaled only if they saw other traffic in the vicinity. Furthermore, the suburban intersection in Ann Arbor had traffic from only one secondary road, whereas the other intersections had traffic from two directions. Therefore, the effect of other traffic was computed by each turn direction at urban intersections. The results revealed that drivers tended to signal more frequently if no traffic was in the vicinity. However, none of the four pairwise differences was significant.

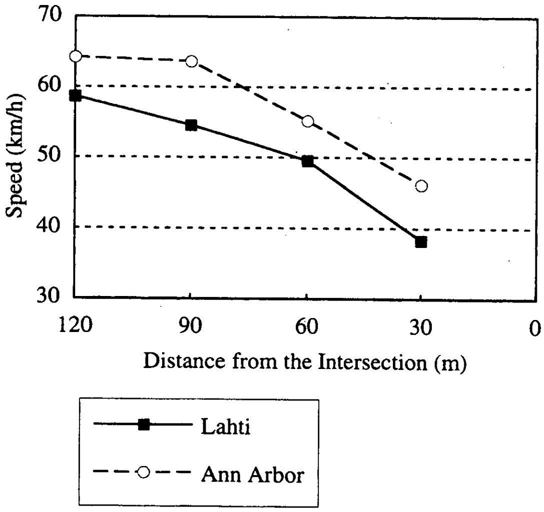


FIGURE 2 **Mean speed while approaching an intersection from a secondary road.**

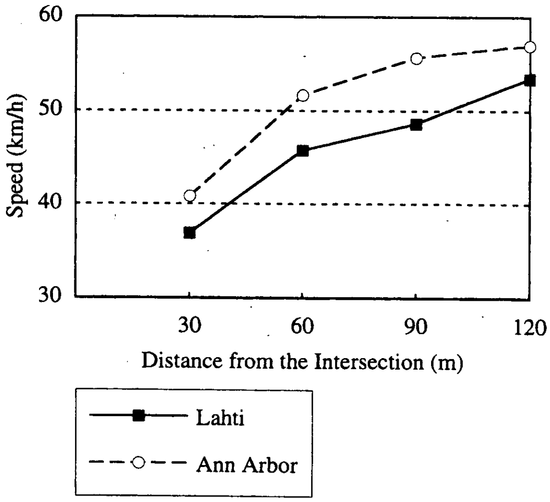


FIGURE 3 **Mean speed after turning onto secondary road.**

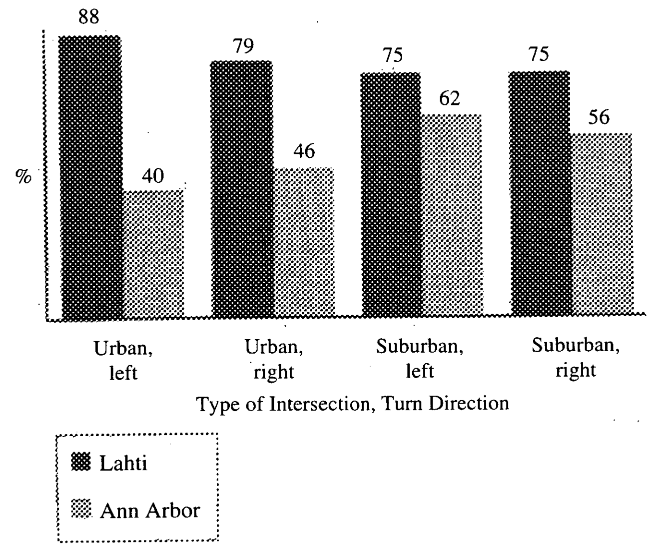


FIGURE 4 Percentage of drivers using turn signals before turning, by turn direction.

Stopping Behavior at Intersections with a Stop Sign

Stopping behavior by intersection and turn direction is presented in Figure 5. The following two hypotheses were tested for each turn direction and between each pair of intersections.

Is there a cross-national difference in the proportion of full stops? A comparison of the corresponding traffic streams showed that at each intersection right-turning drivers in Lahti came to a full stop more frequently than those in Ann Arbor [for urban intersections, $\chi^2(1) = 11.1$ and $p < .001$; for the first suburban intersections, $\chi^2(1) = 5.39$ and $p < .03$; for the second suburban intersections, $\chi^2(1) = 23.5$ and $p < .00001$]. The through-traveling drivers were investigated at urban intersections only. As in the case of the right-turning drivers, drivers in Lahti came to a full stop more frequently than those in Ann Arbor [$\chi^2(1) = 13.6, p < .001$]. Comparison of the left-turning drivers showed that at the second suburban intersections the difference was significant [$\chi^2(1) = 7.77$, $p < .01$], but the difference was not significant at the urban intersection. (At the first suburban intersection the test of significance was not performed because of the small number of left-turning drivers.)

Is there a cross-national difference in the proportion of no stops? Only the difference for the across-traveling drivers at urban intersections was significant [$\chi^2(1) = 4.26$, $p < .04$], with drivers in Lahti coming to no stops more frequently than those in Ann Arbor.

Gap Acceptance When Entering a Main Road

Overall, there was no significant difference between the cities in the durations of the first gap (accepted or rejected). Critical gaps, that is, the gap size that 50 percent of the drivers accepted, were computed for drivers who rejected the first gap (only the durations of the first rejected and accepted gaps were measured). In each case the critical gap was longer in Lahti than in Ann Arbor (Figure 6). Unfortunately, the results based on the long time separations (because of the small numbers of drivers) allowed only the order of the critical gaps, but not their absolute values, to be defined.

The durations of the first gaps that were accepted were submitted to an analysis of variance by using city and turn direction as factors. The two main effects and their interaction were not statistically significant.

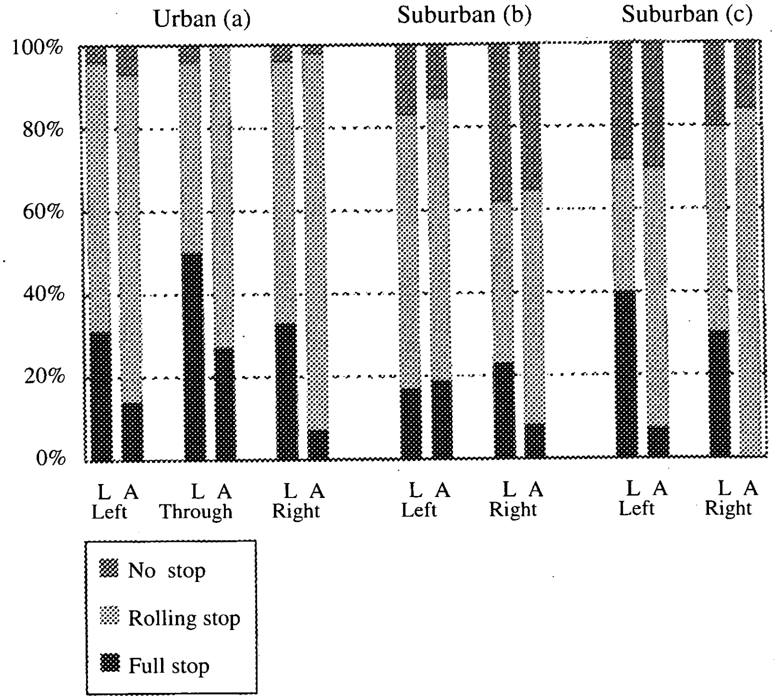


FIGURE 5 Stopping behaviors at three intersections in each city, by turn direction (L = Lahti, A = Ann Arbor).

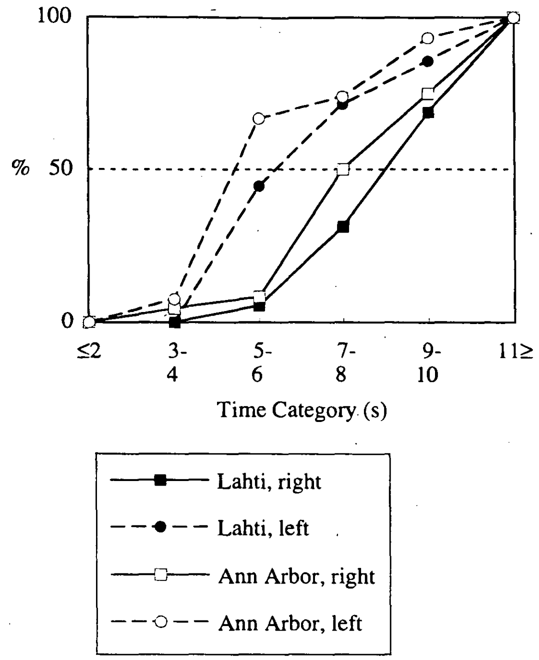


FIGURE 6 **Critical gaps at both intersections, by turn direction.**

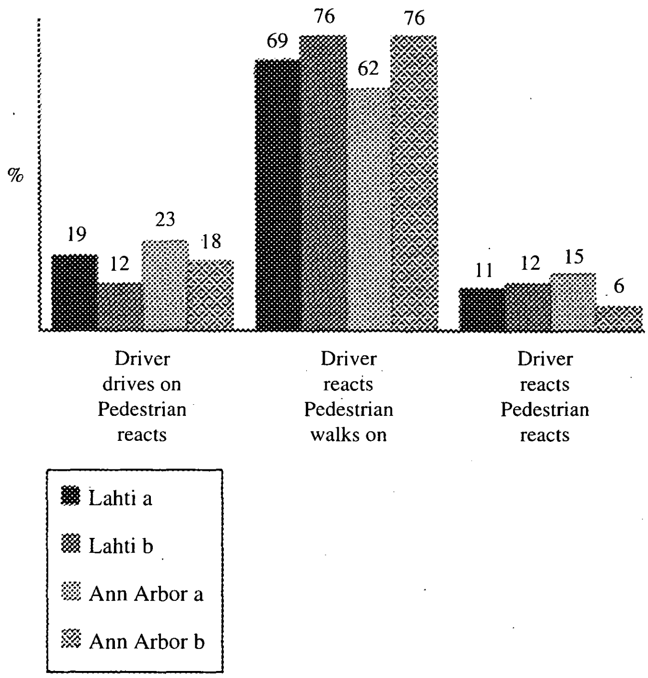


FIGURE 7 **Proportions of different type of interactions.**

Yielding to Pedestrians at Intersections

The results in Figure 7 indicate no significant difference in the proportions of the interactions between cities or between intersections in each city. Driver behavior, cross-tabulated by the number of pedestrians in the interaction, showed that if one to two pedestrians were in interaction, there were some instances of drivers in each city who did not yield to pedestrians (the difference between the cities was not significant). However, when there were three to four pedestrians, that happened only in Ann Arbor [18 and 23 percent; $\chi^2(1) = 8.66, p < .01$].

DISCUSSION OF RESULTS

The goals of the present study were (a) to develop a set of measures that could be used to observe driver behavior in different countries, different parts of a country, or longitudinal studies and (b) to compare driver behavior in Finland with driver behavior in Michigan. The results of the study will be discussed in terms of these two goals.

Development of Set of Measures

The comparison made between Finland and Michigan showed that each of those measures was usable. Given that, the following are essential questions: How reliable and valuable are the results that these measures provide? The problem of reliability includes three broad areas: (a) techniques of data collection and interpretation of the data, (b) matching of environmental factors and specific rules, and (c) numbers of drivers.

Techniques of Data Collection and Interpretation of Data

The speed and headway data collected by the traffic counter are reliable (*6*). Data collection by the video recording technique provides the possibility of examining every event carefully and replaying it if necessary. In addition, the video recording technique provides the possibility of counting actual traffic volumes. However, the data have room for interpretations. The most evident interpretation problem of the present study was the separation of rolling and no stops at intersections with a stop sign. The separation was based on the comparison of walking and driving speeds, but it may involve systematic bias if several people estimate the speed. In the present study, however, one person interpreted all of the data.

Matching of Environmental Factors and Specific Rules

The lists of selections and definitions show the limited scope of the study. However, it is emphasized that these selections and definitions must be explicit to show how much caution must be taken while generalizing the results. On the other hand, the limitations of the study do not invalidate the results, because the goal of the study was to collect basic information on driver behavior.

If one estimates how well the match succeeded in the comparisons that were performed, some shortcomings can be seen. First, in the comparison of speed behavior and headways, the second site in Ann Arbor seemed to be different because there were intersections. In addition, the similarity of the 85th percentile of the speed but a lower percentage of drivers exceeding the speed limit suggests that the speed limit was relatively high. The second major concern was that the traffic volumes on the main road were substantially different for Measures 4 and 8a (Table 1). However, this presumably did not influence the comparisons unduly because the results showed that drivers who accepted or rejected the first gap decreased their speeds similarly in both cities. The comparison of gap acceptance behavior was based on critical gaps, ruling out a major influence of different volumes on the main road.

Number of Drivers

In studies like the present one the question about the number of drivers is always connected to the amount of time that is allowed for data collection. Especially in the case of measures such as speed while approaching an intersection and gap acceptance, attention should be paid to this question, because only a small amount of data is usable for further analysis. Perhaps it would be possible to collect these data with loop detectors, providing possibilities for collecting a substantial amount of data.

The study was not based on data collection at a random sample of sites in two cities, because the goal was to develop and test the set of measures. If the goal is to obtain a general picture from one or two countries, special attention must be paid to the selection of the sites to cover the target area without any bias.

Value of Results

The question of the value of the comparison of driver behavior is difficult to answer on the basis of this individual comparison. However, the question is worth discussing. It was assumed that the measures would show different aspects of driver behavior that have potential safety effects. As indicated earlier, it is not known whether this is the case. It is hoped that further research will show the possible connections by comparisons of driver behavior and traffic accidents. The main difficulty of this approach has been the lack of a sound methodology for investigating driver behavior. Therefore, the study produced a research tool. Also, this kind of tool may be necessary when similar technical applications are intended for use in different countries or within a large country. Of course, it is possible to add other measures to this set of measures: the use of safety belts, the use of motorcycle and bicycle helmets, and the proportion of drunk drivers, for example (*6,8*).

Comparison of Driver Behavior in Finland and Michigan

The results of the study suggest that, overall, driver behaviors are rather similar in Lahti and Ann Arbor, and most of the differences are minor. This main finding was expected because the patterns of road accidents in Finland and the United States are relatively similar (*9*). Although no comparison of road accidents in Finland and Michigan has been conducted, there is no reason to believe that the patterns are more different than those between Finland and the United States. In addition, there is evidence that drivers assess risks connected to traffic similarly (*10*).

However, there were substantial differences in the proportion of drivers who used turn signals and in the proportion of full stops at intersections with a stop sign. These differences suggest that Finnish drivers obey specific traffic rules more often than Michigan drivers, which may reflect differences in societal values and norms. This conclusion is supported by two arguments. First, it is well known that individual freedom is more emphasized in the United States than in Europe and, especially, in Scandinavia (*11*). In traffic research this aspect has been frequently evident in the rates of usage of safety belts, with safety belt use rates being higher in Europe than in the United States (*12*). This difference also applies for Finland versus Michigan (*7*). Second, given that minimum driver training is more extended in Finland than in Michigan (*7*) it can be assumed that both the perceived behavioral differences and the mandatory duration of driver training reflect the values and norms of the society. In addition, the training may directly improve the frequency of obeying specific rules, because the applicants usually concentrate on learning rules (*13*). On the other hand, the conclusion concerning obeying traffic rules cannot be generalized to all specific rules, because drivers in Lahti tended to exceed the speed limit more frequently than drivers in Michigan, and yielding to pedestrians and the proportion of full stops were not different. Also, differences in behavior may be influenced by general differences in traffic control or trip length. For example, Michigan drivers may be more unwilling to come to a full stop than Finnish counterparts because of more frequent stop signs instead of yield signs and longer average trips.

The present results did not support the assumption that some positive indication of the longer history of motorized transportation in the United States would be detected. In contrast, compared with drivers in Lahti, drivers in Ann Arbor tended to decrease speed later while approaching the intersection from a secondary road, to accelerate more rapidly after turning onto the secondary road, and to accept slightly smaller gaps while entering the road.

ACKNOWLEDGMENTS

This study was conducted as a joint project of the Technical Research Centre of Finland (VTT) and the University of Michigan Transportation Research Institute (UMTRI). Appreciation is extended to the Academy of Finland and the Henry Ford Foundation for support of this research. The author thanks the following people for their contributions: Veli-Pekka Kallberg, Risto Kulmala, and Erkki Ritari (VTT); Martti Niskanen (city of Lahti, Finland); Masami Aoki, Michael Sivak, and Eric Traube (UMTRI); Bob Stegink (University of Michigan); and Nancy Greenleaf Gibson and Richard Watkins (city of Ann Arbor).

REFERENCES

1. Smeed, R. J. Research into Driver Behaviour. *Proc., Meeting of the International Statistics Institute*, Vienna, 1974.
2. Benjamin, T. *A Fifteen-Country Study of Some Factors Influencing the Number and the Severity of Road Accidents—Drivers' Risk-Taking, Speed Behaviour and Accidents on Motorways.* International Drivers' Behaviour Research Association, Arcueil, France, 1980.
3. Leutzbach, W., et al. *Vergleich der Verkehrssicherheit in der Bundesrepublik Deutschland und Grossbritannien* [Comparison of Traffic Safety in Federal Republic of Germany and Great Britain]. Lfd. 183. Bundesanstalt für Strassenwesen, Bergisch Gladbach, Germany, 1988.
4. Nagayama, Y. International Comparison of Traffic Behavior and Perceptions of Traffic. *IATSS Research*, Vol. 13, No. 1, 1989, pp. 61–69.
5. Slätis, A. *Hur Mäter Man Trafikmoralen—Studier av Trafikbeteenden i Stocholm* [How to Measure Traffic Responsibility—Studies on Traffic Behavior in Stockholm]. Report 57. Traffic Safety Association in Stockholm, Stockholm, 1990.

6. Heino, J. *Liikennekäyttäytymisen Seuranta 1993* [Follow-up Study on Road-User Behavior 1993]. Research Notes 69. The Central Organization for Traffic Safety, Helsinki, 1994.
7. Luoma, J. *Toward a Methodology for Field Observations of Driver Behavior: A Comparison of Finland and Michigan.* Report UMTRI-94-16. The University of Michigan Transportation Research Institute, Ann Arbor, 1994.
8. Streff, F. M., D. W. Eby, L. J. Molnar, H. C. Joksch, and R. R. Wallace. *Direct Observation of Safety Belt and Motorcycle Helmet Use in Michigan: Fall 1993.* Report UMTRI-93-44. The University of Michigan Transportation Research Institute, Ann Arbor, 1993.
9. Luoma, J., and M. Sivak. *Traffic Accidents in Finland and the U.S.A: A Cross-Cultural Comparison of Associated Factors.* Report UMTRI-92-26. The University of Michigan Transportation Research Institute, Ann Arbor, 1992.
10. Luoma, J. Driver Risk Assessment in Finland and Michigan. Report UMTRI-95-15. The University of Michigan Transportation Research Institute, Ann Arbor, 1995.
11. Hofstede, G. Cultures and Organizations. Software of the Mind. Intercultural Cooperation and Its Importance for Survival. McGraw-Hill Book Company, London, 1991.
12. Campbell, B. J., and F. A. Campbell. Seat Belt Law Experience in Four Foreign Countries Compared to the United States. AAA Foundation for Traffic Safety, Falls Church, Va., 1986.
13. Luoma, J. Autokoulu ei Opeta Tunnistamaan Vaaroja [Driving school does not improve recognition hazards]. *Liikennevilkku*, Vol. 31, No. 4, 1984, pp. 22–24.

Publication of this paper sponsored by Committee on Vehicle User Characteristics.

Effect of Wide-Base Tires on Rollover Stability

ANDREW D. ST. JOHN AND WILLIAM D. GLAUZ

One of the most cited advantages of using wide-base tires is to increase the stability relative to rollover of a truck or trailer. The rollover thresholds of a two-axle full trailer when it is equipped with conventional dual-wheel axles and when it is equipped with wide-base tire–axle assemblies are compared, and the stability benefits of design changes to trailers enabled by the wide-base configuration are evaluated. Because it is important in such analyses to account for the locations of the forces exerted by the pavement on the tires, it was necessary to include the effects of tire deflections in the analyses. The results showed that the tire deflection effects reduced the rollover stabilities over those determined by a rigid tire analysis by 5 to 6 percent. The increases in the rollover stabilities of trailers with wide-base tires rather than dual tires were relatively small (up to 6 percent). However, if the user of wide-base tires takes advantage of the ability to spread the suspension springs and lower the center of gravity of the trailer, increases in rollover stability of 18 to 21 percent are possible.

Wide-base tires, also referred to as super singles, are truck tires that are typically between 380 and 460 mm (15 and 18 in.) in nominal width, in contrast to normal truck tires that have a nominal width of 280 mm (11 in.). Certain segments of the trucking industry find a number of safety and economic advantages to substituting a wide-base tire for each pair of dual standard tires on the axles. Perhaps the major safety advantage claimed is a reduced propensity for the truck or trailer to roll over. Improved rollover stability is believed to be enhanced even further if, in addition to simply mounting the wide-base tires in place of the dual tires, one also takes advantage of the resulting wider wheel stance to spread the truck suspension springs, enabling the roll-restoring moment due to the springs to increase. In addition, with the wider stance on a tank trailer, one can lower the tank and its center of gravity and reduce the distance between the center of gravity and the roll center. The purpose of this paper is to investigate these claims analytically. More specifically, the purposes of the paper are to

1. Compare the rollover thresholds of a two-axle full trailer when it is equipped with conventional dual-wheel axles and when it is equipped with wide-base tire–axle assemblies, and
2. Evaluate the stability benefits of design changes enabled by the wide-base configuration.

All analyses presented deal with static or steady-state conditions. Two rollover situations are treated: side slopes and cornering on a pavement with zero superelevation.

Trailer data were taken from Fancher and Mathews (*1*), in which data typical of those for gasoline tankers used in California are provided. The trailer considered is a full trailer, with no vertical loads or moments shared with the tractor. Also, the front and rear axle suspensions both have the same characteristics. The tire data are taken from Fancher et al. (*2*) and from manufacturers. The duals are 11R22.5 tires, and the wide-base tires are 425/65R22.5 tires.

Table 1 lists the data used. The data are for a single axle; the terms are described in more detail later in this paper, along with their associated equations.

The use of wide-base tires instead of dual tires enables potential trailer design improvements. They include a 15 percent increase in the suspension system roll stiffness K_r, a decrease of 150 mm (6 in.) in the sprung weight roll arm ($h_s - h_r$), and a reduction of 150 mm (6 in.) in the height of the sprung weight h_s. These modifications would apply to the unmodified values in Table 1.

TIRE DEFLECTIONS

In analyzing units with dual wheels it has been conventional to assume that the load on the two wheels can be taken to act midway between the wheels. This is important here because where the forces from the pavement act on the trailer has a major effect on the moments that resist rollover. Furthermore, the only way to decide where these forces act is to include the effects of tire deflections in the analyses. The inclusion of tire deflections is especially important here because the wide-base tires are more compliant; they deflect more than comparable dual tires. This compliancy factor, by itself, is destabilizing.

The work of Fancher et al. (*2*) indicates that the vertical stiffness, or the spring rate, of a truck tire is not always linear. When a truck tire is lightly loaded its deflection is not a linear function of the load. Above certain loads, however, the deflection does become a linear function of the load. If the rated load of a tire is denoted by F_R, the deflection Z of the tire is a linear function of the applied vertical force F, if F is greater than $F_R/3$. F and Z are both measured perpendicular to the pavement surface. At a load of $F_R/3$ the tire deflection is about 0.55 Z_R, where Z_R is the deflection at a load of F_R. This is illustrated in Figure 1. Over the linear range the slope of the curve dZ/dF is $1/S$, where S is defined as the vertical stiffness of the tire. The subsequent analyses can be simplified to the case of rigid tires (no tire deflections) by taking the limit as $S \to \infty$.

From Figure 1 it can be seen that the slope of the linear portion of the curve is

$$1/S = 0.45Z_R/(2/3)F_R \tag{1}$$

so that the tire vertical stiffness is $S = 1.48\ F_R/Z_R$.

The equation for the linear portion of the curve is

$$Z = \frac{1}{S}\ (0.48F_R + F) \tag{2}$$

A. D. St. John, 8470 East Amethyst Place, Tucson, Ariz. 85715. W. D. Glauz, Midwest Research Institute, 425 Volker Boulevard, Kansas City, Mo. 64110.

TABLE 1 Data for Each Single Axle of a Two-Axle Trailer

Parameter	Symbol	Units[a]	Dual Tires	Wide Base Tires
Height, unsprung weight	h_u	mm	495	495
Height, sprung weight	h_s	mm	1990	1980
Height, roll center	h_r	mm	740	740
Unsprung weight	W_u	kg	680	600
Sprung weight[b]	W_s	kg	1,020	1,020
Total weight[c]	W	kg	1,700	1,620
Roll stiffness for the sprung weight	K_r	mm-kg per degree	923,000	923,000
Single tire vertical stiffness	S	kg per mm	89	120
Single tire rated load	F_R	kg	2,450	3,090
Track width,[d] outer duals	T_o	mm	2160	—
Track width,[d] inner duals	T_i	mm	1500	—
Track width,[d] wide base singles	T	mm	—	2010

[a] 1 mm = 0.03937 in., 1 kg = 2.203 lb.

[b] Empty weight. Weight when fully loaded is 7,950 kg (17,500 lb).

[c] Empty weight. Weight when fully loaded is 8,630 kg (19,000 lb) for duals, 8,550 kg (18,825 lb) for wide base tires.

[d] Center-to-center.

Dash (—) indicates data not applicable.

The nonlinear portion of the deflection/force curve can be fitted with a quadratic function passing through the origin, meeting the linear portion at $F_R/3$, and having a slope of $1/S$ at $F_R/3$. The result is

$$Z = \frac{1}{S}\,(3.89F - 4.33F^2/F_R) \qquad (3)$$

AXLE ROLL ANGLE

The axle roll angle ϕ_A is the angle between the axle and the surface of the pavement. When the loads applied to all of the tires on the axle are equal, it is assumed that ϕ_A equals zero. When the trailer rolls, the load is shifted to the tire(s) on one side of the axle, and the roll angle ϕ_A is determined from the changes in the vertical deflections of the tire. The expressions for the roll angle depend on the axle configuration (i.e., whether there are dual tires or wide-base single tires) and whether the tire deflections are linear or nonlinear, which depends on the load ranges.

Dual Tires on a Side Slope

Figure 2 illustrates the roll angle for an axle with dual tires on a side slope that has an angle of θ. Figure 2 also illustrates the vertical force on one tire (Tire 2) along with its components parallel and perpendicular to the pavement surface.

For the trailer to roll over it is necessary that all of the load be shifted off of Tires 1 and 2 and onto Tires 3 and 4. Attention is first directed to the situation in which F_1 and F_2 have been reduced to zero but in which Tire 2 still remains barely in contact with the pavement. Then, all of the trailer weight is carried by Tires 3 and 4, but not equally. However,

$$F_3 + F_4 = W \qquad (4)$$

where W is the sum of the unsprung and sprung weights W_u and W_s, respectively.

If Z_n is the deflection of the nth tire and Tire 2 is touching the pavement, the axle roll angle is given by either of the equations,

$$\sin\phi_A = (Z_3 - Z_2)/T_i$$

or

$$\sin\phi_A = (Z_4 - Z_2)/\left(\frac{T_o + T_i}{2}\right) \qquad (5)$$

Alternatively, if Tire 2 carries no load,

$$\sin\phi_A = (Z_4 - Z_3)/\left(\frac{T_o - T_i}{2}\right) \qquad (6)$$

It is convenient to consider that when the entire axle load is carried by Tires 3 and 4 (that is, when Tire 2 carries no load), these two tires each carry half of the entire axle load plus or minus a difference or load shift of ΔF. That is, one can write the loading on Tires 3 and 4 as

$$F_3 = \frac{W}{2} - \Delta F \qquad (7)$$

and

$$F_4 = \frac{W}{2} + \Delta F \qquad (8)$$

Now, if it is assumed that $F_3\cos\theta$ and $F_4\cos\theta$, the loads on Tires 3 and 4 normal to the pavement, respectively, are such that the tire

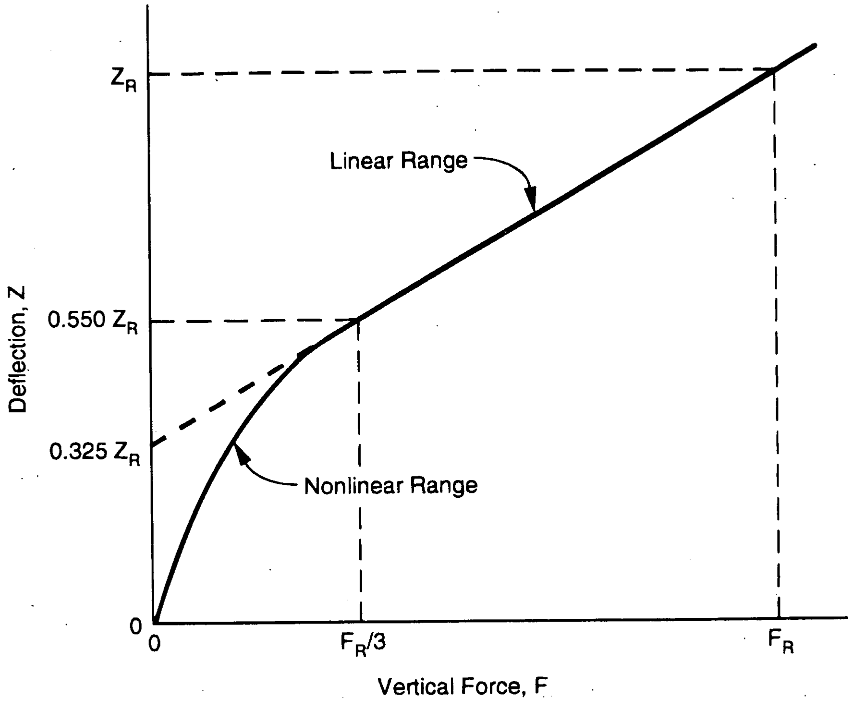


FIGURE 1 **Tire deflection versus vertical force curve.**

deflections Z_3 and Z_4 are in the linear range, Equation 2 can be used and Equations 7 and 8 can be substituted into the two forms of Equation 5 (with $Z_2 = 0$) to obtain

$$\Delta F = \frac{(T_o - T_i)}{(3T_i + T_o)} \left(\frac{0.48F_R}{\cos\theta} + \frac{W}{2} \right) \tag{9}$$

Equation 9 defines the load shift between Tires 3 and 4 when Tire 2 is just barely touching the pavement, assuming that the deflections of both tires are in the linear range. The assumption that both tires are in the linear deflection range means that the (lighter) load on Tire 3 be at least $F_R/3$. That is,

$$\left(\frac{W}{2} - \Delta F \right) \cos\theta \geq F_R/3 \tag{10}$$

If Equation 10 is not satisfied, Equation 3 must be used instead of Equation 2 for Z_3 and Equation 9 is replaced by

$$A(\Delta F)^2 + B(\Delta F) + C = 0 \tag{11}$$

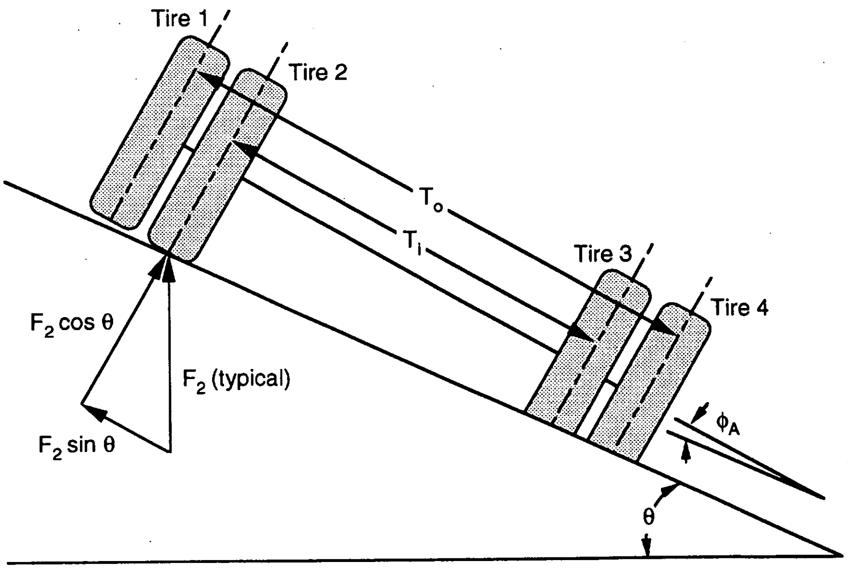


FIGURE 2 **Dual tires on a side slope.**

where

$$A = \frac{4.33}{F_R}\left(\frac{T_o + T_i}{2}\right)\cos\theta$$

$$B = 3.89\left(\frac{T_o + T_i}{2}\right) - \frac{4.33}{F_R}\left(\frac{T_o + T_i}{2}\right) W\cos\theta + T_i$$

$$C = -3.89\left(\frac{W}{2}\right)\left(\frac{T_o + T_i}{2}\right) + \frac{4.33}{F_R}\left(\frac{W}{2}\right)^2\left(\frac{T_o + T_i}{2}\right)\cos\theta$$

$$+ \frac{0.48T_i}{\cos\theta} + T_i\left(\frac{W}{2}\right) \qquad (12)$$

Equation 11 is to be used when Equation 10 cannot be satisfied. The load shift ΔF can be found by solving Equation 11 for a specified value of θ when Tire 2 is just barely touching the pavement and Tire 3 is in the nonlinear range or by using Equation 9 if Tire 3 is in the linear range. (The case in which both Tires 3 and 4 are in the nonlinear range is not addressed because the weight of the trailer would have to be unrealistically low.) The axle roll angle ϕ_A, is then found by using the value of the load shift ΔF in Equations 7 and 8 to get the resultant tire loads, using these (times $\cos\theta$) in Equation 2 or 3, as appropriate, to get the tire deflections, and then using the deflections in Equation 6.

With the dual tire configuration on a side slope the load shift ΔF can be found corresponding to θ, which provides static equilibrium of the trailer with Tire 2 barely touching the pavement. However, it has not yet been shown whether this state of balance is stable or unstable. If the roll axle angle ϕ_A is increased a slight additional amount, the overturning moment due to gravity will increase and the heights of the centers of gravity will also increase, both of which are destabilizing. However, the load will be shifted from Tire 3 to Tire 4, which will tend to resist rollover. To determine the net effect it is necessary to determine the relationships between the additional load shift and the increase in the axle roll angle for a fixed side slope angle.

Starting with a side slope angle θ, the load shift ΔF, and the axle roll angle θ_A, determined by the method described previously, it is then necessary to determine the additional load shift δF corresponding to an additional amount of axle roll, $\delta\phi_A$.

If Tires 3 and 4 are both deflected in their linear range, substitution of Equation 2 into Equation 6 yields

$$\sin(\phi_A + \delta\phi_A) = \frac{4}{S(T_o - T_i)}(\Delta F + \delta F)\cos\theta \qquad (13)$$

If ϕ_A and $\delta\phi_A$ are small angles (as they would typically be), $\sin(\phi_A + \delta\phi_A) \doteq \sin\phi_A + \sin\delta\phi_A$, Equation 13 becomes

$$\sin(\delta\phi_A) = \frac{4}{S(T_o - T_i)}(\delta F)\cos\theta \qquad (14)$$

However, if Tire 3 is deflected in its nonlinear range, Equation 3 is used in place of Equation 2 and equations parallel to Equation 13 and 14 can be developed. For the small angle assumption, Equation 14 would be replaced with

$$\sin\delta\phi_A = \frac{2\cos\theta}{S(T_o - T_i)}\left\{(\delta F)\left[1 + 3.89 - \frac{8.67}{F_R}\left(\frac{W}{2} - \Delta F\right)\cos\theta\right]\right.$$

$$\left. + \frac{4.33}{F_R}(\delta F)^2\cos\theta\right\} \qquad (15)$$

Wide-Base Single Tires on a Side Slope

Figure 3 illustrates the geometry for the wide-base tire situation. The tire deflections and the axle roll angle are related by

$$\sin\phi_A = (Z_2 - Z_1)/T \qquad (16)$$

In the following, ΔF again represents the amount of the axle load transferred between tires, but in this case it is the load transferred between Tires 1 and 2. If both tires are in the nonlinear range, then substituting from Equation 3 for both Z valves in Equation 16 gives

$$\sin\phi_A = \frac{2\Delta F}{ST}\left(3.89\cos\theta - \frac{4.33}{F_R} W\cos^2\theta\right) \qquad (17)$$

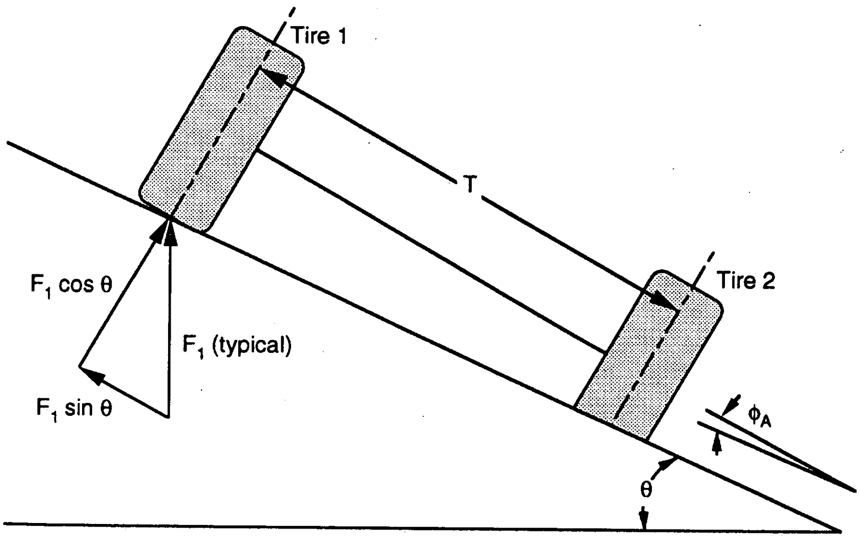


FIGURE 3 Wide-base single tires on a side slope.

whereas if Tire 1 is in the nonlinear range and Tire 2 is in the linear range,

$$\sin\phi_A = \frac{1}{ST}\left[0.48F_R + \left(\frac{W}{2} + \Delta F\right)\cos\theta - 3.89\left(\frac{W}{2} - \Delta F\right)\cos\theta\right] + \frac{1}{ST}\left[\frac{4.33}{F_R}\left(\frac{W}{2} - \Delta F\right)^2\cos^2\theta\right] \quad (18)$$

In the discussion of dual tires on a side slope, it was noted that for the side slope that produces equilibrium when Tire 2 carries no load but barely touches the pavement, it is necessary to determine whether the equilibrium is stable or unstable. With the wide base tires this equilibrium situation is always unstable, because any increase in ϕ_A will increase the overturning moments due to gravity as well as increase the heights of the centers of gravity, but no additional load shift is possible, so there will be no compensating restoring moment.

Cornering Acceleration with Zero Superelevation

All of the earlier discussion about the geometrics of the tire–axle systems is equally valid during a cornering maneuver except if the pavement has no superelevation (the simplest case), in which cos θ is equal to 1.0.

MOMENT EQUILIBRIUM EQUATIONS

If the trailer is to not roll over the overturning and stabilizing moments on the trailer must be in equilibrium (that is, be balanced). The trailer will roll over when the moments caused by the forces acting on the tires from the pavement can no longer balance the moments caused by gravity or lateral acceleration. In this subsection the equilibrium equations are developed for dual and wide-base tires for both side slope and cornering situations.

Moment Equations for Dual Tires on a Side Slope

Here and in the following, discussion the convention that counterclockwise moments are stabilizing (resist rollover) and are positive is adopted. Clockwise moments are negative, and if the net moment is negative, the trailer will roll over. The key angles and dimensions are shown in Figure 4. Figure 4 also shows the origin, 0, about which moments are summed. The selection of an origin is arbitrary, but this location, which is at the pavement surface along a perpendicular to the center of the axle, is convenient because the forces due to friction parallel to the pavement can be ignored because they create no moment.

The following symbols are further defined; typical values for the fixed quantities were given in Table 1:

h_s = the height of the center of gravity of the sprung mass, which is that portion of the trailer supported by the suspension springs;
h_u = the height of the center of gravity of the unsprung mass; the remainder of the trailer is not supported by the suspension, such as the tires and axles;
h_r = the height of the roll center, a point in space about which the sprung mass rotates;
ϕ_s = the roll angle of the sprung mass relative to a perpendicular to the axle; and
K_r = the roll stiffness of the trailer (per axle) created by the springs; as the sprung mass is rotated relative to the axle these springs create a moment resisting the roll.

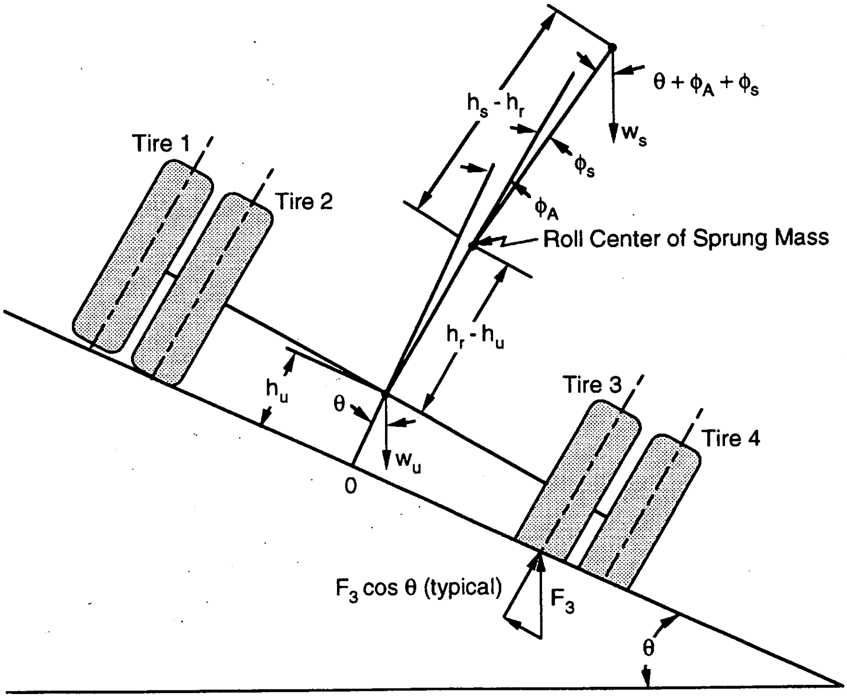


FIGURE 4 Key angles and dimensions for an axle with dual tires on a side slope.

First, it is convenient to consider an approximation in which the tires and the suspension are all rigid. That is, the trailer is treated as a rigid body. Summing moments about the origin for the case in which θ is just great enough to cause all of the trailer weight to be placed on Tire 4 gives

$$\frac{T_oW}{2}\cos\theta - (W_uh_u + W_sh_s)\sin\theta = 0 \tag{19}$$

For any lesser angle some weight would be on the other tires and no rollover will occur. If the angle is any greater, the (rigid) trailer will roll over. The critical side slope is thus

$$\tan\theta = \left(\frac{T_oW}{2}\right)/(W_uh_u + W_sh_s) \tag{20}$$

This approximation will prove useful later as a starting point for the iterative computation of the critical angle θ when the assumption that the trailer is rigid is dropped.

Now, dropping the rigidity assumption, if the trailer is allowed to roll against its suspension, the springs create a counterclockwise restoring moment $K_r\phi_s$, but the rolling displaces the sprung mass center of gravity, which creates a clockwise moment. Considering, first, only the sprung mass and taking moments about the roll center, the net moment due to the roll of the sprung mass is

$$M_r = K_r\phi_s - W_s(h_s - h_r)\sin(\theta + \phi_A + \phi_s) \tag{21}$$

If the sprung mass is in equilibrium, the net moment must be zero. If the angle ϕ_s is small one can approximate the sine term in Equation 21 as

$$\sin(\theta + \phi_A + \phi_s) \doteq [\sin(\theta + \phi_A) + \cos(\theta + \phi_A)\sin\phi_s] \tag{22}$$

in which the approximation $\cos\phi_s$ is taken as 1.0. Furthermore, one can set $\sin\phi_s$ equal to $(2\pi/360)\phi_s$, in which the factor $2\pi/360$ enables the angle to be written in degrees rather than radians. Making this substitution in Equation 22 and setting M_r to zero gives

$$\phi_s = \frac{\left[\left(\frac{W_s}{K_r}\right)(h_s - h_r)\sin(\theta + \phi_A)\right]}{\left[1 - \left(\frac{2\pi}{360}\right)\left(\frac{W_s}{K_r}\right)(h_s - h_r)\cos(\theta + \phi_A)\Big)\right]} \tag{23}$$

Next, considering the entire trailer depicted in Figure 4 and taking the situation in which all of the load is on Tires 3 and 4 and Tire 2 is barely in contact with the pavement, the sum of moments about the origin, 0, gives

$$\begin{aligned} M = \left(F_3\frac{T_i}{2} + F_4\frac{T_o}{2}\right)\cos\theta - W_uh_u\sin\theta - W_s[h_u\sin\theta \\ + (h_r - h_u)\sin(\theta + \phi_A) + (h_s - h_r)\sin(\theta + \phi_A + \phi_s)] \end{aligned} \tag{24}$$

Use of Equations 7 and 8 for the tire forces in Equation 24 gives

$$\begin{aligned} M = \left[W\left(\frac{T_o + T_i}{4}\right) + \Delta F\left(\frac{T_o - T_i}{2}\right)\right]\cos\theta - Wh_u\sin\theta \\ - W_s[(h_r - h_u)\sin(\theta + \phi_A) + (h_s - h_r)\sin(\theta + \phi_A + \phi_s)] \end{aligned} \tag{25}$$

When this equation is solved for M equal to 0, it gives the value of θ at which the trailer is in static equilibrium, with the trailer's weight totally on Tires 3 and 4 and with Tire 2 barely touching the pavement.

When a solution of Equation 25 for M equal to 0 is found, it indicates a condition of static equilibrium, but the condition may be stable or unstable. If it is in stable equilibrium a small increase in the axle roll ϕ_A will cause the net moment M to become positive (counterclockwise), and the trailer will tilt back to the equilibrium position. If it is unstable the small change will result in a negative net moment and the trailer will roll over.

If the axle roll angle is increased by a small amount, $\delta\phi_A$ (given in the previous subsection as a function of δF), each of the heights, h_u, h_r, and h_s will be increased by the amount

$$\delta h = \left(\frac{T_o + T_i}{4}\right)\sin(\delta\phi_A) \tag{26}$$

Under these circumstances the roll moment on the sprung mass M_r is increased by an amount δM_r. Replacing M_r with $M_r + \delta M_r$ in Equation 21 as well as replacing ϕ_A and ϕ_s with their incremental equivalents, making the usual small angle approximations for the incremental angles, and subtracting Equation 21 from the result yields

$$\delta\phi_s = \frac{\left[\frac{W_s}{K_R}(h_s - h_r)\left(\frac{2\pi}{360}\right)\cos(\theta + \phi_A + \phi_s)\right](\delta\phi_A)}{\left\{1 - \left[\frac{W_s}{K_R}(h_s - h_r)\left(\frac{2\pi}{360}\right)\cos(\theta + \phi_A + \phi_s)\Big)\Big]\right]\right\}} \tag{27}$$

Thus, for a small increase in δF the axle roll angle is increased by $\delta\phi_A$ and the sprung weight angle is increased by $\delta\phi_s$, which is given in Equation 27. Inserting the total amounts of the forces, heights, and angles in Equation 25 and then subtracting the M equal to 0 terms from the equilibrium solution, the incremental moment after using small angle approximations is expressed as

$$\begin{aligned} \delta M = \delta F\left(\frac{T_o - T_i}{2}\right)\cos\theta - W(\delta h)\sin\theta \\ - W_s[(h_r - h_u)\cos(\theta + \phi_A)]\sin(\delta\phi_A) \\ - W_s[(h_s - h_r)\cos(\theta + \phi_A + \phi_s)]\sin(\delta\phi_A + \delta\phi_s) \end{aligned} \tag{28}$$

This equation gives the incremental moment that would be produced if, from the static equilibrium position with Tire 2 barely touching the pavement, a small added rotation raised Tire 2 slightly. If δM is positive the trailer will return to the equilibrium position; if it is negative the trailer will roll over.

Moment Equations for Wide-Base Single Tires on Side Slope

Figure 4 is applicable for wide-base single tires on a side slope except that there are only two tires, as shown in Figure 3. Equation 23, for the angle ϕ_s, when the sprung mass is in equilibrium, is applicable here as well as for the dual tire case. For the moment on the overall unit, Equation 25 is replaced by

$$\begin{aligned} M = (\Delta F)T\cos\theta - Wh_u\sin\theta - W_s[(h_r - h_u)\sin(\theta + \phi_A) \\ + (h_s - h_r)\sin(\theta + \phi_A + \phi_s)] \end{aligned} \tag{29}$$

where ΔF is the load transferred from Tire 1 to Tire 2.

Dual Tires Subjected to Cornering Acceleration

Figure 5 illustrates the key dimensions, angles, and forces acting on a trailer, viewed from the rear, when it is turning left on a pavement with zero superelevation. To turn left horizontal forces on the tires (not shown) must be equal to the mass times the lateral acceleration (V^2/R, where V is the trailer speed and R is the radius of the turn) required to turn the trailer. The equal reacting forces tending to overturn the trailer are shown at the centers of gravity of the sprung and unsprung masses. These are $A(W_s/g)$ and $A(W_u/g)$, respectively, where A is the lateral acceleration and g is the acceleration due to gravity.

Proceeding as before, first consider the sprung mass. Writing the equation for the moments about the roll center for the sprung mass, using the small angle assumptions for ϕ_s, setting the net moment equal to zero for equilibrium, and solving for ϕ_s yields

$$\phi_s = C_1(\sin\phi_A + A/g) \qquad (30)$$

where

$$C_1 = \frac{W_s(h_s - h_r)}{\left[K_r - W_s(h_s - h_r)\left(\frac{2\pi}{360}\right)\right]} \qquad (31)$$

The moment acting on the overall unit for the case in which Tire 2 is just barely touching the pavement but carrying no load is

$$M = \frac{F_3T_i}{2} + \frac{F_4T_o}{2} - (W_uh_u + W_sh_s)\frac{A}{g} - W_s[(h_r - h_u)\sin\phi_A + (h_s - h_r)\sin(\phi_A + \phi_s)] \qquad (32)$$

Using Equations 7 and 8 for F_3 and F_4, the small angle assumptions for ϕ_s, and Equation 30 for ϕ_s yields

$$M = \frac{W}{4}(T_o + T_i) + \frac{\Delta F}{2}(T_o - T_i) - W_s(h_s - h_u)\sin\phi_A - \frac{2\pi}{360}W_sC_1(h_s - h_r)\left(\sin\phi_A + \frac{A}{g}\right) - (W_uh_u + W_sh_s)\frac{A}{g} \qquad (33)$$

Setting M equal to 0, for equilibrium, and solving Equation 33 for A/g gives

$$\frac{A}{g} = \left\{\frac{W}{4}(T_o + T_i) + \frac{\Delta F}{2}(T_o - T_i) - W_s\sin\phi_A\left[(h_s - h_u) + \frac{2\pi}{360}(h_s - h_r)C_1\right]\right\}\Big/\left[\frac{2\pi}{360}W_s(h_s - h_r)C_1 + (W_uh_u + W_sh_s)\right] \qquad (34)$$

Equation 34 gives the lateral acceleration that corresponds to the trailer being in equilibrium when Tire 2 carries no load but is barely in contact with the pavement.

The acceleration given by Equation 34 may or may not cause the trailer to overturn, because the equilibrium condition may be unstable or stable. A small additional amount of acceleration may cause the trailer to overturn, or it may just cause Tire 2 to raise from the pavement, causing a small increase in the angle ϕ_A, and thus transferring additional load from Tire 3 to Tire 4. The calculation procedure is analogous to that for the side slope situation. Equation 34 is modified by replacing ΔF with $\Delta F + \delta F$, ϕ_A with $\phi_A + \delta\phi_A$ from Equation 13 or 15, and h_i with $h_i + \delta F$ from Equation 26. The final solution is found as the maximum possible value of A/g corresponding to a value of δF in the range $0 \leq \delta F \leq F_3$.

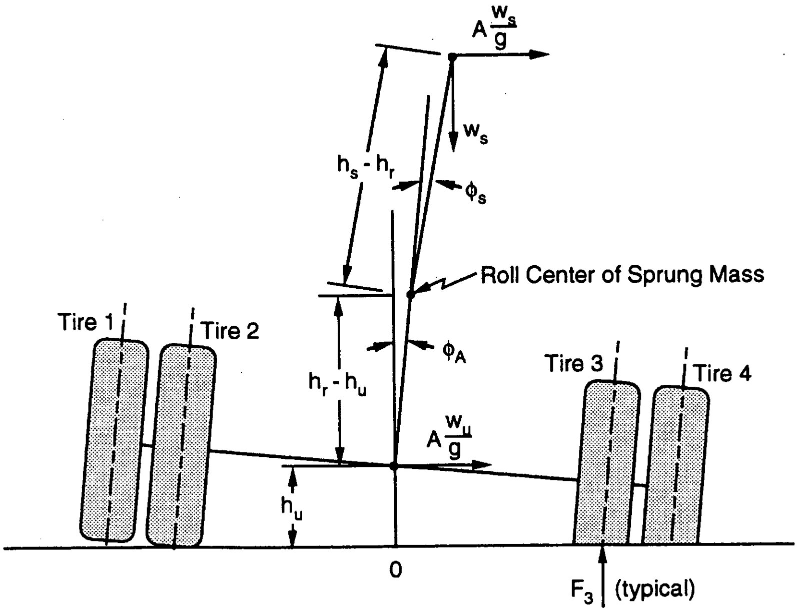


FIGURE 5 Dual wheel axle subject to cornering acceleration.

Wide-Base Tires Subject to Cornering Acceleration

Figure 5 is applicable for wide-base tires subject to cornering acceleration except that there are only two tires, as shown in Figure 3. Equation 30, which gives ϕ_s as a function of A/g for dual tires, is applicable for wide-base singles as well. Denoting the load shift from Tire 1 to Tire 2 as ΔF, the acceleration as a function of the load shift is given by

$$\frac{A}{g} = \frac{T(\Delta F) - W_s \sin\phi_A \left[(h_s - h_u) + \frac{2\pi}{360}(h_s - h_r)C_1 \right]}{\frac{2\pi}{360} W_s(h_s - h_r)C_1 + (W_u h_u + W_s h_s)} \tag{35}$$

NUMERICAL RESULTS

Dual Wheels on a Side Slope with an Empty Trailer

The general aim of the calculations is to determine the side slope angle θ for which the trailer is in equilibrium with Tire 2 unloaded but barely touching the pavement. Equilibrium is the condition when the net moment is zero. Having established the equilibrium condition, it is then necessary to determine if that condition is stable. If so, the calculations continue with larger side slope angles until enough load is transferred from Tire 3 to Tire 4 to render the trailer unstable.

The calculations must be conducted iteratively. The general approach is to choose a trial value of θ. Then, assuming that Tire 3 is loaded in the linear range, calculate ΔF by using Equation 9. Next, compare the normal load on Tire 3 with $F_R/3$ by using Equation 10. If the load is insufficient to place Tire 3 in the linear range, recompute ΔF by using Equation 11. Now, calculate ϕ_A from Equation 6, ϕ_s from Equation 23, and M from Equation 24. If M is positive, the trailer is stable and Tire 2 has not yet become unloaded, so a larger value of θ can be tried. If M is negative, Tire 2 has lifted from the pavement, so a smaller θ should be tried. This process is repeated until a result of M equal to 0.0 is found.

To start, recall that a first approximation can be found by assuming that the trailer and tires are rigid by using Equation 20. By using the data in Table 1, the solution is θ equal to 37.78 degrees. This is an upper bound on the side slope that will produce rollover. Suppose the iteration is started at 35 degrees. From Equation 9 it can be found that the trial value of ΔF is 648 kg (1,427 lb). Equation 7 indicates that the normal component of the load on Tire 3 is only 167 kg (367 lb), which is substantially less than $F_R/3$. Thus, the tire is not in the linear range, so ΔF is found to be 155 kg (342 lb) by solving Equation 11.

The loads on Tires 3 and 4 can then be found from Equations 7 and 8, respectively, as 570 and 824 kg (1,256 and 1,816 lb). Thus, Tire 3 is in the nonlinear range and Tire 4 is in the linear range. By using Equations 3 and 2, for Tires 3 and 4, respectively, their deflections are determined to be 18.4 and 22.4 mm (0.724 and 0.883 in.), respectively. The axle roll angle is then calculated from Equation 6 as 0.703 degrees. Next, the sprung mass roll angle ϕ_s is determined from Equation 23 as 0.827 degrees. Finally, the moment is calculated from Equation 25 to be −73 260 mm-kg (−6,353 in.-lb). This is negative, indicating that Tire 2 has lifted and the trailer will roll over on this side slope, so a smaller slope must be tried.

Table 2 shows the results of iterative calculations. At 34 degrees it will also roll over, but at 33 degrees the moment is positive, indicating that Tire 2 will not lift. For small changes in the side slope angle, the moment is nearly linear with the side slope angle. A simple interpolation between 33 and 34 degrees indicates that an angle of 33.549 degrees should be close. Table 2 bears this out by showing that the moment for an angle of 33.549 degrees is approximately zero.

Examination of this equilibrium status by using Equations 15, 26, 27, and 28 indicates that the trailer is in stable equilibrium; it will not roll over on this side slope. Therefore, larger side slope angles are examined by allowing load to be transferred to Tire 4. Carrying out further iterations by using Equations 23 and 25, but replacing h_u with δh in the latter, ultimately leads to the solution that θ is equal to 34.16 degrees. This is the angle at which all load is transferred to Tire 4, the equilibrium becomes unstable, and the trailer will roll over.

Dual Wheels on a Side Slope with a Loaded Trailer

By using the data from Table 1 for a loaded trailer, Equation 20 provides a starting estimate of θ equal to 29.95 degrees. Proceeding as before, a trial value of ΔF of 555 kg (1,223 lb) is found for θ equal to 23 degrees. Equation 10 confirms that Tire 3 is in the linear range (as is Tire 4), so the tire forces given by Equations 7 and 8 can be used with Equation 2 to find the tire deflections. Proceeding as with the unloaded trailer case, the moment M is −813 mm-kg (−9,373 in.-lb) for this side slope. Proceeding iteratively, a solution of 22.67 degrees is found for θ, which produces a negligible moment, and it can be shown to be an unstable equilibrium condition.

Wide-Base Tires on a Side Slope with an Empty Trailer

Examination of Equation 10 shows that for the empty trailer the tires are both in the nonlinear range, even if θ is large enough to transfer the entire weight to Tire 2. Therefore, Equation 17 is used to calculate ϕ_A, where ΔF is taken as $W/2$, corresponding to transferring all of the load from Tire 1 to Tire 2. Then, Equation 23 is used to find ϕ_s and Equation 29 is used to find M.

Proceeding in this fashion, iteration yields a solution of θ equal to 34.34 degrees as the equilibrium solution with M nearly equal to 0. This is a position of unstable equilibrium, as noted earlier.

TABLE 2 Calculations for Empty Trailer with Dual Tires

θ (°)	ΔF (kg)[a]	ϕ_A (°)	ϕ_s (°)	M (mm-kg)[a]
35	155.1	0.7028	0.8271	-551
34	157.2	0.7069	0.8072	-171
33	159.1	0.7109	0.7870	208
33.549	158.1	0.7087	0.7981	0

[a] 1 mm = 0.03937 in., 1 kg = 2.203 lbs.

Wide-Base Tires on a Side Slope with a Loaded Trailer

In the case of wide-base tires on a side slope with a loaded trailer, Tire 2 will be in the linear range but Tire 1 will be in the nonlinear range for the conditions of interest; that is, all or nearly all of the load is shifted to Tire 2. The calculations are conducted just as for the case describe in the previous section, except that Equation 18 instead of Equation 17 is used to calculate ϕ_A. The final solution is that unstable equilibrium will occur at θ equal to 23.95 degrees.

Wide-Base Tires on a Side Slope with a Loaded and Improved Trailer

Several factors combine to give tanker trucks with wide-base tires a potentially greater resistance to rollover. If the truck is fitted with axles with increased track width, the wider stance increases stability. In fact, these analyses already have assumed that this increased track width was used. When using wide-base tires on an axle designed for them it is also possible to widen the spring mount width, which also increases roll stability. Finally, with the wider wheel spacing it is possible to lower the tank on the frame; the lower center of gravity causes an additional improvement in roll stability. The one negative factor is that the wide-base tires have a lower vertical stiffness than a set of dual tires because a wide-base tire has only two side walls, compared with four side walls for a set of duals; this somewhat reduces their resistance to rollover.

Calculations were carried out to incorporate three types of trailer design improvements. The three improvements consist of widening the spring spacing on the frame by 15 percent, lowering the center of gravity of the tank and its contents by 150 mm (6 in.), and decreasing the moment arm from the center of gravity to the trailer's roll center (the point in space about which the trailer rotates) by raising the height of the roll center by 150 mm (6 in.) without changing the center of gravity. The latter difference is typical of what is experienced in the industry; the 15 percent increase in the spring spacing is illustrative, but an even greater increase is possible.

Changing K_r and h_s to reflect these improvements and carrying out the calculations in the same way as was done earlier, a solution of θ equal to 26.72 degree is found for the case of a loaded trailer on a side slope. This represents an angle that is nearly 12 percent greater than that associated with the loaded trailer without these improvements.

Cornering with Dual Tires on an Empty Trailer

For cornering with dual tires on an empty trailer and in other situations described later, the goal is to find the largest cornering acceleration possible that still provides an equilibrium situation. This maximum acceleration corresponds to the equilibrium being unstable, and the trailer will roll over. There is no superelevation, so θ is taken as 0 degrees. The case in which Tire 2 becomes unloaded but remains in contact with the pavement is examined first.

First, Equation 11 is used to determine ΔF, assuming that Tire 3 is in the nonlinear range, which it is. The value of ΔF is found to be 181 kg (398 lb). The value of ϕ_A is found from Equation 6, after calculating the tire deflections, Z_3 and Z_4 from Equations 3 and 2, respectively. Equation 34 is used to find A/g, and Equation 30 can then be used to find ϕ_s, if desired. The value of A/g, the rollover threshold for lateral acceleration, is found to be 0.6638. These results and others are given in the first row of Table 3.

The remainder of Table 3 provides the results of assuming that additional load can be shifted from Tire 3 to Tire 4. [The total amount of the load shift cannot exceed 851 kg (1,875 lb), half of the total weight on the axle.] The added load shift is designated by δF, the added axle roll is calculated with Equation 15, and δh is found from Equation 26. The modified Equation 34 is used for A/g. As shown in Table 3, increasing the load shift from Tire 3 to Tire 4 increases the cornering rollover threshold slightly, but A/g reaches a maximum of 0.6715 with a load shift of about 545 kg (1,200 lb). However, the difference between this acceleration and that at which Tire 2 begins to leave the pavement (0.6638) is not of great practical interest.

TABLE 3 Cornering Calculations for Empty Trailer with Dual Tires

$\Delta F + \delta F$ kg[a]	$\phi_A + \delta\phi_A$	δh mm[a]	A/g
181	0.7757	0	0.6638
454	2.3689	25.4	0.6712
545	3.0122	35.7	0.6715
635	3.7126	46.9	0.6708

[a] 1 mm = 0.03937 in., 1 kg = 2.203 lbs.

Cornering with Dual Tires on a Loaded Trailer

For cornering with dual tires on a loaded trailer Tires 3 and 4 can be shown to be in the linear range when Tire 2 is unloaded but barely touching the pavement. By using the appropriate equations, Equations 9, 2, 6, and 34, the acceleration producing an equilibrium condition is A/g equal to 0.4083. As before the effect of additional load transfer from Tire 3 to Tire 4 can be examined; in this case A/g does not increase, so this equilibrium condition is unstable and the trailer will roll over with this cornering acceleration.

Cornering with Wide-Base Tires on an Empty Trailer

As the entire load on Tire 1 is shifted to Tire 2, Tire 2 is in the linear range but Tire 1 is in the nonlinear range. Equation 18 is used for ϕ_A and Equation 35 is used for A/g. The solution is A/g equal to 0.6806.

Cornering with Wide-Base Tires on a Loaded Trailer

The solution for cornering with wide-base tires on a loaded trailer is found in the same way as the previous case. The solution is A/g equal to 0.4326, with the entire load on Tire 1 shifted to Tire 2.

TABLE 4 Effects of Improvements to Trailer with Wide-Base Tires

Improvement	A/g	$\delta A/g$
None	0.4326	—
K_r	0.4421	0.0095
h_s-h_r	0.4480	0.0154
h_s	0.4661	0.0335
Sum	—	0.0585
All	0.4923	0.0598

Dash (—) indicates data not applicable.

TABLE 5 Summary of Results from Trailer Rollover Analyses

Tire type	Loading	Design improvements	Critical side slope (degree)	Critical cornering accel. (g's)
Dual tires	Empty	None	34.16	0.672
Wide base	Empty	None	34.34	0.681
Dual tires	Loaded	None	22.67	0.408
Wide base	Loaded	None	23.95	0.433
Wide base	Loaded	Increased spring spacing width	—	0.442
Wide base	Loaded	Decreased center of gravity	—	0.466
Wide base	Loaded	Decreased moment arm	—	0.448
Wide base	Loaded	All three improvements	26.72	0.492

Dash (—) indicates data not calculated.

Cornering with Wide-Base Tires on a Loaded and Improved Trailer

A number of calculations were carried out with different improvements made to the trailer. These were

1. Widening the spring mount width to increase the sprung weight roll stiffness by 15 percent, to 1 061 000 mm-kg (92,000 in.-lb) per degree;
2. Raising the roll center h_r by 150 mm (6 in.) with no change in h_s, which reduces the moment arm $h_s - h_r$ by a like amount;
3. Lowering the center of gravity of the sprung mass h_s by 150 mm (6 in.) with no other change;
4. Summing the effects of changes 1, 2, and 3; and
5. Implementing changes 1, 2, and 3 simultaneously.

The results are given in Table 4, along with the case with no changes. Making all of the changes together causes a slightly greater improvement in rollover stability than the sum of the individual improvements, indicating that the improvements slightly aid each other. Overall, the rollover threshold with all of the improvements is improved about 13.8 percent over the case with none of these improvements.

Importance of Including Tire Deflections in the Calculations

As indicated early in this paper the analyses could be simplified by making the assumption that the trailer tires are rigid, that the tires do not deflect. One could make this assumption initially and rederive the various equations. Alternatively, one could repeat the (computerized) calculations with values of S, the tire stiffness, asymptotically approaching infinity. That is the approach used in the present study.

For illustration, in the case of a loaded trailer with rigid dual tires, the critical side slope is 23.93 degrees, a 5.6 percent increase over the more accurate calculation. Similarly, use of the assumption of rigid wide-base tires on a fully loaded trailer will increase the critical side slope by 6.1 percent to 25.40 degrees.

CONCLUSIONS

Improved rollover stability is probably the greatest potential safety advantage to the use of wide-base tires. This paper presented the basic methods of analysis of this stability question from the viewpoints of resistance to rolling over while resting on a side slope and of a trailer's ability to undergo a cornering maneuver without rolling over. The results of a series of calculations are presented in Table 5.

When conventional dual tires and axles are replaced with wide-base tires and axles designed for them and no other changes are made to the trailer, stability is improved, but by relatively small amounts. If the trailer is empty, in particular, the increases in the critical side slope and in the critical cornering acceleration are barely perceptible. If the trailer is loaded, however, the improvement in stability that can be realized by switching from dual tires to wide-base tires is more noticeable (about 6 percent), but still small.

If, in addition to switching from dual tires to wide-base tires, the trailer is modified to take advantage of the wider stance of the tires, further stability increases are possible. If all three trailer improvements considered here are incorporated, the critical cornering acceleration can be increased by an additional 13.6 percent over that achievable by replacing wide-base tires with dual tires but with no other changes. Furthermore, the allowable cornering acceleration that is possible is increased by nearly 21 percent over that with dual tires when the dual tires are replaced with wide-base tires and all three trailer improvements are made. The increase in critical side slope is similar (nearly 18 percent) when these same changes are made.

ACKNOWLEDGMENT

The research reported here is a portion of a Midwest Research Institute study sponsored by the Western States Petroleum Association.

REFERENCES

1. Fancher, P. S., and A. Mathews. *A Vehicle Dynamics Handbook for Single-Unit and Articulated Heavy Trucks.* Report DOT HS 807 185. NHTSA, U.S. Department of Transportation, May 1987.
2. Fancher, P. S., et al. *A Factbook of the Mechanical Properties of the Components for Single-Unit and Articulated Heavy Trucks.* Report DOT HS 807 125. NHTSA, U.S. Department of Transportation, Dec. 1986.

Publication of this paper sponsored by Committee on Motor Vehicle Technology.

External Viewing of Vehicle Contents Under Various Tinting and Illumination Conditions

DENNIS R. PROFFITT, JANE E. JOSEPH, MUKUL BHALLA, MARCO BERTAMINI, FRANK H. DURGIN, CHERYL W. LYNN, AND JACK D. JERNIGAN

The purpose of the study was to determine the degree to which motor vehicle window tint films impede a police officer's ability to see clearly into a stopped vehicle. Three hundred twenty subjects were asked to view the contents and occupants of one of four experimental cars. One car had no aftermarket tint film, and the other three had aftermarket tint films tinted to various degrees. Although similar experiments have been conducted, all of those experiments yielded equivocal results because of methodological flaws. The present experiment was an attempt to correct some of those problems and to simulate standard procedures used during traffic stops by the Virginia State Police. In general, the study found that the abilities of subjects to detect occupants and objects in vehicles were substantially diminished as the degree of window tinting increased. However, at night the detrimental effects were substantially reduced when headlights and a spotlight were shone on the vehicle, as would be the case in a traffic stop.

The aftermarket tinting of motor vehicle window glass raises at least three distinct traffic safety concerns. First, the driver of an automobile may encounter situations in which visibility is impeded by tinted windows. Second, visual communication between drivers and pedestrians, cyclists, or other drivers may be impaired. Third, the safety of police officers who must approach a stopped car on foot may be endangered. Tinting may impede an officer's ability to detect weapons, contraband, or threatening acts by the driver or passengers. This last concern motivated the current experiment.

Three studies and one demonstration have investigated the influence of tinting on the ability to identify objects within a parked car. One study was sponsored by tint film manufacturers (*1*), and the others were conducted by police departments (*2–4*). The study sponsored by manufacturers found no detrimental effects from window tinting, even for tinting films with transmittance values as low as 20 percent. The studies sponsored by police departments found that window tinting greatly reduced the abilities of officers to identify objects inside the experimental cars. These conflicting results are likely due to differences in the study designs. More important, all of the studies suffered from serious design flaws that make it difficult to draw any generalizations.

PURPOSE AND SCOPE

The purpose of the study described here was to determine to what degree tinted window films impede an officer's ability to see clearly into a stopped vehicle. Similar experiments have all yielded equivocal results because of serious methodological flaws. The present experiment was an attempt to correct those methodological problems. Every attempt was made to make the procedure for approaching the vehicles used in the experiment as similar to standard police procedures as possible. The experiment was limited to testing the procedures used by police in Virginia.

Although the experiment was designed to control for differences other than transmittance, the generality of the results is somewhat limited. First, only three levels of tinting were tested. Second, ambient light differed somewhat on different testing days and nights. More important, light reflectance, which could not be measured and controlled, affects the ability to see clearly into a vehicle. Reflectance is so situation specific that it is impossible to make broad conclusions that apply to all conditions. In the current experiment assessments were made at five different testing locations to minimize effects that might be specific to a particular location or a vehicle's orientation to a reflection source such as the sun.

METHODS

In the present study observers attempted to identify the interior contents of vehicles tinted to have various levels of light transmittance. Standard police procedures were simulated, and viewing occurred during daytime, dusk, and nighttime hours and during nighttime hours with the use of auxiliary lighting. Under standard police procedures officers park their vehicle about 6.1 m (20 ft) behind the vehicle being stopped, with the headlights pointing slightly more toward the driver side door. The officer then attempts to ascertain the number of occupants. The officer then walks toward the vehicle, stopping at the rear of the driver side of the vehicle, and determines if the trunk is latched. The officer then approaches the driver side of the vehicle at a distance of about 0.5 m (1.5 ft), stopping just behind the driver side door so that it cannot be opened into him or her. During the approach the officer scans the back seat and its passengers but concentrates on the driver. At night the officer rests a spotlight on his or her shoulder away from his or her dominant hand and uses it to see into the interior of the vehicle.

The 160 male and 160 female volunteers who participated in the study were primarily undergraduate students who were passing by the testing locations on the grounds of the University of Virginia. Others were students enrolled in an introductory psychology course who participated to fulfill a course requirement. All subjects were asked if they needed eyeglasses to drive. If they stated that their driver's license stipulated that they wear glasses, they were asked to

---

D. R. Proffitt, J. E. Joseph, M. Bhalla, M. Bertamini, and F. H. Durgin, University of Virginia, Charlottesville, Va. 22903. C. W. Lynn and J. D. Jernigan, Virginia Transportation Research Council, 530 Edgemont Road, Charlottesville, Va. 22903.

wear them during the testing. If they did not have their glasses with them, they were excluded from the study.

All four test vehicles were 1987 Dodge Aries K four-door sedans. All had identical blue exterior paint, dark blue interiors, and black dashboards. The only difference among the vehicles was the degree of window tinting on the side and rear windows.

Table 1 shows the target and actual transmittance values for the four test vehicles. Actual levels differ from the prescribed levels to some degree. Because tint film is applied over factory glass tinted to different transmittance levels, the resulting level of light transmittance is multiplicative. For instance, a 50 percent aftermarket film applied over an 82 percent factory tinted window theoretically results in a total transmittance of 41 percent ($0.50 \times 0.82 = 0.41$). Since the same aftermarket film is applied to factory glass with different levels of tinting, different results are achieved with different vehicles. In this case the prescribed and actual transmittance values were reasonably close, with the actual transmittance slightly higher than the prescribed transmittance except for Vehicle 3, in which the actual transmittance values were quite a bit lower than ordered. No aftermarket tint film was applied to Vehicle 0, Vehicle 2 represented the maximum reduction in transmittance allowed by Virginia law, and the transmittance values chosen for Vehicle 1 represented intermediate levels. The prescribed levels of tinting for Vehicle 3 represented the maximum reduction of transmittance allowed by any state in the nation, that being Florida.

The objects placed inside the vehicle were arranged in the same way for each test episode. Three mannequins were seated upright in the vehicle: one in the driver's seat, one in the front passenger's seat, and one in the back right passenger's seat. The mannequin in the driver's seat held a pair of scissors in the left hand. Its hands were arranged such that the right hand was placed by its right side and was covered slightly by the right pant leg. The left hand, holding the scissors, was positioned at the bottom part of the steering wheel. Five common objects of various colors and sizes were arranged on the back left seat and back left floor: a black flashlight, a yellow highlighter pen, and a red soda can were placed on the seat, and a pink spiral notebook and a white tennis shoe were placed on the floor.

A stopwatch was used for all timing. Tests without auxiliary lighting took place during the months of October through December 1993. Testing occurred at three different times of day (midday, 2:00 to 3:30 p.m.; dusk, 5:00 to 5:45 p.m.; and night, 6:00 to 7:30 p.m.) and in five different locations on the grounds of the University of Virginia. During April 1994 additional tests were conducted with auxiliary lighting at night to simulate a police stop in which the police cruiser's headlights and the officer's flashlight would shine light into the stopped vehicle. Headlights were shone into the vehicle being viewed, and observers used a handheld spotlight to look into the vehicle. Each observer saw only one car, and the identification rate was determined by examining the performances of different groups of subjects.

Six performance measures were taken:

1. Mannequin detection: detection of the number of mannequins seated upright in the vehicle from a viewing distance of 6.1 m (20 ft) behind the vehicle;
2. Confidence rating: the level of certainty reported about the number of mannequins detected;
3. Distance at certainty: how close to the vehicle the subject needed to be to state with confidence that there were three mannequins;
4. Detection of driver's hand positions: detection of the position of the driver's hands when looking into the front side window and standing approximately 0.5 m (1.5 ft) from the front window;
5. Detection of object in driver's hand: detection of the object that the driver was holding in the left hand when looking into the front side window and standing approximately 0.5 m (1.5 ft) from the front window; and
6. Rear seat object detection: detection of the five objects that were on the backseat and back floor of the vehicle on the driver's side when looking into the back side window approximately 0.5 m (1.5 ft) from the vehicle.

The main independent variables of interest in the study were level of tinting (factory, 50 percent/50 percent, 50 percent/35 percent, and 35 percent/20 percent transmittance windows) and viewing condition (midday, dusk, night, and night with auxiliary lighting). Although testing occurred in different locations, this variable was introduced only to minimize situation-specific effects. Preliminary analyses suggested that there were differences in locations for the different tasks, but these differences were not systematic. That is, one particular testing location was not associated with systemati-

TABLE 1 Transmittance Values for Test Vehicles

	Target [a] (%)			Actual [b] (%)	
Vehicle	Windshield	Front Side	Rear/ Rear Side	Front Side	Rear Side
0	*	*	*	88	88
1	*	50	50	53	53
2	*	50	35	53	38
3	*	35	20	40	13

[a] Target transmittance is the intended transmittance aftertint film is applied over a factory-tinted window.

[b] Actual transmittance for the rear window could not be determined because the measuring device can be used only on an automobile window that will open.

* No aftermarket tint film was applied.

cally better performance than another location across all of the tasks. Thus, the analyses presented exclude the location variable.

Each of the six dependent measures was submitted to a 4 (tinting level: Vehicles 0, 1, 2, and 3) × 4 (viewing condition: midday, dusk, night, and night with auxiliary lighting) analysis of variance (ANOVA). Both factors were manipulated between subjects.

RESULTS AND DISCUSSION OF RESULTS

Mannequin Detection

Figure 1 shows the effect of tinting on detecting the number of mannequins seated upright in the vehicle from 6.1 m (20 ft). For viewing at night with auxiliary lighting there was a decline in performance as the tinting level increased; however, this trend was not statistically significant [$F(3, 79) = 1.6, P = .19$].

The ANOVA for comparison of the effect of tinting for all four viewing conditions revealed a main effect of tinting [$F(3, 319) = 15.3, P < .001$]. As shown in Figure 1, the trend was a decline in performance as tinting level increased. Dunnett's one-tailed *t* test (*5*) compared each level of tinting with that of the control level (Vehicle 0) and revealed that performance with Vehicles 2 and 3 was different from performance with Vehicle 0 but that performance with Vehicle 1 was not different from performance with Vehicle 0. The main effect of viewing condition was also significant [$F(3, 319) = 7.2, P < .001$]. Each of the six performance measures was submitted to a preliminary ANOVA. Dunnett's test compared each of the viewing conditions with that of the control condition, that is, midday viewing, and revealed that performance was worse at dusk, night, and night with auxiliary lighting compared with that at midday. However, the significant tinting level × viewing condition interaction [$F(9, 319) = 2.4, P < .05$] qualified this effect further. Performance declined as transmittance level decreased from 35 percent (Vehicle 2) to 20 percent (Vehicle 3) when viewing occurred at night without auxiliary lighting and at dusk. However, there was no decrement in performance from 35 to 20 percent transmittance when viewing occurred at midday and at night with auxiliary lighting. In other words performance declined when viewing the most heavily tinted car at dusk and at night, but performance was not worse with the most tinted car when viewing occurred at midday and at night with auxiliary lighting. Thus, the use of auxiliary lighting at night overcame the effect of heavy tinting. The overall model explained 23 percent of the variance in responses [$F(15, 319) = 5.9, P < .001$].

Confidence Ratings for Mannequin Detection

Unlike mannequin detection performance, confidence or certainty about the number of mannequins reported from a viewing distance of 6.1 m (20 ft) behind the vehicle was affected by tinting when viewing occurred at night with auxiliary lighting. Figure 2 shows that confidence in reports dropped significantly as tinting level increased [$F(3, 79) = 8.7, P < .001$]. Dunnett's test compared each level of tinting with that of the control level (Vehicle 0) and revealed that confidence when viewing the most tinted vehicle was significantly lower compared with the control level but that confidence with the other two tinting levels was not different from that with the control level. The overall model explained 33 percent of the variance in responses [$F(7, 79) = 5.1, P < .001$].

The analysis that compared the effect of tinting for the four viewing conditions revealed a main effect of tinting [$F(3, 319) = 28.7, P < .001$] and a main effect of viewing condition [$F(3, 319) = 8.4, P < .001$], but no interaction. As shown in Figure 2, confidence decreased as tinting level increased, and confidence was generally greater at midday and lower at night. Dunnett's test revealed that confidence ratings with Vehicle 2 and Vehicle 3 were lower than

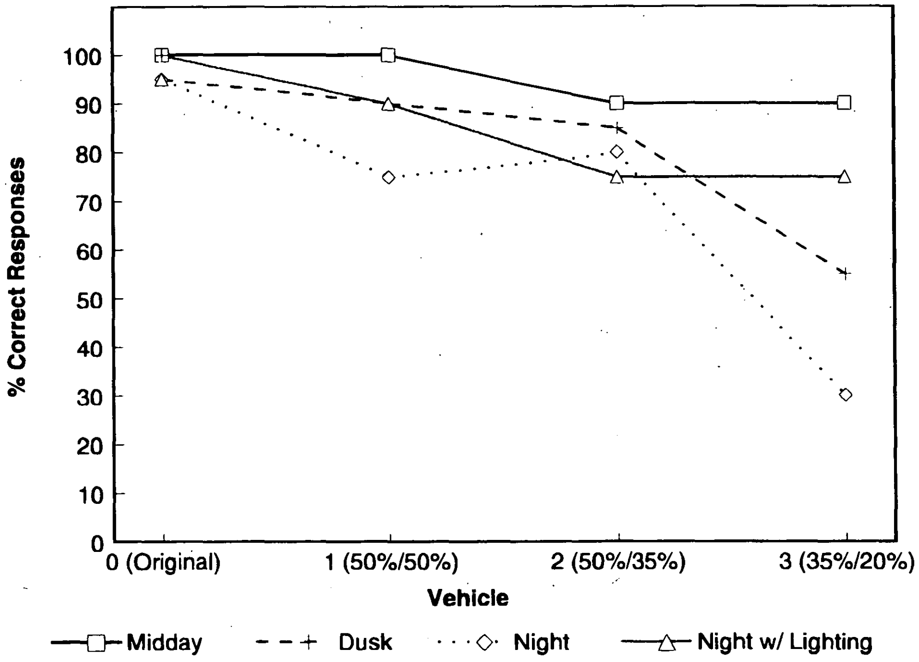


FIGURE 1 Effect of window tinting and viewing condition on mannequin detection.

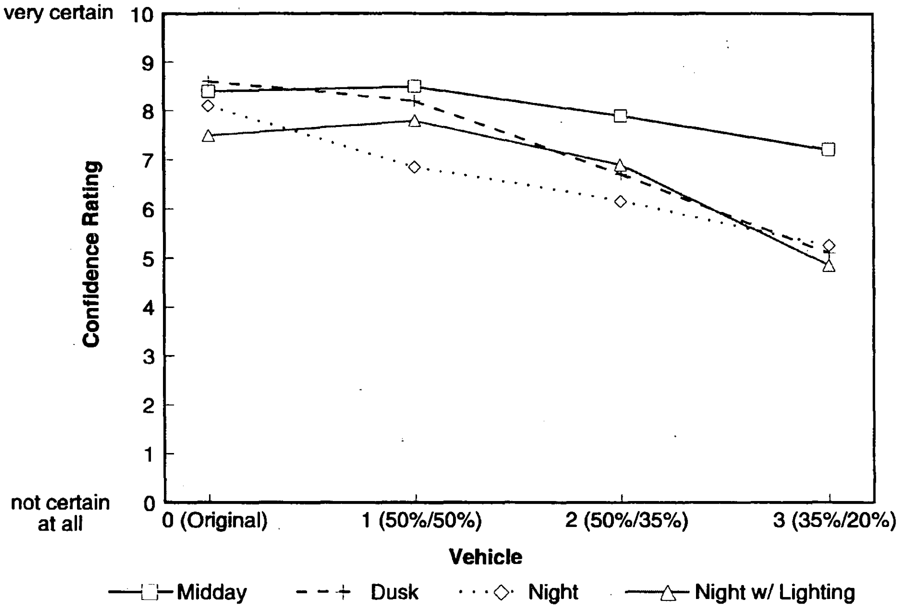


FIGURE 2 Effect of window tinting and viewing condition on confidence ratings.

those with Vehicle 0. Ratings with Vehicle 1 were not different from ratings with Vehicle 0. The overall model explained 29 percent of the variance in responses [$F(15, 319) = 8.3$, $P < .001$]. Dunnett's test also revealed that confidence at midday was higher than that at dusk, night, or night with auxiliary lighting.

Distance at Certainty

The analysis that compared the effect of tinting for all four viewing conditions revealed a main effect of tinting level [$F(3, 299) = 22.8$, $P < .001$]. As shown in Figure 3, the trend for all four viewing conditions was a decrease in distance as tinting level increased. That is, subjects needed to be closer to the more tinted vehicles to be certain about the number of mannequins. Dunnett's test revealed that in comparison with the control vehicle (Vehicle 0), subjects needed to be significantly closer to all of the other vehicles, implying that even lower levels of tinting can lower confidence about reports of the number of occupants inside a vehicle. The main effect of viewing condition was also significant [$F(3, 299) = 2.9$, $P < .05$]. That is, the distance at which subjects became certain about the number of mannequins was generally greater when viewing occurred at mid-

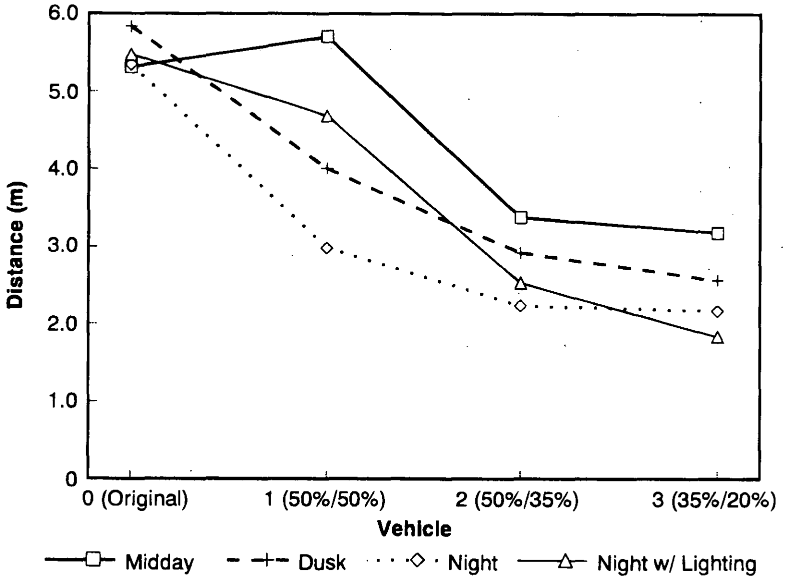


FIGURE 3 Effect of window tinting and viewing condition on distance at certainty.

day and was lower at night. In fact, Dunnett's test revealed that distance at certainty at night was significantly lower than distance at certainty at midday. However, this distance at dusk or at night with auxiliary lighting was not significantly different from this distance at midday. The fact that the tinting level × viewing condition interaction was not significant, however, implies that greater tinting lowered confidence in all viewing conditions. The overall model explained 23 percent of the variance in responses [$F(15, 299) = 5.7$, $P < .001$].

Detecting Positions of Driver's Hands

Subjects were good at detecting the positions of the driver's hands with the use of auxiliary lighting regardless of tinting level (Figure 4). The effect of tinting was not statistically significant [$F(3, 79) = 0.8, p = .49$].

The ANOVA that assessed the effect of tinting for all four viewing conditions revealed a main effect for viewing condition [$F(3, 319) = 33.0$, $P < .001$], but no main effect for tinting and no interaction. As shown in Figure 4, detection of the positions of the driver's hands was much worse at night without auxiliary lighting compared with that under the other three viewing conditions. Dunnett's test revealed that only performance at night without auxiliary lighting was worse than performance at midday. The other two viewing conditions were associated with levels of performance similar to that achieved at midday. The overall model explained 27 percent of the variance in responses [$F(15, 319) = 7.4$, $P < .001$].

Detecting Object in Driver's Hand

With the use of auxiliary lighting subjects were very good at reporting the object that the driver was holding in the left hand. Only one subject failed to detect the object. Tinting had no effect on performance [$F(3, 79) = 1.0$, $P = .4$].

Figure 5 shows the effect of tinting on performance for the four viewing conditions. Detecting the object held in the driver's left hand was not affected by the level of tinting but was affected by viewing condition [$F(3, 319) = 35.1$, $P < .001$]. Dunnett's test revealed that the use of auxiliary lighting at night yielded performance comparable to that at midday. However, performance at dusk and at night without auxiliary lighting was significantly worse than that at midday. The overall model explained 27 percent of the variance in responses [$F(15, 319) = 7.6$, $P < .001$].

Detecting Objects in Backseat and Back Floor

The use of auxiliary lighting allowed subjects to report, on average, 3.45 of 5 objects on the backseat and back floor of the vehicle. As shown in Figure 6, tinting had no effect on performance. The use of auxiliary lighting overcame the effects of window tinting, even with the most tinted vehicles [$F(3, 79) = 0.6$, $P = .65$].

Figure 6 shows the effect of tinting on object detection for the four viewing conditions. The main effect for tinting was significant [$F(3, 319) = 56.3$, $P < .001$], indicating that fewer objects were detected with the more heavily tinted vehicles. In fact, Dunnett's test revealed that performance was worse with all levels of tinting compared with that with the control level (Vehicle 0). The main effect for viewing condition was also significant [$F(3, 319) = 232.0$, $P < .001$], indicating that the fewest objects were detected at night with no auxiliary lighting and the most objects were detected at night with auxiliary lighting. The significant tinting level × viewing condition interaction [$F(9, 319) = 13.6$, $P < .001$], shown in Figure 6, indicates that tinting affected object detection at midday and at dusk so that fewer objects were detected when the windows were more tinted. Object detection was consistently high at night with auxiliary lighting and consistently poor at night without it. The overall model

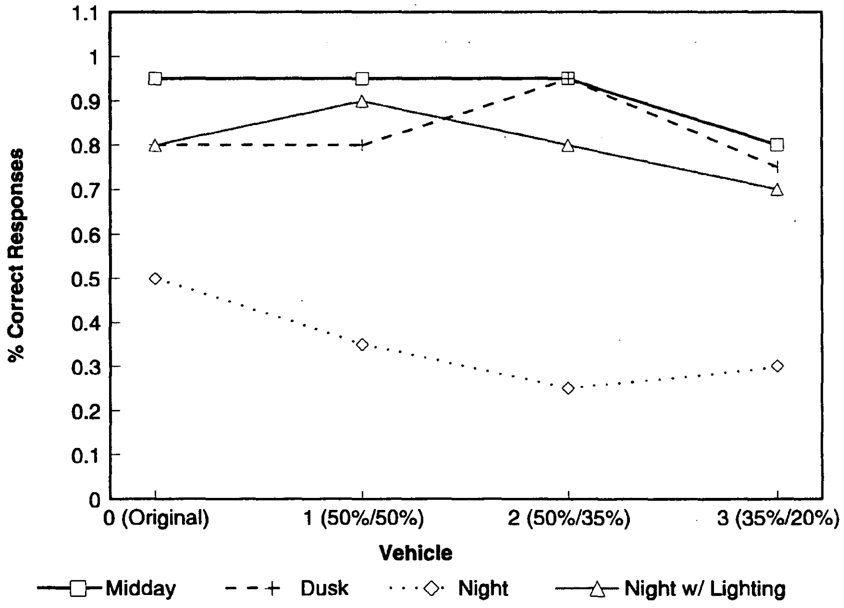


FIGURE 4 Effect of window tinting and viewing condition on detecting positions of driver's hands.

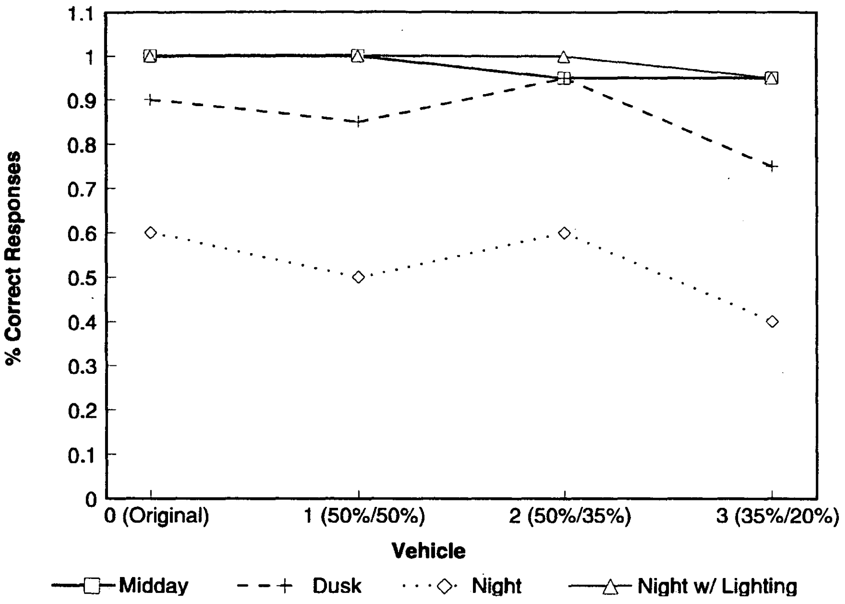


FIGURE 5 Effect of window tinting and viewing condition on detecting object in driver's hand.

explained 76 percent of the variance in responses [$F(15, 319) = 65.8$, $P < .001$].

SUMMARY AND CONCLUSIONS

In general, higher degrees of window tinting made seeing inside a vehicle more difficult. Window tinting impaired performance of four of the six tasks in the study (mannequin detection, certainty in mannequin detection, distance at certainty, and object detection). For all of these tasks the heaviest tinting level (Vehicle 3 representing the maximum level for any state) significantly impaired performance relative to that with no tinting. The maximum legal level allowed in Virginia (Vehicle 2) also significantly impaired performance of these four tasks relative to that with no added tinting. An intermediate level of tinting (Vehicle 1) impaired performance of only two of these four tasks (distance at uncertainty and object detection) relative to that with no tinting. The four tasks that were

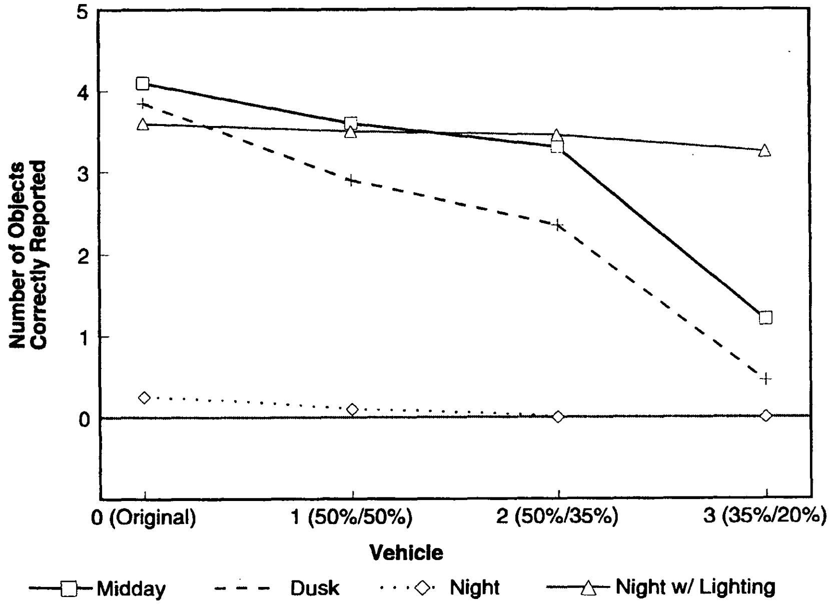


FIGURE 6 Effect of window tinting and viewing condition on object detection in backseat and back floor.

impaired by window tinting all involved looking into the vehicle through the rear window or the rear side windows. The two tasks that were not affected by window tinting (detecting the position of the driver's hands and detecting the object held in the driver's hand) involved looking into the vehicle through the front side window, which had the least amount of tint film applied relative to the amount applied to the other windows of the same vehicle (Table 1). In sum tinting affected looking into a vehicle through the back windows but not through the front side window. Also, the legal limits of window tinting allowed in Florida and Virginia significantly impaired an individual's ability to see inside those vehicles through the rear windows.

Not surprisingly, poor viewing conditions (night and dusk viewing) impaired subjects' ability to see inside vehicles. All six of the tasks used in the study were negatively affected by poor viewing conditions. Viewing the vehicle contents at night without the use of auxiliary lighting was significantly worse than viewing the contents during midday for all of the tasks. Many of the tasks were also more difficult when viewing occurred at dusk than when viewing occurred at midday (mannequin detection, certainty in mannequin detection, detecting the object held in the driver's hand, and object detection). However, only two of the tasks were more difficult when viewing occurred at night with the use of auxiliary lighting than when viewing occurred at midday (mannequin detection and confidence in mannequin detection). In other words the use of auxiliary lighting at night significantly improved performance relative to viewing at night without auxiliary lighting for four of the tasks. In fact, for two of the tasks (mannequin detection and object detection) the use of auxiliary lighting overcame the effects of heavy window tinting.

The experimental results indicate that the detrimental effects of window tinting on viewing people and objects within a stopped vehicle at night are greatly reduced by the use of auxiliary lighting. Performance at night with the use of auxiliary lighting was not affected by window tinting in terms of the accuracy of any of the judgments about what was inside the vehicle. The level of window tinting did not influence mannequin detection, reporting the positions of the driver's hands, reporting the object held in the driver's hand, or reporting the objects on the backseat and back floor.

Window tinting levels did affect the subjects' confidence in their judgments about the number of occupants in the vehicles. At the initial 6.1-m (20-ft) viewing distance, subjects' confidence in the accuracies of their judgments about the number of mannequins in the vehicles decreased with the degree of window tinting. Moreover, the distance from the vehicle at which subjects felt confident in this judgment decreased with the degree of window tinting.

Relative to nighttime viewing without auxiliary lighting, auxiliary lighting significantly improved the accuracy of performance on all of the assessments of vehicular contents: mannequin detection, detecting the positions of the driver's hands, detecting the object in the driver's hand, and object detection. Only confidence and the distance at which subjects felt confident were not affected by the use of auxiliary lighting.

The use of auxiliary lighting to detect objects within the vehicles from close range dramatically improved performance. Subjects who used auxiliary lighting at night to detect objects in the backseat and on the back floor detected more objects than subjects performing under all other viewing conditions, including midday viewing. In addition, nighttime performance with the use of auxiliary lighting was better for the task of detecting the weapon in the driver's hand compared with performance at dusk.

It is tempting to conclude that police officers might do well to use auxiliary lighting during daytime and dusk hours when approaching a stopped vehicle with tinted windows. This conclusion cannot be made, however, without performing an empirical study since the effectiveness of auxiliary lighting may interact with the overall level of ambient illumination that is present at different times of day.

REFERENCES

1. *Visibility into Automobiles.* IIT Research Institute, May 1990.
2. Boyd, P. *Report to Congress on Tinting of Motor Vehicle Windows.* NHTSA, U.S. Department of Transportation, 1991.
3. New York State Police. *The Effect of Aftermarket Tinting on Police Officer Safety.* New York Department of State Police and Motor Vehicles, 1992.
4. Virginia State Police. *Report on Sunscreening Material on Motor Vehicles.* Report of the Departments of State Police and Motor Vehicles, Dec. 1988.
5. Dunnett, C. W. A Multiple Comparisons Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association*, Vol. 50, 1955, pp. 1096–1121.

Publication of this paper sponsored by Committee on Traffic Law Enforcement.

To Belt or Not To Belt: Should Florida Mandate Installation of Safety Restraints in Large School Buses?

MICHAEL R. BALTES

A summary of a report that focused on the many issues related to the installation of safety restraints and various other feasible safety investment options for large school buses in Florida to make a safety investment option recommendation to the Florida legislature is presented. To accomplish this objective the existing literature was reviewed and evaluated to draw conclusions from the accumulated evidence. In addition, two supplemental analyses were performed: a safety investment cost–benefit analysis and a descriptive analysis of Florida school bus accident data. Based on the evidence gathered from the literature review and the results from the two supplemental analyses, it was concluded that the installation of safety restraints will not significantly improve the overall safety performance of large school buses in Florida. The potential benefit to be obtained from the installation and use of safety restraints, quantified by the annual fatalities and injuries prevented per annual dollar invested, was shown to be diminutive and thus not cost-effective. However, other feasible safety investment options were shown to be significantly more cost-effective in terms of their potential to prevent fatalities and injuries per annual dollar invested.

For the last 25 years approximately 40,000 people have died annually as a result of traffic crashes in the United States. Although airplane crashes and train accidents receive a greater portion of the media spotlight, the number of fatalities and life-threatening injuries involving passenger cars greatly exceeds those sustained in all other modes of surface transportation combined. Data compiled by the National Center for Statistics and Analysis reveal that in the United States in 1991

- 6.1 million traffic accidents occurred, a rate of 1 every 5 sec;
- severe or fatal injuries occurred at a rate of 1 every 88 sec; and
- minor or moderate injuries occurred at a rate of 1 every 19 sec.

NHTSA estimates that approximately 50 percent of all traffic fatalities could be prevented annually if all front-seat outboard occupants wore safety restraints. Also, NHTSA estimates that between 1983 and 1990 safety restraints saved nearly 25,000 lives and prevented about 650,000 moderate to critical injuries.

The ability of safety restraints to reduce fatalities and serious injuries to automobile occupants when accidents occur has been recognized (*1–9*), resulting in their mandatory use in all but two states. According to information disseminated by NHTSA, as of July 1994 the states of Maine and New Hampshire do not require the mandatory use of safety restraints in passenger cars. The federal government requires that three-point safety restraints (lap belts with shoulder harnesses) be installed as standard equipment in the front outboard seating positions of automobiles, light trucks, and vans. By 1995 all automobile manufacturers will be required to install three-point safety restraints in the rear outboard seating positions as standard equipment as well; currently, only lap belt safety restraints are required. Although safety restraints in passenger cars, light trucks, and vans have proven to be effective life-saving and injury-mitigating devices, their effectiveness in other heavier vehicles, such as heavy trucks, transit buses, and large school buses has not been substantiated empirically. Large school buses are Type B, C, and D school buses [gross vehicle weight (GVW), >4,540 kg (10,000 lbs)]. Type A [GVW, <4,540 kg (10,000 lbs)] school buses are required by federal law to have seat belts.

This summary paper involves the investigation of the available literature to date from technical reports, journals, and periodicals pertaining to the issue of large school bus safety. It should be made clear to the reader that the scope of the report and this summary paper focuses only on those issues associated with safety on board large school buses and not on the development of programs and safety devices to protect children in the loading and unloading zones. In addition to the comprehensive review of the pertinent literature, two supplemental analyses were performed: a safety investment cost–benefit analysis and a descriptive analysis of Florida school bus accident data. The results of this research effort are summarized in this paper.

PROPONENT AND OPPONENT VIEWS

It is frequently assumed by the general public that since safety restraints have proven their effectiveness in passenger cars and other small vehicles, their availability and use in large school buses will produce the same fatality- and injury-mitigating benefits. The installation of safety restraints and the issue of safety restraint use is a frequent topic of discussion among school transportation professionals and a topic frequently raised by concerned parents of school children. The states of New York and New Jersey, as well as numerous school districts (the number of school districts in the United States that operate all or a proportion of their large school bus fleet with safety restraints is unknown) across the United States, have recently implemented legislation or policies mandating the installation and use of safety restraints (lap belts only) in their large school buses. Interestingly, the state of New York requires that lap belts be installed in all newly purchased large school buses but does not mandate their use. Controversy exists, however, regarding just how effective the provision of safety restraints and mandatory safety restraint use laws would be in reducing fatalities and injuries

---

College of Engineering, University of South Florida, 4202 East Fowler Avenue, ENB 118, Tampa, Fla. 33620-5350.

to occupants of large school buses. The debate is heated, and both sides make strong appeals in support of their views.

Proponents of safety restraints in large school buses concede that the current practice of compartmentalization is effective in reducing fatalities and injuries, but they argue that when combined with safety restraint use, fatality and injury rates could be reduced even further. Compartmentalization, as set forth in Federal Motor Vehicle Safety Standard (FMVSS) 222, requires that seats must be spaced no more than 60.96 cm. (24-in.) apart, as measured from the seating reference point (the point at which the human torso and thigh pivot), and seat-back height must be a minimum of 50.8 cm (20-in.) to the top of the seat back, as measured from the seating reference point. Also, limitations are placed on the amount of seat-back deflection both forward and backward. By adhering to these specifications, a compartment is created that is intended to restrain the school bus occupant, thereby limiting the severity of injuries in the event of an accident. They contend as well that requiring safety restraints in large school buses will reinforce the habit of buckling up in young children when they ride with their parents, and as a consequence safety restraint usage will carry over to other vehicles through adulthood. Also, they assert that safety restraint use will improve on-board school bus occupant behavior and decrease driver distractions, translating into the possible avoidance of accidents. Lastly, proponents argue that the cost of installing safety restraints (lap belts) is minimal, that is, no more than $1,000 to $1,500 per large school bus.

Opponents of safety restraints in large school buses maintain that because of their weight and large size, distinct yellow color, carefully selected routes for pick-up and drop-off, governed operating speed, lighting features, and unique FMVSSs, 220, 221, and 222, they are inherently safer than passenger cars, vans, and light trucks and, consequently, do not need safety restraints to improve occupant safety. FMVSS 220, School Bus Rollover Protection (49 CFR 571.220), specifies performance requirements for the structural integrity of the passenger compartment of school buses when subjected to forces that may be encountered in rollover crashes. FMVSS 220 applies to all school buses (Types A, B, C, and D). FMVSS 221, School Bus Body Joint Strength (49 CFR 571.221), requires interior and exterior body panel joints to prevent or reduce panel separation in a crash. FMVSS 221 applies only to large school buses, those with GVW ratings greater than 4,540 kg (10,000 lbs). FMVSS 222, School Bus Seating and Crash Protection (49 CFR 571.222), sets occupant protection standards for passengers and establishes passive barriers to prevent or reduce injuries from the impact of school bus occupants against structures within the vehicle during crashes and sudden driving maneuvers. Large school buses must meet all of the requirements of FMVSS 222; however, Type A school buses, those with GVWs less than 4,540 kg (10,000 lb), must meet all of the specified requirements except the 50.8-cm (20-in.) maximum distance between the seating reference point and seat back or passive barrier in front of it.

Opponents also contend that, in the case of serious accidents, safety restraints may actually increase the likelihood of injury and can imperil occupants of large school buses, especially young occupants, in accidents involving fire and rollovers. Also, they assert that if school bus drivers do not insist that occupants wear the safety restraints, the potential carry over effect will be lost. And could cause the children to become desensitized to safety restraint usage and could carry over the wrong message, that is, that it is not important to wear safety restraints in other modes of transportation. Lastly, opponents question the cost-effectiveness of safety restraints, arguing that the funds would be better spent on other, more effective safety investment options such as improved driver training, higher seat backs, crossing control arms, increased enforcement of laws against passing stopped school buses, and adult school bus monitors.

STATISTICAL SAFETY RECORD OF SCHOOL BUSES

An analysis of the crash performance of large school buses in 1987 led the National Transportation Safety Board (NTSB) to state that "poststandard large school buses are an extremely safe form of transportation when compared to other modes of transportation" (*10*) (*poststandard* refers to school buses manufactured for sale in the United States after the implementation of FMVSSs 220, 221, and 222 on April 1, 1977). NTSB's contention is supported by 1986 data pertaining to national occupant fatality and fatality rates by vehicle type compiled by the TRB committee that investigated large school bus safety (*11*). It was estimated by the TRB committee that passenger cars had a fatality rate of 1.9 and that large school buses had a fatality rate of 0.5 per 161 million km (100 million mi) traveled, statistically making large school buses four times safer than passenger cars on a vehicle mile basis, and accordingly, many more times safer on a passenger mile basis, because of the higher occupancy of large school buses.

Furthermore, Gutoskie reports that in Canada, for the period 1982 through 1985, "motor vehicle occupants were approximately 16 times more likely than school bus occupants to be injured in road accidents per passenger kilometer of travel" (*12*), and Farr concluded that "a student is 8 times more liable to be injured while travelling to or from school in a vehicle other than a school bus" (*13*). In an analysis of California accident data, Urcell deduced that "school buses without seat-belts are 16.2 times more safe than automobiles" with front and rear seat-belts (*14*).

REVIEW OF SCHOOL BUS ACCIDENT STUDIES

Recognizing the need for and the importance of studying school bus accidents, the Texas Transportation Institute (TTI) (*15*) and NTSB (*10*) each conducted comprehensive studies that investigated real-world school bus accidents.

TTI's case-by-case evaluation included 13 school bus accidents that involved 19 fatalities. That analysis suggested that 12 of the 19 fatalities would have been prevented had safety restraints (lap belts) been available to those who were fatally injured and that an additional 4 deaths might have been prevented had safety restraints (lap belts) been available or proper student disciplinary procedures exercised. In the remaining three cases Hatfield and Womack (*15*) discerned that the effect of safety restraints, had they been available, could not be determined "based on the limited data available and the fact that real world collisions are extremely difficult to evaluate on the basis of laboratory tests or subjective opinions."

TTI also assessed "accident characteristics and/or injury patterns which might be related to the seat belt issue in all injury-producing [Texas] school bus accidents" (*15*). That analysis produced insight into impact modes, which are relevant to assessing the effectiveness of safety restraints (rear-end, side, and frontal impacts and rollovers). The results rendered that approximately 46 percent of all fatal injury-causing school bus accidents in Texas (over the 12-year period of study) were accounted for in either side impact or rollover collisions. Moreover, although rollover accidents represented a small share (6 percent) of all injury-causing school bus accidents in Texas, they accounted for a much higher proportion of all fatal and incapacitating injuries to Texas school bus occupants in Texas, 15 and 18 percent, respectively. It is particularly important to emphasize this finding, because safety restraints generally are considered to improve occupant safety in accidents involving either a side impact or a rollover.

NTSB reviewed 43 accidents that involved 1,119 unrestrained occupants "to evaluate the real-world performance of school buses built to the 1977 Federal school bus standards" (*10*). The objective of the study was to focus primarily "on events *during* the crash: how well did the bus perform; how did occupants sustain their injuries, if any; and how serious were the injuries" (*10*). NTSB also examined the question of whether lap belts are needed for the occupants of large school buses manufactured for sale in the United States after the implementation of FMVSSs 220, 221, and 222 on April 1, 1977.

Based on the evidence accumulated from the investigation, NTSB concluded that FMVSS 222, which provides for compartmentalization, worked well in the NTSB-investigated crashes in protecting occupants of poststandard large school buses from injury in all accident types. They also recommended that federal safety standards not be amended to require that all newly purchased large school buses be equipped with safety restraints and that such actions (requiring safety restraints), in terms of reduced fatalities and injuries to the occupants of large school buses, have not been empirically proven.

SCHOOL BUS CRASH AND SLED TESTS

In 1984 Transport Canada (TC) performed full-scale crash testing of three different-sized school buses to evaluate the effect that safety restraints might have on improving school bus occupant protection and to assess whether current Canadian school bus standards provide a passable level of occupant safety (*13*). Data were collected on the relative severities of injuries to occupants both with and without safety restraints and with the use of three different seat spacings. The TC researchers concluded that compartmentalization affords occupants ample protection in frontal collisions and that the use of lap belts may result in more serious head and neck injuries to restrained occupants of large school buses in frontal collisions.

The University of California at Los Angeles (UCLA) in 1967 conducted three crash tests in which different impact modes: a frontal, a rear-end, and a side impact (90 degree), were fabricated "using research techniques and engineering methodology designed to provide realistic and objective findings relating to [large] school bus passenger safety" (*16*). On the basis of the test data gathered, the UCLA research team concluded that "the greatest single contribution to school bus passenger collision safety is the high-strength, high-back safety seat. Next in importance is the use of a three-point belt, a lap-belt or other form of effective restraint" (*16*).

In 1972 UCLA conducted a second series of crash tests, the Series II tests. The second series of tests involved two types of collisions, a head-on and a side impact (90 degree) absent the rear-end impact collision scenario. The school bus seat types, safety restraints, anthropomorphic testing devices (ATDs), and data-gathering procedures were similar to those of the Series I tests. In addition to the similarities to the Series I tests, however, a rearward-facing seat without a lap belt and a seat positioned sideways along the school bus wall were evaluated as well. Based on the accumulated evidence, the UCLA researchers concluded, "For buses provided with safety seats having a performance profile comparable to the UCLA design, seat-belts [lap-belts] will contribute a significant measure of safety, especially during severe upset [rollover] collision exposures" (*17*). It is important to note that the Series I and Series II UCLA tests were conducted many years (1967 and 1972) before the issuance of FMVSSs 220, 221, and 222 on April 1, 1977.

In 1985 Thomas Built Buses, Inc., conducted three crash impact tests: a frontal impact into a fixed barrier, a right-side impact by a moving barrier, and a left-side impact by a moving barrier. Based on the results of the three crash impact tests, the Thomas Built research group concluded that compartmentalization performs as it was designed in frontal and side impacts. They also found that in the case of the side impacts, very little difference exists between the restrained and unrestrained ATDs relating to the severity of head and chest injuries (*18*).

In 1978 NHTSA conducted sled tests to evaluate the restraint performances of various production school bus seats designed to satisfy the requirements of FMVSS 222 (*19*). They concluded that the use of lap belts did not reduce head injury criteria (HIC) values but, in fact, actually caused an increase in them. The data showed that the average restrained (lap belt) ATD HIC value was 278 and the average unrestrained ATD HIC value was 157.5, a measured difference of 120.5. They attributed these higher HIC values to the fact that the contact point for the ATD's head is moved upward as a result of using the lap belt. Also, the results indicated that compartmentalization worked as intended and that there were no additional benefits that could be derived by the use of lap belts.

ALTERNATIVE SEAT AND RESTRAINT SYSTEMS

To further investigate the issue of large school bus safety, TC conducted several tests that used five alternative seat types, each of which incorporated a restraint system (*20*). These five seat types included:

- contoured and padded seat back with lap belt,
- less aggressive (more easily collapsible) seat back with lap belt,
- rearward-facing seat with a lap belt,
- three-point restraint system (passenger car-type lap and shoulder belt restraint system), and
- multiple-point restraint system (harness-type restraint system that consisted of a lap belt and dual torso restraint).

In addition to the five alternative seat and restraint types, an unmodified standard school bus seat (39-in. bench seat) affixed with manual lap belts was tested for the express purpose of providing an experimental control mechanism used to make simple comparisons with the five alternative seating systems.

The TC tests evidenced that contoured and padded and less aggressive seats fitted with lap belts are not the answer for increased occupant safety. With respect to three-point restraint systems, the TC researchers concluded that they possess the necessary potential to increase occupant safety but that further reflection must be given to testing and design before they can become a viable large school bus safety investment option. However, the TC test results revealed that rearward-facing seats fitted with lap belts can significantly augment school bus occupant safety. This finding contests the results of the Series II tests performed at UCLA in 1972. The TC team again emphasized that continued research is required if rearward-facing seats fitted with lap belts are to become standard issue in Canadian school buses. In the United States FMVSS 222 would have to be amended before rearward-facing seats fitted with lap belts could become a tangible safety investment option for large school buses.

Like their TC counterparts, the 1972 UCLA researchers also investigated the effectiveness of numerous alternative restraint systems and a single alternative seat system in conjunction with testing the effectiveness of lap belts (*17*). Conclusions similar to those of TC regarding the three-point restraint system were reached by the UCLA researchers. UCLA test results also established that restraint bars, gate-bar lap restraints, armrests, airbags, and airseats do not have the capability of offering increased protection to the occupants of large school buses.

SUMMARY OF LITERATURE REVIEW

Collectively, the body of literature reviewed provided inconclusive and at times contradictory evidence relating to the effectiveness or potential effectiveness of safety restraints in large school buses. In addition, it raised the prospect that safety restraint use might result in harmful epidemiological consequences to school bus occupants in certain accident types, particularly if the accident type is frontal in nature (*13*). Moreover, the results from some studies reviewed have been criticized regarding their methodological soundness (*21*). Specifically, the authors criticized the 1984 TC test. Also, the relevance of results inferred (UCLA Series I and II tests) before the issuance of FMVSSs 220, 221, and 222 on April 1, 1977, should be seriously questioned.

SAFETY COST–BENEFIT ANALYSIS

Even if the literature review allowed one to draw the conclusion that safety restraints, primarily lap belts, are beneficial in terms of their ability to consistently mitigate the fatalities and injuries sustained by occupants of large school buses, their installation and the installation of other safety investment options such as crossing control arms and external loud speaker systems must involve weighing the costs against the overall potential benefits in absolute terms, that is, a quantifiable number of children's lives preserved and injuries prevented or lessened by the installation of each safety investment option.

Although the concept of placing a monetary value on the life of a child is offensive and unthinkable to most of the general public, public investment decisions ranging from roadway safety improvements, police and fire protection, sanitation, and large school bus occupant safety involve an implicit financial trade-off between spending and the benefits to be gained. Thus, it is crucial to determine the relative worth of each safety investment option by calculating a quantifiable number of lives saved and injuries prevented for a specified capital outlay and unit of time.

Method

The safety cost–benefit analysis examined nine different safety investment options to determine a quantifiable number of fatalities and injuries that might be reduced in Florida per safety investment option per year. The safety investment options evaluated included

- lap belts,
- lap-shoulder belts (three-point restraints),
- lap-dual shoulder belts (multiple-point or four-point restraints),
- higher seat backs ("New York" seats),
- adult school bus monitors,
- electrically operated crossing control arms,
- dual stop signal arms,
- external loudspeaker systems, and
- rearward-facing seats with a lap belt.

The installation cost estimates for the nine safety investment options investigated are only for newly purchased large school buses. The cost of retrofitting large Florida school buses with the nine safety investment options was not explored. The nine safety investment options were analyzed by using several parameters. The parameters are identified in Table 1.

Because so few deaths or serious injuries are sustained by occupants of large school buses in a typical school year in Florida or the

TABLE 1 Safety Cost–Benefit Analysis Parameters

Safety Investment Option	Annual Investment	Installation Cost/Bus	Annual Maintenance Cost/Bus	Discount Rate	Residual Value	Economic Life Span (years)
Lap-belt	$1,000,000	$1,500	$35	7%	$0	15
Lap/shoulder belts	$1,000,000	$3,800.77	$40	7%	$0	15
Lap/dual shoulder belts	$1,000,000	$4,643.94	$40	7%	$0	15
Higher seat-backs	$1,000,000	$250	$0	7%	$0	15
Adult school bus monitors	$1,000,000	$3,825	$0	n/a	n/a	n/a
Crossing control arms	$1,000,000	$350	$25	7%	$0	15
Stop signal arms	$1,000,000	$475	$15	7%	$0	15
Rearward-facing seats w/lap-belt	$1,000,000	$3,113.26	$35	7%	$0	15
External loud speaker systems	$1,000,000	$350	$15	7%	$0	15

TABLE 2 TRB Safety Investment Option Effectiveness Rates (*11*)

Safety Investment Option	Percent Effective at Reducing Fatalities and Injuries
Lap-belts	0-20%
Higher seat-backs	0-20%
Adult school bus monitors	25-75%
Crossing control arms	5-25%
Stop signal arms	0-30%
External loud speaker systems	0-20%

United States (the TRB committee that investigated the issue estimated that 10 school bus occupants are fatally injured each year in the United States), a dearth of comprehensive accident data is available to determine empirically the effectiveness of restraint systems and other safety investment options for large school buses. On the other hand, however, there is a prodigious amount of statistical and empirical literature that demonstrates quite conclusively that safety restraints save lives and reduce injury severity in passenger cars, vans, and light trucks (*1–9*). Again, because of the infrequency of death and serious injury to the occupants of large school buses, estimates of how much supplemental protection that might be provided by safety restraints or other safety investment options are, by circumstance, purely conjectural. Therefore, because of this problem the identical effectiveness rate assumptions used by the TRB committee that investigated large school bus safety were used in the cost–benefit analysis calculations (Table 2).

Results of Safety Cost–Benefit Analysis

Lap belts would potentially prevent up to 0.032 fatalities per $1 million annual investment (Table 3). This translates into one child's life being saved approximately every 31 years. The annual fatality and injury reduction results for the other eight safety investment options are summarized in Table 3.

The marginal improvement in safety to the occupants of large school buses in Florida associated with the use of the three safety restraint types resulted in their falling outside of the range of cost-effectiveness that would provide a compelling basis for an investment recommendation. However, the results of the cost–benefit analysis indicate that higher seat backs [60.96 cm (24-in.), as measured from the seating reference point] are the safety investment option that could offer the most benefits in terms of fatalities prevented and injuries reduced to occupants of large school buses in Florida. They have the potential to prevent up to 0.3 fatalities, 7 incapacitating injuries, 42 nonincapacitating injuries, and 104 possible injuries per year. Similar conclusions were reached by the TRB committee that investigated large school bus safety (*11*).

DESCRIPTIVE ANALYSIS OF FLORIDA SCHOOL BUS ACCIDENT DATA

The descriptive analysis of Florida school bus accident data is based on the supposition that impact mode (i.e., direction of impact) is directly related to the severity of the injury sustained. The available research on frontal and rear impact collisions does not produce a result favorable to the installation and use of safety restraints in large school buses. In actuality, the literature points to the fact that an occupant of a large school bus restricted by a safety restraint (lap belt) in a frontal collision would have a tendency to bend forward (jackknife) and strike the top of the seat back ahead with the head, face, and chest, thereby increasing the forces and thus the injuries to these parts of the body. In contrast, an occupant of a large school

TABLE 3 Annual Fatality and Injury Severity Reduction per Safety Investment Option

Safety Investment Option	Injury Severity Reduction			
	Fatal	Incapacitating[1]	Non-incapacitating[2]	Possible[3]
Lap-belts	0.032	0.73	4.5	11
Lap/shoulder belts	0.028	0.634	3.9	9.6
Lap/dual shoulder belts	0.03	0.66	4.05	10
Higher seat-backs	0.3	7	42	104
Adult school bus monitors	0.006 - 0.018	0.09 - 0.3	0.58 - 1.8	1.44 - 4.3
Crossing control arms	0.0148 - 0.074	0.0221 - 0.111	0.0289 - 0.145	0.067 - 0.333
Dual stop signal arms	0.042	0.063	0.082	0.2
External loud speaker systems	0.035	0.05	0.07	0.16
Rearward-facing seats w/lap-belts	0.0169 - 0.047	0.38 - 1.06	2.37 - 6.5	5.8 - 16.03

[1] Severe lacerations, broken or distorted limbs, skull, chest or abdominal injuries, unconscious at or when taken from the accident, or unable to leave the scene without assistance.

[2] Lump on head, abrasions, minor lacerations, etc.

[3] Momentary unconsciousness, complaint of pain, nausea, hysteria, etc.

TABLE 4 Florida Large School Bus Accident Frequency by Impact Mode, 1986–1991

	Impact Mode					
	Frontal	Rear-end	Side	Rollover	Other	Total
Number of Accidents	63 (1.3%)	1,482 (31.3%)	1,334 (28.2%)	15 (.32%)	1,838 (38.8%)	4,732 (100%)

Note: Percent does not equal 100 due to rounding error.

bus not restricted by a safety restraint (lap belt) during a frontal impact collision would have a tendency to slide forward in the seat and strike the seat back ahead with the upper torso and knees, causing the force of impact to be more evenly spread across the upper torso (the intent behind compartmentalization). In side impact collisions, however, the available research indicates that the use of safety restraints (lap belts) would be only slightly beneficial, contingent on the occupants not being seated in the direct impact zone during an accident. In rollovers the available research tends to be based more on conjecture than on fact. Deductions favoring lap-belt use in rollovers are grounded on the benefits of diminished tossing about and the elimination of partial ejection and, in extremely rare instances, the full ejection of an occupant. Therefore, to determine the potential effectiveness of safety restraints in large school buses in Florida, two objectives were defined: determine the frequency and distribution of accidents by four primary impact modes (frontal, rear-end, side, and rollovers) and determine occupant injury severity by the same four impact modes.

In Table 4 the frequencies of large school bus accidents in Florida by frontal impact, rear-end impact, side impact, and rollover are identified. As reported in Table 4, 63 (1.3 percent of all large school bus accidents in Florida during the 6-year period of study were frontal in nature. Table 4 also illustrates that a higher proportion of accidents involving large school buses in Florida were either rear-end or side impact collisions: 1,482 (31.3 percent and 1,334 (28.2 percent), respectively. Not surprisingly, school bus accidents that resulted in rollovers constituted the smallest proportion of all accident possibilities. Of the 4,732 reported accidents between 1986 and 1991, only 15 (0.32 percent) involved an overturned school bus. One plausible explanation for this result, in part, may be the flat nature of Florida's topography.

Table 5 provides a frequency distribution of the level of injury severity sustained by occupants in the 4,732 accidents that involved large school buses in Florida. Over the 6-year period of study, 9 (0.02 percent) fatalities, 202 (0.45 percent) incapacitating injuries, 1,251 (2.8 percent) nonincapacitating injuries, 3,091 (7 percent) possible injuries, and 39,878 (89.7 percent) no injuries (none) were reported. A total of 7 (0.015%) injuries sustained by occupants of large Florida school buses were of unknown injury severity.

The distributions of fatalities, incapacitating injuries, nonincapacitating injuries, possible injuries, and no injury (none) are identified according to impact mode for the 6-year period of study (Table 6). These data represent the actual number of school bus occupants in Florida who sustained a particular level of injury as determined by one of the four primary impact modes. The figures for side impact collisions were derived by using the aggregate of angle and sideswipe impacts as provided by the state of Florida school bus accident data base.

The nine reported deaths to occupants of large school buses in Florida were the result of three separate accidents. Because of this small number of accident cases, it was possible to obtain and review the actual accident reports for each of the accidents. Excerpts written by the reporting officers as well as accident circumstances were taken from the accident report narratives to better judge whether the deaths could have been prevented had safety restraints been available to those occupants. It was concluded that the three accidents were so catastrophic and freakish in nature that it is doubtful that the presence of safety restraints of any configuration or design would have altered the tragic outcomes.

By comparing the frequencies in Table 4 and Table 6, it becomes apparent that frontal impact collisions represent the second smallest number (63 or 1.3 percent) of all injury-producing school bus accidents but account for a disproportionate number (11.15 to 1 ratio) of incapacitating injuries to school bus occupants. Rollover accidents represent the smallest number (15 or 0.32 percent) of all injury-producing school bus accidents but account for the second highest proportion (9.375 to 1) of incapacitating injuries to school bus occupants. In contrast, rear-end and side impact collisions

TABLE 5 Large Florida School Bus Occupant Injury Severity by Year

	Year						
Injury Severity	1986	1987	1988	1989	1990	1991	Total
Fatal	0	5	1	0	0	3	9 (0.02%)
Incapacitating	17	29	33	19	57	47	202 (0.45%)
Non-incapacitating	163	295	166	271	235	121	1,251 (2.8%)
Possible	499	543	363	454	776	456	3,091 (7%)
None	5,976	6,724	6,868	6,428	7,096	6,786	39,878
Unknown	0	2	0	2	0	3	7 (0.015%)
Total	6,655	7,598	7,431	7,174	8,164	7,416	44,438 (100%)

TABLE 6 Large Florida School Bus Occupant Injury Severity by Impact Mode

Impact Mode	Injury Severity						
	Fatal	Incapacitating	Non-incapacitating	Possible	None	Unknown	Total
Frontal	0	24	55	108	398	0	585
Rear-end	0	85	370	1,207	15,774	6	17,442
Side	5	52	504	721	10,505	1	11,788
Rollover	0	5	10	95	71	0	181
Total	5	166	939	2,131	26,748	7	29,996

Note: Four of the deaths to occupants of large Florida school buses were not the result of one of the four primary impact modes utilized in the descriptive analysis of large Florida school bus accident data.

account for 1,482 (31.3 percent) and 1,334 (28.2 percent) of all injury-producing school bus accidents, respectively, but result in the smallest proportion (1 to 1 and 1.81 to 1, respectively) of incapacitating injuries sustained by occupants of large Florida school buses.

DESCRIPTIVE ANALYSIS OF FLORIDA SCHOOL BUS ACCIDENT DATA RESULTS

The descriptive analysis of Florida school bus accident data does not provide compelling evidence that safety restraints are needed in large Florida school buses to improve occupant safety. Rather, the considerable number of occupants of large Florida school buses who were either uninjured or received minor or moderate injuries (44,220, or 99 percent; 7 injuries of unknown severity were omitted) simply reiterates the fact that large school buses in Florida are an extremely safe mode of transportation. Furthermore, based on an extensive review of the accident reports, the availability of safety restraints to those occupants of large school buses in Florida who were fatally injured in 1986, 1989, and 1990 was rendered moot, since the nine fatalities most likely would have occurred even if the occupants would have been held securely in place by a safety restraint. The fact that only 9 (0.02%) fatalities (eight of the fatalities were the result of two accidents) and 202 (0.45%) incapacitating injuries were sustained by the 44,438 Florida school bus occupants involved in the 4,732 large school bus accidents reported in the statewide accident data base for the years 1986 through 1991 substantiates the effectiveness of the safety investment options already being used in and on large school buses in Florida and the reality that serious accidents involving large school buses in the state of Florida are infrequent.

CONCLUSION

This paper includes a summary compilation and review of pertinent literature pertaining to the issue of large school bus safety and two supplemental analyses, a safety investment cost–benefit analysis and a descriptive analysis of Florida school bus accident data.

Based on the evidence from the literature review, it was recommended that safety restraints of any configuration or design not be installed in large school buses in Florida until adequate testing and further design modifications are completed to justify an investment recommendation. To date it is the opinion of the author that neither adequate testing and research nor empirical evidence exists to justify an investment recommendation.

In addition, the conclusion to recommend that the state of Florida not mandate the installation of safety restraints in their large school buses was also based, in part, on the results from the two supplemental analyses. When solely considering the economics (benefits versus costs) of school bus safety, several questions must be asked: What price do we place on large school bus occupant safety? How do we measure safety? Who is going to pay for the safety? The results of the cost–benefit analysis demonstrated that the marginal improvement in safety to occupants of large school buses in Florida associated with the use of the three safety restraint designs tested resulted in their falling outside the range of cost-effectiveness, thus making them secondary safety items at best.

The statistical safety record of large school buses in Florida is impressive. The fact that only 9 (0.02 percent) fatalities and 202 (0.45 percent) incapacitating injuries were sustained by the 44,438 large school bus occupants during the 6-year period of study points to the inherent safety of large school buses in Florida.

This exemplary safety record of large school buses in Florida, however, should not permit the state of Florida to rest on its laurels. Accidents involving large school buses in Florida will continue to occur, and as long as one child is fatally or severely injured, the pursuit of increased safety must be a constant process.

REFERENCES

1. Campbell, B. J. *The Effectiveness of Rear-Seat Lap-Belts in Crash Injury Reduction.* Highway Safety Research Center, University of North Carolina, Chapel Hill, 1986.
2. Campbell, B. J. Seat Belts Effectiveness. *Proc., International Symposium on Seat Belts,* Tokyo, Japan, 1979.
3. Huelke, D. F. Effectiveness of Occupant Restraints in Reducing Serious Injuries and Fatalities. Presented at the International Symposium on Occupant Restraint, Toronto, Ontario, Canada, 1981.
4. Maghsoodloo, S., et al. A Quantification of the Impact of Restraining Systems on Passenger Safety. *Journal of Safety Research, Vol. 20, No. 3, 1989, pp. 115–128.*
5. McGee, D. L., and P. Rhodes. Estimating Trends in the Effectiveness of Seat Belts in Saving Lives, 1975–1985. *Statistics in Medicine, Vol. 8, No. 3, 1989, pp. 379–385.*
6. Evans, L. The Effectiveness of Safety Belts in Preventing Fatalities. *Accident Analysis & Prevention Journal,* Vol. 18, No. 3, 1986, pp. 229–241.
7. Kerwin, E. M., et al. Seat Belt Effectiveness in Injury-Producing Accidents: The Colorado Matched Pairs Study. University of Colorado School of Medicine. Presented at the American Public Health Association Annual Meeting, Washington, D.C., 1985.

8. Frazier, R. G. Effectiveness of Seat Belts in Preventing Motor Vehicle Injuries. *New England Journal of Medicine*, Vol. 264, 1961, p. 1254.
9. Partyka, S. C. *Belt Effectiveness in Fatal Accidents.* DOT Report HS 807 285. U.S. Department of Transportation, 1988.
10. National Transportation Safety Board. *Safety Study—Crashworthiness of Large Poststandard School Buses. Bureau of Safety Programs, Washington, D.C., 1987.*
11. *Special Report 222: Improving School Bus Safety. TRB, National Research Council, Washington, D.C., 1989.*
12. Gutoskie, P. A. *An Analysis of Canadian School Bus Accident Records 1982-83–1984-85.* Transport Canada, Ottawa, Ontario, Canada, 1986.
13. Farr, G. N. *School Bus Safety Study, Vol. I.* Traffic Safety Standards and Research, Transport Canada, Ottawa, Ontario, Canada 1984.
14. Urcell, C. R. *A Study Relating to Seat Belts for Use in Buses.* Southwest Research Institute, San Antonio, Tex. 1977.
15. Hatfield, N. J., and K. N. Womack. *Safety Belts on School Buses: The Texas Experience.* Texas Transportation Institute, Texas A&M University, College Station, 1986.
16. Severy, D. M., H. M. Brink, and J. D. Baird. *School Bus Passenger Protection.* Institute of Transportation and Traffic Engineering, University of California at Los Angeles, 1967.
17. Wojcik, C. K., and L. R. Sandes. *School Bus Seat Restraint and Seat Anchorage Systems.* Institute of Transportation and Traffic Engineers, School of Engineering and Applied Science, University of California at Los Angeles, 1972.
18. *Seat Belts in School Buses.* Thomas Built Buses, Inc., High-Point, N. C., 1992.
19. Bayer, A. R. *School Bus Passenger Seat and Lap Belt Sled Tests.* NHTSA, U.S. Department of Transportation, 1978.
20. Farr, G. N. *School Bus Seat Development Study.* Traffic Safety and Standards and Research, Transport Canada, Ottawa, Ontario, Canada, 1987.
21. Weber, K., and J. M. Melvin. Memorandum to Colleagues Concerned About Child Passenger Safety. Department of Mechanical Engineering and Applied Mechanics, University of Michigan, Ann Arbor, 1986.

Publication of this paper sponsored by Committee on Transportation Safety Management.

Simulation Tool for Evaluating Effectiveness of Freeway Incident Response Operations

TETI G. NATHANAIL AND KOSTAS G. ZOGRAFOS

A methodology for evaluating the effectiveness of freeway incident management (FIM) systems is presented. Evaluation of the effectiveness of a FIM system requires the establishment and estimation of the measures of effectiveness. For the estimation of the measures of effectiveness of a freeway incident response system, a simulation model that focuses on the operations of freeway emergency response units (FERUs) has been developed. Various aspects of the system are covered. First, an analysis of the arrival rates of mobile servers to the incidents to be serviced is done. These incidents occur randomly, follow discrete temporal and spatial distributions, and are of different natures and severities. Then, on the basis of the number of the mobile servers, the area in which these servers provide service to emergency calls is divided into smaller districts and the mobile servers are assigned to each one of the districts. The proposed simulation model generates freeway incidents according to their spatial, temporal, and severity characteristics. Alternative dispatching policies have been integrated in the model for the assignment of FERUs to incidents. The proposed model has the ability to estimate the total time that has elapsed between the occurrence of the incident and the completion of the service and to provide alternative emergency response policies. The value of alternative measures of effectiveness is calculated as a function of the total service time, and the performance of these policies is evaluated.

In many real-life events an emergency response is required. Examples of such events are accidents involving vehicles or pedestrians on the road network, medical emergencies, fire, failure of electric and other utility networks, and others. These events occur at a time and place that are not known in advance. These events are referred to as incidents. Depending on the nature of the incident, specialized authorities are called to provide assistance and service the incident.

An emergency response system is considered efficient when it provides fast and appropriate service to the incidents that need to be serviced. The main attributes of an emergency response system are the appropriate equipment, specialized personnel, and the time required for service to be provided to the incident. Incident restoration time is considered the time that elapses between the incident occurrence and the completion of the provided service. The restoration time can be divided into four components, as shown in Figure 1 (*1*):

- Time T_1 is the time from the incident occurrence until the notification of the response authority about the incident, or the detection time. Identification of the incident's characteristics, such as the type of incident and its location and severity, is included in the detection time.
- Time T_2 is the time that has elapsed between the notification of the responding authority about the incident and the allocation or assignment of the most appropriate server to the restoration of the incident, or the dispatch time.
- Time T_3 is the time that has elapsed between the assignment of the response unit and the arrival of the unit at the site of the incident, or the travel time to the incident scene.
- Time T_4 is the time required for all the necessary activities undertaken by the response unit to completely restore the incident, or the service time. Depending on the nature of the incident, service time can be divided into on-scene time (if special actions are required at the scene of the incident) and removal time (if transfer of the involved person or equipment to some other location is necessary).

The study described here presents a tool for evaluating alternative policies of dispatching and allocating emergency response units. More specifically, the proposed simulation tool has the capacity to examine alternative dispatching and assignment strategies of freeway emergency response units (FERUs) and to estimate the results of the incident restoration time. For this purpose the study focuses on the response operations as they are conducted in an area where incident occurrences generate a demand for service. A microscopic simulation model has been developed. The model assesses the roadway network characteristics and the response operations according to the selected policy to be applied. Thus, it generates the calls for service in time and space and identifies their priority for service. Furthermore, it selects the most appropriate server, according to a dispatching policy, and assigns the incident to the server, providing routing instruction to the FERUs. Finally, it estimates the time to the completion of the service and directs the server to the next incident location if there is another incident waiting to be serviced. The restoration time for each incident and the total restoration time are the measures of effectiveness applied in a case study to compare the different response policies tested.

EMERGENCY RESPONSE OPERATIONS

Emergency response systems can be divided into two major categories:

- Response systems in which the service is provided at a specific place, such as a hospital, and
- Response systems in which the servers travel to the point of the demand and either provide service at the site of the call or transfer the involved person or equipment to a servicing facility.

T. G. Nathanail, TRD International S.A., 18, Navarinou Square, Thessaloniki, 54622, Greece. K. G. Zografos, Department of Operations Research and Management, University of Patras, Rion-Patra, 26500, Greece.

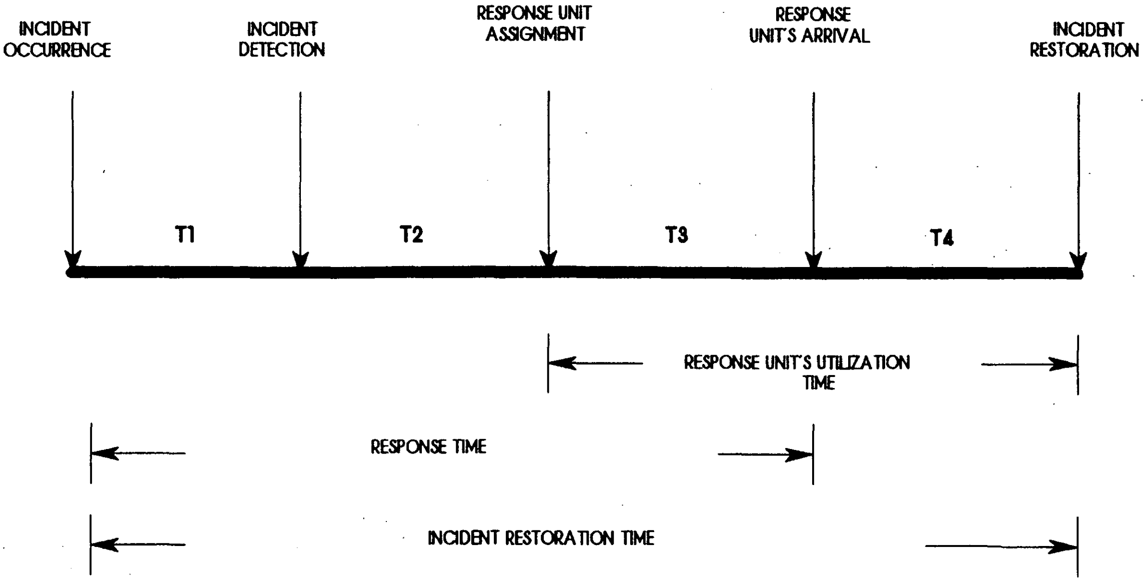


FIGURE 1 **Time components of incident duration.**

In the present study emphasis is given to the second category, in which equipped vehicles arrive at the scene of an incident to provide the necessary assistance. The sequence of the servers' operations for the incident clearance is the main attribute of a dispatching policy. The first in–first out (FIFO) and the nearest neighborhood (NN) policies are the ones tested in the present case study (*1,2*).

FIFO dispatching policy dispatches the server to the point where demand for service occurs first. Two disadvantages characterize the FIFO policy. The first is that under high workloads there is an excess delay to the servicing of other calls because of the travel time spent by the server. Also, sites with demand for service may be overpassed by the server because there is another call for service that arrived first.

The NN policy seems to result in less waiting time for service under high workloads because it dispatches the server to the closest location with a need for assistance, regardless of the time of occurrence of the calls.

Variations of these dispatching policies require that an area be divided into districts and that each district be assigned to one and only one server or require the accumulation of demand before a tour is being scheduled for a server (traveling salesman policy) (*2*).

Finally, a different priority for servicing the calls may be given on the basis of the type of call and the urgency and the severity of the incidents (*1,3*).

EMERGENCY RESPONSE METHODOLOGY: A GENERAL APPROACH

Simulation of incident restoration operations is a complicated process that includes the generation of incidents and their characteristics, the representation of response unit operations, and the estimation of the total incident duration and its components. A flow diagram of the model simulating the operations of the incident response units is shown in Figure 2.

The first module of the simulation model is the generation of incidents by their characteristics, such as time, location, and type of incident. The following assumptions have been considered for the incident generation module:

- Incidents follow a Poisson distribution in terms of time of occurrence, which is described by an interarrival time.
- The distribution of incidents in terms of location of occurrence is assumed to follow a uniform distribution. The locations of the area under study that have a higher probability of incident occurrences can also be considered.
- The incidents are generated by type and attributes. An indicator of their priority order is also given.

In the second module the incidents are assigned to servers depending on the jurisdictions of the possible servers. The jurisdictions of the servers are defined in advance by the control center or may vary depending on the type of emergency. Districting models for areas with incident occurrences to minimize the overall servicing time have been proposed by Toregas et al. (*4*), Marlin (*5*), Nathanail (*6*), and Zografos et al. (*1*). If the server is available the incident response process commences; otherwise, the incident is placed in the server's queue and waits until the server is available.

In the third module, the appropriate server assigned to incidents is dispatched to the incident site. The travel time of the server to the location of the incident is then computed as a function of the traffic volume at the time of the assignment and the geometric characteristics of the sections of the roadway network that constitute the fastest path from the location of the server to the incident site. The attributes of the roadway network (geometric characteristics and traffic volumes at certain times of the day to estimate the actual travel speed of the mobile server) exist in a data base of the network used for the incident response process.

After the arrival of the server at the location of the incident, the on-site and off-site response operations are described by the fourth module of the simulation model. Depending on the type of incident, the on-site response time is estimated, assuming an average response time for each type of incident. If the vehicles involved in the incident must be transferred to a servicing facility (e.g., hospi-

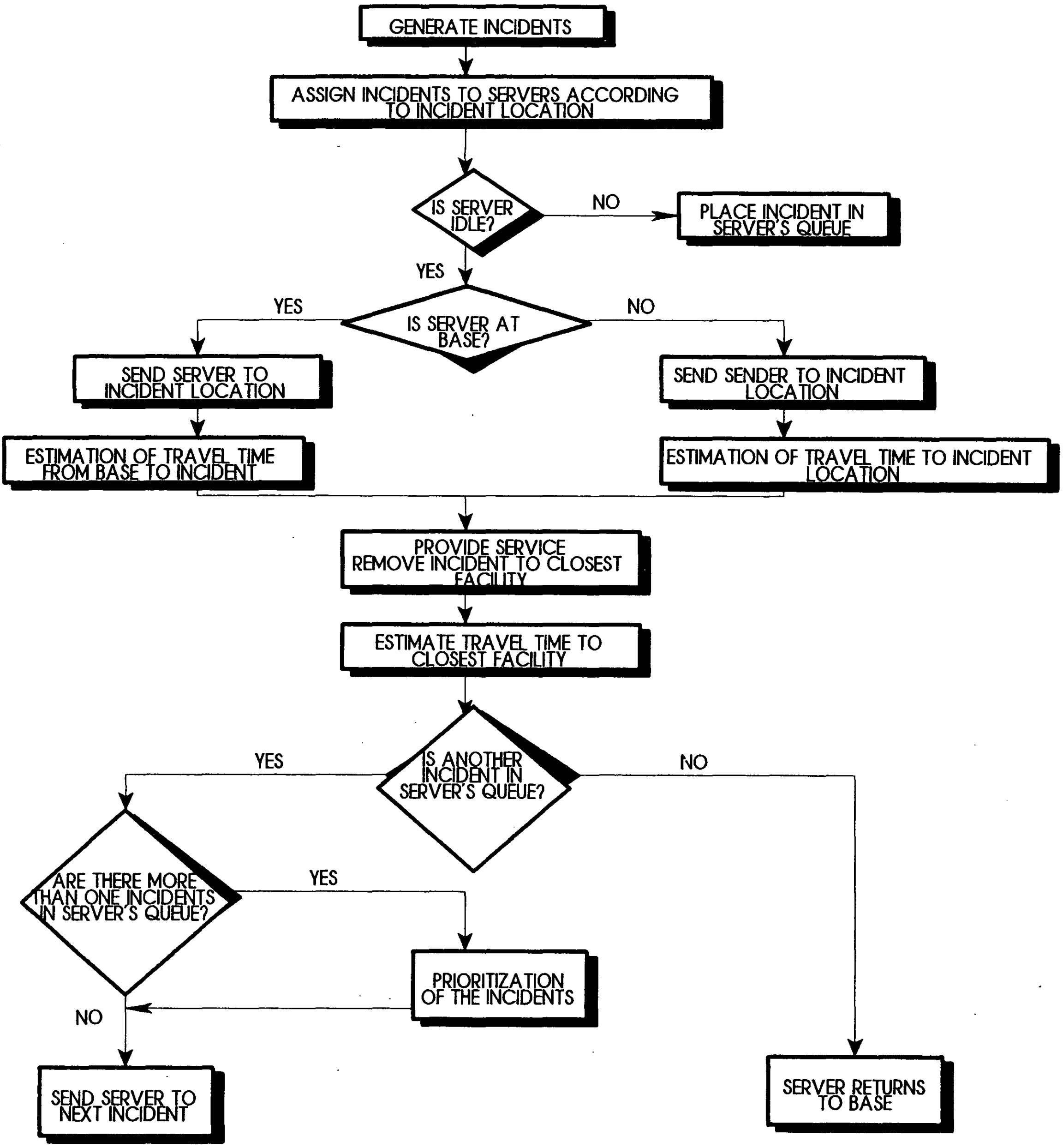


FIGURE 2 Simulation of incident response unit operations.

tal, garage, and accident investigation site), the facility is chosen among the available ones in the area and according to the dispatching policy on the basis of the criterion of the shortest path in terms of travel time from the incident site. The on-site response time and the travel time to the closest facility are then computed.

The server is considered available to service another incident after the completion of the required operations for the restoration of the previous incident. If an incident is in the server's queue, then the server is directed to the incident site and the server follows the same procedures described earlier. If more than one incident is in the server's queue, then these incidents are given a priority index according to their type and the dispatching policy, and the server is directed to the incident with the higher priority. If no incidents are waiting for service in the server's queue, then the server is directed to its initial position.

ROADWAY INCIDENTS AND THEIR CHARACTERISTICS

The main attributes of incidents are the time of occurrence, the location, and the severity. Incidents appear randomly in terms of their characteristics, which implies that incident response is based on online information.

For the purposes of applying the methodology described earlier, a case study in which an incident occurs on a roadway network is considered.

TABLE 1 Probability of Incident Occurrences on Four-Lane Highways with Adequate Shoulders

LOCATION	TYPE OF INCIDENT	NUMBER OF BLOCKED LANES	PROBABILITY OF OCCURRENCE
Lane	Accident	one	.72%
		two	.11%
		three	.02%
	Disablement	one	3.12%
		two	.03%
Shoulder	Accident	-	4.03%
	Disablement	-	91.97%

Incidents occurring on the roadway network can be divided into two categories in terms of severity, accidents, and other incidents. Accidents are incidents that involve personal injury or death. Incidents are minor events such as vehicle disablements (breakdowns) and spills. Furthermore, incidents are divided into categories that are related to the size of traffic blockage that they cause, which indicates an index of severity as well. This is expressed as the share of the roadway that they block, that is, the number of lanes. Incidents may occur on the shoulder, or they may occupy one, two, or more lanes of the highway. Finally, the time required for the clearance of the incident scene is another factor that differentiates incidents by severity. Incident clearance time is a direct derivative of the type of service required by the incident so that the situation is restored. Two types of service can be outlined. The first type of service is provided on site. The server is released after the completion of on-site restoration of the incident and is available to respond to another incident. The second type of service requires transfer of the vehicle involved in the incident to a service facility (i.e., accident investigation site, garage, etc.). Although the type of service required characterizes the incidents, it is mainly an attribute of the dispatching policy and depends on the structure of the response system and the equipment of the response unit.

In the present case study the incidents occurring on the roadway network were categorized by the number of lanes that they blocked. For a highway network consisting of four-lane highways with adequate shoulders, Lindley (*7*) has done an incident analysis in which the probability of incident occurrences by location and severity is estimated (Table 1).

In the same study (*7*) mean detection and clearance times were estimated by type of incident for incidents occurring on a highway network (Table 2). Because no data on the clearance times required by the shoulder accidents and disablements were available, for the purposes of the case study these times can be assumed to be 10 min for shoulder accidents and 5 min for shoulder disablements. Note that the detection and clearance times were considered to be given, whereas the dispatch and travel times were estimated by the simulation model, assuming a certain dispatching policy.

The incident arrival rate was estimated to follow a Poisson distribution, and a generation rate of 20 incidents per 1 million vehicle miles was assumed. This number is only an indication and may be estimated in each case on the basis of historical incident and traffic data collection.

CASE STUDY AND RESULTS

The methodology for evaluating alternative policies for dispatching emergency response units to incidents occurring on a roadway was tested on a hypothetical roadway network for the FIFO and NN dispatching policies and for the allocation of one, two, three, and four incident response units. The network consisted of a highway connected to an arterial by entrance and exit ramps. The time frame of the simulation was a 3-hr period during the evening peak. The generation of the incidents in time and space and their characteristics followed the distributions given earlier. Setting the starting time of simulation at 0:00 hr and the beginning of the corridor on which incidents occur at 0.00 m, the incident generation module resulted in the incident characteristics described in Table 3.

Because of the complexity of the incident restoration problem on a highway network, the applied policy should also consider the prevailing situation on the network, such as traffic volumes and the time of incident occurrence (peak versus off-peak hour) to provide adequate service to the persons involved in the incident and the rest of the users of the network by minimizing travel delays. However, consideration of the minimization of motorist delay because of lane blockage caused by incidents is not described here and can be found in other reports by the authors (*1,6*).

The average incident duration and the time components for the two applied dispatching policies and for different numbers of response units are given in Figures 3 and 4. In all cases it was

TABLE 2 Detection and Clearance Times by Type of Incident on Highways

TYPE OF INCIDENT	DETECTION TIME (min)	CLEARANCE TIME (min)
one-lane accident	10	10
two-lane accident	10	15
three-lane accident	10	20
one-lane disablement	10	5
two-lane disablement	10	10
shoulder accident	10	N/A
shoulder disablement	10	N/A

TABLE 3 Result of Simulation Module for Incident Generation Simulation

Incident identification number	Time of incident appearance (hour:min:sec)	Distance of incident from beginning of corridor	Type of incident
1	0:00:00	31485.84	shoulder disablement
2	0:26:41	4852.416	shoulder disablement
3	0:39:17	34686.24	shoulder disablement
4	0:56:00	34015.68	shoulder disablement
5	0:58:23	32125.92	shoulder disablement
6	1:05:30	34046.16	shoulder disablement
7	1:10:13	6769.61	shoulder accident
8	1:12:40	42885.36	two-lane accident
9	1:35:26	32796.48	one-lane disablement
10	2:00:24	8436.86	one-lane disablement
11	2:01:12	7336.54	shoulder disablement
12	2:02:00	33863.28	shoulder disablement
13	2:02:18	32583.12	shoulder disablement
14	2:14:24	4346.45	shoulder disablement
15	2:20:24	20141.18	shoulder disablement
16	2:26:54	3154.68	shoulder disablement
17	2:36:18	22792.94	shoulder disablement
18	2:43:48	3602.74	shoulder disablement
19	2:46:48	2847.14	shoulder disablement
20	2:57:24	32461.20	shoulder accident
21	2:59:42	35783.52	shoulder disablement

assumed that the detection and clearance times are not affected by the dispatching policy, and their values were obtained from the literature, (Table 2). The general trend shown in Figures 3 and 4 indicates that the total incident duration decreases as the number of response units increases. Application of the FIFO policy resulted in a 73 percent reduction in incident duration when the number of response units was increased from one to two, a 22.1 percent reduction when the number of units was increased from two to three, and an 18 percent reduction when the number of units was increased from three to four. Similarly, when the NN policy was applied, there was a 66 percent reduction in incident duration when the number of units was increased from one to two, a 17.2 percent reduction when the number of units was increased from two to three, and a 17.8 percent reduction when the number of units was increased from three to four. Furthermore, use of FIFO policy resulted in a 32.6 percent longer incident duration than use of the NN policy, resulting in a high utilization rate of the response unit by FIFO policy, as indicated in the study of Bertsimas (*2*).

Since application of the different dispatch policies mainly affects dispatch and travel times, an analysis of the share of these two time components was conducted. Results of such an analysis are given in Figures 5 and 6, which indicate the percentages of incidents for which the travel and dispatch times were within a certain range of the total incident duration. The general trends of the analysis show that with both dispatching policies, dispatch and travel times constituted an important portion of the total incident duration. In more than 50 percent of the observations, the dispatch and travel times constituted 60 percent of the total incident duration when one unit was assigned to respond to the incident. For more than one response unit the dispatch and travel times constituted an average of 20 percent of the total incident duration in most observations.

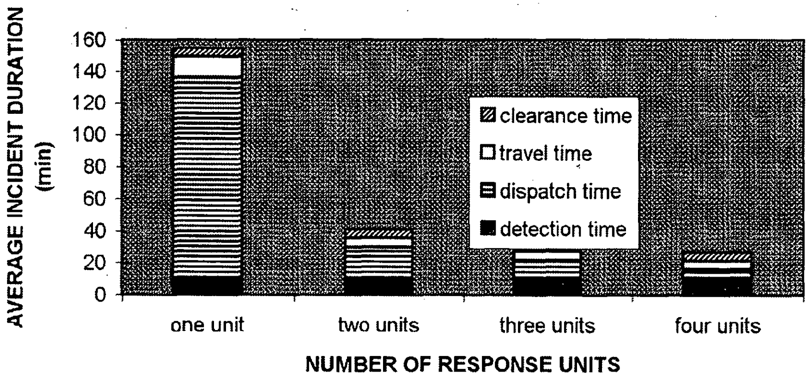


FIGURE 3 Average incident duration for FIFO dispatching policy.

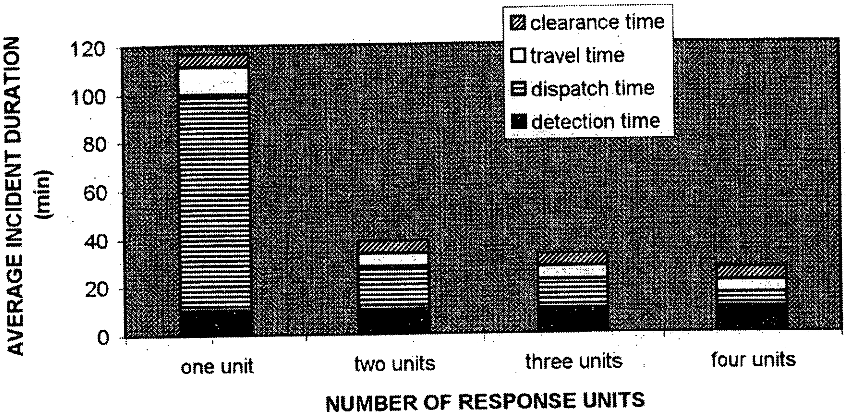


FIGURE 4 Average incident duration for NN dispatching policy.

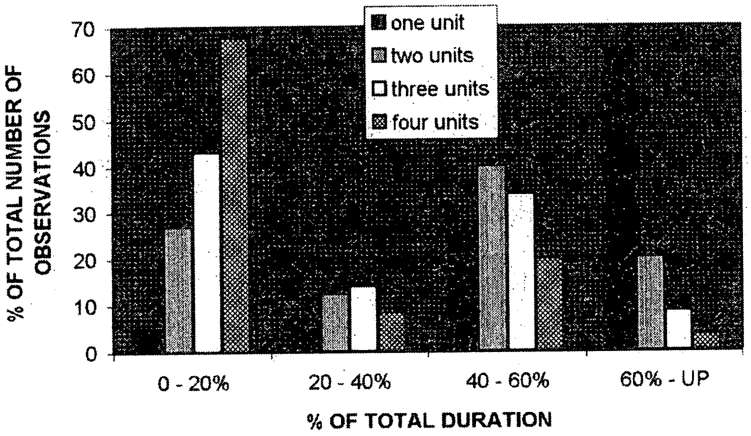


FIGURE 5 Average dispatch and travel time shares for FIFO policy.

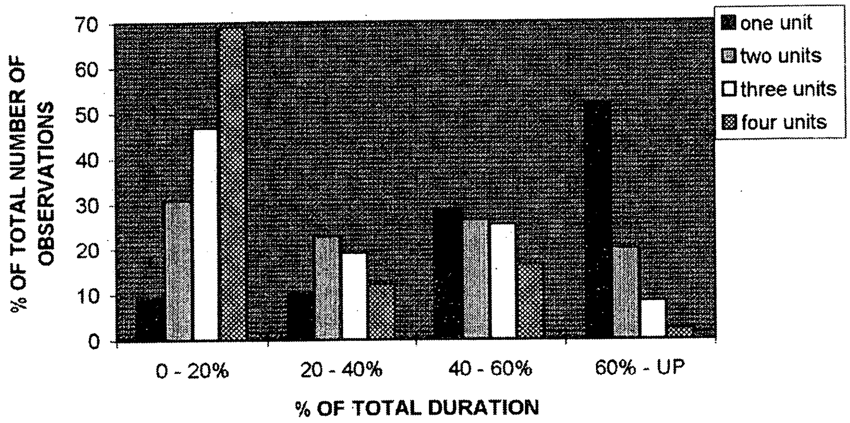


FIGURE 6 Average dispatch and travel time shares for NN policy.

CONCLUSIONS

A methodology for evaluating alternative policies of dispatching FERUs (and emergency vehicles in general) has been described. A simulation model that generates the operations of the FERU has been developed on the basis of a selected dispatching policy. The simulation model estimates the total incident duration as a measure of effectiveness of the dispatching policy and provides decision makers with a useful tool for selecting the optimum policy. Work in progress involves the development of a cost–benefit model that will be able to quantify the total cost of the freeway emergency response system for alternative levels of emergency response resources and to estimate the corresponding benefits in user costs, that is, fuel consumption, travel, and social impacts.

REFERENCES

1. Zografos, K. G., T. Nathanail, and P. Michalopoulos. Analytical Framework for Minimizing Freeway-Incident Response Time. *Journal of Transportation Engineering,* ASCE, 1993.
2. Bertsimas, V. R. Stochastic and Dynamic Vehicle Routing in the Euklidean Plane: The Multiple-Server Capacitated Vehicle Case. Working paper. Operation Research Center. Massachussetts Institute of Technology, Cambridge, 1990.
3. Jarvis, S. A Simple Procedure for Ambulance Allocation in Areas of Low Demand. In *Urban Public Safety Systems*, Vol. 5. Massachussetts Institute of Technology, Cambridge 1977.
4. Toregas, C., R. Swain, C. ReVelle, L. Bergman. The Location of Emergency Service Facilities. *Operations Research*, Vol. 19, 1970.
5. Marlin, P. G. Application of the Transportation Model to a Large-Scale "Districting" Problem. *Operations Research*, Vol. 8, 1987.
6. Nathanail, T. *Freeway Incident Delay Minimization Through the Optimum Deployment of Traffic Restoration Units.* M.S. thesis. University of Miami, Miami, Fla., 1991.
7. Lindley, J. A. *Transportation Research Circular 344: The Urban Freeway Congestion Problem.* TRB, National Research Council Washington, D.C., 1989.

Publication of this paper sponsored by Committee on Transportation Safety Management.

Meta-Analysis of Evaluations of Public Lighting as Accident Countermeasure

RUNE ELVIK

A meta-analysis of 37 studies evaluating the safety effects of public lighting is reported. The 37 studies contain a total of 142 results. The studies included were reported from 1948 to 1989 in 11 different countries. The presence of publication bias was tested by the funnel graph method. It was concluded that there is no evidence of publication bias and that it makes sense to estimate a weighted mean safety effect of public lighting on the basis of the 142 individual results. This is done by the log-odds method of meta-analysis. The validity of the combined results was tested against a number of rival hypotheses. It was concluded that the results are unlikely to have been caused by regression-to-the-mean and secular accident trends. The results were robust with respect to research design, decade of study, country of study, and type of traffic environment studied. The safety effects of public lighting were, however, sensitive to accident severity and type of accident. It was concluded that the best current estimates of the safety effects of public lighting are, in rounded values, a 65 percent reduction in nighttime fatal accidents, a 30 percent reduction in nighttime injury accidents, and a 15 percent reduction in nighttime property-damage-only accidents.

Public lighting of roads is widely accepted as an effective road accident countermeasure. Numerous studies have been done to determine the effects of public lighting on the number of accidents. In a synthesis of safety research related to traffic control and roadway elements, Schwab et al. (*1*) summarized the results of research by stating that "night accidents can be substantially reduced in number and severity by the use of good road lighting." This interpretation of the evidence from evaluation studies is not accepted by Vincent (*2*). In a critical review of 29 publications on road lighting and accidents, he concludes that "All of the studies claiming statistically significant accident reductions resulting from road lighting are deficient in any or all of: site selection, types of comparison, accident measures, measures of lighting and statistical evaluation techniques."

In nonexperimental accident research numerous threats to the validity of results exist. It is rarely possible to deal with all of them in a fully satisfactory way. Most literature surveys do not discuss the threats to validity at all or treat them informally, as Vincent (*2*) did. This paper argues that some issues that arise in studies attempting to summarize and interpret evidence from a number of evaluation studies can be resolved by quantitative meta-analysis. Three issues lend themselves to treatment by quantitative meta-analysis:

1. Is it meaningful to summarize the results of a number of studies of the effects of a certain accident countermeasure into an estimate of the mean effect on safety of the countermeasure? If yes, what is the best estimate of mean safety effects?
2. Which are the most and least valid and reliable results of studies that have evaluated the effects of an accident countermeasure? How can the most valid results be identified?
3. Why do the results of different evaluation studies concerning the same countermeasure vary? What are the most important sources of variation in study results?

This paper reports the results of a quantitative meta-analysis of evaluation studies concerning the safety effects of public lighting. Those studies were designed to address the three issues raised in the preceding paragraph. The studies have evaluated the effects on safety of public lighting on any type of road, including residential streets, rural highways, and freeways and covered both rural and urban areas and lighting of intersections as well as continuous roadway segments.

EVALUATION STUDIES INCLUDED IN META-ANALYSIS

Thirty-seven studies evaluating the effects of public lighting on road safety are included in the meta-analysis. The 37 studies contained a total of 142 results concerning the effects of road lighting on road safety; these results were expressed in terms of either changes in the number of nighttime accidents or changes in the nighttime accident rate per million vehicle kilometers of travel. The studies were retrieved by a systematic literature survey. A detailed description of how the literature survey was conducted is given elsewhere (*3*). The final sample consisted of evaluation studies that satisfied the following requirements:

1. The study contained one or more numerical estimates of the effects of public road lighting on the number of accidents or the accident rate.
2. The study primarily assessed the effects of introducing lighting at unlit locations. Studies that primarily assessed the effects of changing the level of existing lighting were not included.
3. The study presented the number of accidents on which estimates of the effects of lighting were based. Studies giving only accident rates, without stating the number of accidents used to estimate those rates, were not included.
4. The study was published. Unpublished studies were not included.

In the meta-analysis each estimate of safety effect was used as the unit of analysis. A total of 142 results were included. The results that were included in the analysis are provided in a later section (see Table 4). For each result, data concerning the following variables were collected:

Institute of Transport Economics, P.O. Box 6110 Etterstad, N-0602 Oslo, Norway.

1. Author or authors of study,
2. Year of publication,
3. Country to which each result refers,
4. Study design (coded variable with seven categories),
5. Type of traffic environment studied (coded variable with three categories),
6. Type of accident studied (coded variable with five categories),
7. Accident severity (coded variable with four categories),
8. Number of nighttime accidents before or without lighting,
9. Number of nighttime accidents after or with lighting,
10. Number of daytime accidents before or without lighting,
11. Number of daytime accidents after or with lighting, and
12. Estimate of the effect of lighting on road safety.

Table 1 describes in more detail how the variables included in the analysis were coded.

In terms of study design a broad distinction can be made between various forms of before-and-after studies on the one hand and various forms of comparative studies on the other. Conforming to the language of epidemiology [see, e.g. Hennekens and Buring (*4*)], the comparative studies will be referred to as case–control studies, in which one or more lit locations constitute the cases, whereas one or more unlit locations constitute the controls. The two main groups of research design differ in terms of the criterion of safety (CS) effect generally adopted. In before-and-after studies the basic CS effect is the odds ratio, commonly defined as

CS effect = (no. of nighttime accidents after/no. of nighttime accidents before)
÷ (no. of daytime accidents after/no. of daytime accidents before)

If this ratio is less than 1.0 lighting reduces the number of nighttime accidents. If it is more than 1.0 lighting increases the number of nighttime accidents. In some before-and-after studies, as well as in all case–control studies, the odds ratio is expressed in terms of accident rates rather than the number of accidents. If the introduction of public lighting does not affect exposure, the odds ratio of accident rates will be identical to the odds ratio of accident frequencies. The comparability of the two measures of safety effect is discussed in a subsequent section of the paper.

TECHNIQUES OF META-ANALYSIS

Meta-analysis can be done by several techniques (*5–9*). The simplest kind of meta-analysis is the vote counting method, which consists of compiling a frequency distribution of results by safety effect. A vote count of the 142 results concerning the safety effects of road lighting included in the present study shows that 115 results (81 percent) indicate that safety has improved and 27 results (19 percent) indicate that safety has deteriorated. Since the majority of results indicate that safety has improved, it is concluded that road lighting is likely to improve safety in most cases.

TABLE 1 Variables Included in Meta-Analysis

Variable	*Categories of the variable*
Author	Listed alphabetically
Year of publication	1948 through 1989
Country of origin	11 different countries represented
Study design	(1) 22 = before-and-after study with nighttime accidents on unlighted road sections as comparison group (2) 23 = before-and-after study with daytime accidents as comparison group (3) 2223 = before-and-after study with daytime accidents as comparison group and an additional comparison group of unlighted road sections (4) 2331 = before-and-after study with daytime accidents as comparison group and data on traffic volume by time of day before and after lighting (5) 26 = case-control study where comparisons between cases and controls are stratified according to one or more confounding variables (6) 27 = case-control study where cases and control have been matched according to one or more confounding variables (7) 33 = simple case-control study; cases and controls are compared directly with no control for confounding variables
Traffic environment	(1) Urb = urban; (2) Rur = rural; (3) Mwy = Motorway (freeway)
Type of accident	(1) All = all accidents; (2) Ped = pedestrian accidents; (3) Veh = accidents involving just vehicles; (4) Junc = accidents at junctions; (5) Sec = accidents between junctions
Accident severity	(1) Du = Fatal accidents; (2) Psu = injury accidents, (3) Msu = property-damage-only accidents (4) All = accidents of unspecified severity; all accidents included
Number of accidents	Recorded directly, in the following four categories: (1) NL = nighttime, lit road; (2) = NU = nighttime, unlit road; (3) DL = daytime, lit road; (4) DU = daytime, unlit road
Effect of lighting	Defined in terms of the odds ratio = O = (NL/NU)/(DL/DU), which may be equivalently expressed in terms of accident rates (number of accidents per million vehicle kilometres of travel)

A simple vote count is, however, not very informative. A refinement of the vote counting method consists of grouping results according to their statistical significance. Applied to the 142 results concerning the safety effects of road lighting, this version of the vote counting method shows that 45 results indicated a statistically significant safety improvement at the 5 percent level of significance. Ninety-seven results did not show any statistically significant changes in safety at this level of significance (*5*). This result illustrates the point raised by Hauer (*10*) about the danger of relying on tests of statistical significance alone in summarizing the results of several evaluations of a safety measure. Evidence of safety effects typically comes in small doses that are not always statistically significant. When a large number of studies are put together, however, their combined evidence can be very strong indeed.

The basic idea in more sophisticated techniques of meta-analysis is to combine statistically the evidence from several studies by computing a weighted mean result. Weighting can be done by several techniques, depending on the statistical properties of the results that are combined. In the present study the log-odds method described by Fleiss (*5*) was used.

Once a method for combining the results of different studies has been chosen, it is possible to study the effects of several variables on the combined result of case studies. Does, for example, the combined safety effect of public lighting vary according to the research design used in different studies? In meta-analysis this question can be answered by defining a variable describing study design (Table 1), combining evidence from all studies that use the same design, and comparing the combined evidence from studies that use different designs. In this paper the effects of several variables on the results of evaluation studies have been analyzed in this manner.

IS THERE A GENERAL EFFECT OF PUBLIC LIGHTING ON ROAD SAFETY?

Vincent (*2*) argues that it does not make sense to estimate a mean safety effect of public lighting, because the locations studied have not been sampled at random from a known sampling frame. Besides, the safety effect of public lighting is likely to vary substantially from one case to another, depending, inter alia, on luminance levels, traffic environment, and predominant type of accident at the location. In meta-analysis three requirements must be fulfilled for a weighted mean estimate of safety effect to make sense: (a) there should not be publication bias, (b) the assumption that all results belong to a distribution having a well-defined mean value should be reasonably well supported, and (c) all studies should use comparable measures of safety effect.

Testing for Publication Bias

The term *publication bias* refers to the tendency not to publish results that are unwanted or believed not to be useful, for example, because they show an increase in accidents or because they are not statistically significant (*6*).

Light and Pillemer (*6*) have developed a graphical technique of testing for publication bias called the *funnel graph method*. It relies on visual inspection of a diagram in which each study result is plotted in a coordinate system. The horizontal axis shows each result. The vertical axis shows the sample size on which each result is based. The idea is that if there is no publication bias the scatter plot of study results should resemble the form of a funnel turned upside down. The dispersion of points in the diagram should narrow as sample size increases, since large sample sizes provide more precise estimates of effects than small sample sizes. If the tails of the scatter plot are symmetrical and the density of points is the same in all areas of the diagram, this indicates that there is no publication bias.

Figures 1 to 4 show funnel graph diagrams of study results for studies of the effects of public lighting on fatal accidents (Figure 1), injury accidents (Figure 2), property-damage-only accidents (Figure 3), and accidents of unspecified severity (Figure 4). The latter category presumably includes accidents at all levels of severity. Statistical weight is used as a measure of sample size. The statistical weight of a result is proportional to the inverse of the variance of that result. For example, for a result based on 45 (dark, before), 25 (dark, after), 90 (day, before) and 85 (day, after) accidents, the statistical weight is 1/(1/45 + 1/25 + 1/90 + 1/85). Accidents of different degrees of severity were treated separately, because both safety effects and sample sizes are likely to differ across severity levels.

Inspection of Figures 1 to 4 does not give any indication of a clear publication bias. There is, however, a considerable amount of spread in the results. This indicates that statistically aggregating the results in terms of a weighted mean estimate of safety effect may be problematic.

Is There a True Mean Safety Effect?

The shape of scatter plots in funnel graph diagrams indicates if it makes sense to estimate a weighted mean safety effect. If the funnel graph is bimodal (has two humps) or multimodal or if there is no clear pattern in the scatter plot, a weighted mean will not be very informative. If a funnel pattern is clearly visible, estimating a weighted mean safety effect will be informative and will indicate the size of the effect that studies tend to converge to as sample size increases. In Figures 1 to 4 the funnel pattern is visible and a weighted mean value of the safety effects of lighting has been estimated.

In addition, Fleiss (*5*) describes a formal test of the homogeneity of the results. This test indicates that the results referring to fatal accidents and property-damage-only accidents are homogeneous, whereas there is a statistically significant heterogeneity in the results referring to injury accidents and accidents of unspecified severity. It was nevertheless decided to combine evidence from the various studies referring to injury accidents and accidents of unspecified severity to explore some of the sources of heterogeneity in the results.

Comparability of Measures of Effect

As pointed out earlier two measures of safety effect have been used in studies evaluating the safety effects of public lighting: changes in the odds ratio based on the number of accidents and changes in the odds ratio based on accident rates. In the funnel graph diagrams these two measures of safety effect have been mixed, relying on the assumption that neither the total amount of exposure nor its distribution between daytime and nighttime is affected by road lighting.

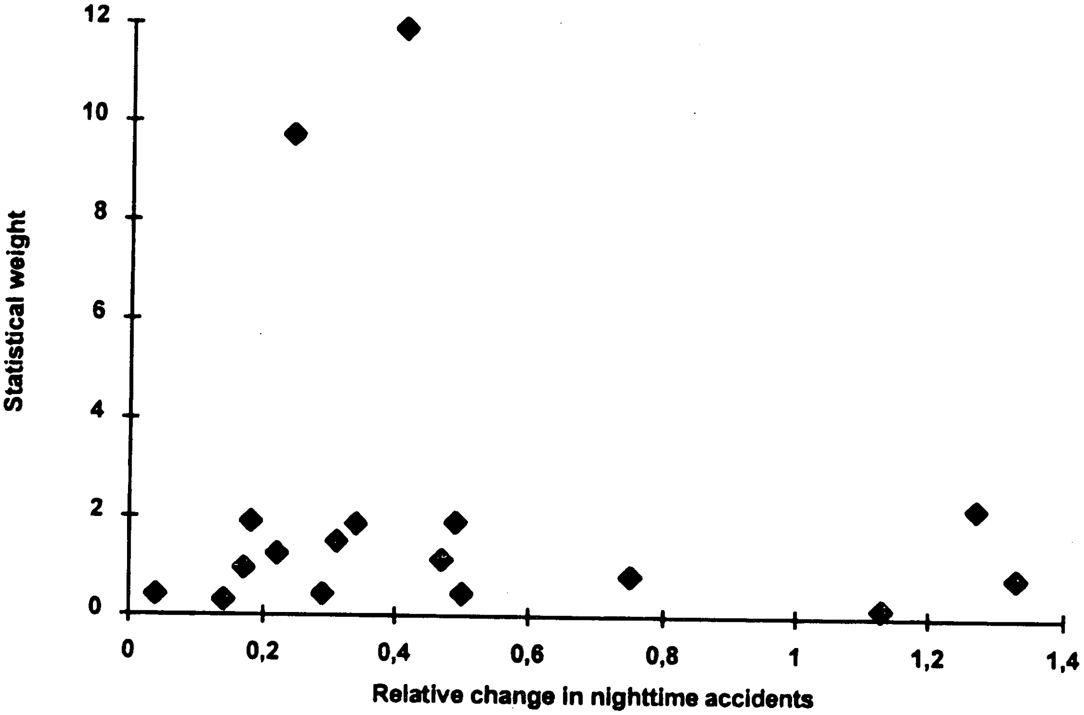


FIGURE 1 Funnel graph diagram for fatal accidents.

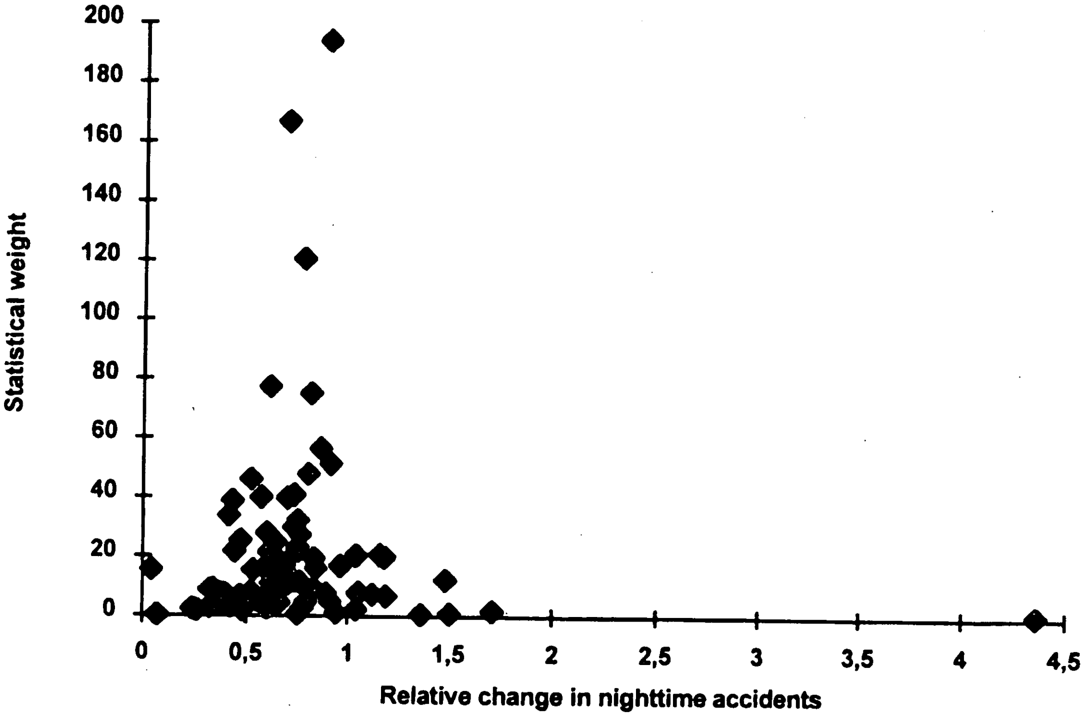


FIGURE 2 Funnel graph diagram for injury accidents.

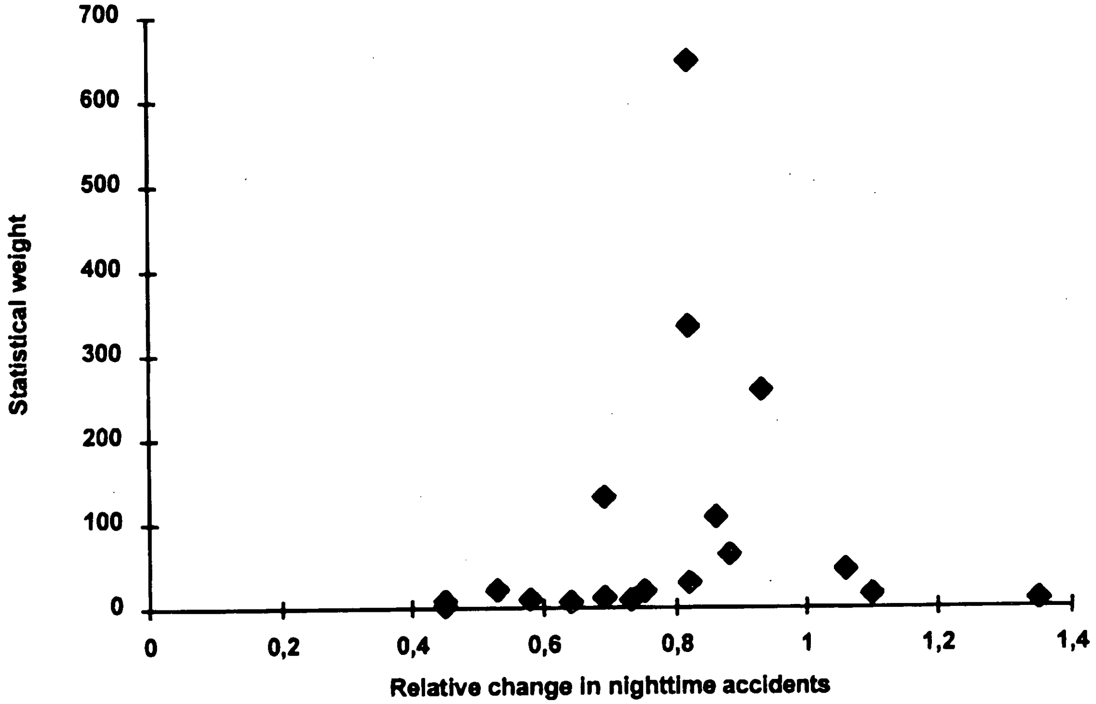


FIGURE 3 Funnel graph diagram for property-damage-only accidents.

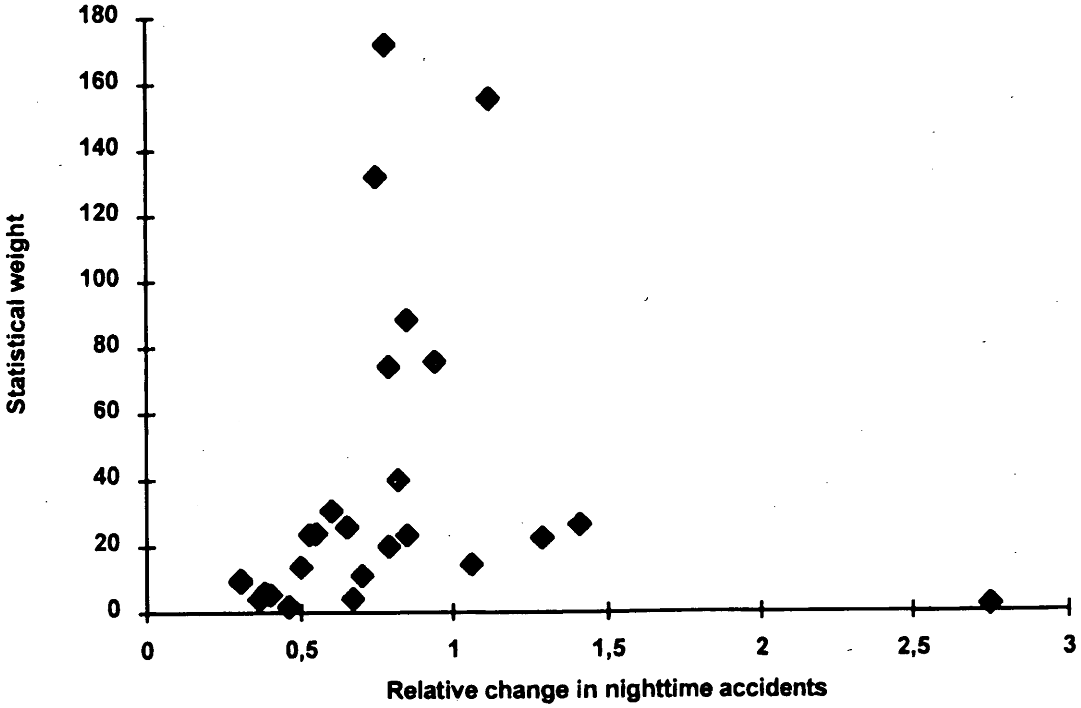


FIGURE 4 Funnel graph diagram for accidents of unspecified severity.

The validity of this assumption can be tested by relying on before-and-after studies in which both measures of safety effect can be estimated and compared. This can be done in all before-and-after studies in which exposure data are available for both the before and after periods. The studies of Tamburri, et al. (*11*), Box (*12*), Lipinski and Wortman (*13*), Walker and Roberts (*14*), Jørgensen (*15*), and Lamm et al. (*16*) allow this kind of comparison to be made. The combined estimate of the safety effect of lighting, based on these studies and measured by means of the number of accidents, is a 30 percent reduction in the number of nighttime accidents (lower 95 percent confidence limit, 21 percent reduction; upper limit, 38 percent reduction). If the safety effect is measured by means of accident rates, the combined estimate is a 33 percent reduction in nighttime accident rate (lower 95 percent confidence limit, 25 percent reduction; upper limit, 41 percent reduction). These values are very close to each other. It is concluded that changes in accident rates and changes in accident frequency can be interpreted as equivalent measures of the changes to be expected in the number of accidents with the introduction of road lighting.

VALIDITY OF EVALUATIONS OF PUBLIC LIGHTING

All of the evaluation studies included in this meta-analysis are nonexperimental. In this section, a number of threats to the validity of these studies will be discussed, including

1. Regression to the mean,
2. Secular accident trends, and
3. Contextual confounding variables.

Regression to the Mean

The most common research design in evaluation studies concerning the safety effects of public lighting is a before-and-after design, in which nighttime accidents form the experimental group and daytime accidents are used as a comparison group. In this kind of research design, regression to the mean (*17,18*) may jeopardize the validity of the results. In particular, if road lighting is introduced because of an abnormally high recorded number of accidents in the before period, a subsequent decline in the number of accidents must be expected even if lighting has no effect.

The use of daytime accidents as a comparison group in before-and-after studies will take care of the regression-to-the mean effect, provided that this effect affects daytime accidents to the same extent as nighttime accidents. This is not likely to be the case if road lighting was introduced because an abnormally high proportion of all accidents occurred in darkness. In that case one might expect the percent decline in nighttime accidents because of regression to the mean to be greater than the corresponding percent decline in daytime accidents, thus creating an apparent effect of road lighting.

On the other hand, a high percentage of nighttime accidents could indicate a real problem. In that case one would expect the true effect of road lighting to be greater when the percentage of nighttime accidents is high than when it is low. By juxtaposing the results of before-and-after studies and case–control studies made at locations with various percentages of nighttime accidents, it is possible to get an indication of whether a greater effect of road lighting at locations with a high percentage of nighttime accidents reflects regression to the mean or a genuine accident problem in the darkness.

If the regression-to-the-mean hypothesis is correct, one would expect the apparent effect of lighting to vary according to the percentage of all accidents occurring at night in before-and-after studies but not in case–control studies. If the real-darkness-problem hypothesis is correct, one would expect the effect of road lighting to vary according to the percentage of all accidents occurring at night in both before-and-after and case–control studies.

Table 2 presents data that are relevant for the two hypotheses. Study locations have been grouped according to the percentage of all accidents occurring at night (in the before period in before-and-after studies). In both before-and-after studies and case–control studies the effect of road lighting on the number of nighttime accidents is found to be greater at locations where more than 50 percent of all accidents occur at night than at locations where fewer than 50 percent of all accidents occur at night. This result weakens the regression-to-the-mean hypothesis and strengthens the real-darkness-problem hypothesis. However, the validity of the assumptions underlying the comparison cannot be tested directly. Hence, the comparison is just an indication, not a stringent test.

Secular Accident Trends

Over time the percentage of all accidents occurring at night may change. Changes in traffic distribution by hour of the day, improved vehicle headlights, and changes in the driver population are some of the factors that could generate such changes. In before-and-after studies with just one before period and just one after period and no comparison group consisting of locations where road lighting was

TABLE 2 Results of Before-and-After Studies and Case–Control Studies by Proportion of Nighttime Accidents: Weighted Mean Effect of Public Lighting on Nighttime Accidents

			Per cent change in nighttime accidents		
Study design	Percentage of accidents at night	Proportion of statistical weights	Lower 95%	Best estimate	Upper 95%
Before-and-after designs (cf table 1)	> 50%	0.089	-28	-35	-41
	33-50%	0.326	-17	-21	-25
	< 33%	0.231	-17	-22	-26
Case-control designs (cf table 1)	> 50%	0.071	-24	-32	-39
	33-50%	0.136	-7	-15	-21
	< 33%	0.147	-14	-21	-27
All designs	All	1.000	-20	-23	-25

TABLE 3 Weighted Mean Effect of Public Lighting on Nighttime Accidents According to Potential Confounding Variables

			Per cent change in nighttime accidents		
Variable	Category	Proportion of statistical weights	Lower 95%	Best estimate	Upper 95%
Accident severity	(1) Fatal	0.008	-52	-65	-75
	(2) Injury	0.387	-26	-29	-32
	(3) PDO	0.381	-13	-17	-21
	(4) Unspecified	0.224	-13	-18	-23
Study design (cf table 1 for fuller description)	(A) Fatal accs				
	(2) Design 23	0.798	-48	-63	-74
	(3) Design 2223	0.161	-40	-73	-88
	(5) Design 26	0.041	+95	-59	-91
	(B) Injury accs				
	(1) Design 22	0.036	-5	-26	-32
	(2) Design 23	0.526	-25	-30	-34
	(3) Design 2223	0.080	-16	-29	-39
	(4) Design 2331	0.007	-32	-60	-77
	(5) Design 26	0.154	-17	-26	-35
	(6) Design 27	0.044	-24	-39	-51
	(7) Design 33	0.153	-15	-25	-34
	(C) PDO accs				
	(2) Design 23	0.868	-11	-16	-20
	(4) Design 2331	0.008	+35	-19	-51
	(5) Design 26	0.038	+9	-15	-33
	(7) Design 33	0.086	-17	-30	-40
	(D) Unspec accs				
	(2) Design 23	0.024	-25	-50	-66
	(4) Design 2331	0.217	-18	-29	-37
	(5) Design 26	0.593	-1	-8	-15
	(6) Design 27	0.166	-17	-28	-38
Decade of publication	(1) 1940s	0.125	-8	-15	-22
	(2) 1950s	0.052	-21	-30	-39
	(3) 1960s	0.174	-14	-19	-25
	(4) 1970s	0.523	-19	-22	-26
	(5) 1980s	0.126	-25	-31	-37
Country	(1) Australia	0.198	-14	-19	-25
	(2) Denmark	0.024	-0	-17	-31
	(3) Finland	0.015	-1	-22	-38
	(4) France	0.017	-24	-39	-51
	(5) Germany	0.010	+1	-24	-43
	(6) Great Britain	0.123	-27	-32	-38
	(7) Israel	0.003	-8	-46	-68
	(8) Japan	0.005	-32	-56	-71
	(9) Sweden	0.063	-14	-24	-32
	(10) Switzerland	0.015	+0	-21	-38
	(11) United States	0.527	-17	-20	-23
Traffic environment	(1) Urban	0.593	-19	-22	-25
	(2) Rural	0.117	-19	-26	-32
	(3) Motorways	0.290	-20	-23	-25
Type of accident	(1) Not stated	0.478	-18	-21	-24
	(2) Pedestrian	0.045	-45	-52	-58
	(3) Vehicles only	0.312	-13	-17	-21
	(4) Junctions	0.125	-24	-30	-36
	(5) Midblocks	0.040	-0	-14	-25
All	All	1.000	-20	-23	-25

Note: The statistical weights sum to 1.000 for each variable (each severity level for the variable study design)

not introduced, the possibility that secular accident trends are confounded with the effects of road lighting cannot be ruled out. However, in all other research designs that have been used in evaluations of the safety effect of public lighting, this particular source of error can be ruled out.

Table 3 compares the results of evaluations that have relied on different research designs. With a few exceptions the weighted mean safety effect of lighting is virtually identical in all research designs. It is therefore highly unlikely that the results of before-and-after studies with only daytime accidents as a comparison group could be explained in terms of secular accident trends alone. The study results that were included in the analysis are listed in Table 4.

Contextual Confounding Variables

To what extent do variables related to study context affect the results of evaluations of the safety effects of public lighting? Table

TABLE 4 Data from 37 Studies of Safety Effects of Public Lighting

Study	Year	Country	Design	Environ-ment	Type of accident	Accident severity	Night before/without	Night after/with	Day before/without	Day after/with	Effect
(19)	1948	USA	23	Urb	All	Du	3	1	3	2	0,500
			23	Urb	All	Psu	45	34	47	57	0,623
			23	Urb	All	Msu	201	200	324	365	0,883
			23	Urb	All	Du	17	5	10	6	0,490
			23	Urb	All	Psu	210	135	172	152	0,727
			23	Urb	All	Msu	828	789	1411	1443	0,932
			23	Urb	All	Du	8	4	3	2	0,750
			23	Urb	All	Psu	96	51	75	59	0,675
			23	Urb	All	Msu	323	340	547	672	0,857
			23	Urb	All	Psu	67	86	80	99	1,037
			23	Urb	All	Psu	173	82	126	99	0,603
			23	Urb	All	Psu	43	23	45	23	1,047
			23	Urb	All	Psu	72	28	31	36	0,335
(20)	1955	GB	23	Urb	Ped	Du	6	1	1	4	0,042
			23	Urb	Ped	Psu	31	19	73	71	0,630
			23	Urb	Kjt	Du	4	2	8	3	1,333
			23	Urb	Kjt	Psu	120	98	283	330	0,700
(21)	1958	CH	23	Urb	All	Psu	70	65	159	231	0,639
(22)	1958	GB	23	Urb	Ped	Du	15	6	5	11	0,182
			23	Urb	Ped	Psu	144	85	314	323	0,574
			23	Urb	Kjt	Du	13	9	11	6	1,269
			23	Urb	Kjt	Psu	333	303	918	1086	0,769
(23)	1960	USA	26	Mwy	All	All	52	168	71	177	1,291
(24)	1962	USA	26	Mwy	All	Psu	8	2	13	2	4,361
			26	Mwy	All	All	27	108	42	316	0,500
(25)	1962	GB	23	Mwy	All	Psu	8	7	13	19	0,599
			23	Mwy	All	Psu	41	3	71	22	0,236
(26)	1962	USA	26	Mwy	All	All	184	1004	172	997	0,943
			26	Mwy	All	All	401	1004	514	997	1,120
(27)	1965	S	23	Urb	All	Psu	14	13	41	69	0,552
			23	Urb	All	Msu	48	52	96	95	1,095
			23	Rur	All	Psu	23	15	35	42	0,543
			23	Rur	All	Msu	27	20	85	86	0,732
(28)	1966	GB	23	Urb	Ped	Psu	7	0,5	1	1	0,071
			23	Urb	Kjt	Psu	2	3	5	5	1,500
			23	Rur	All	Psu	40	22	37	39	0,522
			23	Mwy	All	Psu	82	54	123	132	0,614
(29)	1966	USA	33	Mwy	All	Psu	588	706	547	950	0,691
			33	Mwy	All	Msu	395	576	430	911	0,688
(11)	1968	USA	2331	Urb	Junc	All	75	27	39	39	0,304
			2331	Urb	Junc	All	25	11	31	34	0,396
			2331	Urb	Junc	All	33	13	31	34	0,377
			2331	Urb	Junc	All	37	15	12	12	0,355
			2331	Urb	Junc	All	11	5	7	8	0,455

* Number of nighttime accidents on unlit roads before and after.

(*continued on next page*)

3 presents results that shed light on this question for the variables (a) definition of accident severity, (b) study design, (c) decade of publication of study, (d) country where the study was performed, (e) traffic environment where the study was performed, (f) type of accident studied.

The effects of road lighting vary significantly with respect to accident severity. Nighttime fatal accidents are reduced by about 65 percent, nighttime injury accidents are reduced by about 30 percent, and nighttime property-damage-only accidents are reduced by about 15 percent. This means that studies that do not specify the severity of accidents are less informative than studies that specify accident severity. The observed weighted mean safety effect in studies of accidents of unspecified severity is an 18 percent reduction in nighttime accidents. This indicates that most of the accidents probably were property-damage-only accidents.

These results hold when controlling for study design. In general, study design appears to have a minor effect on study results. As argued earlier the robustness of the results with respect to study design indicates that the results are valid and not just the product of various confounding factors that are left uncontrolled by the various research designs. Different research designs take different confounding factors into account. Therefore, agreement of results across research designs indicates that uncontrolled confounding factors are not major sources of variation in the results of different studies.

The oldest study included was reported in 1948; the most recent was reported in 1989. Studies performed in different decades have yielded similar results. There is no indication that the safety effects of road lighting have diminished over time. Eleven different countries are represented in this analysis. Studies performed in different countries have also yielded similar results. It should be noted, however, that most studies have been performed in the United States, Great Britain, and Australia. Studies performed in other countries have been on a smaller scale, as indicated by their contribution to the statistical weights.

Three types of traffic environment have been identified: urban, rural, and freeways. The results of evaluation studies are the same for all three environments. This holds when controlling for accident severity. With respect to type of accident, studies can be divided into three groups. The first and largest group consists of studies that do not specify the types of accident studied. A second group consists of studies in which a distinction is made between pedestrian accidents and other accidents. A third group consists of studies in which a distinction is made between accidents at junctions (inter-

TABLE 4 (*continued*)

Study	Year	Country	Design	Environment	Type of accident	Accident severity	Night before/without	Night after/with	Day before/without	Day after/with	Effect
(30)	1969	USA	23	Urb	All	All	13	9	37	38	0,674
(31)	1969	USA	26	Urb	All	Du	4	14	2	15	0,468
			26	Urb	All	Psu	203	309	295	551	0,811
			26	Urb	All	Msu	83	240	220	592	1,062
(32)	1970	CH	23	Rur	All	Psu	64	77	92	94	1,178
			23	Mwy	All	Psu	10	5	25	12	1,042
			23	Mwy	All	Psu	18	5	41	27	0,422
			23	Mwy	All	Psu	4	6	25	22	1,705
			23	Urb	All	Psu	36	14	104	36	1,123
(33)	1971	AUS	23	Urb	Ped	Psu	10	6	23	18	0,767
			23	Urb	Ped	Psu	16	10	18	19	0,592
			23	Urb	Ped	Psu	15	6	16	20	0,320
			23	Urb	Ped	Psu	17	6	28	20	0,494
		USA	23	Urb	Ped	Psu	175	122	221	294	0,524
			23	Urb	Kjt	Psu	152	317	176	427	0,860
			23	Urb	Kjt	Msu	983	1674	1069	2215	0,822
			23	Urb	Ped	Du	84	22	42	46	0,239
			23	Urb	All	Du	60	38	40	62	0,409
			23	Urb	All	Psu	48	37	52	53	0,756
			23	Urb	All	All	38	30	62	70	0,699
(34)	1971	DK	2223	Urb	Ped	Psu	20	21	58	93	1,047
(12)	1972	USA	2331	Urb	All	Psu	23	10	15	17	0,384
			2331	Urb	All	Psu	52	20	30	30	0,385
			2331	Urb	All	Msu	23	4	25	12	0,448
			2331	Urb	All	Msu	53	33	75	49	0,687
			26	Mwy	Rear	All	176	198	614	863	0,794
			26	Mwy	Kjt	All	69	48	142	123	0,786
			26	Mwy	Ped	All	11	11	11	4	2,750
			26	Mwy	Off	All	102	102	132	92	1,410
			26	Mwy	All	All	356	697	888	2184	0,784
			26	Mwy	All	All	270	72	428	192	0,822
(35)	1972	GB	2223	Urb	All	Du	4	0,5	2	2	0,140
			2223	Urb	All	Psu	85	60	134	138	0,750
			2223	Rur	All	Du	11	3	5	6	0,220
			2223	Rur	All	Psu	73	56	121	145	0,620
			2223	Mwy	All	Du	8	2	4	6	0,170
			2223	Mwy	All	Psu	54	50	99	95	0,960
			2223	Urb	All	Du	1	1	0,5	0,5	1,130
			2223	Urb	All	Psu	18	9	37	32	0,660
			2223	Rur	All	Du	11	3	9	8	0,310
			2223	Rur	All	Psu	84	56	132	118	0,750
			2223	Mwy	All	Du	13	4	10	9	0,350
			2223	Mwy	All	Psu	110	75	186	175	0,730
(36)	1972	AUS	23	Urb	Ped	Psu	32	13	57	58	0,399

* Number of nighttime accidents on unlit roads before and after.

(*continued on next page*)

sections) and accidents at road sections (midblock accidents). On the basis of these classifications, road lighting appears to have a greater effect on pedestrian accidents than on other types of accidents and a greater effect at junctions than at other locations.

The general impression is that the contextual variables have a rather small impact on the results of evaluation studies. It is particularly reassuring that results are robust with respect to study design. Study decade, the country where the study was performed, and type of traffic environment hardly affect study results. On the other hand, accident severity and type of accident seem to be of some importance for study results. These variables are not directly related to study design. However, any good study should specify clearly the severity of the accidents that are studied and indicate clearly the types of accidents that are studied.

DISCUSSION OF RESULTS

The analysis presented here shows that the results of studies that have evaluated the effects of public lighting on road safety are quite robust with respect to a number of potentially confounding variables. These results cannot be dismissed as merely showing the vagaries of poor data, inadequate research design, or peculiarities of the locations that have been investigated. There is little to support the misgivings voiced by Vincent (*2*) with respect to these and related points.

On the other hand, the present analysis did not consider every conceivable source of error in previous studies. In particular, errors that may arise from an inappropriate choice of comparison groups in case–control studies or from the use of an inappropriate statistical technique in analyzing data were not considered. Most studies provide few details concerning the sampling of cases and controls. It is therefore difficult to know whether biased sampling is found and how it may have affected evaluation results. As far as statistical techniques for data analysis are concerned, most studies have relied on quite simple techniques, like estimating an odds ratio and testing it for statistical significance. More advanced multivariate analyses, in which the choice of statistical techniques is more important, are not found in this area.

The effect of public lighting on road safety was found to vary with respect to accident severity and type of accident. There are no doubt a large number of other variables with respect to which the effects of public lighting might be expected to vary. It would, for example, be of interest to know whether lighting satisfying current

TABLE 4 *Continued*

Study	Year	Country	Design	Environ-ment	Type of accident	Accident severity	Night before/without	Night after/with	Day before/without	Day after/with	Effect
(37)	1973	GB	22	Urb	All	Psu	44	26	*12532	*8785	0,840
			22	Urb	All	Psu	23	16	*3924	*3286	0,830
			22	Rur	All	Psu	23	27	*3381	*2681	1,480
			22	Rur	All	Psu	93	35	*9027	*7245	0,470
(38)	1976	GB	23	Mwy	All	All	52	24	34	53	0,296
(13)	1976	USA	2331	Rur	Junc	All	356	438	656	1022	0,748
(14)	1976	USA	2331	Rur	Junc	All	90	46	225	207	0,551
(39)	1977	DK	33	Mwy	All	Psu	91	434	191	1006	0,905
			33	Mwy	All	Psu	91	289	191	759	0,799
(40)	1977	AUS	23	Urb	Ped	Psu	162	87	219	276	0,426
			23	Urb	Kjt	Psu	779	762	746	820	0,890
			23	Urb	Kjt	Msu	1908	1840	3854	4510	0,824
(41)	1977	JPN	23	Mwy	All	Psu	95	52	109	135	0,442
		USA	26	Mwy	All	Du	6	36	0,5	14	0,288
			26	Mwy	All	Psu	38	639	28	804	0,533
			26	Mwy	All	Msu	45	1372	41	2454	0,533
(42)	1978	SF	23	Urb	Sec	Psu	104	67	181	153	0,762
			23	Urb	Sec	Msu	112	75	187	153	0,818
			23	Urb	Junc	Psu	19	12	36	25	0,909
			23	Urb	Junc	Msu	26	15	43	39	0,636
(43)	1978	ISR	2223	Urb	Ped	Psu	79	34	77	61	0,623
(15)	1980	DK	2331	Urb	All	Psu	8	5	10	13	0,480
(44)	1981	S	26	Rur	Junc	Psu	58	11	90	36	0,474
			26	Rur	Junc	Psu	27	3	26	11	0,263
			26	Rur	Junc	Psu	153	34	306	82	0,829
			26	Rur	Junc	Psu	104	48	194	77	1,163
			26	Rur	Junc	Psu	19	20	58	69	0,885
			26	Rur	Junc	Psu	1	3	4	16	0,750
			26	Rur	Junc	Psu	31	13	102	36	1,188
			26	Rur	Junc	Psu	21	9	57	31	0,788
(1)	1982	D	23	Urb	Ped	Psu	51	19	44	51	0,321
		CH	23	Urb	Ped	Psu	34	15	52	60	0,382
		F	27	Urb	Junc	Psu	290	209	389	459	0,611
		USA	23	Mwy	Junc	Psu	76	43	83	80	0,587
(45)	1985	S	23	Rur	Junc	Psu	58	19	137	64	0,701
(16)	1985	D	2331	Mwy	All	All	30	77	61	148	1,062
			2331	Mwy	All	All	46	121	102	316	0,845
(46)	1986	S	27	Rur	Junc	All	114	63	258	236	0,604
			27	Rur	Junc	All	449	157	1256	517	0,849
			27	Rur	Junc	All	93	43	251	218	0,532
			27	Rur	Junc	All	119	43	390	218	0,646

* Number of nighttime accidents on unlit roads before and after.

Study	Year	Country	Design	Environ-ment	Type of accident	Accident severity	Night before/without	Night after/with	Day before/without	Day after/with	Effect
(47)	1987	GB	26	Mwy	All	Psu	669	212	264	51	0,412
			26	Mwy	All	Psu	71	57	256	35	1,037
			26	Mwy	All	Psu	58	24	267	44	0,733
			26	Mwy	All	Psu	61	35	301	116	0,681
			26	Mwy	All	Psu	59	144	218	398	0,749
(48)	1989	USA	33	Urb	Junc	Psu	1	42	3	93	1,355
			33	Urb	Junc	Msu	15	160	19	447	0,453
			33	Urb	Sec	Psu	36	2	51	3	0,944
			33	Urb	Sec	Msu	133	19	218	23	1,354
			23	Urb	Junc	Psu	21	15	29	36	0,575
			23	Urb	Junc	Msu	51	57	117	174	0,752
			23	Urb	Sec	Psu	15	8	28	31	0,482
			23	Urb	Sec	Msu	29	23	84	114	0,584
							15879	18769	54272	57940,5	0,737

* Number of nighttime accidents on unlit roads before and after.

warrants is more effective than lighting not satisfying current warrants. However, few studies provide information concerning this. The availability of data limits the topics that can be included in a meta-analysis.

CONCLUSIONS

The following conclusions summarize the results of the research reported in this paper.

1. A meta-analysis of 37 evaluation studies of the safety effect of public lighting containing 142 results has been performed. The log-odds method was applied.

2. The presence of publication bias was tested. No evidence of publication bias was found.

3. Changes in accident rate were found to predict accurately changes in the number of accidents associated with the introduction of public lighting. These two measures of safety effect were therefore treated as equivalent in the meta-analysis.

4. The validity of research results was tested with respect to (a) regression to the mean; (b) secular accident trends; and (c) contextual confounding variables, including definition of accident severity, study design, decade of publication, country where the study was performed type of traffic environment, and type of accident studied. It was concluded that regression to the mean and secular accident trends are unlikely to have affected the results of evaluation studies materially. As far as confounding variables are concerned, accident severity and type of accident studied were found to affect study results. The other confounding variables did not affect study results.
5. The best current estimate of the safety effects of road lighting in rounded values is a 65 percent reduction in nighttime fatal accidents, a 30 percent reduction in nighttime injury accidents, and a 15 percent reduction in nighttime property-damage-only accidents.

REFERENCES

1. Schwab, R. N., N. E. Walton, J. M. Mounce, and M. J. Rosenbaum. Roadway Lighting. *Synthesis of Safety Research Related to Traffic Control and Roadway Elements*, Vol. 2. Report FHWA-TS-82-233. FHWA, U.S. Department of Transportation, 1982.
2. Vincent, T. Streetlighting and Accidents. Paper 17. In *Traffic Accident Evaluation* (D. C. Andreassend and P. G. Gipps, eds.), Papers presented at Esso-Monash Civil Engineering Workshop, Normanby House, Monash University, February 15 to 17, 1983. Department of Civil Engineering, Monash University, Australia, 1983.
3. Elvik, R. *Metaanalyse av Effektmålinger av Trafikksikkerhetstiltak*. TØI-Rapport 232. Transportøkonomisk Institute, Oslo, Norway 1994.
4. Hennekens, C. H., and J. E. Buring. *Epidemiology in Medicine*. Little, Brown & Co, Boston, 1987.
5. Fleiss, J. L. *Statistical Methods for Rates and Proportions*, 2nd ed. John Wiley and Sons, New York, 1981.
6. Light, R. J., and D. B. Pillemer. *Summing Up. The Science of Reviewing Research*. Harvard University Press, Cambridge, Mass., 1984.
7. Hedges, L. V., and I. Olkin. *Statistical Methods for Meta-Analysis*. Academic Press, San Diego, Calif., 1985.
8. Hunter, J. E., and F. L. Schmidt. *Methods of Meta-Analysis. Correcting Error and Bias in Research Findings*. Sage Publications, Newbury Park, Calif., 1990.
9. Rosenthal, R. M. Meta-Analytic Procedures for Social Research. *Applied Social Research Methods Series*, Vol. 6. Sage Publications, Newbury Park, Calif., 1991.
10. Hauer, E. Should Stop Yield? Matters of Method in Safety Research. *ITE-Journal*, Sept. 1991, pp. 25–31.
11. Tamburri, T. N., C. J. Hammer, J. C. Glennon, and A. Lew. Evaluation of Minor Improvements. In *Highway Research Record 257*, HRB, National Research Council, Washington, D.C., 1968, pp. 34–79.
12. Box, P. C. Freeway Accidents and Illumination. In *Highway Research Record 416*, HRB, National Research Council, Washington, D.C., 1972, pp. 10–20.
13. Lipinski, M. E., and R. H. Wortman. Effect of Illumination on Rural At-Grade Intersection Accidents. In *Transportation Research Record 611*, TRB, National Research Council, Washington, D.C., 1976, pp. 25–27.
14. Walker, F. W., and S. E. Roberts. Influence of Lighting on Accident Frequency at Highway Intersections. In *Transportation Research Record 562*, TRB, National Research Council, Washington, D.C., 1976, pp. 73–78.
15. Jørgensen, E. *Eksempler på Effektstudier fra SSV*. Vejdirektoratet, Sekretariatet for Sikkerhedsfremmende Vejforanstaltninger (SSV), Næstved, 1980.
16. Lamm, R., J. H. Klöckner, and E. M. Choueiri. Freeway Lighting and Traffic Safety—A Long-Term Investigation. In *Transportation Research Record 1027*, TRB, National Research Council, Washington, D.C., 1985, pp. 57–63.
17. Hauer, E. Bias-by-Selection: Overestimation of the Effectiveness of Safety Countermeasures Caused by the Process of Selection for Treatment. *Accident Analysis and Prevention*, Vol. 12, 1980, pp. 113–117.
18. Hauer, E. On the Estimation of the Expected Number of Accidents. *Accident Analysis and Prevention*, Vol. 18, 1986, pp. 1–12.
19. Seburn, T. C. Relighting A City. *Proc., Institute of Traffic Engineers Nineteenth Annual Meeting*, 1948, pp. 58–72.
20. Tanner, J. C., and A. W. Christie. Street Lighting and Accidents—A Study of Some New Installations in the London Area. *Light and Lighting*, Vol. 48, 1955, pp. 395–397.
21. Borel, P. Accident Prevention and Public Lighting. *Bulletin des Schweitzerisches Elektrotechnisches Verbands*, Vol. 49, No. 1, 1958, pp. 8–11.
22. Tanner, J. C. Reduction of Accidents by Improved Street Lighting. *Light and Lighting*, Vol. 51, 1958, pp. 353–355.
23. Taragin, A., and B. M. Rudy. Traffic Operations as Related to Highway Illumination and Delineation. *Bulletin 255* HRB, National Research Council, Washington, D.C., 1960, pp. 1–22.
24. Billion, C. E., and N. C. Parsons. Median Accident Study—Long Island, New York. *Bulletin 308*, HRB, National Research Council, Washington, D.C., 1962, pp. 64–79.
25. Christie, A. W. Some Investigations Concerning the Lighting of Traffic Routes. *Public Lighting*, Vol. 27, 1962, pp. 189–204.
26. Ives, H. S. Does Highway Illumination Affect Accident Occurrence? *Traffic Quarterly*, Vol. 16, 1962, pp. 229–241.
27. *Väg-och Gatubelysningens Inverkan på Trafik-Säkerheten*. Meddelande 60. Transportforskningskommissionen, Stockholm, Sweden, 1965.
28. Christie, A. W. Street Lighting and Road Safety. *Traffic Engineering and Control*, Vol. 7, 1966, pp. 229–231.
29. Institute of Traffic Engineers and Illuminating Engineering Society. Joint Committee of Public Lighting. *Public Lighting Needs*. Special Report to U.S. Senate, 1966.
30. Cleveland, D. E. Illumination. *Traffic Control and Roadway Elements—Their Relationship to Highway Safety*, Revised. Automotive Safety Foundation, Washington, D.C. 1969, Chapt. 3.
31. Tennessee Valley Authority. A Study of the Benefits of Suburban Highway Lighting. *Illuminating Engineering*, April 1969, pp. 359–363.
32. Walthert, R., F. Mäder, and P. Hehlen. *Données Statistiques sur la Proportion des Accidents le Jour et la Nuit, leurs Causes et Conséquences*. La Conduite de Nuit, Automobil Club de Suisse, 1970.
33. Fisher, A. J. *A Review of Street Lighting in Relation to Road Safety*. Report 18. Australian Department of Transport, Australian Government Publishing Service, Canberra, 1971.
34. Jørgensen, N. O., and Z. Rabani. *Fodgængeres Sikkerhed i og ved Fodgænger-Overgange*. RfT-Rapport 7. Rådet for Trafiksikkerhedsforskning, Copenhagen, Denmark, 1971.
35. Cornwell, P. R., and G. M. Mackay. Lighting and Road Traffic. Part 1. Public Lighting and Road Accidents. *Traffic Engineering and Control*, Vol. 13, 1972, pp. 142–144.
36. Pegrum, B. V. The Application of Certain Traffic Management Techniques and Their Effect on Road Safety. National Road Safety Symposium, Department of Transport, Canberra, Australia 1972.
37. Sabey, B. E., and H. D. Johnson. *Road Lighting and Accidents: Before and After Studies on Trunk Road Sites*. TRRL Report LR 586. Transport and Road Research Laboratory, Crowthorne, Berkshire, United Kingdom, 1973.
38. Austin, B. R. Public Lighting—The Deadly Reckoning. *Traffic Engineering and Control*, Vol. 17, 1976, pp. 262–263.
39. Andersen, K. B. *Uheldsmønsteret på Almindelige 4-Sporede Veje*. RfT-Rapport 20. Rådet for Trafiksikkerhedsforskning, Copenhagen, Denmark, 1977.
40. Fisher, A. J. Road Lighting as an Accident Countermeasure. *Australian Road Research*, Vol. 7, No. 4, 1977, pp. 3–15.
41. Ketvirtis, A. *Road Illumination and Traffic Safety*. Road and Motor Vehicle Traffic Safety Branch, Transport Canada, Ottawa, Ontario, Canada, 1977.
42. National Board of Public Roads and Waterways. *Traffic Safety Effects of Road Lights*. Väg-och Vattenbyggnadsstyrelsen, Helsinki, Finland, 1978.
43. Polus, A., and A. Katz. An Analysis of Nighttime Pedestrian Accidents at Specially Illuminated Crosswalks. *Accident Analysis and Prevention*, Vol. 10, 1978, pp. 223–228.
44. Brüde, U., and J. Larsson. *Vägkorsningar på Landsbygd inom Huvudvägnätet. Olycksanalys*. VTI-Rapport 233. Statens Väg-och Trafikinstitut, Linköping, Sweden, 1981.
45. Brüde, U., and J. Larsson. *Korsningsåtgärder Vidtagna inom vägförvaltningarnas Trafiksäkerhetsarbete. Regressions-och åtgärdseffekter.*

VTI-Rapport 292. Statens Väg-och Trafikinstitut, Linköping, Sweden, 1985.

46. Brüde, U., and J. Larsson. *Trafiksäkerhetseffekter av Korsningsåtgärder.* VTI-Rapport 310. Statens Väg-och Trafikinstitut, Linköping, Sweden, 1986.
47. Cobb, J. Light on Motorway Accident Rates. *The Journal of the Institution of Highways and Transportation*, Oct. 1987, pp. 29–33.
48. Box. P. C. Major Road Accident Reduction by Illumination. In *Transportation Research Record 1247*, TRB, National Research Council, Washington, D.C., 1989, pp. 32–38.

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

Evaluation of Advanced Surveying Technology for Accident Investigation

KENNETH R. AGENT, JOHN A. DEACON, JERRY G. PIGMAN, AND NIKIFOROS STAMATIADIS

The objective was to evaluate the use of advanced technology for the investigation of traffic accidents. Substantial time and manpower are sometimes required to properly investigate serious vehicular crashes and document on-scene data. An alternative to the traditional coordinate method for on-scene data collection is the use of electronic total stations. This is a form of digital surveying equipment that can be used to obtain detailed measurements, with the option of plotting the stored data by computer. The study compared the use of advanced surveying technology with the traditional coordinate method. Detailed estimates of data collection activity were documented for a selected accident sample, and accident clearance times were documented for a much larger sample. The analysis indicated that the investigation of traffic accidents by using total-station surveys provides a substantial improvement over the traditional coordinate procedure. The number of measurements obtained at an accident scene increased by a factor of approximately 2 when the total-station equipment was used. The time to collect the data decreased by about 33 percent, and the man-hours decreased by about one-half. Computer plotting by the total-station procedure also resulted in a significant time savings. Decreased data collection time resulted in significant time and fuel savings to the driving public. Recommendations were made for continued use of the equipment and the purchase of more equipment when funds become available. It addition, policies for using the total-station equipment at all fatal and serious injury traffic accidents are recommended.

Serious traffic accidents traditionally require substantial time for on-scene investigation and, as a result, often seriously inconvenience and delay traffic. Electronic surveying equipment now provides an alternative to the traditional coordinate procedure. Instead of using a tape measure, digital surveying equipment is used to obtain detailed measurements, and the accident diagram is plotted by computer by using data automatically stored during the measurement process.

The study described here explores the use of advanced technology for obtaining and processing accident data. Potential benefits include decreasing the data collection time, increasing the accuracy of the data, reducing risks to investigators, and making analysis of the data easier. Moreover, decreasing the amount of time required to collect on-scene data reduces traffic delays, decreases fuel consumption and the emission of pollutants, and reduces the potential for secondary accidents. However, this could mean increased equipment costs, increased training requirements, weather-related delays, and limited availability of equipment and trained operators.

The traditional method of collecting field data at a traffic accident site is based on a coordinate procedure (*1*). Distance measurements are made with either a measuring tape or a wheel. By this procedure a reference line and a reference point on that line are identified, and measurements are made relative to those references. Two measurements are required to locate a spot. One is the distance (usually the shortest) from the spot to the reference line. The second is the distance from that location on the reference line to the reference point. In addition to distance, direction must be specified. The coordinate method typically requires a minimum of two people plus those necessary for traffic control. Although it can produce accurate results, a substantial amount of time is required to collect the data and manually prepare a scale drawing, especially if a large number of points is necessary.

An alternative to the coordinate procedure is computerized surveying equipment, or electronic total stations. The total station not only replaces the tapes and measuring wheels but also stores data digitally rather than in the traditional field book. The data are subsequently downloaded into a computer to produce detailed scale maps. This provides the possibility of a procedure that is more efficient than preparing drawings manually from the data collected by the coordinate method.

The objective of the present study was to evaluate the use of advanced surveying technology for the investigation of traffic accidents on the basis of the experiences of the Lexington–Fayette Urban County Police (LFUCP) and the Kentucky State Police (KSP). The only other known formal evaluation of total stations was performed in the state of Washington (*2*). The study compared the investigation of traffic accidents by using surveying technology with that by using the traditional coordinate method. Measures of merit included, among others, the time required for data collection, the amount of data collected, and the time required to prepare the accident diagram. The cost-effectiveness of total-station surveys was also assessed.

EQUIPMENT

The total station is a combination of an electronic distance meter, which uses infrared light to measure distance, and a digital theodolite. The instrument transmits an invisible infrared beam that reflects back from a prism placed at the measurement spot. The infrared light replaces the typical measuring tape or wheel. The theodolite measures the horizontal angle from the baseline and the vertical angle from the horizon. The investigator places the total station at a position where most or all of the measurement locations can be observed. The prism is mounted on a tall rod so that measurements can be taken over objects such as passing traffic.

The total station calculates three-dimensional coordinates for each point relative to a reference point. Distance, horizontal angle,

---

Kentucky Transportation Center, College of Engineering, University of Kentucky, Lexington, Ky. 40506-0281.

and vertical angle are obtained. A code is entered for each datum point, and the data are stored for later downloading into a computer. The computer interprets the field codes and prepares a map of the scene. Either a rough drawing made by using the data directly from the total station or a more detailed drawing made by using a drafting program can be obtained.

This type of equipment is available from several manufacturers. The Sokkia surveying system was used in this evaluation. A list of the necessary equipment and their costs is given in Table 1. The equipment included the surveying system as well as the computer software and plotters. Five total-station units were purchased: two units were used by LFUCP and three were used by KSP. The three KSP units were placed in separate posts located across the state. The total cost of the equipment in 1993 was approximately $100,000, about $20,000 per unit. Since this cost included plotters and other office equipment, the cost of purchasing only the surveying equipment was considerably less. A 1-week training class entitled Incident Management with Accident Reconstruction/Mapping Technology was included in the total cost.

DATA

Although the total-station procedure has many potential advantages, the one of greatest interest was the reduction in the amount of time necessary for on-scene data collection. Estimates of data collection time (by both coordinate and total-station procedures) were documented for a selected accident sample by using accident investigation work logs. In addition, accident clearance times were documented for a much larger sample of accidents to determine the extent to which total-station procedures would likely prove advantageous.

Field Investigation Information

An accident investigation work log was completed by both KSP and LFUCP after each accident investigation with the total-station equipment (Figure 1). For each accident general information relating to the road type, type of accident, numbers and types of vehicles involved, and accident severity was entered into the work log. Also entered into the log was information concerning the data collection process including the time to collect the data, the number of measurements, the number of officers and man-hours required to collect the data, and the road closure time. Finally, the time to prepare the accident diagram was recorded.

KSP investigated 32 accidents using total-station surveys. For seven of these data collection was duplicated by the coordinate procedure, and an accident investigation work log was filed for each of these coordinate surveys as well. LFUCP investigated 16 accidents using total-station surveys. Also, a review of LFUCP accident records of past investigations done by the coordinate procedure was performed.

Accident Scene Clearance Times

The time required for on-scene data collection was among the items reported on the accident investigation work log. Because the sample of accidents for which work logs were prepared was relatively small, independent estimates of the time to clear the accident scene were desired. Fortunately, the uniform accident report identifies both the time that the police arrive at an accident scene and the time that the scene is cleared, that is, the time that all traffic lanes are opened and traffic movement is unrestricted. It should be noted that the time that the scene is cleared does not necessarily coincide with termination of data collection activities. In fact, data collection typically concludes before the scene is officially cleared. However, these two times, the data collection time and the accident clearance time, helped to gauge the potential effects of any improvements associated with the use of total-station equipment.

By using a computerized file of traffic accidents for the time period of 1991 through 1993, summaries were obtained for both Lexington–Fayette Urban County and statewide. Average clearance times were obtained as a function of variables such as accident severity, highway type, vehicle type, number of vehicles, and type

TABLE 1 Total-Station Equipment and Cost

	Number		Cost[a]	
Equipment	KSP	LFUCP	KSP	LFUCP
Leitz Total Station and Supplies	3	2	$31,052	$20,702
Notebook Computer and Software	4	1	10,000	2,200
Various Surveying Supplies			6,370	4,140
Hewlett Packard Plotters	1	1	5,840	5,840
Digitizer Boards	1	1	1,676	1,676
Digitizer Table Stands	1	1	453	453
Computer Modem	1		1,163	
Map Software	4		3,702	
Autosketch Computer Program	2		350	
Laser Printer	1		725	
Computer Equipment	1	1	1,525	1,525
Printing Supplies			554	734
Total			$62,685	$37,995

[a] Purchase price in 1993.

ACCIDENT INVESTIGATION WORK LOG

DATE __________

LOCATION ____

CASE NUMBER _________________

GENERAL INFORMATION:

ROAD TYPE: Interstate ____ Two-Lane ____ Four-Lane ____
TYPE OF ACCIDENT: Collision ____ Non-Collision ____
NUMBER OF VEHICLES: ____
TYPE OF VEHICLE: Vehicle 1 ___________
Vehicle 2 ___________
Vehicle 3 ___________
Vehicle 4 ___________
SEVERITY: Property-damage-only ____ Injury ____ Fatal ____
ALCOHOL INVOLVED: Yes ____ No ____

FIELD INVESTIGATION:

TIME TO COLLECT MEASUREMENTS _______
TOTAL NUMBER OF MEASUREMENTS _______
ROADWAY-RELATED MEASUREMENTS _____
ACCIDENT-RELATED MEASUREMENTS _____
NUMBER OF OFFICERS COLLECTING DATA _____
MAN-HOURS TO TAKE MEASUREMENTS _____
TOTAL ROAD CLOSURE TIME ______
COMMENTS ____
____
____

ACCIDENT DIAGRAM:

TIME TO PREPARE DIAGRAM _____
MAN-HOURS TO PREPARE DIAGRAM _____
COMMENTS ____
____
____

FIGURE 1 Work log.

of accident. The number of accidents requiring excessive clearance times indicates the potential for the effective use of total-station equipment.

RESULTS

Field Investigation

In a 1-year time period the total-station procedure was used a total of 32 times by KSP and 16 times by LFUCP. The benefits of using the total-station procedure compared with the coordinate procedure are summarized in Table 2. The mean number of measurements per accident was approximately doubled for total-station surveys, whereas the time and man-hours required to collect the data decreased substantially. Data collection time, for example, decreased from an average of more than 3 hr to about 2 hr. The mean number of measurements per man-hour increased dramatically when the total-station equipment was used. Finally, KSP (but not LFUCP) reported a slight reduction in the average number of on-scene investigators with total-station surveys. However, for either coordinate or total-station surveys, the practical minimum number of investigators is two.

By using past LFUCP accident reports, additional data, summarized in Table 2, were available for 124 accidents in which the coordinate procedure was used. The times given for the coordinate method were the clearance times obtained from the LFUCP accident reports. The number of measurements was also recorded on the police report. A comparison showed that the measurements per man-hour increased by a factor of slightly more than 3 by the total-station procedure compared with those by the coordinate procedure.

At seven locations where KSP investigated an accident, both coordinate and total-station surveys were used. In these situations the scene was first measured by the coordinate procedure; measurements were repeated later with the total-station equipment. At these sites 3 times as many measurements were taken by the total-station procedure in approximately one-half the time.

TABLE 2 Comparison of Coordinate and Total-Station Data Collection Methods

	KSP		LFUCP	
Coordinate Method				
Number of Accidents	7[a]		124[b]	
Mean Number of Measurements	36.1	(22.1-50.2)[c]	47.9	(41.9-53.9)
Mean Data Collection Time (Min)	214	(153-275)	188	(175-201)
Mean Number of Man Hours	11.9	(7.9-15.8)	10.8	(9.5-12.0)
Mean Measurements per Man Hour	3.3	(2.1-4.5)	5.2	(4.6-5.8)
Mean Number of Investigators	3.3	(2.8-3.7)	3.3	(3.1-3.5)
Total Station Method				
Number of Accidents	32		16	
Mean Number of Measurements	83.6	(71.0-96.2)	96.9	(74.8-118.9)
Mean Data Collection Time (Min)	111	(89-134)	124	(93-154)
Mean Number of Man Hours	4.5	(3.6-5.4)	6.7	(5.0-8.5)
Mean Measurements per Man Hour	21.8	(18.2-25.4)	17.7	(12.1-23.2)
Mean Number of Investigators	2.5	(2.2-2.7)	3.2	(2.9-3.6)

[a] The KSP collected total-station data at the seven locations where data were collected using the coordinate method. At these seven locations, an average of 109 measurements were collected in an average of 4.1 man hours giving an average of 26.6 measurements per man hour. Data collection times and number of investigators averaged 100 minutes and 2.4 people, respectively.
[b] These data are based on information given on the police accident report.
[c] Numbers in parentheses represent the 95-percent confidence interval for the mean.

The differences in the means between coordinate and total-station surveys in Table 2 were analyzed by Student's *t*-test to ascertain if they were statistically significant at a significance level of 5 percent. Separate analyses were conducted for the KSP data (both 7 by 32 and 7 by 7) and the LFUCP data (124 by 16). For the 7-by-7 KSP data, paired *t*-tests were also performed. With one exception, that is, the mean number of investigators used by LFUCP, all differences between total-station and coordinate surveys were found to be statistically significant at a 5 percent or better significance level.

Factors that might affect data collection at an accident site, including accident severity, road type, and numbers and types of vehicles, were also summarized. The results of this summary are given in Table 3. The data for coordinate surveys were limited to those obtained by LFUCP because of the rather large sample of data. Data from both KSP and LFUCP were used to examine the total-station procedure.

The average accident clearance times for the coordinate surveys were largest for Interstate highways, for accidents involving more than three vehicles, and for accidents involving trucks. Also, the average time to clear fatal accident sites was substantially longer than that to clear injury accident sites.

Similar trends for road type, number of vehicles, and vehicle type were found for total-station surveys, but the man-hours required to collect the data were dramatically reduced. The data show that total-station equipment was used primarily in the investigation of fatal accidents. The time to investigate a small number of property-damage-only accidents was substantially less than that to investigate injury or fatal accidents.

Discussions were held with officers experienced with total-station surveys to obtain their opinions and observations. All comments were positive: the consensus was that total stations allowed more accurate and more detailed measurements in a shorter period of time. Although a short training period was required, total stations were not found to be difficult to operate. One drawback was that total stations could not be used during certain weather conditions. However, in these instances points could often be marked with paint at the time of the accident, and measurements could be delayed until the weather was more favorable.

Accident Diagram

Measurements taken at the accident scene are used to prepare an accident diagram. Although only a rough diagram of the accident scene is prepared in many cases, it is sometimes necessary or desirable to prepare a scaled accident diagram. A limited amount of information was available from the police agencies to compare the time required to prepare a scaled drawing of the accident scene by using data from either the coordinate or the surveying procedures.

This information showed that the time savings associated with preparing the accident diagram by the total-station procedure compared with that by the coordinate procedure was of a magnitude similar to that for data collection. Discussions with the police officers indicated that the time to prepare detailed diagrams was decreased by at least 50 percent by the total-station procedure. The total-station procedure also provided a more automated method with substantially more datum points. This resulted in the preparation of a more accurate diagram in a shorter time period.

Benefit–Cost Analysis

A comparison was made between the cost of purchasing total-station equipment and the primary benefit thought to be associated

TABLE 3 Variables Affecting Data Collection

	Number of Accidents[a]		Average Number of Measurements		Average Time (Min)		Average Man-Hours	
Variable	Coord[b]	TStation	Coord	TStation	Coord	TStation	Coord	TStation
Road Type								
Two-Lane	69	30	48	85	190	110	10.4	4.5
Four-Lane	42	11	45	113	180	130	9.9	7.1
Interstate	8	6	72	65	270	145	21.0	6.8
Number of Vehicles								
One	66	7	40	87	175	95	8.9	5.0
Two	46	29	55	87	200	120	12.0	5.2
Three	7	8	68	85	220	110	13.8	4.3
Over Three	5	3	65	125	245	160	20.7	9.1
Vehicle Type								
Cars Only	117	38	46	88	180	110	9.9	5.1
Trucks Involved	7	8	78	85	295	138	24.8	5.8
Severity								
Fatality	62	32	54	94	205	125	12.8	5.5
Injury	60	8	41	100	170	115	8.6	6.4
Property Damage		7		57		80		3.2
All	124	38	48	88	185	120	10.8	5.2

[a] Inconsistent subtotals indicate missing information.
[b] Excludes two property-damage-only accidents.

with its use. Although the police agency benefits by the reduction in man-hours required to investigate a traffic accident and prepare an accident diagram, the major benefits are realized by motorists in terms of reduced delay (time cost) and fuel consumption (fuel cost). To obtain a first-order approximation of benefits, therefore, the focus was placed on travel-time delay and fuel savings.

A review of the literature was conducted to determine the appropriate costs of time and fuel. Time costs ranging from $6.25 to $7.00 for each vehicle hour of delay have been reported (*2–4*). Fuel cost was estimated by assuming that each vehicle hour of delay uses about 1.1 gallons of fuel (*4*), which costs about $1.00 per gallon. A cost of $8.00 per vehicle hour of delay was therefore used in the analysis as the estimate of the combined cost of time and fuel.

The cost of the field equipment necessary for a total-station survey was determined to be about $15,000 per unit. An additional cost of about $10,000 would be necessary for office equipment, but the office equipment would generally support multiple field units. Nevertheless, given a maximum cost of $25,000 to implement a total-station procedure and time and fuel savings valued at $8.00 per vehicle hour of delay saved, a reduction of about 3,125 vehicle hours in delay would be necessary to offset the original equipment costs.

The potential vehicle hours saved must be estimated to compute the associated delay and fuel savings. Kentucky experience, summarized earlier, suggests that the time required to investigate severe accidents was reduced by about 60 min when using total-station surveys. The typical investigation took about 180 min by the coordinate method, compared with about 120 min by the total-station procedure.

Various scenarios were used to estimate the reduction in vehicle delay and fuel costs associated with a reduction in data collection time from 180 to 120 min. During some accident investigations the road is blocked for a portion of the time, whereas in other cases traffic flow continues but at a slower rate and speed. In all cases the travel time is increased during the accident investigation. The analysis assumed either a complete blockage, a partial blockage, or an intermittent blockage. A graphical description of the assumed departures during the investigation is provided in Figure 2. The stopped delay (in vehicle hours) is represented by the shaded areas in Figure 2.

For the complete blockage scenario the road was assumed to be blocked for the entire data collection time. For the partial blockage scenario traffic was assumed to continue to flow during the investigation, but at a reduced rate. For the intermittent blockage scenario traffic was assumed to be blocked for a portion of the blockage time and was allowed to flow at a reduced rate during the remainder of the blockage time.

Estimates were made for a four-lane freeway (one direction of travel) and for a two-lane roadway. Regardless of the scenario, it was assumed that the departure rate following completion of the accident investigation was 3,600 vehicles per hour (v/hr) in one direction for the freeway and 1,800 v/hr in both directions for the two-lane, two-way roadway.

Only stopped time delay was considered. The additional delay associated with stopping and starting and with any reduced speed in approaching the accident zone and departure from the queue was not included. Moreover, the calculations did not differentiate between the obstructed flow rates by the coordinate and total-station procedures. Since the total-station procedure is less disruptive to

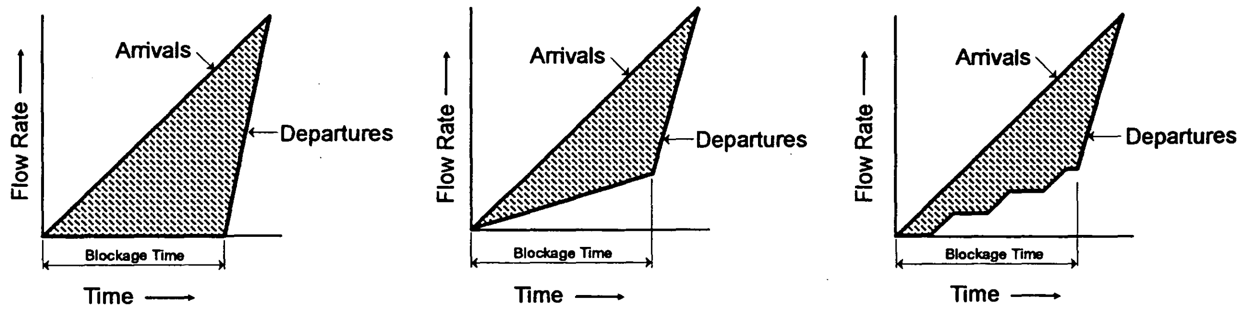


FIGURE 2 Assumed departures for complete (*left*), partial (*middle*), or intermittent (*right*) blockages.

traffic, an additional savings in delay would likely be gained from this factor. Thus, the benefits attributed here to total-station surveys are considered to be conservative.

The two-lane scenario involved the complete closure of one lane, with the second lane being used to accommodate one-way travel in alternating directions. The assumption was made that 3 min was alternately allocated to each direction, followed by a 30-sec dead interval. This yielded a 7-min cycle time for use of the open lane.

Accidents occur on all types of roads under a variety of traffic volume. The average annual daily traffic (ADT) on state-maintained, rural, two-lane roads in Kentucky is about 1,500 vehicles per day (vpd), compared with about 6,900 vpd in urban areas (*5*). The average ADT on four-lane divided (non-Interstate or parkway) roadways varies from 8,800 vpd in rural areas to 21,500 vpd in urban areas. Average ADTs on rural Interstates are 21,400 vpd, compared with 53,700 vpd on urban Interstates. On the basis of these average ADTs, hourly volumes on the majority of Kentucky roadways are typically less than 1,000 v/hr on two-lane highways and 2,000 v/hr (one direction) on four-lane freeways. These hourly maxima defined the upper limits of the hourly volumes investigated here.

The vehicle hours of delay associated with an accident investigation can be estimated by using the length of a typical investigation for both the coordinate and the total-station procedures, the type of traffic interruption created by the investigation, and the traffic volume during the investigation period. The macroscopic analysis assumed that flow was deterministic and at steady state. Results of the analysis, presented in terms of the savings in stopped-time delay by reducing on-scene data collection time from 180 to 120 min, are summarized in Figures 3 and 4.

The freeway analysis, presented in Figure 3, shows the reduction in vehicle hours of delay assuming complete blockage as well as partial blockage (obstructed one-way flow rates of 500, 1,000, and 1,500 v/hr). For example, if an approach flow rate of 1,000 v/hr is reduced to 500 v/hr during the investigation, approximately 1,500 vehicle hours of delay can be eliminated by reducing the investigation time from 180 to 120 min.

The two-lane, two-way road analysis, presented in Figure 4, examines complete blockage as well as intermittent blockage (one-way flow rates of 250, 500, and 750 v/hr). For example, if an approach two-way flow rate of 500 v/hr is reduced to an intermittent, obstructed one-way flow rate of 250 v/hr during the investigation, a savings of about 800 vehicle hours of delay can be realized.

By using time and fuel costs of $8.00 per vehicle hour of delay and the reduction in stopped delay shown in Figures 3 and 4, the estimated dollar savings from use of the total-station procedure can be estimated. When this savings is compared with the cost of $25,000 for total-station equipment and training, the breakeven point requires a savings of 3,125 hours of delay.

During high-volume periods the savings associated with total-station surveys pays for the equipment almost immediately: only one or two accident investigations are required. Even for low-volume conditions, the savings would pay for the equipment after as few as about 10 accident investigations. Given the potential use of the equipment, these savings can likely be realized in no more than a few months and possibly within a few weeks (depending on the types and locations of accidents investigated).

This analysis understates the dollar benefits of the total-station procedure because it neglects benefits resulting from reduced police manpower requirements and from improved accuracy in documenting and reconstructing traffic accidents. Moreover, traveler benefits can be quite substantial even with time savings considerably less than the 1 hr that has been assumed, particularly when traffic volumes are large. For example, traveler benefits exceed total-station purchase costs after only three accident investigations that reduce 60-min blockages by 5 min on four-lane freeways carrying approximately 3,000 v/hr in the blocked direction.

Accident Clearance Times

The time necessary to clear an accident scene can be determined by using information given on the police accident report and recorded in the statewide accident data base. An analysis was performed to determine the averages and distributions of clearance times in Lexington–Fayette Urban County as well as statewide during the period from 1991 through 1993. Average clearance times were also compared with several contributing factors.

The number of accidents with long clearance times gives an indication of the potential frequency of highly productive applications of the survey equipment. However, the critical factor affecting clearance times is not always the data collection time. Cleaning up spills and removing vehicles are examples of other reasons for long clearance times. In those instances reducing the investigation time is not expected to significantly reduce the clearance time.

The average clearance time for Lexington–Fayette Urban County accidents was 70 min. This compares with an average of 51 min statewide. The distribution of clearance times is given in Table 4. In Lexington–Fayette Urban County almost one-half of the clearance times were 15 min or less and approximately two-thirds were 30 min or less. Statewide statistics also show that about two-thirds of the clearance times were 30 min or less.

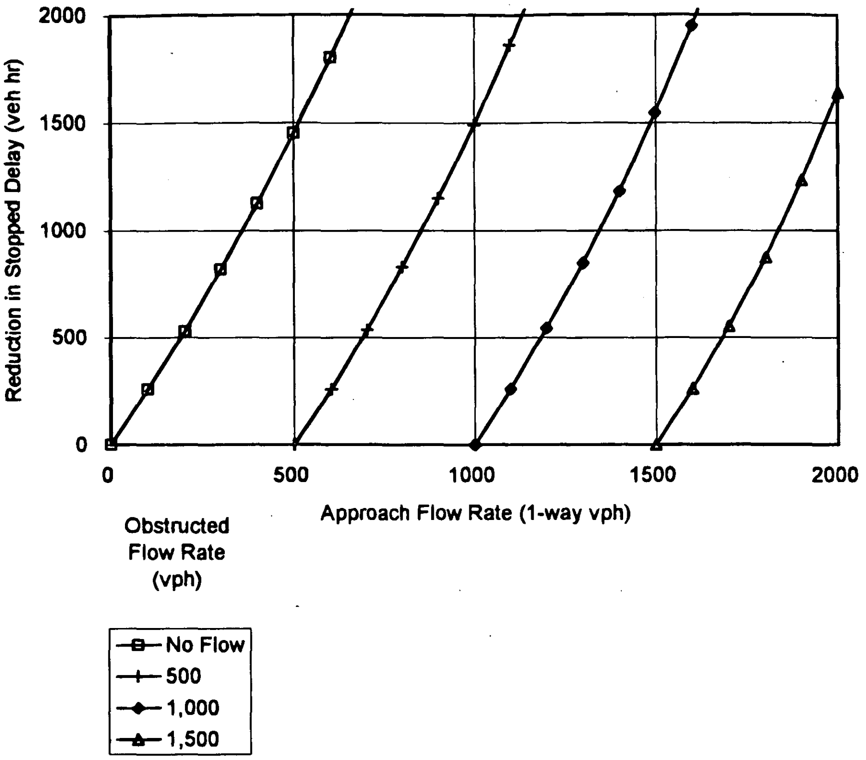


FIGURE 3 Reduction in stopped delay on a freeway associated with use of total-station equipment (freeway obstruction: 2-hr instead of 3-hr blockage).

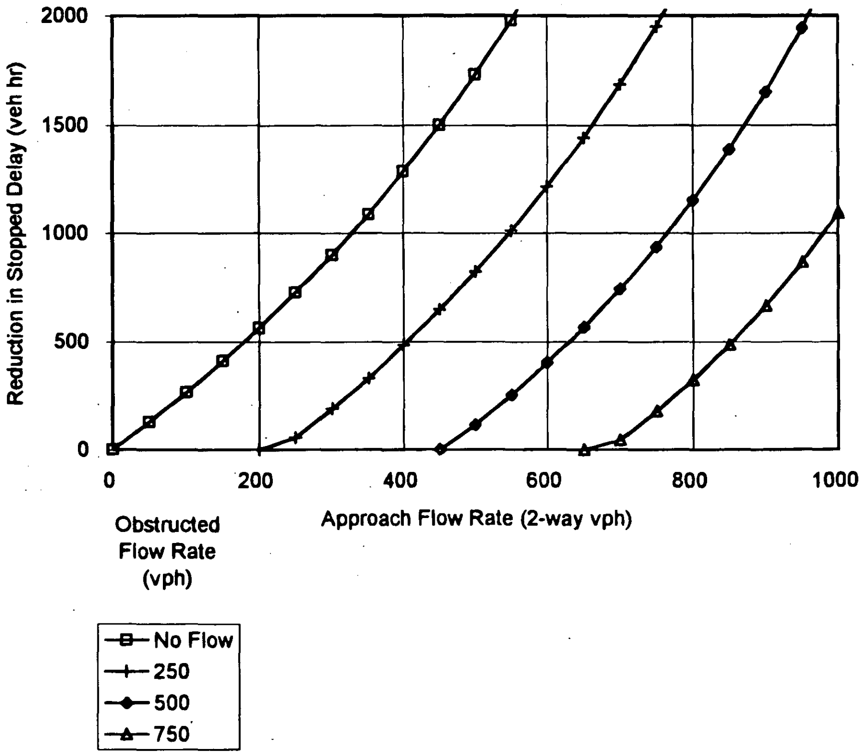


FIGURE 4 Reduction in stopped delay on a two-lane highway associated with use of total-station equipment (two-lane, two-way, road obstruction: 2-hr instead 3-hr blockage).

TABLE 4 Distribution of Accident Clearance Times (1991–1993)

Time Range (Min)	Percent in Time Range	
	Lexington-Fayette Urban County	Statewide
15 or less	46.3	35.2
16 - 30	21.2	32.4
31 - 45	13.3	15.5
46 - 60	6.2	6.6
61 - 90	3.5	4.2
91 - 120	0.9	1.2
121 - 180	0.8	0.8
181 - 240	0.5	0.4
241 - 300	0.4	0.2
Over 300	7.0	3.5

About 8 percent of all accidents in Lexington–Fayette Urban County had clearance times longer than 180 min, compared with about 4 percent statewide. These are typical of the types of accidents for which total-station equipment can be most beneficial. In 1 year this represents about 950 accidents in Lexington–Fayette Urban County and about 5,500 accidents statewide. Thus the number of accidents for which total-station equipment could likely be used in a highly effective manner is substantial.

Analysis of each of Kentucky's 120 counties found a large range in average accident clearance times, with 11 counties having average times longer than 120 min and 3 counties with averages longer than 180 min; 5 counties had average times shorter than 30 min. Despite large variations among the counties, a significant number of long-clearance-time accidents occur in almost all counties. During the 3-year analysis period, for example, all counties experienced at least 10 accidents for which the clearance time was 120 min or longer (Figure 5). Thirty counties experienced at least 100 accidents for which the clearance time was much longer (300 min or more). Although not all of these long-clearance-time accidents occur when traffic is heavy, the potential for cost-effective use of total-station equipment during accident investigations appears to be distributed rather widely throughout the state.

As indicated in Table 5, clearance times are related to several accident characteristics. It appears that accidents for which the savings resulting from total-station surveys are likely to be greatest include fatal accidents, nighttime accidents, truck accidents, single-vehicle accidents, and Interstate or parkway accidents. Road user savings, such as reduced delays and reduced fuel consumption, are related directly to traffic volumes, and as a result, high-volume roads and streets are priority locations for expedited accident investigations.

CONSISTENCY OF FINDINGS

Researchers in the state of Washington (*2*) have documented an investigation of the benefits of total-station surveys for accident investigation similar to the present one. One facet of their investi-

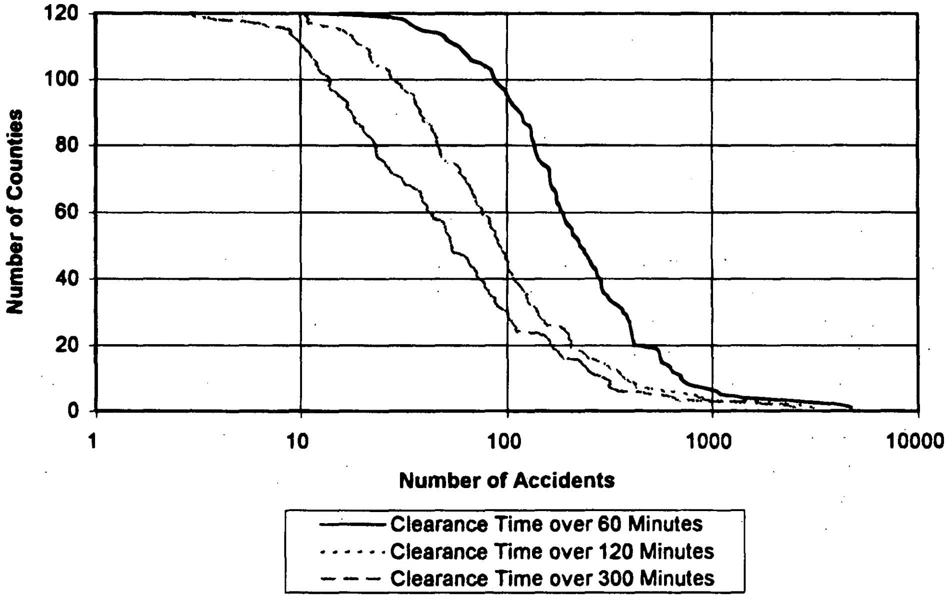


FIGURE 5 Cumulative distribution of county accidents by clearance time.

TABLE 5 Influence of Accident Characteristics on Clearance Time

		Average Clearance Time (Min)	
Variable	Category	Lexington-Fayette Urban County	Statewide
Severity	Fatal	188	120
	Injury	54	57
	Property Damage Only	69	47
Time of Day	Midnight - 5:59 am	117	91
	6:00 am - 11:59 am	78	59
	Noon - 5:59 pm	59	42
	6:00 pm - 11:59 pm	52	42
Highway Type	Interstate/Parkway	91	77
	Other State Maintained	56	50
	Local/County	72	45
Light	Daylight	62	47
	Dawn/Dusk	58	52
	Darkness	73	58
Vehicle Type	Truck Involved	79	70
	No Truck Involved	65	48
Number Vehicles	One	83	74
	Two	65	42
	Three	44	41
	Four or More	48	53
Type Accident	Other Vehicle	63	42
	Fixed Object	87	76
	Non-Collision	93	83
	Pedestrian or Bicycle	66	50
Location	Intersection	53	37
	Non-intersection	66	57
Land Use	Rural	79	75
	Business	55	37
	Residential	60	43
	Limited Access	76	68

gation was a detailed examination of three accident locations, all on two-lane highways, for which data were collected by both the coordinate method and the total-station method. The total-station surveys produced an average of about 70 percent more measurements per hour and reduced the average investigation time from 130 to 60 min. Table 6 compares the Washington and Kentucky summaries. Although the differences between states were large, the total-station surveys were always superior: they produced at least 21 more measurements each hour and reduced data collection times by a minimum of 60 min.

A second facet of the Washington investigation was a before-and-after study of clearance times of urban freeway accidents in Seattle. In the before period 20 accidents had been investigated by using coordinate surveys. In the after period 15 similar accidents had been investigated by using total-station surveys. As a result of the use of total stations the average clearance time was reduced from 182 to 131 min, a statistically significant average time savings of 51 min. Although similar before-and-after data are not available for Kentucky, the 51-min savings is compatible with estimates obtained from other types of comparisons in Kentucky.

A third major aspect of the Washington investigation was a benefit–cost analysis of the investigation of a single, peak-period accident. Peak-period volumes on the five-lane facility were approximately 9,090 v/hr. It was assumed that two lanes were blocked and that a total-station survey reduced the blockage time from 182 to 131 min. Under these conditions the savings attributed to the use of the total station were estimated to include 7,000 vehicle hours of delay. The researchers concluded that the approximately $15,000 cost for the total-station equipment would be more than recovered after just one use in investigating a major, peak-period accident on an urban freeway. Analyses in Kentucky not only support this conclusion but also suggest a favorable benefit–cost ratio after total-station equipment is used to investigate just a few severe accidents on less heavily trafficked facilities.

Finally, according to estimates by the Washington State Patrol, the automated drafting used with the total-station surveys reduced

TABLE 6 Comparison of Kentucky and Washington Data

	Washington	KSP	LFUCP
Measurements per Hour			
Coordinate Method	28.8	10.0	15.5
Total Station Method	49.8	44.2	46.1
Percent Increase	73	342	197
Average Investigation Time (Min)			
Coordinate Method	130	216	186
Total Station Method	60	114	126
Percent Decrease	54	50	32

the time for preparing drawings of each accident site from about 8 hr to about 2 hr. Investigators in Kentucky reported a more modest improvement of about 50 percent.

CONCLUSIONS

The analysis shows that the investigation of traffic accidents with total-station (survey) equipment provides a substantial improvement over investigation by the traditional coordinate procedure. The number of measurements obtained at an accident scene increased (by a factor of about 2) when total-station equipment was used. The time to collect the data decreased by about 33 percent, and the number of man-hours decreased by about one-half. The increase in the number of measurements results in a more accurate and detailed investigation and accident diagram. The use of computer plotting in the total-station procedure results in a significant time savings when a detailed accident diagram is needed.

More important, the decreased time required to collect field data results in significant time and fuel savings to the driving public. Estimates of the savings in delay demonstrated that the total-station equipment would result in savings that would pay for its cost after only a few investigations. This indicates that the use of this type of equipment for accident investigation is economically justified.

An analysis of clearance times at accident sites showed that there was potential for substantial use of total-station equipment throughout the state of Kentucky.

RECOMMENDATIONS

The analysis results in the recommendation that the use of total-station equipment be expanded as funds become available for equipment purchase. Specifically, each KSP post should have this equipment, as should other police departments with officers with advanced training in accident investigation. Proper training must be provided to ensure that the equipment is used properly. A policy should be established that use of the total-station equipment should be considered at all fatal and serious injury traffic accidents to ensure that optimum use is made of the equipment.

REFERENCES

1. Agent, K. R., and J. G. Pigman. *Traffic Accident Investigation.* Report KTC-93-10. Transportation Center, University of Kentucky, 1993.
2. Jacobson, L. N., B. Legg, and A. J. O'Brien. Incident Management Using Total Stations. In *Transportation Research Record 1376*, TRB, National Research Council, Washington, D.C., 1992, pp. 64–70.
3. *Traffic Congestion: Trends, Measures, and Effects.* Report GAO/PEMD-90-1. U.S. General Accounting Office, Nov. 1989.
4. Lindley, J. A. Urban Freeway Congestion Problems and Solutions: An Update. *ITE Journal*, Vol. 59, No. 12, Dec. 1989, pp. 21–23.
5. Agent, K. R., and J. G. Pigman. *Analysis of Traffic Accident Data in Kentucky (1988–1992).* Report KTC-93-23. Transportation Center, University of Kentucky, Sept. 1993.

The contents of this report reflect the views of the authors, who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the University of Kentucky or the Lexington-Fayette Urban County Government. This report does not constitute a standard, specification, or regulation. The inclusion of manufacturer names and trade names is for identification purposes and is not considered an endorsement.

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

Microscopic Accident Potential Models for Two-Lane Rural Roads

BHAGWANT N. PERSAUD AND KORNEL MUCSI

Fundamental to the present research is the use of hourly traffic volumes in regression models for estimating accident potential on two-lane rural roads. By using data from Ontario, Canada, a simple model form, and a regression package that allows the assumption of a negative binomial error structure, regression models were calibrated for the different combinations of time periods (24 hr, day hours, and night hours) and geometric (roadway and shoulder width) characteristics. It is shown that the effect of day/night conditions is different for single-vehicle and multivehicle accidents. For single-vehicle accidents the accident potential is higher during the night, whereas for multivehicle accidents the opposite is true. This indicates the importance of differentiating between single-vehicle and multivehicle accidents and day/night conditions. The refinement of the regression predictions by the empirical Bayesian (EB) estimation procedure for individual road sections is illustrated. It is shown through a validation exercise that the EB procedure provides better estimates of accident potential than the conventional method only on the basis of the short-term accident count for a section.

Estimation of the accident potentials of road sections usually requires the use of a relationship between accidents and a measure of traffic volume, traditionally, the average daily traffic (ADT). If this relationship is nonlinear, ADT-based models would be unsuitable for use in estimating safety during portions of a day, for example, specific hours, peak periods, and nighttime. Such estimates might be required to evaluate strategies that affect traffic volumes or safety during certain parts of the day and to identify potentially hazardous traffic operating conditions. The fundamental premise of the research on which this paper is based is that for these estimates it is preferable to use microscopic models with hourly volumes as the measure of traffic intensity. The increasing availability of hourly traffic volume data, taken together with recent advances in accident modeling, prompted a fresh look at developing models for estimating the accident potential of two-lane rural roads. These constitute a substantial portion of the North American road network.

The procedure for developing these microscopic models follows on that used in other recent work (*1,2*) and differs from the early attempts at accident modeling in two important aspects. First, like the early work, regression models are developed to relate accident occurrence and traffic volume, but the models are developed by using a generalized linear modeling package (GLIM) (*3*) that allows errors in accident counts to be more properly described by the negative binomial distribution rather than the traditional normal distribution assumed in conventional regression packages. Second, the procedure allows for the refinement of the regression estimate of accident potential by using an empirical Bayesian (EB) procedure.

The remainder of the paper presents the results of research aimed at estimating microscopic accident prediction models for two-lane rural roads in Ontario, Canada. A summary of the theory, data, results, and validation along with a detailed example application are presented. Every attempt has been made to minimize repetition from related earlier publications, but a certain amount has been unavoidable in the interest of making this a paper that can reasonably stand alone.

THEORETICAL FOUNDATIONS

The fundamental estimator for $E(m)$, the expected number of accidents during T hours on a section of length L km, is given by Equation 1, where F represents the traffic volume, and a and b are parameters to be estimated in a regression model.

$$E(m) = a\,L\,T\,F^{b} \qquad (1)$$

Reviews by Satterthwaite (*4*) and Hauer (*5*) indicate that when geometric and environmental conditions are appropriately controlled, this model form is quite common. Other reasons for its selection are its simplicity, the parsimony of the independent parameters, and the correspondence with logic that the predicted number of accidents would be zero for a traffic volume of zero. In a special case in which b is equal to 1, a linear relationship is indicated, but this need not be assumed a priori as has been done in several studies.

Regarding the error distribution for accident counts, several studies (*2,6*) have indicated that it is more appropriate to describe the accident count in a population of entities with the negative binomial distribution than with the Poisson or normal distributions.

Following recent work by Persaud (*1*) and others the generalized linear interactive modeling (GLIM) software package (*3*) was applied for the parameter estimation. GLIM allows the specification of different error structures, including the negative binomial, and it uses an algorithm for the parameter estimation by the method of maximum likelihood.

According to the theory the variance of m is related to $E(m)$ as follows:

$$\mathrm{VAR}(m) = E(m)^2/k \qquad (2)$$

where k can be estimated by using a maximum likelihood procedure that assumes that each squared residual of the regression model is an estimate of $\mathrm{VAR}(m)$ and that each count comes from a negative binomial distribution with mean $E(m)$ and variance given by Equation 2.

$E(m)$ is an average over sites that are similar in the values of their independent variables, and if the variance of the m values is large,

---

B. N. Persaud, Department of Civil Engineering, Ryerson Polytechnic University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada. K. Mucsi, Department of Civil Engineering, University of Toronto, Ontario M5S 1A4, Canada.

TABLE 1 Parameter Estimates for Level 1 Models

	Model Group			Model Parameters			Sample Size
Code	Acc. Type	Severity Group	Time Period	ln(a) (std. error)	b (std. error)	k	Number of Accidents
101	S.V.	F.& I.	24 Hr	-12.97 (0.03)	0.430 (0.006)	1.5	2340
102	M.V.	F.& I.	24 Hr	-16.88 (0.05)	1.137 (0.009)	0.7	1677
103	All	F.& I.	24 Hr	-13.59 (0.03)	0.674 (0.007)	1.3	4018
104	S.V.	F.& I.	Day	-13.49 (0.05)	0.502 (0.010)	2.1	1124
105	M.V.	F.& I.	Day	-17.12 (0.69)	1.180 (0.012)	0.7	1109
106	All	F.& I.	Day	-14.30 (0.05)	0.793 (0.010)	1.2	2234
107	S.V.	F.& I.	Night	-13.02 (0.04)	0.491 (0.009)	1.1	1028
108	M.V.	F.& I.	Night	-16.61 (0.06)	1.080 (0.011)	1.1	430
109	All	F.& I.	Night	-13.33 (0.04)	0.643 (0.009)	1.5	1459
110	S.V.	Total	24 Hr	-11.80 (0.03)	0.444 (0.005)	2.2	8033
111	M.V.	Total	24 Hr	-15.93 (0.04)	1.123 (0.008)	0.8	3992
112	All	Total	24 Hr	-12.26 (0.03)	0.627 (0.005)	1.8	12026
113	S.V.	Total	Day	-12.30 (0.04)	0.490 (0.008)	1.7	3470
114	M.V.	Total	Day	-16.21 (0.06)	1.173 (0.011)	0.8	2663
115	All	Total	Day	-13.02 (0.04)	0.741 (0.008)	1.5	6134
116	S.V.	Total	Night	-11.97 (0.03)	0.557 (0.007)	3.6	3878
117	M.V.	Total	Night	-15.70 (0.05)	1.071 (0.010)	0.7	1007
118	All	Total	Night	-12.14 (0.03)	0.650 (0.007)	2.7	4886

the value of $E(m)$ by itself is not very useful, particularly if it is applied to a specific site. Therefore, VAR(m) must also be estimated and used in the refinement of estimates for a specific site. For such a case and in general, a refined estimate of accident potential can be obtained from an EB procedure used by others in recent years (*1,2*). This procedure combines the regression estimate, $E(m)$, of sites similar in all independent variables and the short-term accident count (x) of the site.

For reasonable assumptions on the distributions of x and m (*7*), the EB estimate of accident potential is

$$E(m|x) = wE(m) + (1 - w)x \tag{3}$$

where

$$w = [1 + \mathrm{VAR}(m)/E(m)]^{-1} = [1 + E(m)/k]^{-1} \tag{4}$$

It can also be shown (*2,7*) that the variation in (m/x) can be estimated by

$$\mathrm{VAR}(m|x) = (x + k)/[1 + (k/E(m)]^2 \tag{5}$$

As discussed in some of the earlier work, a large value of k is consistent with a sound regression estimate, which is given a relatively large weight w. Conversely, for a small k, indicated by a regression estimate with a large variance, substantial weight is given to the accident count.

DATA

Three groups of data pertaining to traffic flow, geometry, and accidents were obtained from the Ministry of Transportation, Ontario (MTO), for each of 2,014 two-lane rural road sections. Hourly traffic volumes on these sections were estimated on the basis of the Annual ADT and the seasonal and hourly variation factors at permanent counting stations (PCSs). It was assumed that the variations in traffic volume (both seasonal and hourly) at a section were similar to those at the associated PCS. The accident data file provided information on all reported accidents on two-lane roads in Ontario for 1988 and 1989. Each accident record was associated with a unique road section and was characterized by the date and time of occurrence, class, severity, number of vehicles involved, and other relevant information. The raw data were used to assemble a regression data set stratified by section, hour, weekday/weekend, season, and day/night condition.

MODEL CALIBRATION AND ANALYSIS

Two levels of models were developed. In Level 1 models all two-lane roads are placed in the same category; Level 2 models

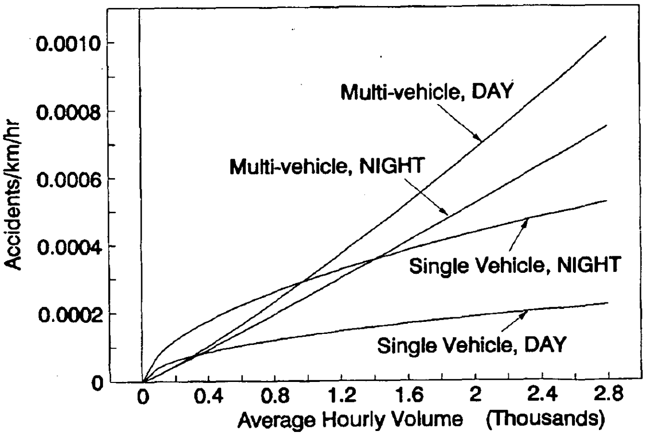


FIGURE 1 Plots of selected Level 1 models (all severity levels combined).

were calibrated separately for roads grouped according to geometric and other factors that may be associated with accident causation.

Level 1 Models

The availability of estimated hourly volumes for each road section made it possible to estimate models for various time periods. Three different periods were identified: 24 hr, daytime hours only, and nighttime hours only. For each period the models were calibrated separately for groupings on the basis of severity and whether the accident involved a single vehicle or multiple vehicles. The values of k and the model parameters are given in Table 1, and the graphs of selected models are depicted in Figure 1. Note that GLIM uses the linear form of Equation 1, resulting in estimates of $\ln(a)$ rather than a and that the model, as calibrated, is for estimating the number of accidents per hour per kilometer.

Examination of the curves in Figure 1 and the model parameters in Table 1 reveals two main points of interest. First, the relationship for single-vehicle accidents is described with a convex curve ($b < 1$), whereas that for multivehicle accidents is characterized by a concave curve ($b > 1$). Second, it appears that the potential for single-vehicle accidents is higher during nighttime hours, whereas for multivehicle accidents the opposite is true. These observations serve to emphasize the importance of analyzing single-vehicle and multivehicle accidents separately and also the usefulness of differentiating between nighttime and daytime accidents.

Level 2 Models

As indicated earlier the Level 1 regression models apply for all two-lane rural roads in the province of Ontario. It is natural that within this group there are sections with different geometric characteristics. Level 2 models are the result of an attempt to create more homogeneous groups of road sections. The importance of second-level models is twofold. First, models built for subgroups provide better predictions for member sections; second, comparison of the accident potentials on sections with different geometric features is facilitated.

To facilitate the development of the Level 2 models, subgroups of road sections with similar lane and shoulder width combinations were formed after exploratory analysis revealed that these features were important in explaining differences in accident occurrence on sections with similar traffic volumes. Tables 2 and 3 contain the estimated Level 2 model parameters and estimates of the coefficient k for four different geometric groups, whereas Figures 2 and 3 are plots of selected models.

In analyzing these results it is important to stress that accident prediction models do not necessarily explain accident causation; they only represent an estimate of the accident potential by taking into consideration factors associated with accident occurrence. Nevertheless, the following points are worth noting:

1. For the 24-hr period the highest accident potential for both single-vehicle and multivehicle accidents is on roads with total lane widths of 6.7 m and shoulders with widths of 2.4 m (narrow lanes and wide shoulders).
2. The lowest potential for single-vehicle accidents is on roads with total lane widths of 7.3 m and shoulders with widths of 3.0 m (wide lanes, wide shoulders).
3. By contrast, the lowest potential for multivehicle accidents is on roads with total lane widths of 6.7 m and shoulders with widths of 1.8 m (narrow lane, narrow shoulders).
4. Similar to what has been observed for Level 1 models, parameter b for multivehicle accidents is almost always greater than 1, indicating a convex relationship between accidents and traffic volume, whereas for single-vehicle accidents it is always less than 1.
5. By comparing the accident potential for day and night hours on a homogeneous group of sections in terms of lane and shoulder widths, it can be observed that the accident potential for single-vehicle accidents is higher during nighttime hours (Figure 3), whereas for multivehicle accidents the accident potential is higher during daytime hours.

TABLE 2 Parameter Estimates for Level 2 Models (Fatal and Injury Accidents)

	Model Group					Model Parameters			Sample Size	
Code	Lane Width	Shldr. Width	Acc. Type	Severity Group	Time Period	ln(a) (std. error)	b (std. error)	k	# of Accs.	Total km
201	6.7	1.8	S.V.	F.& I.	24 Hr	-12.40 (.12)	0.291 (.027)	2.0	225	1899
202	6.7	1.8	M.V.	F.& I.	24 Hr	-17.36 (.15)	1.212 (.029)	0.8	116	1899
203	6.7	1.8	All	F.& I.	24 Hr	-13.29 (.12)	0.589 (.027)	4.0	342	1899
204	6.7	2.4	S.V.	F.& I.	24 Hr	-13.30 (.12)	0.491 (.026)	1.8	274	1953
205	6.7	2.4	M.V.	F.& I.	24 Hr	-17.85 (.17)	1.321 (.033)	1.1	201	1953
206	6.7	2.4	All	F.& I.	24 Hr	-14.29 (.13)	0.786 (.026)	1.9	476	1953
207	7.3	2.4	S.V.	F.& I.	24 Hr	-11.63 (.14)	0.214 (.027)	2.1	376	1694
208	7.3	2.4	M.V.	F.& I.	24 Hr	-17.47 (.21)	1.236 (.036)	1.5	259	1694
209	7.3	2.4	All	F.& I.	24 Hr	-12.92 (.15)	0.568 (.027)	1.7	636	1694
210	7.3	3.0	S.V.	F.& I.	24 Hr	-11.75 (.13)	0.201 (.023)	0.9	330	1649
211	7.3	3.0	M.V.	F.& I.	24 Hr	-15.72 (.16)	0.956 (.027)	0.8	441	1649
212	7.3	3.0	All	F.& I.	24 Hr	-13.06 (.13)	0.592 (.023)	1.3	772	1649

TABLE 3 Parameter Estimates for Level 2 Models (All Severity Groups Combined)

	Model Group					Model Parameters			Sample Size	
Code	Lane Width	Shld. Width	Acc. Type	Severity Group	Time Period	ln(a) (std. error)	b (std. error)	k	# of Accs.	Total km
301	6.7	1.8	S.V.	Total	24 Hr	-11.27 (.09)	0.342 (.022)	1.6	861	1899
302	6.7	1.8	M.V.	Total	24 Hr	-15.95 (.14)	1.101 (.027)	1.2	271	1899
303	6.7	1.8	All	Total	24 Hr	-11.72 (.10)	0.509 (.022)	2.0	1133	1899
304	6.7	2.4	S.V.	Total	24 Hr	-11.86 (.11)	0.448 (.022)	4.9	932	1953
305	6.7	2.4	M.V.	Total	24 Hr	-16.92 (.16)	1.323 (.029)	1.0	515	1953
306	6.7	2.4	All	Total	24 Hr	-12.67 (.11)	0.704 (.021)	2.6	1448	1953
307	7.3	2.4	S.V.	Total	24 Hr	-10.82 (.12)	0.284 (.023)	2.4	1203	1694
308	7.3	2.4	M.V.	Total	24 Hr	-16.41 (.18)	1.206 (.032)	1.2	622	1694
309	7.3	2.4	All	Total	24 Hr	-11.78 (.12)	0.552 (.022)	2.6	1826	1694
310	7.3	3.0	S.V.	Total	24 Hr	-10.59 (.11)	0.221 (.020)	3.0	1175	1649
311	7.3	3.0	M.V.	Total	24 Hr	-15.36 (.14)	1.038 (.024)	0.8	1020	1649
312	7.3	3.0	All	Total	24 Hr	-11.81 (.11)	0.558 (.019)	1.5	2196	1649
313	6.7	1.8	S.V.	Total	Day	-11.28 (.18)	0.292 (.037)	1.4	369	1899
314	6.7	1.8	M.V.	Total	Day	-16.20 (.20)	1.140 (.038)	3.3	173	1899
315	6.7	1.8	All	Total	Day	-12.24 (.17)	0.577 (.035)	2.4	543	1899
316	6.7	2.4	S.V.	Total	Day	-13.48 (.21)	0.707 (.041)	16	407	1953
317	6.7	2.4	M.V.	Total	Day	-17.47 (.24)	1.423 (.043)	1.0	359	1953
318	6.7	2.4	All	Total	Day	-14.53 (.20)	1.023 (.037)	2.3	767	1953
319	7.3	2.4	S.V.	Total	Day	-11.39 (.23)	0.349 (.042)	1.5	520	1694
320	7.3	2.4	M.V.	Total	Day	-17.56 (.28)	1.403 (.049)	1.1	409	1694
321	7.3	2.4	All	Total	Day	-13.26 (.23)	0.794 (.040)	1.9	930	1694
322	7.3	3.0	S.V.	Total	Day	-10.18 (.22)	0.105 (.038)	2.4	470	1649
323	7.3	3.0	M.V.	Total	Day	-15.96 (.23)	1.139 (.037)	0.8	677	1649
324	7.3	3.0	All	Total	Day	-12.83 (.21)	0.709 (.034)	1.2	1148	1649
325	6.7	1.8	S.V.	Total	Night	-11.80 (.12)	0.538 (.029)	3.1	405	1899
326	6.7	1.8	M.V.	Total	Night	-15.48 (.17)	0.993 (.037)	2.0	69	1899
327	6.7	1.8	All	Total	Night	-11.89 (.12)	0.600 (.028)	2.7	475	1899
328	6.7	2.4	S.V.	Total	Night	-12.18 (.12)	0.600 (.027)	20	452	1953
329	6.7	2.4	M.V.	Total	Night	-16.19 (.18)	1.179 (.036)	22	119	1953
330	6.7	2.4	All	Total	Night	-12.43 (.12)	0.711 (.026)	25	572	1953
331	7.3	2.4	S.V.	Total	Night	-11.66 (.14)	0.513 (.030)	4.2	594	1694
332	7.3	2.4	M.V.	Total	Night	-15.70 (.21)	1.068 (.040)	3.4	164	1694
333	7.3	2.4	All	Total	Night	-11.96 (.14)	0.627 (.029)	6.2	759	1694
334	7.3	3.0	S.V.	Total	Night	-11.71 (.14)	0.489 (.026)	2.5	583	1649
335	7.3	3.0	M.V.	Total	Night	-14.66 (.17)	0.903 (.030)	0.5	263	1649
336	7.3	3.0	All	Total	Night	-12.08 (.13)	0.619 (.025)	1.5	847	1649

EXAMPLE OF ACCIDENT PREDICTION BY EB METHOD

The following example illustrates how the regression prediction can be refined by the EB procedure to estimate the accident potential of an individual road section.

Suppose that it is desired to estimate the expected number of single-vehicle accidents during an a.m. peak hour (8:00 to 9:00 a.m.) for a 6-month period in 1989 for a 9-km section (T = 183 hr and L = 9 km in Equation 1) having 1.8-m shoulders and a roadway width of 6.7 m. The section had an average volume of 171 vehicles per hr (F = 171 in Equation 1) and recorded two single-vehicle accidents from 8:00 to 9:00 a.m. during the 6-month period of interest. First, the expected number of accidents is calculated by using Equation 1 and the appropriate regression parameters, in this case, for Model 301 in Table 3, for which ln(a) is equal to -11.27, b is equal to 0.342, and k is equal to 1.6. This yields

$$E_{89}(m) = 9 \times 183 \times 0.00007395 = 0.1218 \text{ accidents}$$

with variance given by Equation 2,

$$\text{VAR}_{89}(m) = 3 \times 10^{-9} \times 9^2 \times 183^2 = 0.00927 \text{ accidents}^2$$

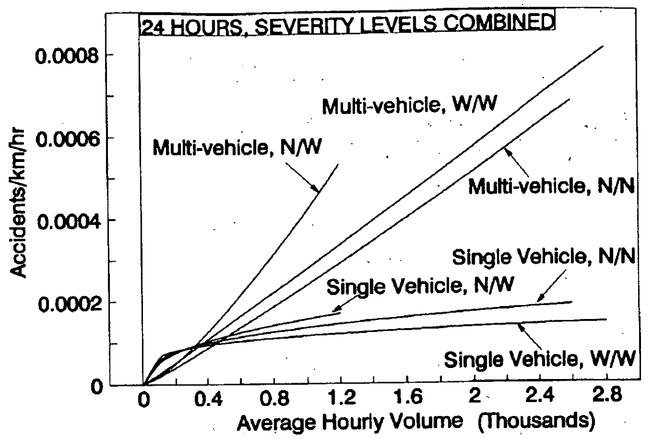


FIGURE 2 Plots of Level 2 models for various lane and shoulder width combinations (n = narrow; w = wide).

In refining this estimate, the weight w to be applied is (by Equation 4)

$$w = 1/[1 + E_{89}(m)] = 1/(1 + 0.1218/1.6) = 0.9293$$

The refined estimate (by Equation 3) is

$$E_{89}(m|x) = w[E_{89}(m)] + (1 - w)x_{89}$$

$$= 0.9293 \times 0.1218 + (1 - 0.9293) \times 2 = 0.2546 \text{ accidents}$$

with variance (Equation 5) given by

$$\mathrm{VAR}_{89}(m|x) = (x_{89} + k)/[1 + k/E_{89}(m)]^2$$

$$= (2 + 1.6)/(1 + 1.6/0.1218)^2 = 0.0180 \text{ accident}^2$$

MODEL VALIDATION

This section shows that the EB method provides estimates of accident potential that are better than the results that would be obtained by a method that is based only on accident counts and that assumes that the number of accidents in one period is a good estimate of the accident potential for another period, after adjusting for traffic volume differences.

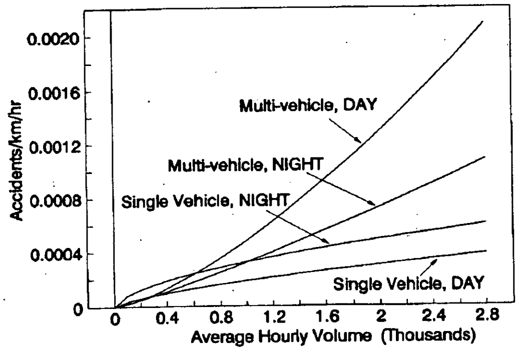


FIGURE 3 Plots of selected Level 2 models illustrating differences between daytime and nighttime accident frequency (severity levels combined: narrow lanes, wide shoulders).

The method involved a comparison of the prediction accuracy resulting from the use of 1989 accident counts as estimates of the 1988 counts, as opposed to using the EB procedure on the basis of 1989 data. In either case the relative measure of the discrepancy between the predicted and observed numbers of accidents was the weighted squared difference between the two values. The weight is the squared product of the section length, hours of exposure, and the average hourly volume. This calculation was done for each hour for each section in the data set, and a mean value was calculated. It is assumed that the method that provides a lower value of this mean weighted squared difference is better. It can be seen from the values in Table 4 that the EB estimate provides lower values for the measure of accuracy than those obtained by using the accident count method, thus confirming its superiority.

ACKNOWLEDGMENTS

The research for this paper resulted from a project with MTO and was supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada. These sources of support are gratefully acknowledged.

TABLE 4 Mean* Squared Differences Between Predicted and Observed Numbers of Accidents Per Kilometer Per Hour

Model	Accident Count Method	EB Estimate
112	8.555	5.775
115	8.200	6.070
118	15.869	0.118
301	10.419	1.389
302	2.358	0.046
303	10.721	1.436
313	5.536	0.910
314	0.605	0.030
315	6.140	0.950
325	11.386	1.602
326	0.621	0.103
327	12.151	1.713

* Sum of weighted squared differences divided by the number of observations.

Values in Table are to be multiplied by 10^{-10}.

REFERENCES

1. Persaud, B. N. *Estimating the Accident Potential of Ontario Road Sections.* Project Report. Department of Civil Engineering, Ryerson Polytechnic University, Ontario, Canada, 1993.
2. Hauer, E. Empirical Bayes Approach to the Estimation of "Unsafety." The Multivariate Regression Approach. *Accident Analysis and Prevention,* Vol. 24, 1992, pp. 457–477.
3. Baker, R. J., and J. A. Nelder. *The GLIM System—Release 3.* Rothamsted Experimental Station, Harpenden, United Kingdom, 1978.
4. Satterthwaite, S. P. A Survey of Research into Relationships Between Traffic Accidents and Traffic Volumes. Supplementary Report SR 692. Transport and Road Research Laboratory, Crowthorne, United Kingdom, 1981.
5. Hauer, E. *Traffic Flow and Safety.* TRB, National Research Council, Washington, D.C., forthcoming.
6. Maycock, G., and R. D. Hall. *Accidents at 4 Arm Roundabouts.* Laboratory Report LR 1120. Transport and Road Research Laboratory, Crowthorne, United Kingdom, 1984.
7. Hauer, E. On the Estimation of the Expected Number of Accidents. *Accident Analysis and Prevention,* Vol. 18, 1986, pp. 1–12.

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

Analysis of Driver Safety Performance Using Safety State Model

EDWARD J. LANZILOTTA

A significant component in the pursuit of safety is estimation of risk probability. In transportation systems virtually all safety-related events and outcomes involve an intermediate event known as an accident. The safety state model is a probabilistic model that is used to estimate the probability of an accident as a function of the human–machine system state. By using a discrete Markov network, the safety state model forms a framework for capturing the human–machine and human–human interactions in a transportation system. The observed data are used to calibrate the model, which is subsequently used to estimate the risk probability performance of other human operators. The theoretical development of this model is reviewed. In addition, motivation and background, as well as advantages and disadvantages with respect to existing quantitative methods of risk probability estimation, are discussed. Finally, the applicability to driver performance analysis is discussed.

A current trend in the automobile industry is an emphasis on safety. Automobile manufacturers are implementing features intended to improve highway safety. The motivation is well justified because the magnitude of damage, injury, and death on highways remains a significant problem.

A key component of safety in the human–machine system is the human driver. As controller of the vehicle the driver must monitor the state of the vehicle (position and speed) and the surrounding environment (including other vehicles, as well as roadside elements), make decisions with regard to control actions, and actuate the vehicle controls to carry out the control decisions. The driver is solely responsible for the state of the vehicle: the control decisions and actuation determine the resulting position and speed of the vehicle in the environment. The control actions of the driver play a significant role in the effectiveness of any designed-in safety device or system. In the worst case a driver can counteract or override the effects of a designed-in safety system because of either a lack of training or a higher level of risk-taking.

In this work the driver–vehicle system is modeled as a closed-loop control system (*1,2*), with the vehicle as the "plant" and the human driver as the "controller." The driver senses the state of the vehicle and environment and provides control input to the vehicle. The control strategy is a time stream of decisions made by the driver, which govern the actuation response to the driver's perception of the system state. These decisions are the result of some combination of rational thought and instinctive response, and the basis for the decisions is typically obtained through a combination of training and experience.

Measuring driver performance is a challenge, especially with respect to safety-related decision behavior. Current driver evaluation methods focus on perception and actuation, because these task skills are observable and testable (e.g., through eye examinations and simplistic road tests). However, the decision-making component of the human–machine control system plays a significant role in governing the interaction between a vehicle and its environment. These interactions ultimately determine the safety of the driver's actions.

This research is focused on modeling system behavior in ground transportation systems. In particular, researchers are interested in evaluating the decision behavior of vehicle drivers. This behavior is evaluated in terms of risk probability as a function of the state of both the vehicle and its environment during operation. A probabilistic model is used to represent the state of the human–machine system. This model, termed the *safety state model,* provides a framework for observing system behavior and driver decision behavior as a function of time. The observed data are used to calibrate the model, which is subsequently used to predict the risk probability as a function of the system state. By using the resultant relationship between system state and estimated risk probability, the safety state trajectory of a human–vehicle system can be transformed into a risk probability trajectory as a function of time. The risk probability trajectory is used as the basis for evaluation of driver performance with respect to safety. In the event of an accident the safety state trajectory provides a chain of events leading to the accident, which may be used for determination of causality.

DISCUSSION OF SAFETY AND RISK ASSESSMENT

Before developing the theory of the safety state model consider the terms used and the meanings associated with them. Lowrance (*3*) defines safety as the "judgment of acceptability of risk." This definition provides a working framework for the pursuit of safety, which includes subjective and objective components. The subjective component, which is the judgment of acceptability, evaluates whether a given level of risk is acceptable to the society that is affected. Policies are set on the basis of that judgment. These policies determine the trade-off between a level of risk and the resources expended to reduce that level of risk. Risk judgment is typically performed by policy makers.

The objective component is risk assessment. A variety of definitions of risk can be found in the literature. Rowe (*4*) defines risk as "the potential for unwanted negative consequences of an event or activity," alluding to the notion of chance. Lowrance (*3*) includes the probabilistic component explicitly, defining risk as the "measure of probability and severity of adverse effects." Rescher (*5*) echoes that idea: "Risk is the chancing of negative outcome. To measure risk we must accordingly measure both of its defining components, the chance and the negativity." Gratt (*6*) specifies the relationship between probability and severity in risk assessment by stating that the "estimation of risk is usually based on the expected result of the

---

Human-Machine Systems Laboratory, Department of Mechanical Engineering, Massachusetts Institute of Technology, Room 3-355, 77 Massachusetts Avenue, Cambridge, Mass. 02139.

conditional probability of the event times the consequences of the event given that it has occurred." Wharton (*7*) offers that "a risk is any unintended or unexpected outcome of a decision or course of action," including both positive and negative outcomes.

In the case of transportation safety risk assessment is most often considered in terms of fatalities, personal injury, and property damage: the results of an accident. Based on the definitions of risk, risk assessment in transportation systems can be divided into two subcomponents: the probability and the severity of an accident. The risk probability of an accident estimates the relative likelihood that such an event will occur. The severity estimates the ultimate outcome of the accident in the terms of interest (i.e., injuries and deaths) for a given set of conditions regarding the accident. In a sense the accident event represents a demarcation point in time: all of the events leading to an accident contribute to risk probability, whereas those that occur after the accident are in the domain of severity. These components of risk assessment parallel the concepts of "active safety" and "passive safety" devices, respectively. "Active safety" is a term typically applied to those devices or systems that assist in preventing accidents (such as anti-lock braking systems and traction control), whereas "passive safety" devices are those that reduce the severity of an accident when it does occur (such as airbags and door guard beams). Risk assessment is typically performed by systems analysts. This research is focused on estimation of risk probability.

Risk assessment is, in effect, a subset of reliability engineering, which is focused on estimating the probability and effects of system failures. When a system failure can result in injury or death to a human it becomes a safety issue. Assessment of system reliability with respect to a failure of this type is risk assessment.

Risk probability, especially in transportation systems, is not a static quantity. Instead, risk probability varies as a function of the state of the system, which includes the state of the vehicle as well as the state of the environment. The system state in transportation systems is quite dynamic with respect to time. The driver is responsible for a constant stream of control decisions, and the actions resulting from those decisions determine the state of the vehicle in relation to the state of the environment. Thus, through these control decisions the driver has a profound impact on the risk probability of the vehicle system. Many accident scenarios are the result of compounding several hazard conditions, each of which may be relatively innocuous when it occurs in isolation. Some of these hazards may be due to driver errors (*8*), whereas others may be due to machinery failures in vehicle or wayside equipment. The collected set of potential hazard conditions leading to a particular accident scenario can be considered a system state. Because this state varies with time and the risk probability is a function of this state, risk probability can also be considered a function of time.

Time is an integral component of risk probability. The risk probability can be modeled by probability theory as the relative likelihood of the occurrence of an accident. However, the risk probability of an accident only makes sense if its occurrence is compared with the alternative event, which is the nonoccurrence of an accident. Since nothing "happens" during the nonoccurrence, the event can only be considered with respect to some fixed metric. The safety state model considers the probability of an accident with respect to a fixed time frame, known as a *time slice*. Thus, the risk probability represents the relative likelihood of an accident in a single time slice. On average, it also represents the percentage of time slices that result in an accident. An alternate form of expression is in terms of the mean time (number of time slices) between the occurrence of accidents. This form is commonly known as the mean time between failures (MTBF) and is used extensively in the field of reliability engineering.

Even if the risk probability for a rare event is very small (alternately, the MTBF is very large), probability theory asserts that the event will eventually occur, given enough opportunity (i.e., time). From this, it can be seen that the only way to avoid a probabilistic event is to "get out of the game" before that event occurs. (In fact, this is what happens to most people with respect to rare catastrophic events—the human lifetime is much shorter than the time period in which one could expect to experience a single occurrence.) Thus, the concept of risk exposure is as follows: given a constant risk probability, the expected number of failures over a prescribed period rises with the size of the period. To reduce the overall risk of an undesirable event, one must reduce either the risk probability or the risk exposure.

Estimating the risk probability of transportation accidents is quite difficult for several reasons. First, accidents are relatively rare occurrences and are difficult to predict. In addition, the events and behavior of highest interest for risk probability estimation are those that immediately precede the accident event: attention needs to be focused on a time period that is identified by an event that occurs at the end of the period. Finally, a compound set of hazards and events typically leads to an accident.

A guiding motivation in this work is the notion that near collisions are far more common than actual accidents. If the capability of identifying near collisions and the conditions that lead to them exists, responses (in either design, operating procedure, or policy) can be formulated to reduce the occurrence of near collisions and in the process reduce the number of accidents. A dynamic estimation of risk probability provides a mechanism for identifying system states corresponding to near collisions.

SAFETY STATE MODEL

The safety state model is an extension of the more familiar event tree and fault tree models. An event tree is a representation of possible scenarios that can occur from a fault-precipitating event (*9*). A fault tree, by contrast, works backward from a system failure to identify the logical combination of all of the potential causes of that failure (*10–12*). The safety state model has been inspired by these methods of system safety analysis and represents a step forward in generality.

Event Tree Analysis

Event tree analysis is used for human reliability analysis. The purpose of the method is to identify the probability of system failure from the occurrence of a precipitating event. From an event tree it is possible to detect points in the failure process where human reliability is problematic and to use that knowledge to suggest improvements to manual or automated procedures.

Event tree analysis starts at a precipitating event. From the occurrence of that event, branches are constructed to all of the possible next events. Each branch has a probability associated with the occurrence of the next events. Then, from each of the next events, tree limbs are constructed for subsequent events, with associated probabilities. Once the tree has been completed the overall probability of each possible event path can be calculated. Swain and Guttman (*9*) explicitly state that there should be no more than two branches from each node, representing a binary decision process; others (*11*) allow for event tree construction with more than two branches from an event node, corresponding to partial failure.

A generalized example of an event tree is shown in Figure 1. Figure 1 shows that at the precipitating event on the left the operator can choose to take Action *A* or not. If Action *A* is not chosen a failure will result. If Action *A* is chosen then a second decision point occurs, at which point the operator can choose to take Action *B* or not. If Action *B* is not taken there is a subsequent decision to take Action *C* or not. If Action *C* is not taken a failure will result. If Action *C* is taken the operator will be at the same decision point as if Action *B* was chosen earlier. Thus, Action *C* is known as a corrective action. The probability of system failure can be expressed as

$$P(\overline{A}) \cup [P(A) \cap P(\overline{B}) \cap P(\overline{C})] \cup [P(A) \cap P(B) \cap P(\overline{D})]$$
$$\cup [P(A) \cap P(\overline{B}) \cap P(C) \cap P(\overline{D})]$$

The failure process is thus transformed into a combination of individual failure probabilities, which are more easily determined. The overall probability of failure can then be evaluated through the mathematical combinations of these individual failures, as determined from the event tree.

Event tree analysis is especially well suited for analysis of systems that are procedural by nature, because it effectively measures deviation from an ordered sequence of events. In this regard it has proven to be very useful in the nuclear power industry. However, the binary decision form is not applicable to systems that offer several choices at each decision point. Even when the nonbinary form is used, the event tree method becomes unwieldy as the number of decision options rises. In addition, unless each decision point explicitly includes a time limit, event tree analysis cannot capture the time relationship between events. For these reasons event tree analysis is limited with regard to estimating the risk probability of drivers.

Fault Tree Analysis

Fault tree analysis is another method commonly used in human reliability analysis. In contrast to event tree analysis, fault tree analysis is considered backward looking. The analysis begins with the occurrence of the failure and works backward to identify the combinations of contributing factors.

An important feature of fault tree analysis is the logical combinations of preceding conditions. Through the use of Boolean operators ("and" and "or") the fault tree describes the combinations of precursor faults leading to a system failure. In addition, fault tree analysis can be used as a quantitative method by assigning probabilities to the various failure events. A generalized example of a fault tree is shown in Figure 2. In this example the system will fail either if Event *A* has occurred or if Event *B* occurs, which results from the combined occurrence of Events *C* and *D*. The probability of this system failure can be expressed as $P(A) \cup P(B)$, which is equivalent to $P(A) \cup [P(C) \cap P(D)]$. Boolean algebra provides a powerful tool for logically combining events and hazards that can lead to a failure scenario. Although fault tree analysis does not explicitly exclude "gate-to-gate" connections (*11*) (which correspond to complex Boolean logic expressions), convention dictates that the use of these constructs is avoided.

As with event tree analysis a weakness of the fault tree is coverage: in the case in which contributing events are independent the analysis becomes quite cumbersome, because all combinations of hazards must explicitly be included. Ansell (*13*) notes that fault tree analysis (as well as FMEA, another risk assessment technique) "suffers from a narrowing of our vision of the system by either limiting the number of failure modes for a component or the types of risks considered. They both implicitly rely on the correctness of the technology or science on which the model is built. This is reinforced by cognate dissonance; only perceived possible risks can be guarded against." Fault tree analysis is also weak with regard to capturing the time relationships between events that contribute to an accident.

Fault tree analysis has been applied to estimation of driver risk probability (*14,15*). These studies have been successful in using fault tree methods to identify a framework of causation. However, in neither case was the research directed toward estimating the risk probability as a function of system state. Based on the weaknesses that have been discussed, fault tree analysis is not well-suited to this purpose.

Structure of Safety State Model

The safety state model, an extension of event tree and fault tree analysis, is now described. By assuming that the conditions contributing to system failure are truly independent, the safety state model can be viewed as a generalization of event and fault tree analyses.

Consider a collection of *n* conditions that could possibly contribute to an accident scenario. These conditions include actions taken by the driver (such as acceleration or braking), the state of the

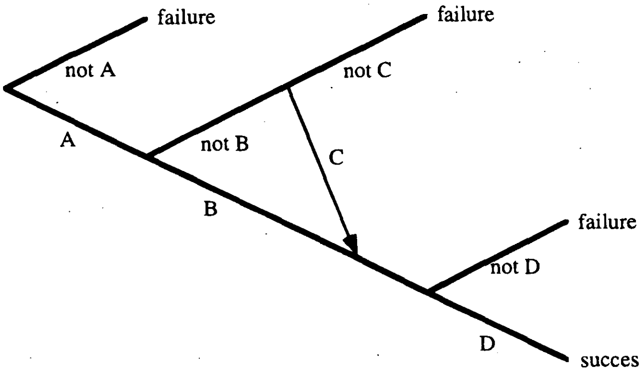


FIGURE 1 Event tree.

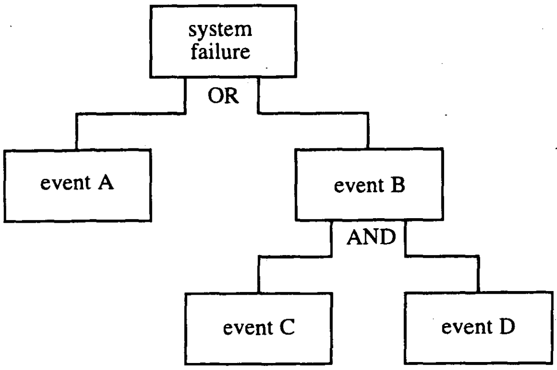


FIGURE 2 Fault tree.

driver (such as fatigue or impairment), the state of another vehicle in the system (such as vehicle ahead braking), and the state of the system environment (such as a red traffic light ahead). These conditions are constrained to be binary conditions. That is, the condition is defined such that it has only two possible values. The complete set of possible combinations of such a set of conditions can be represented by a binary word which is n bits long. The total number of possible combinations is 2^n.

Now consider each of the possible combinations (i.e., each number in the set of possibilities) to represent a state in a Markov network. These states are identified by the associated number, which is within a range of 0 to $(2^n - 1)$, inclusive. The accident scenario is identified as an additional state, labeled 2^n. The state number is termed the *safety state* of the system, and the resultant Markov network is known as the *safety state network.* An example of a three condition safety state network is shown in Figure 3.

The state transition of the Markov network is defined to occur at regular time intervals, with the time period of the interval fixed at h. The value of h is set such that only one condition may change its state (within reasonable probabilistic bounds). At each state transition instant (i.e., at the end of each state transition interval), the model will transition from the current state $S(i)$ to the next state $S(j)$ with probability $p_{i\to j}$. The probability $p_{i\to j}$ represents the holding probability for the state $S(i)$, which is the probability that the state will not change at the next transition. The collection of transition probabilities for a given state ($p_{i\to j}, j = 0, 1, 2, \dots 2^n$) represents a probability distribution, and the sum of these probabilities must be 1 (Equation 1).

$$\sum_{j=0}^{2^n} p_{i\to j} = 1 \tag{1}$$

The safety state corresponding to the accident scenario [state $S(2^n)$] is a trapping state, which means that once it is entered the process can never exit that state. This notion correlates with the reality that the occurrence of an accident is permanent and cannot be undone.

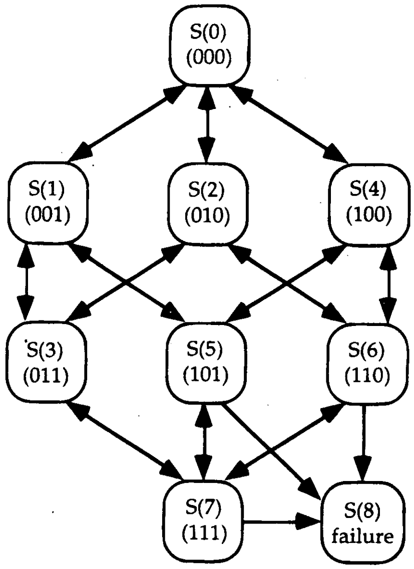


FIGURE 3 Safety state network.

As a trapping state the accident state has a holding probability of 1 and the probability of transition to any other state is zero.

The collection of probability distributions for the entire set of possible safety states can be represented in matrix form (Equation 2). The row of this state transition matrix (P) represents the current state, whereas the column number represents the next state.

$$P = \begin{bmatrix} p_{0\to 0} & p_{0\to 1} & \cdots & p_{0\to 2^n} \\ p_{1\to 0} & p_{1\to 1} & \cdots & p_{1\to 2^n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{2^n\to 0} & p_{2^n\to 1} & \cdots & p_{2^n\to 2^n} \end{bmatrix} = \begin{bmatrix} P_0 & P_F \\ 0 & I \end{bmatrix} \tag{2}$$

Note that because of the structure imposed earlier in the development, the state transition matrix can be partitioned in a convenient manner. The rightmost column ($P_{i\to j}$, $i = 0, 1, 2, \dots, 2^{n-1}$) represents the failure probabilities from any of the nonfailure states and can be represented by the vector P_F. The lowest row is the probability distribution for the failure state, which is all zeros with the exception of the holding probability. The remaining submatrix represents the transitions among the nonfailure states (termed *operational states* because they represent the set of states that are possible during nonfailure operation). The submatrix of operational state transition probabilities is labeled P_O.

When considering this network topology it is important to keep in mind the issues of scale. The number of states in the safety state network grown as a power of 2 with an increasing number of conditions, and the number of elements in the state transition matrix grows as the square of the number of safety states. So, for example, a 10-condition network has roughly 1,000 states and 1,000,000 elements in the state transition matrix. When applying this method to actual systems the analyst must keep in mind the effects of scale and choose the conditions carefully to avoid having an unnecessarily large and unwieldy safety state network.

Estimating Risk

Although the state transition matrix itself is interesting, the ultimate power of this model lies in the ability to estimate the probabilities of future states. Consider the current state $S(i)$ to be represented as a vector $\overline{\theta}(k)$ of dimension 2^{n+1} by 1, in which the ith element is 1 and the remaining elements are zero. (In this notation k represents the transition number as the process progresses in time.) One can calculate the probability distribution of the next state, shown as $\overline{\theta}(k + 1)$, using Equation 3.

$$\overline{\theta}^T(k + 1) = \overline{\theta}^T(k)P \tag{3}$$

Using this strategy one can look beyond the next transition to determine the probability of reaching a given state in any number of transitions in the future (Equations 4). This is a powerful concept, and by using this concept one can evaluate the probabilistic behavior as far in to the future as one would like. Future probabilistic behavior can be summarized in the $\Phi(\tau)$ matrix, which expresses the ability to transition from one state to another in τ transitions (Equation 5). Note that the $\Phi(\tau)$ matrix can be partitioned in exactly the same manner as the state transition matrix (Equation 2).

$$\begin{aligned} \overline{\theta}^T(k + 2) &= \overline{\theta}^T(k + 1)P = \overline{\theta}^T(k)P^2 = \overline{\theta}^T(k)\Phi(2) \\ \overline{\theta}^T(k + \tau) &= \overline{\theta}^T(k)P^\tau = \overline{\theta}^T(k)\Phi(\tau) \\ \overline{\theta}^T(k + \infty) &= \overline{\theta}^T(k)P^\infty = \overline{\theta}^T(k)\Phi(\infty) \end{aligned} \tag{4}$$

$$\Phi(\tau) = \begin{bmatrix} \emptyset_{0\to 0}(\tau) & \emptyset_{0\to 1}(\tau) & \dots & \emptyset_{0\to 2^n}(\tau) \\ \emptyset_{1\to 0}(\tau) & \emptyset_{1\to 1}(\tau) & \ddots & \vdots \\ \vdots & & & \vdots \\ \emptyset_{2^n\to 0}(\tau) & \dots & \dots & \emptyset_{2^n\to 2^n}(\tau) \end{bmatrix} = \begin{bmatrix} \Phi_O(\tau) & \Phi_F(\tau) \\ 0 & I \end{bmatrix} = P^{\tau} \quad (5)$$

The real goal is to determine the probability of reaching the failure state from any given current state. Because the Markov network is a finite single-chain network the ultimate state will be the failure state. (This corresponds to the fatalistic notion that, given enough time, a probabilistic failure will eventually happen.) However, the theory of Markov processes provides a mechanism for calculating the mean time to another state from a known state. In the case of the safety state network, identification of the mean time to the failure state is the point of interest.

To calculate the mean time to failure one first needs to express the probability that the failure will occur on a specific state transition in the future. As shown in Equation 6, $\overline{\Psi}(\tau)$ is a vector quantity that provides the probability that the failure will occur on the τ^{th} state transition in the future as a function of the safety state. The mean time between failure as a function of the safety state is the expected value of the number of transitions until the failure (Equation 7).

$$\overline{\Psi}(\tau) = \Phi_F(\tau) - \Phi_F(\tau - 1) \quad (6)$$

$$MTBF = \overline{M} = \sum_{\tau=1}^{\infty} \tau\overline{\Psi}(\tau) \quad (7)$$

Knowing this, one can estimate the risk probability as the inverse of MTBF (Equation 8).

$$\overline{F} = 1/\overline{M} = \begin{bmatrix} 1/M_1 \\ \vdots \\ 1/M_n \end{bmatrix} \quad (8)$$

In summary, this section details the methodology used to derive the risk probability and mean time between failures given a static state transition matrix. Both the risk probability and mean time between failure are expressed as a function of the system's safety state. Provided there is an obtainable state transition matrix that characterizes average driver behavior, these results can be used to compare the performances of drivers in several ways. Some of these are discussed later.

Model Calibration

As shown, a state transition matrix that is characteristic of average driver behavior is required. To obtain that matrix (i.e., calibrate the model), observations from an existing system are used measuring the binary state of the individual conditions that are combined to form the safety state network. The measurement record marks the points in time at which each individual state change occurs. From these data a safety state trajectory can be constructed as a function of time. This trajectory includes all pertinent data for model calibration, including state occupancy times and state transitions.

By statistical analysis of these data the state transition probabilities are computed. First, state occupancy statistics are used to determine the holding probability, $P_{i\to i}$. The state transition statistics are then used to calculate the individual transition probabilities (Equation 9).

$$p_{i\to i} = (1 - p_{i\to i})\left(\frac{\text{number of transitions to } S(j)}{\text{number of transitions from } S(i)}\right) \quad (9)$$

Driver Performance Measurement

Assuming that a state transition matrix of sufficient quality has been obtained, it is now possible to compare the performances of subject drivers. The safety state model is used to calculate the transformation function between the safety state and the estimated risk probability. This transformation function may be stored in lookup table form, with the safety state as the index in the table.

Data are collected on a subject driver in the same manner used for calibration data: by observing the safety state conditions and recording the changes of state. These data are combined to form a safety state trajectory as a function of time. By using the transformation function the safety state trajectory is then transformed into a risk probability trajectory.

The risk probability trajectory as a function of safety state is used directly for comparing driver performance. The average risk probability for the system average can be computed by taking the weighted average of the risk probability vector elements weighted by the relative time spent in that state (from the calibration data).

Several statistical approaches are useful for comparing the safety-related performances of subject drivers. These include instantaneous risk level, peak risk level, overall average risk level, windowed risk level, and cumulative risk level. In any of these approaches the performance of the subject driver can be compared with either the average risk probability (described in the previous paragraph) or the performance of other subjects.

The instantaneous risk level identifies the current risk probability estimation as a function of time. This measure is independent of history in that the specific safety state trajectory leading to the current state is not identified. (This is a general property of Markov processes.) It is a useful measure for evaluating performance in real time and could be used as feedback information to the driver as well.

The peak risk level identifies the highest level of risk that has occurred. Typically, peak measurements are made within some predefined time period, such as the duration of the test. This measure is useful in identifying the bound on the level of risk that a driver will take.

The overall average risk level provides an overall mean of the risk probability trajectory. To compute this the risk probability trajectory is integrated over time, and the resultant integration is divided by the integration time period. In the overall average the time period continues to grow. As a result the overall average represents a summary of the complete history of safety-related performance.

In contrast, the windowed risk level computes the average over a fixed time interval immediately preceding the current moment. For example, if the time window was defined as 10 min, the windowed average would provide the average safety-related performance for the last 10 min only. This measure "forgets" the past performance that is outside the defined time window and can be used as a means of measuring learning curves or fatigue characteristics.

Finally, the cumulative risk level represents the total amount of risk that has been taken. It is computed by taking the time integral of the risk probability trajectory and represents the expected number of accidents (in a Bernoulli sense). This measure is not intended to predict the occurrence of accidents—the expected value represents an average. However, it can provide useful insight into risk exposure. This measure would not be appropriate for subject feedback, because subject knowledge of this information could significantly affect the subject's performance.

Discussion of Results

The safety state model provides a method for estimating the dynamic risk probability as a function of system state. It has several strengths when compared with event and fault tree methods; however, it has weaknesses as well.

The safety state model is fundamentally different from both event and fault tree analyses in the conceptual definition of the nodes. In both event and fault tree methods the nodes of the tree represent events. In the safety state model the nodes represent system states, and the transitions between the nodes represent events. This is significant, because any state can be reached by several different paths, which represent the occurrence of different events. Once a state has been reached (that is, once a given set of conditions is true), the path to that state (the order of events) does not matter. This is a fundamental concept in Markov process modeling.

Significant among the strengths of the safety state model is its generality. No assumed dependencies or interactions exist between any of the contributing conditions. All combinations of conditions are considered equivalently. Thus, the analyst needs to specify only the conditions that define the safety state, without defining the relationship between them. This serves to improve the robustness of the analysis by reducing the chance of human error (which might be due to either prejudice or oversight). This becomes more significant as the number of conditions increases.

As a result of its generality the method of safety state analysis is easily automated. This frees the analyst from any direct contact with the state transition matrix, which will be of formidable size in any significant problem. The ease of automation allows safety state analysis to be applied to problems that are too large and unwieldy for other methods. However, this strength is tempered by memory and computation requirements, which are significant.

One of the most significant weaknesses is in the area of data collection. In an operational system, the data required for calibration would be quite difficult to gather. For driver performance evaluation a broad range of in-car data would be required. Systems in current operation do not collect appropriate data, and there is debate whether measurements of this nature are even feasible within commonly accepted notions of personal privacy. An alternative approach is to use data collected from simulation systems. Simulation systems can be configured to provide a plethora of appropriate data, and networked simulators provide a mechanism for collecting data on interactions between two or more vehicles. Although there are debates in the driver performance community regarding the applicability of simulator-based results to operational systems, the author believes that simulation-based experiments provide the only currently available means for exploring the viability of this method.

AN EXAMPLE

To illustrate the method of safety state analysis, consider the following example: a frontal impact scenario in a rail system. In this example the goal is to identify the risk probability of striking another vehicle (or obstruction) with the front of the vehicle. The first step is to select the set of conditions that are considered contributory to this failure event. The following set of conditions are used:

- Condition 0: throttle actuated. This condition is true if the throttle is applied and false if the throttle is not actuated.
- Condition 1: brake applied. This condition is true if the brake is applied and false if the brake is not actuated.
- Condition 2: brake failure. This condition is true if the braking system of the train has failed in any way (even if the driver is not aware of this condition) and false if the braking system is functioning properly.
- Condition 3: overspeed. This condition is true if the train is traveling at a speed greater than that allowed by either the static speed limit or the signaling system and false if the train is being operated within the allowable speed bounds; the overspeed condition includes the case of passing a stop signal (entering an occupied block).
- Condition 4: obstacle. This condition is true if there is an obstruction on the track (possibly another vehicle) within the stopping distance of the train and false if the track is clear; for the purposes of this example the stopping distance is defined as the distance required to bring the train to a full stop from the current speed by using full-service braking and is a function of speed.

This set of conditions includes measurement of human control actions (throttle and brake actuation), vehicle state as a result of human control actions (speed condition), on-board equipment failures (brake failure), and external conditions (obstacle). Since each condition is binary, this set of conditions can be combined into a single number. The resulting set of safety states is given in Table 1, with the operational states numbered from 0 to 31. The check marks indicate the conditions that are true for each state. For example, the system is in state 14 when there is no obstacle present but the vehicle is over the speed limit, the brakes are applied, and the brakes have failed. An additional state is included in Table 1, state 32. This state represents the occurrence of the failure event, which in this example is a collision.

To calibrate the model data are collected from an operational system. Once they are collected these data are processed to determine the state transition matrix, which is used to calculate the risk probability as a function of the safety state. Subjects are evaluated by recording the safety state information, converting the resultant safety state trajectory into a risk probability trajectory, and comparing the risk probability trajectories, as discussed previously.

SUMMARY

The research described here is a response to a need for methods in which the safety-related decision performances of vehicle operators can be evaluated. Some of the difficulties in this area include event rarity, compound and interacting errors, and related difficulty in determining causality. This work identifies the human operator as a key component in the safety of transportation systems. The driver-vehicle system is modeled as a closed-loop control system, and a probabilistic model of system behavior is presented.

Based on the work of Lowrance (*3*) an organization for the efforts involved in the pursuit of safety is identified. Safety-related work is divided into subjective and objective components. Risk assessment, the objective component, is further divided into two components. One component, risk probability, measures the probability of occurrence of a necessary intermediate event, whereas risk outcome measures the outcomes of these events in terms of the ultimate risks. In the case of transportation systems, the ultimate undesirable outcome is damage, injury, and death, and the intermediate event is

TABLE 1 Safety State Description

Safety State	Binary Form	Obstacle	Overspeed	Brake Failure	Brake Applied	Throttle Applied	Failure Possible?
0	00000						
1	00001					√	
2	00010				√		
3	00011				√	√	
4	00100			√			
5	00101			√		√	
6	00110			√	√		
7	00111			√	√	√	
8	01000		√				
9	01001		√			√	
10	01010		√		√		
11	01011		√		√	√	
12	01100		√	√			
13	01101		√	√		√	
14	01110		√	√	√		
15	01111		√	√	√	√	
16	10000	√					√
17	10001	√				√	√
18	10010	√			√		√
19	10011	√			√	√	√
20	10100	√		√			√
21	10101	√		√		√	√
22	10110	√		√	√		√
23	10111	√		√	√	√	√
24	11000	√	√				√
25	11001	√	√			√	√
26	11010	√	√		√		√
27	11011	√	√		√	√	√
28	11100	√	√	√			√
29	11101	√	√	√		√	√
30	11110	√	√	√	√		√
31	11111	√	√	√	√	√	√
32	100000						

an accident. Based on this organizational description, the focus of this research is in the area of risk probability assessment.

A probabilistic model for system behavior (the safety state model) is developed. This model is based on finite Markov processes, with event tree and fault tree techniques used as inspirations. From the safety state model a method for determining MTBF and risk probability is developed. Both of these quantities are expressed as a function of system state. A method for calibrating the safety state model is presented and is based on experimental data. Finally, a method for measuring individual driver performance and comparing it with average driver behavior and the behaviors of other individual drivers is presented.

In conclusion, the safety state model represents a unique method for assessing the safety-related decision performances of vehicle drivers. Future developments include experimental verification of the usefulness of the model with data gathered from a human-in-the-loop high-speed rail simulation system.

ACKNOWLEDGMENTS

The author extends his deepest appreciation to the many colleagues who took time to review this paper and to contribute their constructive commentary. This work was prepared with support from the Volpe National Transportation Systems Center, Cambridge, Massachusetts.

REFERENCES

1. Sheridan, T. *Telerobotics, Automation, and Supervisory Control.* MIT Press, Cambridge, Mass., 1992.
2. Forbes, T. W. (ed.). *Human Factors in Highway Traffic Safety Research.* Wiley-Interscience, New York, 1972.
3. Lowrance, W. W. *Of Acceptable Risk.* William Kaufmann, Inc., Los Altos, Calif., 1976.
4. Rowe, W. D. *An Anatomy of Risk.* John Wiley and Sons, Inc., New York, 1977.
5. Rescher, N. *Risk: A Philosophical Introduction to the Theory of Risk Evaluation and Management.* University Press of America, 1983.
6. Gratt, L. B. Risk Analysis or Risk Assessment: A Proposal for Consistent Definitions. In *Uncertainty in Risk Assessment, Risk Management, and Decision Making*, Plenum Press, New York, 1987.
7. Wharton, F. Risk Management: Basic Concepts and General Principles. In *Risk Analysis, Assessment, and Management* (J. Ansell and F. Wharton, eds.), John Wiley and Sons, Inc., New York, 1992.
8. Reason, J. *Human Error.* Cambridge University Press, Cambridge, 1990.
9. Swain, A. D., and H. E. Guttman. *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications.* Sandia

National Laboratory Report NUREG CR-1278. U.S. Nuclear Regulatory Commission, Washington, D.C., 1983.

10. Lewis, E. E. *Introduction to Reliability Engineering.* John Wiley and Sons, Inc., New York, 1987.
11. McCormick, N. J. *Reliability and Risk Analysis.* Academic Press, 1981.
12. Gertman, D. I., and H. S. Blackman. *Human Reliability & Safety Analysis Data Handbook.* John Wiley and Sons, Inc., New York, 1994.
13. Ansell, J. Reliability: Industrial Risk Assessment. In *Risk Analysis, Assessment, and Management* (J. Ansell and F. Wharton, eds.), John Wiley and Sons, Inc., New York, 1992.
14. Joshua, S., and N. Garber. A Causal Analysis of Large Truck Accident Through Fault Trees, *Risk Analysis,* Vol. 12, No. 2, 1992.
15. Kuzminski, P., J. S. Eisele, N. Garber, R. Schwing, Y. Y. Haimes, D. Li, and M. Chowdhury. Improvement of Highway Safety I: Identification of Causal Factors Through Fault-Tree Modeling. *Risk Analysis,* 1995.

Publication of this paper sponsored by Committee on Simulation and Measurement of Vehicle and Operator Performance.

Analysis of State Department of Transportation Safety Expenditures and Highway Safety

PARVIZ A. KOUSHKI, SALEH YASEEN, AND JOHN L. HULSEY

A review of the FHWA annual *Highway Statistics* report indicates that variations in motor vehicle traffic fatality and injury rates, as well as safety and administrative expenditures, are quite significant among the state departments of transportation (SDOTs), even for those in states with similar regional, physical size, population, and other socioeconomic characteristics. The extent of the causal relationships that may exist between highway fatality or injury rates and SDOT upper-management structures, percent budgetary allocations, traffic variables, and network pavement conditions is examined and evaluated. The statistical test of significance and variance component analysis were used to quantify the magnitudes of these relationships. The findings indicated that (a) the states with the lowest highway fatality rates suffered from the highest injury rates and vice versa, (b) increases in the annual average daily traffic per km and in expenditures on safety were both significantly associated with reductions in fatality rates, and (c) no common pattern could be found among the SDOTs in the allocation of budgetary funds between the competing activities of safety and administration. Areas for further research are recommended.

The recent shift in emphasis from primarily federal financial support of transportation infrastructures to greater local and private-sector self-sufficiency has created an enormous challenge for state departments of transportation (SDOTs). The combined effects of deteriorating infrastructures, increasing needs for system expansion in areas experiencing growth, and the rising costs of project implementation at all levels require financial resources that will have to come from a variety of sources. Improving the allocation efficiency of the limited resources of SDOTs may be a key factor in meeting the challenge. This paper aims to analyze SDOT budgetary allocations for safety and administration and highway fatalities.

The minimization of road traffic accidents has always been a top priority of the U.S. Department of Transportation and SDOTs. Despite this strong emphasis, however, roughly one-half million people have lost their lives on U.S. highways over the last decade or so (*1*). Motor vehicles kill more Americans from ages 1 to 34 than any other source of injury or disease (*2*), and the cost of highway collisions for the year 1986 alone was estimated by the National Safety Council to be $57.8 billion (*3*).

Road accident injuries are often measured in relation to the total miles traveled. This indicator, the number of fatalities per 100 million miles of travel, has been reduced significantly over the years to, for example, 1.91 in 1991, compared with 3.5 in 1980. However, when viewed from a public health perspective, which uses deaths per 100,000 population, the reduction is less significant: 16.5 fatalities per 100,000 population in 1991 compared with 23 fatalities per 100,000 population in 1980.

Despite the overall improvement in road safety, however, there has been a growing concern in recent years that both federal and state administration of highway safety-related projects may suffer from serious shortcomings (*4*). Safety expenditures have been demonstrated to be unsystematic (*5*), based on inadequate data and evaluation procedures (*6–8*), and often allocated to projects with questionable safety effectiveness (*9*).

To complicate the matter further, the approach to decision making in the SDOTs has also begun a process of change due to several important and complex factors. These include a significant increase in the public's concern for social and environmental issues, a significant and sudden decline in financial resources, rapid increases in costs, the emergence of new technologies and tools (*10–12*), and strategies for change (*13*).

A review of *Highway Statistics*, the annual report published by FHWA, indicates that motor vehicle traffic fatality and injury rates (fatalities or injuries per 100 million vehicle miles) vary significantly from state to state (*1*). The 1991 fatality rate, for example, varies from a low of 1.16 in Connecticut to a high of 2.83 in Nevada, a 244 percent difference. The rate for personal injuries differs from 82.6 in Vermont to 260.30 in New York, a difference of 315 percent. Several socioeconomic, environmental, and geographical factors affect these highway safety statistics significantly. What may also have a pronounced effect on highway safety, however, is the distribution and allocation of budgetary resources among the competing capital, maintenance service, and safety-related activities of SDOTs. It is generally agreed that highway safety performance is closely correlated with safety expenditures, especially if those expenditures are made on safety-related activities with proven records of effectiveness (*14*).

Again, a review of FHWA's *Highway Statistics* shows that variations in expenditures on safety and administration are quite significant among SDOTs. For example, when viewed in terms of the percentages of the total SDOT budget, expenditures on safety and administration differ from state to state by as much as 15 and 9 times, respectively.

In light of these statistics and the accelerated rate of change in SDOT organizational structures and policy decisions for meeting the changing socioeconomic environment and adapting to the requirements of the Intermodal Surface Transportation Efficiency Act of 1991, several logical questions deserve attention:

1. To what extent do factors of organizational structure, traffic, network condition, and expenditures on safety and SDOT administration affect roadway safety?

P. A. Koushki and S. Yaseen, Department of Civil Engineering, Kuwait University, P. O. Box 5969, Kuwait 13060. J. L. Hulsey, Department of Civil Engineering, University of Alaska, Fairbanks, Alaska 99701.

2. Are differences in budgetary allocations on safety and administration statistically significant for the states with the lowest and highest fatality rates?

3. Are SDOT expenditures on highway safety made in relation to certain physical, geographical, traffic, or SDOT organizational structures?

This paper attempts to provide answers to these questions.

FINDINGS

The data for the 50 states, taken from the FHWA's annual *Highway Statistics* (*1*), include the following: population of driving age (age 16 years and over); numbers of licensed drivers (per 1,000 total resident population), registered vehicles, and vehicle miles of travel; annual average daily traffic (AADT) per mile; volume service flow ratio (V/SF; percentage of total mileage with $V/SF > 0.7$); present serviceability rating $PSR \geq 3.5$ to 5.0; computed as percentage of total urban and rural road mileages; expenditures on highway law enforcement and safety and on administration/miscellaneous (as percentage of total expenditures); and fatality and injury rates (per 100 million vehicle miles of travel). Tables 1 and 2 present these data for the 15 states with the lowest and highest fatality rates, respectively.

The organizational structures of SDOTs, taken from a research report by Koushki et al. (*15*), were also added to the data list. The state-level mean statistics of the data are presented in Table 3.

Factors Affecting Road Safety

In this section the extent to which factors of SDOT upper-management organizational structure, traffic, network condition, expenditures on safety, and expenditures on SDOT administration affect fatality and injury rates are examined.

In their 1991 study of SDOT management information systems, Koushki et al. (*15*) also found that among the 50 SDOTs there were six types of upper-management organizational structures. These organizational types were used to examine the potential relationship that may exist between the upper-management decision-making structure and highway fatality and injury rates. Forty-two states comprised three organizational types, Types 2, 3, and 4. The other three organizational types, which represented those of eight states, were excluded from the analysis because of insufficient sample size (less than five) in each category.

Interesting and important points were revealed (Figure 1). First, the analysis showed that states with the highest fatality rates (13 states) experienced the lowest injury rates, and those with the lowest fatality rates suffered from the highest injury rates (21 states). The remaining eight states were somewhere in between these two extremes, with fatality and injury rates approximating the overall mean rates for the 50 states. It seems that achieving reductions in both the fatality and injury rates simultaneously are similar to performing tasks that are mutually exclusive. The reason is that in severe accidents, individuals who may be saved from death end up with injuries, unless accidents can be prevented from happening altogether.

Second, states with SDOTs with the Type 2 organizational structure experienced the highest fatality rates (2.195; standard deviation, 0.457) and the lowest injury rates (134.0; standard deviation, 27.13). States with SDOTs with the Type 4 organizational structure were the opposite: they had the lowest fatality rates (1.766; standard deviation, 0.450) and the highest injury rates (154.7; standard deviation, 46.13). States with SDOTs with the Type 3 organizational structure were in the middle [fatality rate, 1.97 (standard deviation, 0.295); injury rate, 141.8 (standard deviation, 33.76)]. The results of a statistical test of significance indicated that the difference in the mean fatality rates for states with SDOTs with the Type 2 organizational structure and those with SDOTs with the Type 4 organizational structure were statistically significant at the 99 percent level of significance ($\alpha = 0.01$).

Why do states with SDOTs with the Type 4 organizational structure have the lowest rates of fatalities in the nation? The only apparent difference between the two upper-management organizational hierarchies (Figure 1) is the absence of a group of board or commission members in the decision-making hierarchy. This may mean (a) a shorter chain in the upper level of the decision-making process,

TABLE 1 Traffic, Travel, and Budgetary Allocations for 15 States with Lowest Fatality Rates[a]

Rank/State	Fatality Rate[b]	AADT/ Km	% of Hwys. with $\frac{V}{SF} > 0.7$	VKT (10^6)	Expenditures (% of Total)	
					Safety	Admini.
1. Connecticut	0.721	2252	0.80	42871	3.3	2.8
2. Massachusetts	0.739	2307	0.59	74924	7.5	11.7
3. Rhode Island	0.764	1989	0.51	11515	4.9	3.9
4. New Jersey	0.820	2944	1.30	95455	4.2	4.8
5. Minnesota	0.839	516	0.18	63199	5.3	5.0
6. New Hampshire	0.894	1137	0.78	15995	9.1	18.7
7. Washington	0.913	989	0.15	74783	6.8	6.5
8. Delaware	0.944	2075	1.10	10821	8.2	11.6
9. Virginia	0.957	1528	0.50	98369	5.0	8.0
10. North Dakota	0.981	117	0.00	9581	4.2	5.3
11. Hawaii	1.031	3378	1.60	13110	1.2	6.4
12. Maryland	1.044	2428	1.10	66572	9.9	6.1
13. Illinois	1.050	1068	0.03	137542	5.5	6.7
14. Maine	1.068	898	0.97	19077	7.8	4.6
14. Michigan	1.068	1186	0.41	131915	9.2	9.4
15. Wisconsin	1.087	701	0.20	73187	4.2	5.4

[a] Source: Ref. (1)

[b] Fatality per 100 million vehicle-kilometers of travel

TABLE 2 Traffic, Travel, and Budgetary Allocations for 15 States with Highest Fatality Rates[a]

Rank/State	Fatality Rate[b]	AADT/ Km	% of Hwys. with $\frac{V}{SF} > 0.7$	VKT (10^6)	Expenditures (% of Total)	
					Safety	Admini.
1. Nevada	1.76	393	0.10	16921	9.0	5.0
2. Mississippi	1.75	584	0.12	40084	5.6	4.2
3. New Mexico	1.74	509	0.02	27004	4.4	10.7
4. Arkansas	1.72	484	0.05	35314	6.1	3.3
5. Alabama	1.61	806	0.25	69108	5.8	7.7
5. West Virginia	1.61	784	0.62	25802	1.7	7.4
6. South Carolina	1.60	915	0.56	55474	9.4	8.0
7. Idaho	1.59	282	0.01	16612	6.4	7.3
8. Alaska	1.56	506	0.28	6474	4.7	6.6
9. Louisiana	1.53	1009	0.10	55883	8.8	4.7
10. Montana	1.50	200	0.00	13386	6.1	3.7
11. Kentucky	1.46	860	0.76	56693	6.1	5.6
11. Tennessee	1.46	948	0.11	76100	4.8	6.6
12. Arizona	1.45	1075	0.05	56236	4.0	4.7
13. Florida	1.35	1766	0.14	182708	6.9	4.2
14. South Dakota	1.32	137	0.00	10805	5.0	8.1
15. Iowa	1.32	348	0.08	37062	6.1	6.6

[a] Source: Ref. (1).

[b] Fatality per 100 million vehicle-kilometers of travel.

(b) a potential reduction in conflicts of interest among the policy makers in upper management, (c) potential variations in the organizational hierarchies of middle management, and (d) none of the above. It may mean, however, that the apparent existence of this relationship is basically a chance phenomenon.

The effects of traffic flow variables on highway safety were examined by analyzing the effects of factors such as AADT per mile and the percentage of the roadway network with a V/SF ratio > 0.7 (an indicator of congestion level and travel speed) on highway fatality and injury rates. An increase in traffic volume was generally accompanied by a decrease in traffic fatalities and, at the same time, with an increase in traffic injuries (Figure 2). This finding is in accordance with the expectations. Increases in traffic volume cause reductions in travel speed. Accidents occurring at lower travel speeds will likely be less severe (injurious rather than fatal). Results of a statistical test of significance indicated that the difference in the mean fatality rates for the states with AADT per km < 1250 and those with AADT per km > 1250 was statistically significant at the 95 percent level of significance ($\alpha = 0.05$).

An analysis of the V/SF ratio and its relationship to highway fatality and injury rates indicated that although the same general trend as that for AADT existed between V/SF ratios and traffic fatality and injury rates, the point elasticities of fatality and injury rates (defined as the slope of the curve at a given point) were much less pronounced compared with those of AADT per mile. The test of significance also failed to demonstrate the existence of any statistical significance in the differences in the mean rates of fatalities and injuries with regard to the V/SF ratio variable.

TABLE 3 State-Level Mean Values of Data

Variable Name	Mean Value	Std. Dev.	Minimum	Maximum
Pop. of Driving Age (1000)	3865.30	4250.80	338.0	22988.0
Licensed Drivers (per 1000 Residents)	682.70	51.40	558.0	812.0
Registered Vehicles (1000)	3763	3914	247.0	22253.0
Vehicle-Kilometers of Travel (10^6)	69835	72967	6474	415341
Annual Ave. Daily Traffic per Kilometer	1054	763	117	3378
% V/SF Ratio > 0.7	0.36	0.38	00.0	1.6
% PSR ≥ (3.5 - 5.0)	3.45	1.71	0.9	9.1
Safety Exp. ($1000)	80703.00	125491.00	4401.0	861451.0
Adm. Exp. ($1000)	81324.00	91302.00	8946.0	506511.0
Fatality Rate	1.22	0.28	0.72	1.76
Injury Rate	88.8	23.8	51.3	161.7

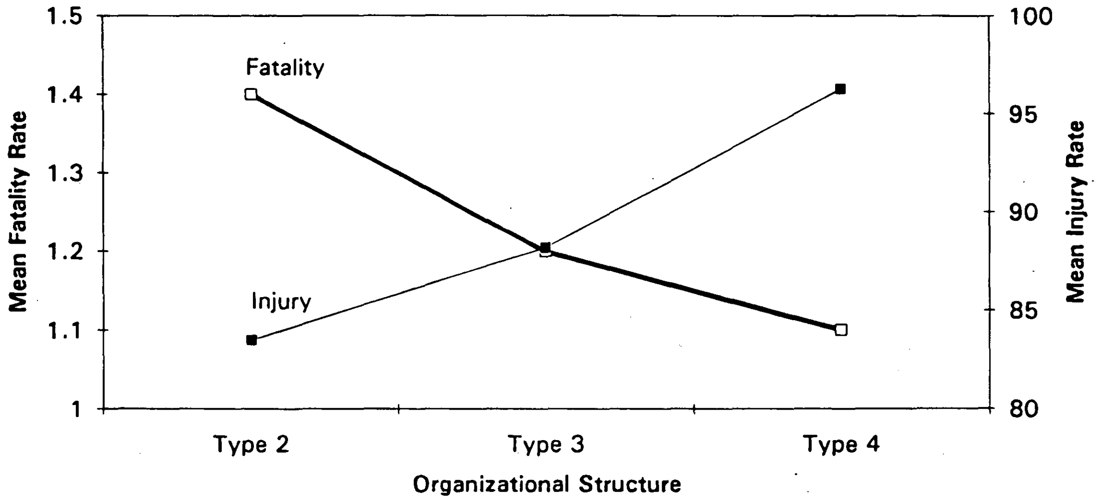


FIGURE 1 **Upper-management organizational structure and highway fatality and injury rates.**

The effects of network conditions on the mean rate of highway fatalities and injuries was examined by using the (PSRs) for the highway network. Only those sections of the network with PSRs of 3.5 to 5.0 (good or excellent), computed as a percentage of the total urban and rural road mileage, were considered in the analysis. The result indicated that as the percentage of highway mileage with high PSRs increased, the fatality rates also increased, whereas the injury rates decreased. This increase in fatality rates may again be caused by a likely increase in travel speed when pavement conditions on a highway network are improved. Higher travel speeds consequently affect the severity of accidents (*16*).

The increase in the fatality rates and the decrease in the injury rates, however, were not strongly correlated with the increase in the percentage of highway mileage with good pavement conditions ($r_{xy} = 0.199$ and -0.131 for fatality and injury rates, respectively). The test of significance also quantified this lack of a strong relationship when no statistically significant difference was found to exist between the mean fatality or injury rates for the states with the lowest and those with the highest percentages of highway mileage with PSRs of 3.5 to 5.0.

An analysis of the impacts of SDOT expenditures on safety on highway fatality and injury rates revealed that as the level of expenditures on safety increased, fatality rates decreased and injury rates increased (Figure 3). The point elasticities of the fatality rates (slope of the fatality curve), however, decreased as the level of expenditures on safety increased. The analysis showed that the reduction in the fatality rate was most elastic with safety expenditures in the range of $30 million or less. Expenditures in the range of $50 million to $100 million on highway safety had the second highest effect on reducing fatality rates. Results of a statistical test of significance also showed that a significant difference in fatality rates existed between the SDOTs that annually spent $50 million or less and

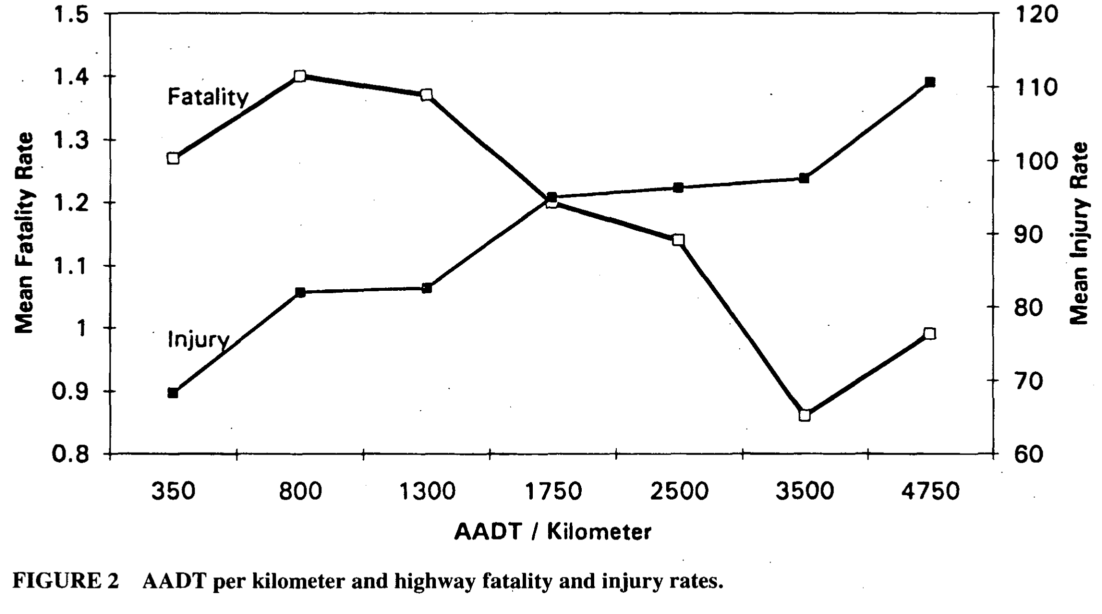


FIGURE 2 **AADT per kilometer and highway fatality and injury rates.**

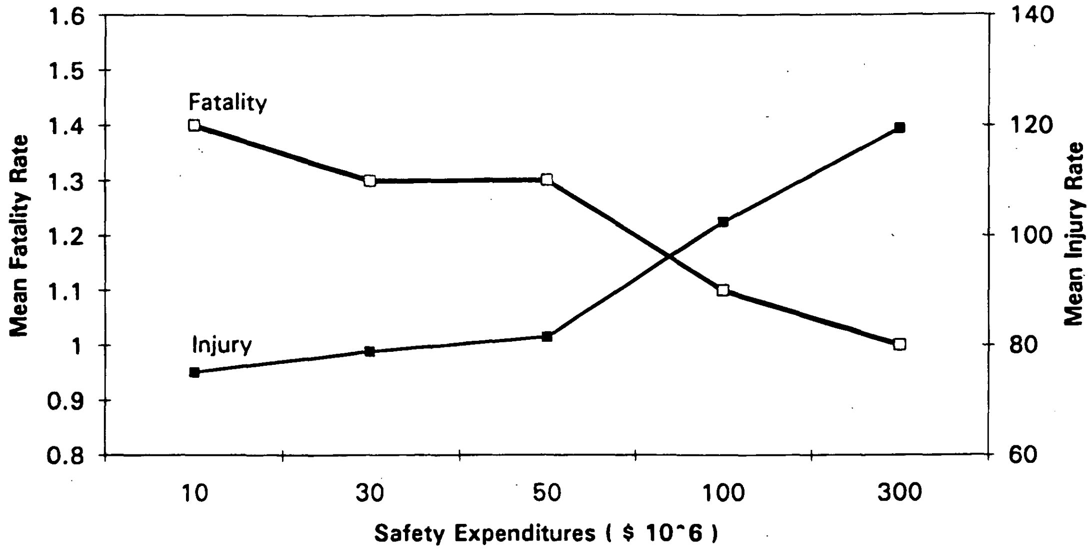


FIGURE 3 Safety expenditures and highway fatality and injury rates.

those that spent $100 million or more on highway safety. The reductions in the fatality rates for the states that spent $100 million or more on safety-related projects were significant at the 95 percent level of significance ($\alpha = 0.05$).

The level of expenditures on SDOT administration, on the other hand, produced no tangible reductions in highway fatality rates. The difference in fatality rates for states that spent $70 million or less and those that spent $100 million or more on administration was not statistically significant, even at the 80 percent significance level. It may therefore be safe to conclude that expenditures on highway safety have a much more pronounced impact on the reduction of highway fatalities than do expenditures on SDOT administration.

However, to assume that safety expenditures alone could improve road safety may also be premature. For example, to examine one point, it was hypothesized that the top 10 states with the lowest fatality rates spent significantly larger percentages of their total annual budgets on highway safety than the bottom 10 states with the highest rates of highway fatalities. The answer was a surprising no. The difference in the mean percentage of expenditures on highway safety for the two groups of states was not statistically significant even at the 30 percent significance level (the computed z value was 0.36, indicating only a 28 percent chance of significance). The reason for this was because of the problem associated with data aggregation. When safety expenditures are aggregated for 10 states (in each group), the largest portion of variations in these expenditures (the within-group variations) is lost to the aggregation process. What remains, the between-group variations, is only a small portion of the total variations (*17,18*). A glance through the data in Tables 1 and 2 confirms this statement.

How, then, do SDOTs distribute their limited budgetary funds among their competing capital, maintenance, administrative, and safety activities? Are the expenditures on highway safety, for example, made in relation to certain physical, geographical, traffic, or the upper-management hierarchy structures of SDOTs?

An exhaustive number of search, sort, and classification procedures failed to demonstrate the existence of any significant pattern in the distribution of budgetary funds among the competing activities by the 50 SDOTs. Among the 15 states with the lowest fatality rates, for example, the allocation of safety expenditures (as a percentage of the state's total budget) ranged from a low of 1.2 percent to a high of 9.9 percent, a difference of 725 percent. Similarly, among the 15 states with the worst highway fatality rates, expenditures on safety ranged between 1.7 and 9.4 percents, a difference of more than 450 percent. Enormously large variations in the physical size, geographical locations, traffic characteristics, and network pavement conditions were also observed among the states within either the group with the lowest fatality rates or the group with the highest fatality rates.

Variance Component Analysis

To determine the relative contributions of organizational structures, traffic, network conditions, expenditures on safety, and expenditures on SDOT administration (independent variables) to variations in fatality and injury rates (dependent variables), the variance component analysis method was used.

This statistical technique estimates the contribution that each of the independent variables makes to the overall variability in the dependent variable. The maximum likelihood procedure is used to compute the estimates of the variance components. The procedure iterates until the log-likelihood objective function converges. The objective function is $l_n(|V|)$, where

$$V = \sigma_0^2 + \sum_{i=1}^{n} \sigma_i^2 \chi_i \chi_i'$$

where

σ_0^2 = residual variance,
n = number of random variables in the model,
σ_i^2 = variance components, and
X_i = part of the design matrix for the random effect of variable i.

The results of the analysis, as presented in Table 4, indicate that the variances in the fatality and injury rates attributable to the factor of AADT per mile are by far larger than those attributable to the

TABLE 4 Maximum Likelihood Variance Component Estimates

Independent Variable	Relative Contribution to Dependent Variable Variance	
	Fatality Rates[a]	Injury Rates[b]
AADT per kilometer	174.631	1008.425
V/SF Ratio	34.257	194.673
Present Serviceability Rating, PSR	--	170.817
Safety Expenditures	7.946	--
Administrative Expenditures	2.471	--

[a] No. of iterations = 8,
[b] No. of iterations = 2.

V/SF ratio, PSR, expenditures on, safety, and expenditures on SDOT administration. The factor of traffic volume (which dictates the travel speed) seems to be the most important variable affecting highway fatality and injury rates. It seems that heavy volumes of traffic, as the AADT per mile indicates, are the main reason behind the high safety performances (low fatality rates) of the small, densely populated states of the Northeast in general and New England in particular. It is also noteworthy that the relative contribution of expenditures on safety to variations in the fatality rates is more than 3 times that of expenditures on SDOT administration.

CONCLUSIONS

The purpose of the present study was to examine the impacts and to evaluate the extent of causal relationships existing between SDOT management structures, traffic variables, network conditions, and budgetary allocations for safety and administration and highway fatality and injury rates. The study's findings indicated that attempts to achieve a reduction in both the highway fatality and injury rates simultaneously are similar to performing tasks that are mutually exclusive. Any successful attempt at saving a life will add one to the list of injuries, unless the accident can be prevented from happening altogether. States that experienced the lowest rates of fatalities suffered from the highest rates of injuries.

States without a board or commission in their upper-management organizational structures experienced the lowest rates of highway fatalities (and the highest injury rates, as was the pattern). Likely reasons may include a shorter link in the decision-making process or a possible reduction in the conflicts of interests among the members of upper management. Chance, however, may be the most likely answer for the existence of this relationship.

Traffic volume, as represented by the AADT per mile, demonstrated the strongest negative association with highway fatality rates. An increase in AADT per mile was accompanied by a significant (statistically) decrease in fatality rates and a significant increase in injury rates. The influence of the V/SF ratio on highway fatality and injury rates, although similar to that of AADT per mile, was less pronounced than that of AADT per mile. PSR, which represents network pavement conditions, was also positively correlated with fatality rates and was negatively correlated with injury rates, but to a much lesser extent.

Analysis of the impact of the level of expenditures on highway safety indicated that increases in safety expenditures resulted in a significant reduction in fatality rates. Increases in the levels of expenditure on SDOT administration, on the other hand, produced no tangible reductions in fatality rates.

Exhaustive search, sort, and classification attempts failed to highlight the existence of any patterns among the 50 states with regard to the expenditure of budgetary funds between the competing activities of safety and administrative.

The application of variance component analysis, performed to determine the relative contribution of the causal factors to variances in the fatality and injury rates, further supported the findings presented earlier.

From the study it can be concluded that although expenditures on highway safety could significantly reduce highway fatalities (which expenditures on administration do not), safety expenditures alone may not be enough to achieve the level of highway safety desired by SDOT management. Highway safety, as a literature search and the data in this paper indicate, seems to be affected by a host of socioeconomic, political, physical, and geographical variables. Factors such as the sparseness of the population and activities both within and between urban areas; lengths of trips; the daytime or nighttime distributions of trips; the percent distributions of inter- and intracity travel; enforcement policies concerning alcohol, drug, and seat-belt use; congestion levels and frequencies; speed limits and speed limit enforcement policies; the homogeneity and age compositions of drivers and the road-user population; the level of road-user education and awareness; and the compliance of road users with traffic rules and regulations may all have significant effects on highway safety, especially at the regional or state levels.

Research is urgently needed to take a microscopic look at the factors that make the small states of the Northeast in general and New England in particular experience the lowest rates of highway fatalities in the nation. Is it because these states are densely populated? Do their small sizes cause intercity trips to be short in length and time? Is it because they have more homogeneous driver populations or stiffer penalties for driving under the influence of drugs and alcohol? Or is it because of a more appropriate distribution of budgetary funds between safety and SDOT administration?

Is it not feasible for the states that share similar regional, physical size, population, and other socioeconomic characteristics to gain from the experiences of those in the same region with the best safety records? Many more logical and unanswered questions may also be posed.

ACKNOWLEDGMENTS

The authors thank the Research Center of the College of Engineering and Petroleum, Kuwait University, for partial support of the study. Thanks are also due to Mrs. Mini and Mrs. Daisy, Civil Engineering Department, Kuwait University, for typing the manuscript.

REFERENCES

1. *Highway Statistics.* FHWA, U.S. Department of Transportation, 1991.
2. Baker, S., B. O'Neill, and R. Karpf. *The Injury Fact Book.* Lexington Press, Lexington, Mass. 1984.
3. *Accident Facts, 1987 Edition.* National Safety Council, March 1987.
4. Donaldson, G. A. Safety Spending: Usually Begrudged, Often Misallocated. Highway Safety at the Crossroads. *Proc., ASCE Specialty Conference,* San Antonio, Tex., ASCE, 1988.
5. Controller General of the United States. *Management Actions Needed to Improve Federal Highway Safety Programs.* Report to the Congress. U.S. General Accounting Office, Washington, D.C., 1976.
6. Hauer, E., and A. S. Hakkert. The Extent and Some Implications of Incomplete Accident Reporting. *Presented at 66th Annual Meeting of the Transportation Research Board, Washington, D.C., 1987.*
7. Hauer, E., and B. Persaud. *The Problem of Identifying Hazardous Locations Using Accident Data.* Safety Research Group, Department of Civil Engineering, University of Toronto, Toronto, Ontario, Canada, 1986.
8. Koushki, P. A., and F. Balghunaim. Determination and Analysis of Unreported Road Accidents in Riyadh, Saudi Arabia. *Journal of Engineering Sciences,* Vol. 3, No. 1, 1991.
9. Ivey, D. L. Highway Safety: Moving from Fantasy to Reality. Highway Safety at the Crossroad. *Proc., ASCE Specialty Conference,* San Antonio, Tex., ASCE, 1988.
10. Sinha, K. C., T. F. Fwa, and I. M. Mouaket. New Tools and Techniques for Highway Maintenance Management. In *Transportation Research Record 1262,* TRB, National Research Council Washington, D.C., 1990.
11. Poister, T. H., L. G. Nigro, and R. Bush. NCHRP Report 163: *Innovative Strategies to Upgrade Personnel in State Transportation Departments.* TRB National Research Council, Washington, D.C., 1990.
12. *NCHRP Research Project Statement 20–24 (7).* NCHRP, FHWA, U.S. Department of Transportation, 1991.
13. *State Department of Transportation—Strategies for Change.* NCHRP Project 20–24 (9). NCHRP, FHWA, U.S. Department of Transportation, 1993.
14. Assimakopoulos, V., B. Stergio, and A. Konida. Trends in Road Safety Process: The Greek Traffic System. *Journal of Safety Research,* Vol. 23, 1992.
15. Koushki, P. A., J. L. Hulsey, and L. Raad. *Revisiting Management Information Systems for the Allocation of Funds for Highway Activities/Projects/Maintenance.* Final Research Report, Project TNW-92-05. Transportation Northwest, University of Washington, Oct. 1991.
16. *Sharing the Main Street.* Road and Traffic Authority, Federal Office of Road Safety, Canberra, Australia, 1993.
17. Theil, G. *Linear Aggregation of Economic Relations.* North-Holland Publishing Company, Amsterdam, 1954.
18. Stopher P. R., and T. E. Lisco. Modeling Travel Demand: A Desegregate Behavioral Approach, Issues and Applications. Transportation Research Forum, 1970.

Publication of this paper sponsored by Committee on Transportation Safety Management.

Concerns of Texas-Mexico Border Communities

RAFAEL F. PEZO

With or without the North American Free Trade Agreement an increase in trade between the United States and Mexico is imminent. This increase in commerce will lead to an increase in traffic, which, in turn, could create problems in terms of highway safety, especially along the Texas-Mexico border. One important issue concerning traffic safety is addressed specifically, public concerns. A questionnaire was prepared and distributed throughout the Texas-Mexico border region to determine the public's main concerns and opinions regarding traffic safety and present driving conditions in the border region. Conclusions were drawn from a total of 724 responses received.

Approximately 70 percent of freight shipped between Texas and Mexico is transported via the highway system (*1*). With or without the North American Free Trade Agreement (NAFTA) trade between the United States and Mexico will increase, and this increase will lead to an increase in traffic, which in turn could create problems in terms of traffic safety, especially along the Texas-Mexico border. To better understand the dynamics of increased trade and its impacts on traffic safety, uniformity and harmonization of U.S. and Mexican regulations in such areas as vehicle safety standards, weight and dimension limitations, highway system and signage, hazardous materials, licensing requirements, and insurance need to be discussed (*2*). Another important traffic safety-related issue that needs to be discussed is law enforcement. The level of safety of the highway system could be jeopardized if adequate measures are not taken in a timely fashion. Maintaining a high level of safety is of great concern, because traffic-related accidents bring consequences (i.e., fatalities, injuries, congestion, and insurance rates) that are major public problems with major economic, social, and political implications (*3*).

Initial steps have been taken to tackle this critical issue. First, public concerns with respect to traffic safety along the border have been determined, and the accident records of the border communities of Texas have been reviewed. To identify public concern a questionnaire was distributed along the border region, and the responses were analyzed.

DESCRIPTION OF QUESTIONNAIRE

A survey consisting of 12 questions was developed in both English and Spanish. The purpose of the survey was to determine public opinions concerning traffic safety on both sides of the Texas-Mexico border. The questionnaire was distributed in shopping malls, ports of entry, university campuses, parks, and truck stops. Trucking companies were visited to target commercial drivers. The University of Texas Pan-American and Laredo State University also distributed hundreds of questionnaires to the general public in their respective border areas. Figures 1 and 2 present the questionnaire as distributed in English and Spanish, respectively.

CHARACTERISTICS OF POPULATION SURVEYED

The population surveyed represented a wide variety of people with different backgrounds, ages, genders, nationalities, and geographic localities within the border region. Of the 724 responses, 51.5 percent were from the El Paso/Ciudad Juarez area, 2.2 percent came from the Del Rio/Ciudad Acuña/Eagle Pass/Piedras Negras area, and 46.3 percent were from the Rio Grande Valley area.

In regards to their language proficiency, 53.1 percent claimed to be bilingual, 26.9 percent claimed to understand only English, and 20.0 percent claimed to understand only Spanish. Additionally, 35.1 percent claimed to be directly involved in the trucking industry, whereas 64.9 percent were not. Texas licensees made up 52.9 percent of those surveyed, 29.0 percent were Mexican licensees, and 10.6 percent had licenses from U.S. states other than Texas. The remaining 7.5 percent of those surveyed had no drivers license.

DATA COLLECTION AND FINDINGS

Upon the return of the questionnaires the responses were entered into a data base by using SAS, the Statistical Analysis Software (*4*). A binary code was used to clearly identify each response. Thereafter, SAS was also used to analyze the data base. In general, the analysis consisted of tabulating the responses and comparing them on the basis of the different characteristics of the population surveyed.

From the analysis of the data base, public concerns were determined. The following are the general results (values are percentage of the survey population) obtained for Questions 8 through 12.

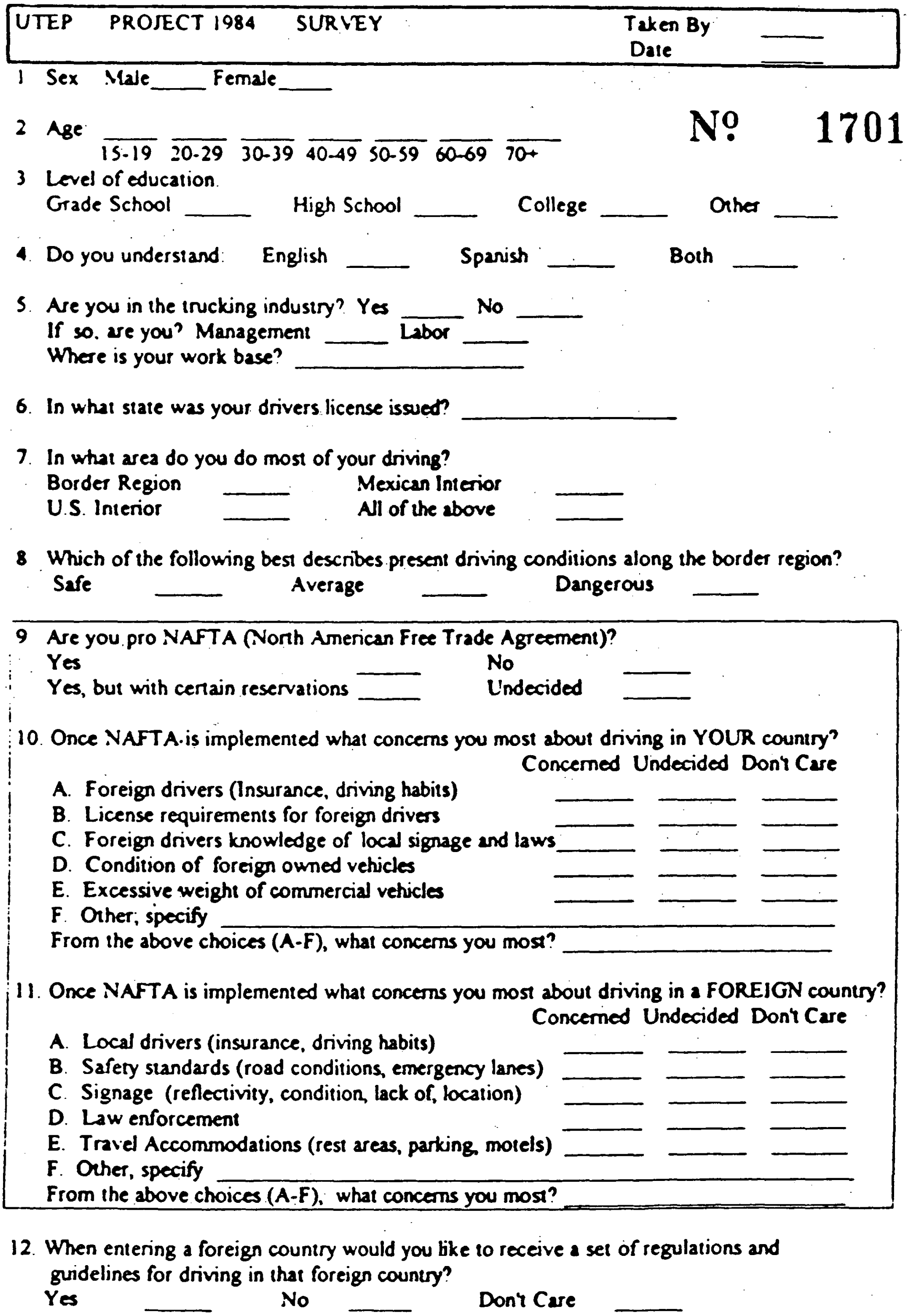
Question 8: What best describes present driving conditions along the border region?

Response	*General Population (percent)*	*Texas Licensees (percent)*	*Mexican Licensees (percent)*
Safe	5.7	7.8	3.8
Average	56.2	58.7	51.9
Dangerous	38.1	33.4	44.3

In general it was found that 56.2 percent of the population surveyed believe that present driving conditions at the border

---

Center for Geotechnical and Highway Materials Research, University of Texas at El Paso, El Paso, Tex. 79968-0516.



UTEP PROJECT 1984 SURVEY

Taken By ____

Date ____

№ 1701

1 Sex Male ____ Female ____

2 Age ____ ____ ____ ____ ____ ____ ____

15-19 20-29 30-39 40-49 50-59 60-69 70+

3 Level of education

Grade School ____ High School ____ College ____ Other ____

4. Do you understand: English ____ Spanish ____ Both ____

5. Are you in the trucking industry? Yes ____ No ____

If so, are you? Management ____ Labor ____

Where is your work base? ________

6. In what state was your drivers license issued? ________

7. In what area do you do most of your driving?

Border Region ____ Mexican Interior ____

U.S. Interior ____ All of the above ____

8 Which of the following best describes present driving conditions along the border region?

Safe ____ Average ____ Dangerous ____

9 Are you pro NAFTA (North American Free Trade Agreement)?

Yes ____ No ____

Yes, but with certain reservations ____ Undecided ____

10. Once NAFTA is implemented what concerns you most about driving in YOUR country?

	Concerned	Undecided	Don't Care
A. Foreign drivers (Insurance, driving habits)	____	____	____
B. License requirements for foreign drivers	____	____	____
C. Foreign drivers knowledge of local signage and laws	____	____	____
D. Condition of foreign owned vehicles	____	____	____
E. Excessive weight of commercial vehicles	____	____	____

F. Other, specify ________

From the above choices (A-F), what concerns you most? ________

11. Once NAFTA is implemented what concerns you most about driving in a FOREIGN country?

	Concerned	Undecided	Don't Care
A. Local drivers (insurance, driving habits)	____	____	____
B. Safety standards (road conditions, emergency lanes)	____	____	____
C. Signage (reflectivity, condition, lack of, location)	____	____	____
D. Law enforcement	____	____	____
E. Travel Accommodations (rest areas, parking, motels)	____	____	____

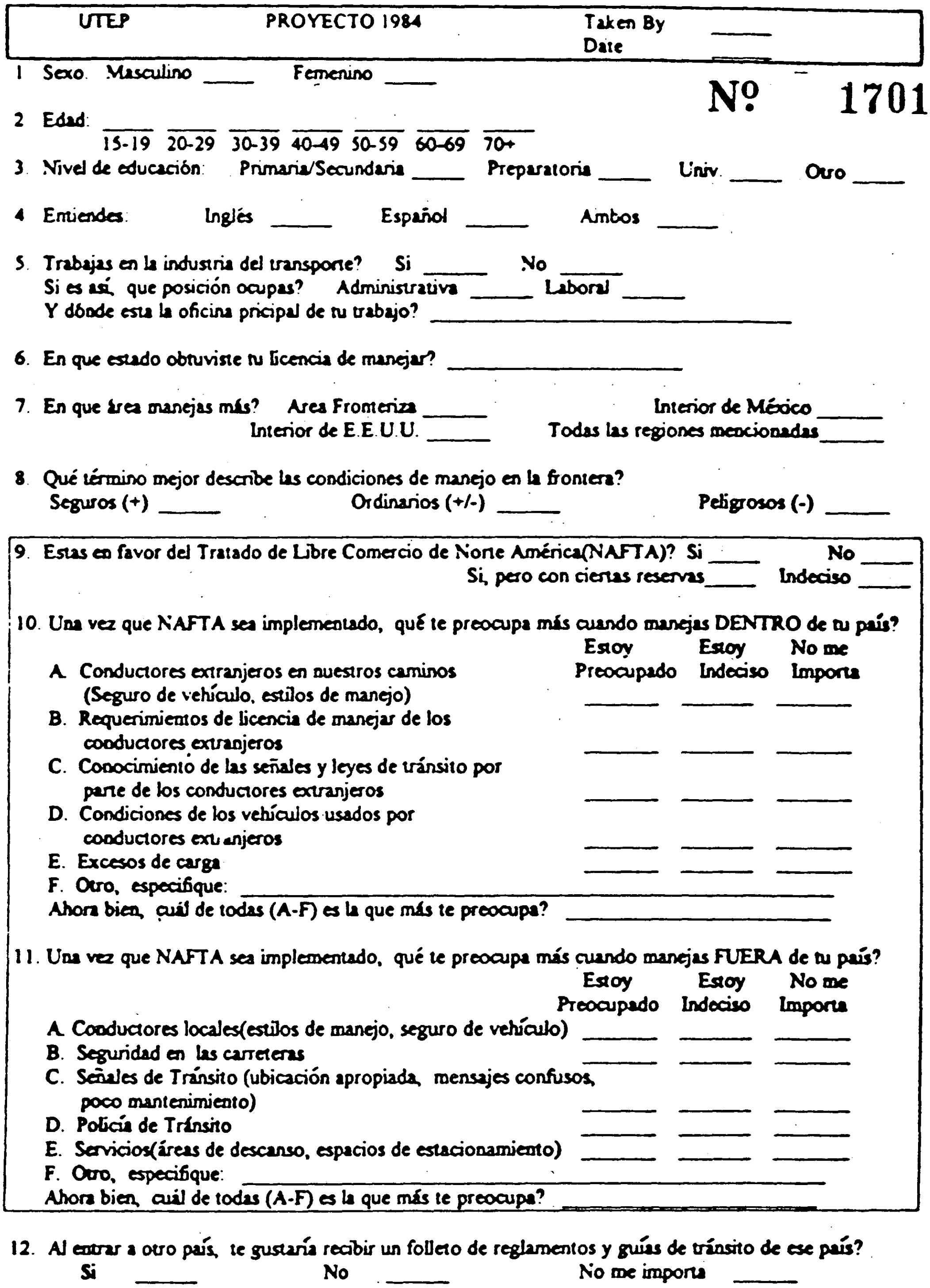
F. Other, specify ________

From the above choices (A-F), what concerns you most? ________

12. When entering a foreign country would you like to receive a set of regulations and guidelines for driving in that foreign country?

Yes ____ No ____ Don't Care ____

FIGURE 1 English version of questionnaire.



UTEP PROYECTO 1984 Taken By ____
Date ____

Nº 1701

1 Sexo. Masculino ____ Femenino ____

2 Edad: ____ 15-19 ____ 20-29 ____ 30-39 ____ 40-49 ____ 50-59 ____ 60-69 ____ 70+

3 Nivel de educación: Primaria/Secundaria ____ Preparatoria ____ Univ. ____ Otro ____

4 Entiendes: Inglés ____ Español ____ Ambos ____

5. Trabajas en la industria del transporte? Si ____ No ____
Si es así, que posición ocupas? Administrativa ____ Laboral ____
Y dónde esta la oficina pricipal de tu trabajo? ____

6. En que estado obtuviste tu licencia de manejar? ____

7. En que área manejas más? Area Fronteriza ____ Interior de México ____
Interior de E.E.U.U. ____ Todas las regiones mencionadas ____

8. Qué término mejor describe las condiciones de manejo en la frontera?
Seguros (+) ____ Ordinarios (+/-) ____ Peligrosos (-) ____

9. Estas en favor del Tratado de Libre Comercio de Norte América(NAFTA)? Si ____ No ____
Si, pero con ciertas reservas ____ Indeciso ____

10. Una vez que NAFTA sea implementado, qué te preocupa más cuando manejas DENTRO de tu país?

	Estoy Preocupado	Estoy Indeciso	No me Importa
A. Conductores extranjeros en nuestros caminos (Seguro de vehículo, estilos de manejo)	____	____	____
B. Requerimientos de licencia de manejar de los conductores extranjeros	____	____	____
C. Conocimiento de las señales y leyes de tránsito por parte de los conductores extranjeros	____	____	____
D. Condiciones de los vehículos usados por conductores extranjeros	____	____	____
E. Excesos de carga	____	____	____

F. Otro, especifique: ____
Ahora bien, cuál de todas (A-F) es la que más te preocupa? ____

11. Una vez que NAFTA sea implementado, qué te preocupa más cuando manejas FUERA de tu país?

	Estoy Preocupado	Estoy Indeciso	No me Importa
A. Conductores locales(estilos de manejo, seguro de vehículo)	____	____	____
B. Seguridad en las carreteras	____	____	____
C. Señales de Tránsito (ubicación apropiada, mensajes confusos, poco mantenimiento)	____	____	____
D. Policía de Tránsito	____	____	____
E. Servicios(áreas de descanso, espacios de estacionamiento)	____	____	____

F. Otro, especifique: ____
Ahora bien, cuál de todas (A-F) es la que más te preocupa? ____

12. Al entrar a otro país, te gustaría recibir un folleto de reglamentos y guías de tránsito de ese país?
Si ____ No ____ No me importa ____

FIGURE 2 Spanish version of questionnaire.

region are average, whereas 38.1 percent believe that conditions are dangerous. Only 5.7 percent believe that conditions are safe. It also appears that Texas licensees have a better perception of the driving conditions than Mexican licensees. Likewise, those involved in the trucking industry have a better perception of the driving conditions on the border than those not in the trucking industry.

Question 9: Are you pro NAFTA?

Response	*General Population (percent)*	*Texas Licensees (percent)*	*Mexican Licensees (percent)*
Yes	36.0	28.0	57.4
Yes, but with certain reservations	27.1	29.5	21.1
No	17.5	20.6	6.2
Undecided	19.4	21.9	15.3

These responses revealed that the majority of the population surveyed is in favor of NAFTA, that is, 63.1 percent. Of the remaining, 19.4 percent are undecided and 17.5 percent are opposed.

In addition, it was found that those who are bilingual and those who only understand Spanish are much more in favor of NAFTA than those who only understand English. Similarly, those involved in the trucking industry show more skepticism concerning NAFTA than those not involved. Those with a Mexican license are by far more in favor of NAFTA than those with a Texas license.

Question 10: Once NAFTA is implemented what concerns you most about driving in your country?

Response	*General Population (percent)*	*Texas Licensees (percent)*	*Mexican Licensees (percent)*
A. Foreign drivers (insurance, driving habits)	44.2	54.4	32.9
B. License requirements for foreign drivers	6.4	6.3	6.0
C. Foreign drivers' knowledge of local signage and laws	19.9	10.5	37.7
D. Condition of foreign-owned vehicles	15.4	19.6	7.8
E. Excessive weight of commercial vehicles	8.9	3.2	10.2
F. Other	5.2	6.0	5.4

When asked about the main concerns while driving in their own country, the general population surveyed stated that foreign drivers, their insurance, and their driving habits concerned them most. These concerns were found to be more profound for the bilingual and English-speaking groups. Of the Spanish-speaking population surveyed, the main concern was foreign drivers' knowledge of local signage and laws. The secondary concern of the English-speaking population was the condition of foreign-owned vehicles. Excessive weights of commercial vehicles were of great concern for the Spanish-speaking population when driving in their own country. In general, Texas licensees show more skepticism of foreign drivers and their vehicles than their Mexican counterparts.

Question 11: Once NAFTA is implemented what concerns you most about driving in a foreign country?

Response	*General Population (percent)*	*Texas Licensees (percent)*	*Mexican Licensees (percent)*
A. Local drivers (insurance, driving habits)	24.9	25.5	25.1
B. Safety standards (road conditions, emergency lanes)	29.6	27.9	31.9
C. Signage (reflectivity, condition, lack of, location)	10.3	2.0	22.7
D. Law enforcement	29.2	41.0	12.9
E. Travel accommodations (rest areas, parking, motels)	2.7	0	4.9
F. Other	3.3	3.6	2.5

When asked about the main concerns while driving in a foreign country, the general population surveyed declared that safety standards and law enforcement concerned them most. These concerns were found to be more profound for both the bilingual and English-speaking groups, whereas the Spanish-speaking population expressed concerns relating to safety standards. The secondary concern for bilingual and English-speaking populations was safety standards; however, the Spanish-speaking population was concerned about signage. Local drivers, their insurance, and their driving habits were also a great concern of bilingual and English-speaking populations. For those involved in the trucking industry and Texas licensees, the biggest concern was by far law enforcement when driving in a foreign country. Mexican licensees claimed to be more concerned about signage when driving in a foreign country.

Question 12: When entering a foreign country would you like to receive a set of regulations and guidelines for driving in that foreign country?

Response	*General Population (percent)*	*Texas Licensees (percent)*	*Mexican Licensees (percent)*
Yes	86.6	82.2	94.3
Don't care	8.2	11.3	2.9
No	5.2	6.5	2.8

When asked if they would like to receive a set of regulations and guidelines for driving in that foreign country, the majority (86.6 percent) replied yes, which shows an overwhelming response to such a practice. In particular, the Spanish-speaking population and Mexican licensees responded positively to this question.

CONCLUSIONS

With or without NAFTA an increase in trade between the United States and Mexico is imminent. This increase in trade will lead to an increase in traffic, which, in turn, could create problems concerning traffic safety, especially along the Texas-Mexico border region. To tackle this critical issue a survey was conducted with the purpose of identifying public concerns with respect to traffic safety. As a result the following conclusions are drawn:

1. The majority believes that the present driving conditions along the border region are average. Drivers with a Texas license show better perceptions of these conditions than those with a Mexican license.

2. The majority is pro NAFTA. The bilingual individuals and those who only understand Spanish are much more in favor of NAFTA than those who only understand English.

3. When driving in their own country drivers with a Texas license are highly concerned about foreign drivers, their insurance and driving habits, and the condition of foreign-owned vehicles. The main concern of Mexican drivers is foreign drivers' knowledge of local signage and laws.

4. When driving in a foreign country law enforcement is by far the main concern of drivers with a Texas license and those involved in the trucking industry, whereas safety standards and signage are the main concerns of those with a Mexican license.

5. When driving in a foreign country the majority would like to receive a set of regulations and guidelines for driving in that foreign country. Drivers with a Mexican license in particular were receptive to this suggestion.

ACKNOWLEDGMENTS

The author expresses appreciation to the members of the Texas Department of Transportation, especially Susan Bryant, for support. Thanks are also extended to Phillip Lane of Laredo State University and Bret Mann of The University of Texas Pan American for their assistance in collecting the required data from their respective areas. Special thanks also go to the personnel of the Texas Department of Public Safety. Additionally, the author acknowledges the assistance of the staff of the Center for Geotechnical and Highway Materials Research at The University of Texas at El Paso, especially Soheil Nazarian, Ceci Garcia, Srihari Krishnaprasad, Linda Avery, Jennifer Eagan, Kelvin Kroeker, Michael Tavares, and Oscar Contreras.

REFERENCES

1. *Planning Activities along the Texas/Mexico Border.* Division of Highway Design, Texas Department of Transportation, Jan. 1993.
2. Michie, D. *Transportation Survey.* Institute for Manufacturing and Materials Management, University of Texas at El Paso, 1988
3. Petrucelli, E. Traffic Injury: A Public Health Concern Worldwide. *ITE Journal,* July 1991, pp. 15–20.
4. Schlotzhauer, S., and R. Littell. *Manual of SAS System for Elementary Statistical Analysis.* SAS Institute, Inc., Cary, N. C., 1987.

Publication of this paper sponsored by Committee on Transportation Safety Management.

Law Enforcement, Pedestrian Safety, and Driver Compliance with Crosswalk Laws: Evaluation of a Four-Year Campaign in Seattle

JOHN W. BRITT, ABRAHAM B. BERGMAN, AND JOHN MOFFAT

Enforcement of pedestrian right-of-way laws at uncontrolled crosswalk locations and its effect on driver compliance were evaluated from 1991 to 1994 through the cooperative efforts of the Harborview Injury Prevention and Research Center and the Seattle Police Department. Citywide, neighborhood and intersection-specific enforcement was evaluated by using a standardized crossing technique to provide drivers with opportunities to stop for a pedestrian. The rates of driver compliance before and after the programs were calculated by an independent observer. The evaluation suggests that targeting small areas may be as effective as citywide campaigns, that brief efforts may be as effective as longer programs, and that benefits to pedestrians from such enforcement in high-volume commuter corridors may be minimal. In light of the often contradictory results, expectations of traffic enforcement to improve pedestrian safety should remain modest. Behavioral and environmental factors that are more salient to the driver than even rather intensive enforcement efforts make it difficult to achieve a consistent positive effect. Continued research is recommended to identify the optimal use of limited traffic enforcement resources in the service of pedestrian safety.

Injuries from pedestrian-motor vehicle collisions were responsible for 5,500 deaths and thousands more injuries in the United States in 1993 (*1*). Elementary school children, older adults over age 65, and those impaired by alcohol are especially vulnerable (*2–9*). The role of law enforcement is one of the least studied of all potential mechanisms for reducing such injuries, yet law enforcement is routinely recommended as one of the essential strategies for prevention (*6,10,11*). Limited traffic enforcement resources, competing departmental priorities, and a lack of awareness of the problem's significance are three common barriers to the enforcement of pedestrian laws. The presence of a strong pedestrian safety program within the Seattle Police Department and its willingness to collaborate with the Harborview Injury prevention and Research Center provided a unique opportunity to investigate the potential safety benefit of one type of enforcement.

WHY CROSSWALKS?

Crosswalks at uncontrolled intersections were chosen as the focal point for the program because they provide a specific target location for law enforcement efforts aimed at increasing the compliance of drivers with pedestrian laws. They were also chosen because drivers' responses to the enforcement effort could be easily measured. In 1992 more than 800 of the 1,900 pedestrian injuries in Washington State occurred at intersections. Although drivers' willingness to stop for pedestrians is certainly not the only factor in collisions, it is a reasonable focus for law enforcement activities.

INTERVENTION

The First Step: Strengthening Crosswalk Law

In 1990 a coalition of safety groups, health professionals, citizen activists, and law enforcement representatives worked together to pass a stronger state crosswalk law. The law focused the attention of the public on pedestrian safety by changing the obligation of the driver from yield to stop when pedestrians were attempting to cross at legal crosswalk locations. The new law set the stage for a change in Seattle Police Department policy with respect to pedestrian law enforcement as well as the initiation of a public information campaign.

Changing Focus from Walkers to Drivers

Where pedestrian enforcement programs exist they often target the pedestrian violator, not the driver. For years Seattle had a well-deserved reputation for strict enforcement of jaywalking laws. A significant problem with such a focus is that the people who are most likely to receive citations, working-age adults, have the lowest risk of injury. It was believed that changing drivers' behaviors, if possible, would afford better protection to all walkers, especially those at higher risk.

In 1990 the Seattle Police Department policy toward pedestrian violations was changed to reflect a new emphasis on increasing the compliance of drivers with the legal rights of pedestrians, especially at intersections. The new policy encouraged traffic officers to issue at least two citations to driver violators for every ticket issued to a jaywalker. This policy led to a sharp reduction in jaywalking citations and a commensurate increase in citations to drivers. Approximately 300 to 500 citations related to pedestrian law enforcement are now written each month. The vast majority of these go to drivers. A jaywalking fine is $38, and a driver citation is $66.

J. W. Britt and A. B. Bergman, Harborview Injury Prevention and Research Center, 325 Ninth Avenue, ZX-10, Seattle, Wash, 98104-2499. J. Moffat, Traffic Safety Commission, 1000 South Cherry, P. O. Box 40944, Olympia, Wash. 98504-0944.

Public Information Campaign

With the passage of the new law came a renewed opportunity to focus attention on pedestrian safety through the media. Several pertinent articles appeared in the major Seattle newspapers, and a request was made for citizens to submit nominations for dangerous intersections. Hundreds of letters came in from residents throughout the city. The Engineering Department installed new warning signs and posters, and signs were distributed on buses throughout the city. Customized pamphlets and posters were produced and distributed. These pamphlets and posters informed the public about the seriousness of pedestrian injuries, the specifics of the new law, and the presence of the enforcement campaign.

Law Enforcement Efforts, 1990 to 1994

Four separate traffic enforcement campaigns were conducted by the Seattle Police Department over the course of the 4 years. Although there were differences between each campaign, they all shared the following design features:

1. A specific area of the city was identified to receive emphasized enforcement by traffic officers. The enforcement consisted of increased officer presence in the designated area, with the purpose of citing drivers who violated the crosswalk law.
2. A time line for the campaign was identified. The shortest campaign lasted 3 weeks; the longest lasted longer than 1 year.
3. Sentinel intersections were identified within the area. These intersections were used to measure the compliance of drivers with stopping for crossing pedestrians. Data on historic traffic volumes and posted speed limits were also available for each location.
4. Baseline measures of driver compliance were conducted before the initiation of the law enforcement efforts.
5. Follow-up measures of driver compliance were obtained after the law enforcement effort stopped.

EVALUATION METHODOLOGY

Compliance as Proxy Measure

Although the expected outcome of a successful safety campaign is injury reduction, direct links of program efforts to changes in injury rates are often difficult. Fluctuations in pedestrian-motor vehicle collisions (Figure 1) may be the result of many factors. Changes in the distributions of vulnerable walking groups, weather that leads to more or less walking, a downturn in the economy resulting in fewer vehicles on the roads, school closures, or increased busing of students may all affect walking, traffic exposure, and pedestrian collisions. For a specific community the actual numbers of collisions may be small and collisions may occur relatively infrequently. As a result other measures that are assumed to be related to injury are often used as proxy measures for a reduction in injuries. There is ample precedent for such observational measures with respect to the use of, for example, seat belts, child car seats, and bicycle helmets (*12,13*). In this case the willingness of drivers to stop for pedestrians in crosswalks was used to evaluate the enforcement effort.

Measuring Driver Compliance

A procedure was adapted from previous research (*14*) to measure driver compliance at crosswalks before the enforcement efforts of the traffic officers. A mock pedestrian approached and entered a crosswalk, attempting to make eye contact with the oncoming drivers. An accomplice stood back from the intersection, out of sight of the drivers, and used a handheld counter to tabulate the number of drivers who stopped so that the pedestrian could cross. To allow for those drivers who were too close to the intersection to safely come to a stop when the pedestrian stepped into the crosswalk, a braking "window" was measured in both directions from the crosswalk [46.6 m (153 ft) for a posted speed limit of 48.3 km/hr 30 (mph)]. This braking distance was calculated assuming no grade, dry pavement, and good tire tread (*15*). Only vehicles outside of this window when the pedestrian entered the crosswalk were considered to have had an adequate opportunity to stop; those vehicles outside of the window were not counted. Because a traffic engineer had screened potential intersections and excluded those with poor sight distances, the vast majority of vehicles had much more than this minimum distance in which to stop. Once able to cross, the pedestrian proceeded to the other side of the street and then repeated the procedure, coming back across the intersection in the same fashion as before. As the pedestrian approached the midpoint of the roadway, the compliance of drivers during the second half of the crossing (far side) could also be measured One hundred such pedestrian

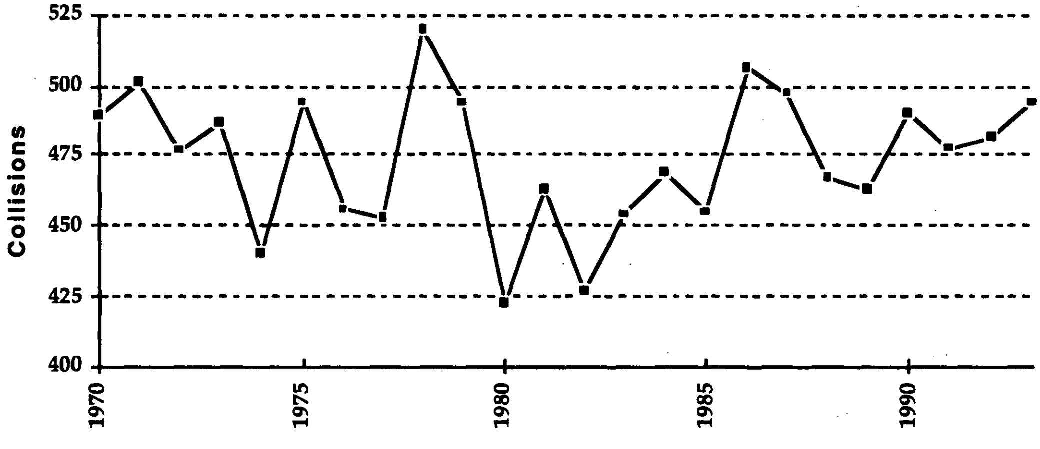


FIGURE 1 Seattle pedestrian—motor vehicle collisions.

motor vehicle conflicts were elicited at each sentinel intersection identified in the campaigns. To account for variability in compliance from other factors that might influence the drivers' willingness to stop, compliance was measured only during clear weather. The same pedestrian, time of day, and crossing technique were used each time. Because of known variations in traffic volumes around the weekend, compliance was measured only on Tuesdays, Wednesdays, and Thursdays.

Methodological Considerations

Pedestrian

A white, male, adult pedestrian was used for all studies. It is assumed that drivers who see pedestrians waiting to cross are likely to respond differently to women, joggers, older adults, children, the disabled, bicyclists, and so forth and the use of the same individual was intended to minimize variability from that effect.

Crossing Technique

A clear and consistent but nonaggressive crossing technique was used. The pedestrian was required to be in the crosswalk and not on the curb, required to attempt to make eye contact with oncoming drivers, and prohibited from taking additional steps into the first traffic lane until the vehicle slowed, indicating that the walker had been seen and would be allowed to pass. Observations of typical crossing techniques used by the public indicate a wide range of crossing behaviors, from waiting on the curb until cars came to a complete stop to very aggressive behaviors in which the pedestrian virtually walked into the paths of oncoming vehicles.

Compliance: Did They Stop?

Judgment as to whether a vehicle stopped entered into the compliance calculations because many vehicles slowed dramatically, sometimes to a crawl, but did not come to a complete stop. If the pedestrian proceeded and the observer judged that the driver would have stopped for the crossing pedestrian, this behavior was recorded as compliance.

Near-Side Versus Far-Side Compliance

In addition to other factors affecting compliance, it was assumed that drivers would behave differently with respect to pedestrians who were close or far away from them as their vehicles approached the crosswalk. Two obvious factors are the visibility and vulnerability of the pedestrian (a protected pedestrian at curbside versus one in the middle of the roadway). For these reasons near-side, far-side, and overall (near-side and far-side combined) compliance are distinguished. For the purposes of the present study near-side compliance represents the compliance of drivers approaching from the crossing pedestrian's left. For a four-lane (two-way) road there are two near-side lanes. For a two-lane road there is only one near-side lane per crossing. Far-side compliance represents the compliance of drivers approaching from the pedestrian's right as the center line of the roadway is crossed. Although compliance was measured in several ways during the campaigns, the near-side compliance at marked crosswalk locations was always reported. This represents the willingness of drivers to allow crossing pedestrians to get an initial "foothold" in the street at locations where the right-of-way of the pedestrian is the most unambiguous.

Choosing Sentinel Crosswalk Locations

The choice of which areas of the city to enforce driver compliance with crosswalk laws was made jointly by the Seattle Police Department and the Harborview Injury Prevention and Research Center. Harborview contacted community leaders throughout the city to determine their interest in having such a program in their community. The purpose of contacting members of the community was not to have citizens choose the target locations but to make certain that the residents would support an emphasized enforcement program should it occur in their community. The police department identified areas of the city where increased officer presence could be justified on the basis of other concerns, such as crime or speeding, in addition to pedestrian safety. It was easier to justify the use of scarce departmental resources if the targeted areas of the city had other needs for police resources as well. Once the areas of the city were identified, city traffic engineers assisted Harborview with identifying uncontrolled crosswalk locations whose configurations and locations would provide adequate visibility for both pedestrians and drivers and not contain any steep grades or curves that might adversely affect the reaction times or stopping distances of the drivers. Initially, a mix of two-lane and four-lane, marked and unmarked, as well as midblock intersections was identified. It was anticipated that higher-speed, multilane roadways in commuter corridors would elicit different levels of compliance than two-lane roads running through residential areas. It was also assumed that marking of the crosswalks, signs, beacons, visibility, and other factors would have an effect. Although actual pedestrian volumes were not measured, prospective sentinel locations were observed beforehand. Locations with a greater volume of pedestrian traffic were preferred.

RESULTS

Campaign 1: Citywide Focus, Summer 1990 to Fall 1991

Table 1 summarizes the first enforcement campaign to go into effect after the new crosswalk law passed the state legislature in 1990. The area identified for the program was the entire city. Twelve sentinel intersections, representing two-lane, four-lane, marked, unmarked, and midblock crosswalks in different parts of the city were identified by city traffic engineers. Although not every marked crosswalk was marked in the same way, all of the sentinel crosswalks were marked with either painted roadway stripes or an overhead beacon. In most cases both were present, as were advance warning signs to drivers. Law enforcement officers were unaware of the locations of the sentinel intersections. Approximately 3,600 citations were written to violating drivers throughout the city during this time. The overall level of compliance did not change (Table 1).

Campaign 2: Neighborhood Focus, September 1992 to January 1993

Table 2 summarizes the second enforcement campaign. As with the first campaign, 12 sentinel crosswalks were identified. No midblock

TABLE 1 Citywide Enforcement Campaign 1, Summer 1990 to Fall 1991

Location	Compliance Before [a]	Compliance After
Near-side (marked)	15%	19%
Near-side overall	15 %	13 %
Far-side overall	52 %	55%
Avg. overall	34%	34 %

[a] Percent is based on 100 opportunities for drivers to stop at each sentinel crosswalk.

crosswalks were included in this second effort. In the second campaign traffic officers were asked to focus their efforts in only five city neighborhoods. Although the boundaries of these neighborhoods are not legally defined, their locations are generally recognized, and street boundaries were set by the police department for the enforcement. It was hoped that this would result in a more concentrated enforcement effort than that during the previous citywide campaign. As in the first campaign traffic officers were unaware of the location of the sentinel intersections. The second campaign was significantly shorter than the first one lasting about 3 months. The largest increases in compliance were seen in the near-side lanes, with the greatest effect at marked intersections. Although there was noticeable improvement with respect to the baseline observation, the majority of vehicles were still not stopping for the pedestrian.

Campaign 3: Neighborhood Focus, July to October 1993

Table 3 summarizes the results of the third enforcement effort. This campaign also identified 12 intersections in five neighborhoods and lasted 3 months, but in contrast to previous efforts, it used only marked sentinel intersections to measure the effect. Only the compliances of drivers approaching the pedestrian's left side (near side) were measured. Getting a foothold in the crosswalk appeared to be the most difficult part of the crossing process, as evidenced from the previous campaign results. In contrast to previous efforts, citation information that allowed an assessment of the strength and distribution of the enforcement effort was available. The citation data indicated that 90 percent of the citations (436 of 487) had been written in only two of the five neighborhoods. Follow-up compliance observations were not performed at the other three locations because they were not considered to have received a significant intervention. The reasons for the skewed distribution of the enforcement effort are not clear but may have been the result of low pedestrian volumes in those neighborhoods.

The results are contradictory, with a dramatic decrease in compliance in Neighborhood 1 and an equally dramatic increase in compliance in Neighborhood 2. Neighborhood characteristics and enforcement patterns may have contributed to the results.

The first neighborhood was actually a portion of the downtown business corridor used heavily by commuters. Traffic volumes are high during afternoon peak volumes, pedestrian traffic at the sentinel crosswalk was sporadic, and none of the 286 citations issued during the campaign were issued at the sentinel intersection where compliance was measured. All citations were issued at surrounding intersections.

The second neighborhood largely comprised multifamily dwellings, such as apartment buildings or condominiums, mixed with small retail businesses. Pedestrian traffic was frequent, vehicular traffic volumes were less, and 78 percent of the citations (117 of 150) issued in the neighborhood were written at the sentinel intersection where compliance was measured. The focus on this crosswalk by traffic officers was unintended but not surprising, since pedestrian activity is known to be frequent here.

Campaign 4: Intersection-Specific Enforcement, May to June 1994

Table 4 summarizes the most recent campaign, which identified two specific intersections for enforcement. Both intersections were on four-lane arterials with the same posted speed limit of 47 km/hr (30 mph) and had similar afternoon peak traffic volumes. Both intersections were marked with painted crosswalks, advance warning

TABLE 2 Neighborhood Enforcement Campaign 2, September 1992 to January 1993

Location	Before	After
Near-side marked	23 %	36 %
Near-side unmarked	3 %	4 %
Near-side overall	11 %	18 %
Far-side overall	47 %	47 %
Compliance Overall	28 %	32 %

TABLE 3 Neighborhood Enforcement Campaign 3, July to October 1993

Neighborhood [a]	Before	# Citations [b]	After
1	19 %	286	7 %
2	9 %	150	30 %

[a] Two neighborhoods accounted for 90% of all citations.

[b] Citations written in the neighborhood, not necessarily at the sentinel crosswalk.

TABLE 4 Intersection Enforcement Campaign 4, April to June 1994

Intersection	Dates	Before	During[a]	After	# Citations
1	4/19-5/9	24%	19%	15%	74
2	5/23-6/10	30%	54%	45%	50

[a] Traffic officers were asked to stop enforcement for one day so mid-program compliance could be measured.

signs to drivers, and an overhead flashing beacon. Near-side lane compliance was used to measure the effect of the 3-week program. Traffic officers were assigned to the intersections each of the 5 weekdays during the hours when traffic volumes were highest (4 to 6 p.m.). Traffic officers were asked to temporarily stop enforcement at the intersection during one afternoon of the 15-day program so that midprogram compliance could be determined.

The results from Intersection 1 suggest that enforcement made no difference in compliance, whereas the results from Intersection 2 seem to suggest the opposite.

The difficulty of linking these inconsistent results with the enforcement effort is compounded by the following: at Intersection 1 drivers received more tickets (74 versus 50), there were more average tickets per day of enforcement (7.4 versus 6.25), and there were more days with traffic officers present (10 versus 8) than at Intersection 2, yet compliance at Intersection 2 increased, whereas compliance at Intersection 1 clearly decreased. During the enforcement of both intersections traffic officers were occasionally called away from their assignments for other competing departmental priorities, leading to these differences in the actual numbers of days that officers were present.

Even more surprising was that Intersection 2, which showed a decrease in driver compliance in this campaign, was the same sentinel intersection for Neighborhood 2 in Campaign 3, which demonstrated a dramatic increase in driver compliance after drivers received 117 citations.

A synthesis of all four enforcement campaigns is provided in Table 5. Near-side compliance measurements are displayed since this measure was obtained for every campaign. Exact citation information became available only for the two most recent campaigns, but this information is included when it is available. Although the programs differed with respect to their duration, location, and concentration of enforcement, all used the same method to measure driver compliance.

DISCUSSION OF RESULTS

First Campaign

The results of the first citywide campaign suggested that although enforcement and other public information efforts may have increased the awareness of drivers about pedestrian safety issues, there is no reason to believe that this campaign improved the safety of people trying to cross the street. Eighty-one of every 100 cars failed to stop for a pedestrian in the near-side lanes of the marked sentinel locations.

Second Campaign

The second campaign suggested that a more focused effort in discrete neighborhoods might be as useful as the initial citywide approach, even though it was conducted for only 3 months. Modest improvements in compliance were seen, especially in the near-side lanes of marked crosswalks. Although legal obligations on the part

TABLE 5 Summary of Crosswalk Enforcement Efforts, 1990 to 1994[a]

Campaign & Focus		Duration	# Citations	Before[b]	Mid-Prog	After
#1	City-wide	1 yr+	3600+ (est)	15%	—[c]	19%
#2	Neighborhood	3 Months	—	11%	—	18%
#3	Neighborhood (1)	3 Months	286	19%	—	7%
	Neighborhood (2)	3 Months	150	9%	—	30%
#4	Intersection (1)	3 Weeks	74	24%	—	15%
	Intersection (2)	3 Weeks	50	30%	—	45%

[a] Size of area enforced and sentinel intersections enforced varied with individual campaigns.

[b] Near-side compliance of marked crosswalks was measured in all campaigns. Compliance for Campaigns 1 & 2 represent averages of 12 sentinel crosswalks. Compliance for Campaigns 3 & 4 represent observations at single sentinel marked crosswalks.

[c] Signifies that the information was not collected for this campaign.

of the driver are the same for unmarked as well as marked crosswalks, unmarked intersections continue to elicit very poor compliance responses from drivers. After the enforcement campaign, drivers were still 9 times more likely to stop for a pedestrian in the near side of a marked intersection than the near side of an unmarked one (36 versus 4 percent).

Third Campaign

The third campaign raised issues about the characteristics of neighborhoods that are likely to benefit from enforcement as well as the practical limitations for conducting such a program from the point of view of traffic enforcement resources. The two sites analyzed varied markedly in their responses to enforcement. This campaign suggests that even intense enforcement conducted in downtown commuter corridors may not result in a safety benefit for the pedestrian. A location where more people live or where people walk frequently appears to predict better results. If drivers routinely encounter pedestrians, perhaps they will be more susceptible to law enforcement reminders.

The skewed distribution of enforcement suggests that extended programs covering several sites may find it difficult to achieve and maintain a high level of enforcement activity because of competing demands for law enforcement resources. Providing a consistent level of enforcement to intervention sites must be an ongoing and important aspect of future evaluations.

Fourth Campaign

The contradictory results of Campaign 4 indicate that success may vary markedly from intersection to intersection and suggest that, despite heavy ticketing, other environmental and perhaps behavioral factors are more salient to drivers than enforcement concerns. This is supported by the fact that the very same intersection where drivers received 117 citations over a 3 month period and that demonstrated a dramatic increase in driver compliance in one campaign (Campaign 3) showed a decrease in driver compliance with enforcement in Campaign 4. The results also suggest that the threshold level of ticketing necessary to achieve even transient changes in driver behavior is quite high.

Pattern of Enforcement

Although it was not a focus of the initial program the citation data available from the two most recent campaigns suggest that the intensity and distribution of the enforcement effort varied dramatically. In Campaign 3 the pattern of citations indicated two significant trends. First, 90 percent of the citations had been written in only two of the five neighborhoods originally identified for enforcement. The second trend that surfaced was that of the distribution of citations over time. As can be seen in Figure 2 the initial enforcement effort in campaign 3 was more intense, tapered off over time, and then increased again near the end of the campaign. Because

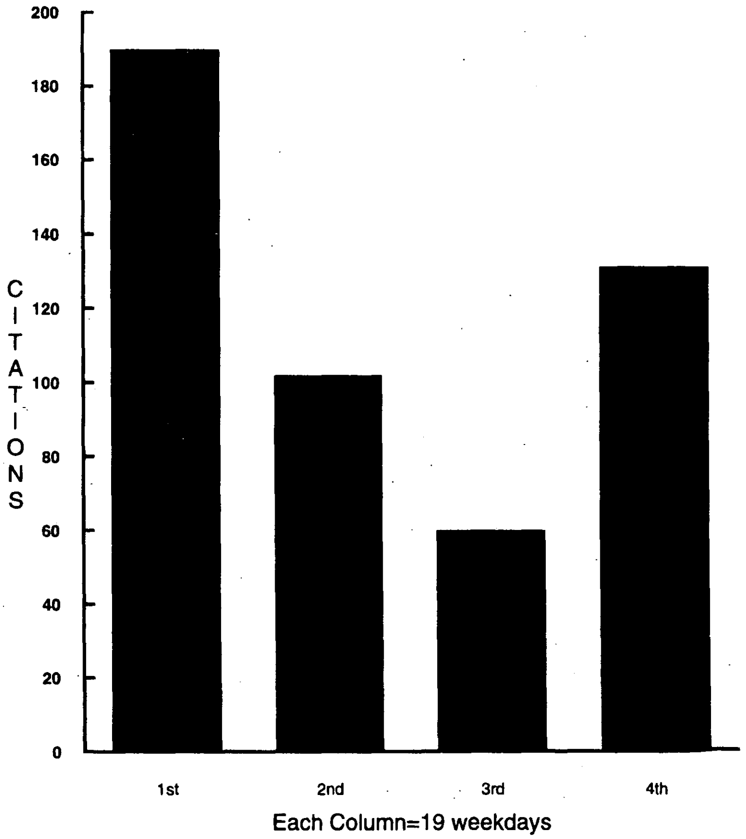


FIGURE 2 Enforcement pattern campaign 3.

Campaign 3 began and ended in the middle of a month, the timeline of Campaign 3 was divided into four equal 19-day periods for ease of comparison. The number of citations issued during each period is displayed.

In the fourth campaign, in which the focus was on daily enforcement at specific intersections, officers were also occasionally unable to be at their assigned locations. For Intersection 2 enforcement occurred on 8 of 14 program days (57 percent), and at Intersection 1 enforcement occurred on 10 of 14 days (71 percent).

Actual citations for the earlier campaigns were not available, but it is likely that such variation in enforcement with respect to location and time occurred in those programs as well. Although they are an expected part of the daily activity of law enforcement agencies, these fluctuations in the intensity and distribution of enforcement must be considered when implementing and evaluating future programs.

LIMITATIONS

Although efforts were made to control as many variables as possible that might affect driver compliance, the following factors may have added to the variability in the results.

- The braking window used to define an adequate braking distance for vehicles and for defining the pedestrian-motor vehicle conflicts assumed average driver reaction time, adequate brake and tire equipment, the posted speed limit, and dry, flat pavement.
- No actual counts of pedestrian volumes were made at the intersections. The regular presence of crossing pedestrians may lead to increased compliance.
- The compliance rates of grouped and single vehicles were not distinguished.
- The intensity of enforcement as well as its distribution varied.
- Streets with posted speed limits of 30 or 35 mph were used for the campaigns. The actual approach speeds of the vehicles were not measured.

CONCLUSION

The authors have been unable to demonstrate that law enforcement efforts directed at motorist violators of crosswalk laws significantly or consistently increase drivers' willingness to stop for pedestrians. It appears that even with a high degree of commitment on the part of law enforcement, the expectations from such programs should remain modest. If intense enforcement efforts aimed at drivers do not elicit a positive effect at marked crosswalks, it is difficult to imagine that they will be effective in locations where the pedestrian right-of-way is more ambiguous. Although there are few standards by which to judge the relative enforcement intensities of these campaigns, the authors are unaware of any law enforcement agency that has conducted and evaluated a more focused effort.

It appears that other uncontrolled factors were responsible for the wide fluctuations in compliance. Day-to-day speed and volume fluctuations and their behavioral effects on drivers may have a greater effect on compliance than even the most aggressive enforcement campaign. Further evaluations should be encouraged. Such evaluations may be able to account for some of this variability and determine whether and to what extent there is a positive effect.

These results, although discouraging, are by no means conclusive and should not be construed as relieving law enforcement agencies from playing an active role in pedestrian safety programs. Rather, they should focus attention on finding the most effective way to make use of such resources as part of a communitywide approach. Communities have legitimate concerns about increasing the safety of the walking public, and all stakeholders must consider carefully what can and cannot be done to provide an increased margin of safety for pedestrians.

ACKNOWLEDGMENTS

The program described here was made possible through the commitment of resources and personnel by the Seattle Police Department as well as the Seattle Engineering Department. The Centers for Disease Control and Prevention and the Office of Safety Countermeasures of NHTSA provided support that allowed the Harborview Injury Prevention and Research Center to conduct the evaluation.

REFERENCES

1. *Traffic Safety Facts 1992: A Compilation of Motor Vehicle Crash Data from the Fatal Accident Reporting System and the General Estimates System.* Report DOT HS 808 022. NHTSA, U.S. Department of Transportation, 1993.
2. Rivara, F. P. Child Pedestrian Injuries in the United States. Current Status of the Problem, Potential Interventions, and Future Research Needs. *American Journal of Diseases of Children* Vol. 144, No. 6, 1990, pp. 692–696.
3. Jonah, A. J., and R. G. Engel. Measuring the Relative Risk of Pedestrian Accidents. *Accident Analysis & Prevention,* Vol. 15, No. 3, 1983, pp. 193–206.
4. *Pedestrian Trip Making Characteristics And Exposure Measures.* Report FHWA/RD-85/074. FHWA, U.S. Department of Transportation, 1983.
5. Malek, M., B. Guyer, and I. Lescohier. The Epidemiology of Child Pedestrian Injury. *Accident Analysis & Prevention,* Vol. 22, No. 4, 1990, pp. 301–313.
6. Zegeer, C. *Synthesis of Safety Research: Pedestrians.* Report FHWA-SA-91-034. NHTSA, U.S. Department of Transportation, 1991.
7. Zegeer, C. V., J. C. Stutts, H. Huang, M. Zhou, and E. Rodgman. Analysis of Elderly Pedestrian Accidents and Recommended Countermeasures. In *Transportation Research Record 1405,* TRB, National Research Council, Washington, D.C., 1993, pp. 56–63.
8. Fell, J. C., and B. G. Hazzard. *The Role of Alcohol Involvement in Fatal Pedestrian Collisions.* Presented at the American Association for Automotive Medicine, 29th Annual Conference, Washington, D.C., Oct. 7 to 9, 1985, National Center for Statistics and Analysis and NHTSA.
9. *Traffic Collisions in Washington 1992: Data Summary and Highway Safety Problem Analysis.* Washington Traffic Safety Commission, 1993.
10. *Pedestrian Safety Law Enforcement Strategies Manual.* Report DOT HS 808 008 NTS-23. NHTSA, U.S. Department of Transportation, 1991.
11. Graham, J. Injuries from Traffic Crashes: Meeting the Challenge. *Annual Review of Public Health,* Vol. 14, 1993, pp. 515–543.
12. DiGuiseppi, C., F. Rivara, T. Koepsell, and L. Pollisar. Bicycle Helmet Use by Children: Evaluation of a Community-Wide Helmet Campaign. *Journal of the American Medical Association,* Vol. 262, No. 16, 1989, pp. 2256–2261.
13. Williams, A., J. Wells, A. Lund, and N. Teed. Observed Use of Automatic Seat Belts in 1987 Cars. *Accident Analysis & Prevention,* Vol. 21, No. 5, 1989, pp. 427–433.
14. Cameron, R. *Crosswalk Compliance Evaluation.* M.S. thesis. University of Washington, 1977.
15. Pline, J. L. (ed.). *Traffic Engineering Handbook,* 4th ed. Institute of Transportation Engineers, Englewood Cliff, N.J., 1992.

Publication of this paper sponsored by Committee on Traffic Law Enforcement.

DISCUSSION

KENNETH TODD
Consultant, Washington, D.C.

Although the study focused on law enforcement and its effectiveness in attaining driver compliance with crosswalk laws, it would be legitimate to ask whether the law is designed to give adequate protection to the pedestrian?

When crossing a road at an unsignalized location, either in midblock or at a marked or unmarked crosswalk, pedestrians face one of three situations. They can wait until conflicting vehicles have passed and a safe, adequate gap is available. When vehicles are at a safe distance or when none is approaching, the issue of right-of-way between driver and pedestrian does not arise.

The crossing less safe is when pedestrians try to pass through a gap that is not entirely adequate. They can walk faster to avoid the conflict, and the driver can slow down without coming to a complete stop.

The most dangerous time to cross is the moment when a vehicle is near and pedestrians exercise their right-of-way and force drivers to stop at the crosswalk. Drivers may be inattentive, they may fail to stop for fear of getting rear-ended, or they may hit a pedestrian who was hidden from view by a vehicle that slowed down or stopped in an adjoining lane (multiple threat).

The study was directed at unsignalized intersection crosswalks where vehicles on side streets are controlled by stop signs. Major roads, like the downtown commuter corridors in the study, have the purpose of carrying relatively large volumes of vehicles at a steady speed, often through a series of coordinated traffic signals. Unsignalized intersections located between signalized ones on a commuter corridor are controlled by stop signs to prevent side-street drivers from interfering with the steady progress of vehicles on the major road. To the pedestrian, however, the law gives the opposite instructions. Subject only to the requirement that they do not suddenly walk or run into the path of a vehicle so close that it constitutes an immediate hazard, pedestrians are not only allowed but encouraged to do what is forbidden to the motorist: interfere with fast-moving vehicles on a major road.

The crosswalk law and the major road concept are incompatible in terms of safety and efficient vehicle movement. To give a pedestrian instructions different from those given to the side-street driver is difficult to justify on operational or safety grounds. A more economical way to get pedestrians safely across a busy street would be to construct a refuge that allows them to cross in two stages.

AUTHORS' CLOSURE

The purpose of the study was not to investigate the "compatibility" of the crosswalk law with other laws designed to regulate motor vehicle traffic, but to determine whether increased enforcement of the crosswalk law would have an effect. If drivers consistently slow or stop their vehicles in response to the presence of pedestrians, it is likely that both the frequency and severity of pedestrian and motor vehicle collisions would decline. The study attempted to determine how effective enforcement might be at eliciting this response under a variety of conditions.

A larger question is whether the cause of pedestrian safety is best served by expending resources in enforcement or in another manner altogether. Our biases, both before and since the study, are that altering the design features of the roadway through traffic calming and other strategies, such as the refuge suggested by discussant, are likely to be a more effective, albeit more expensive way to address the problem. Still, the enforcement community is often called on to reduce pedestrian injuries. Research that will shed light on which enforcement strategies are effective and provide guidance as to the best use of limited traffic enforcement resources should continue to be performed.

Statistical Assessment of Public Opinion Toward Conversion of General-Purpose Lanes to High-Occupancy Vehicle Lanes

FRED MANNERING, JODI KOEHNE, AND SOON-GWAN KIM

Converting general-purpose lanes to high-occupancy vehicle (HOV) lanes is a policy that has been meticulously avoided since the public outcry opposing the lane conversion projects of the 1970s. Now that HOV lanes are firmly established in many metropolitan areas one has to wonder if public sentiments toward such lane conversions have changed. Public opinion of an HOV lane conversion recently completed in the Seattle metropolitan area is assessed. A series of multinomial logit and ordered probit models are estimated to isolate factors that determine commuters' attitudes toward various HOV policies (including lane conversion). The results show that although lane conversions are still strongly opposed by a substantial portion of the population, the intense public resistance encountered in the 1970s appears to be waning. Most of the survey respondents were either neutral or in favor of lane conversion projects.

High-occupancy vehicle (HOV) lanes have become one of the mainstays of urban traffic congestion mitigation in the United States. Although their effectiveness is sometimes questioned by the public, numerous research studies have shown apparent operational, safety, and benefit–cost advantages of HOV facilities. These studies have formed the basis for the continued expansion of HOV lane systems throughout North America.

Two methods have been widely used to provide HOV lanes: (a) constructing new lanes and (b) restriping an existing roadway by reducing lane or shoulder widths to accommodate the addition of an HOV lane. A third method, creating an HOV lane by taking (or converting) an existing general-purpose lane, has considerable potential from a cost reduction perspective. However, this is one approach that the HOV lobby has purposely avoided so as not to arouse public resistance to HOV lane implementation. The concern relating to possible public backlash against converting general-purpose lanes to HOV lanes is well founded. During the early stages of HOV lane implementation (in the 1970s) two infamous projects converted lanes of general traffic to create HOV lanes: (a) the Santa Monica Freeway project in Los Angeles (21.0 km of the inside lane of a four- or five-lane freeway was converted to a three-plus-passenger HOV lane), and (b) the Southeast Expressway project in Boston (13.3 km of a three-lane freeway was converted to a three-plus-passenger HOV lane). Analyses of the impacts of these projects showed that they were generally successful: HOV travel times improved, mode shifts from single-occupant vehicle to carpool or bus were observed, and good levels of safety were found in both cases (*1–3*). Despite these seemingly positive analyses, both the Santa Monica Freeway and Southeast Expressway HOV lanes were terminated after 5 and 6 months of operation, respectively, because of negative public opinion. The lesson seemed clear: in terms of HOV lane viability, operational impacts take a distant back seat to public opinion.

Since the lane conversion attempts of the 1970s the public has had nearly 20 years to warm up to HOV lane projects that have added HOV lanes through either construction or restriping. A natural question to ask is whether or not public opinion toward HOV lanes has swayed sufficiently so that lane conversions are now considered a tolerable policy. In fact, a number of researchers have begun addressing this question (*4,5*). The intent of the study described here was to evaluate public opinion toward a recent HOV lane conversion in the Seattle metropolitan area. It is hoped that the findings can provide valuable information with regard to the current status of public opinion toward lane conversions.

The paper begins with a description of the lane conversion area and the conversion implementation strategy. That section is followed by a discussion of the public opinion survey and sampling procedures. Public opinion is then statistically evaluated by estimating a series of multinomial logit and ordered probit models that address the types of comments that the survey respondents provided and their attitudes toward various HOV policies (including lane conversion). Finally, the implications of the findings are discussed and concluding comments are made.

DESCRIPTION OF STUDY AREA AND IMPLEMENTATION STRATEGY

The Seattle-area HOV lane conversion was made on Interstate 90 near a rapidly growing suburban area (Issaquah, Washington). Interstate 90 is one of two primary east–west routes in the Seattle area. On I-90 three lanes enter and exit Seattle. The limitation to three lanes was the result of a number-of-lanes restriction (i.e., a cap on capacity expansion) that was approved in the 1970s. Specifically, this restriction allowed no more than three lanes of traffic in each direction to cross the I-90 floating bridge that links Seattle with its eastern suburbs. However, east of this floating bridge (toward the suburbs) I-90 included a fourth lane in each direction. During the morning peak period the westbound commute (i.e., into Seattle) encountered considerable delay and queuing at the point where I-90 went from four to three lanes. It was reasoned that converting this fourth general-purpose lane to an HOV lane would ease the delay and queuing at this location; fewer vehicles would be required to merge because travel in the HOV lane would be restricted to carpools, vanpools, and buses. This condition made the fourth lane a strong candidate for conversion because it provided the potential for some immediate operational improvements as well as the usual

---

Department of Civil Engineering, 121 More Hall FX-10, University of Washington, Seattle, Wash. 98195.

HOV lane attributes (e.g., reduced travel times for HOVs, modal shifts, and so on).

Before deciding on a lane conversion the Washington State Department of Transportation (WSDOT) attempted to gain public support for the project. Representatives from local jurisdictions were informed about the project and asked to provide feedback. Public open houses were held, and motorists and local citizens were informed of the project and allowed to voice their concerns in an open forum. On the basis of the input received from both jurisdictional representatives and concerned citizens, a decision was made to pursue the lane conversion.

The new HOV lane was created by using two approaches; about 4 km of HOV lane was created by narrowing existing lanes and restriping the roadway, and another 6 km was created by converting an existing lane. Operation of the 10-km HOV lane was operational in November 1993. The lane is open to vehicles with two or more occupants and is restricted to such vehicles at all times of the day.

SURVEY APPROACH

The objective of the survey was twofold: (a) to determine the impact of the lane conversion on commuting behavior including mode, route, and departure time choices and (*2*) to study commuters' attitudes toward HOV lanes in general and lane conversions in particular. To achieve this objective a carefully constructed survey was designed. The survey was partitioned into three sections. The first section dealt with questions relating to the commute trip. Questions in this section focused on changes (before versus after the HOV lane conversion) in usual mode, route, and departure time. Questions relating to consumers' daily variations in mode, route, and departure times were asked as well. The second section gathered information on commuters' attitudes toward HOV lanes and lane conversion. The third and final section collected socioeconomic information on the commuter and the commuter's household.

The survey was distributed to commuters observed to be traveling in the lane-conversion area. License plate numbers were gathered during the morning commute over a 3-day period in June 1994 (roughly 7 months after the lane conversion). By using the Washington State Department of Motor Vehicles files, the license plate numbers were matched with the addresses of the registered vehicle owners and questionnaires were sent out in late June. In all, surveys were sent out to 1,325 commuters, and 322 responded (a response rate of 24.3 percent). Summary statistics for this sample are presented in Table 1.

Table 1 shows some interesting results with regard to commuters' socioeconomics. For example, roughly 62 percent of the respondents were male. This is a reasonable response because a higher percentage of males is expected in the morning commute. Another interesting finding was the high level of education (more than 16 years) and the high annual household income (more than $75,000). Although Seattle's eastern suburbs are relatively affluent, the $75,000 figure is on the high side. One possible explanation for this finding is that certain socioeconomic groups may have been more likely to respond to the survey. Our subsequent statistical analysis addresses this possibility.

In terms of commuting mode, about 77 percent list single-occupant vehicles (SOVs) as their usual mode of travel. It is also interesting that more than 70 percent of the commuters have used HOV lanes in the Seattle area during peak periods of travel at least once (Table 1). This suggests reasonable familiarity with the HOV system and its potential benefits in terms of saving travel time. Table 1 also shows that the usual mode of HOV lane travel is the two-person carpool, as expected. However, more than 2 percent of the commuters admit to HOV lane violations (i.e., listing SOV as the usual mode of HOV lane travel; Table 1). Given that Seattle-area HOV lane violation rates are very close to this figure, this admission shows an unexpected candor among survey respondents.

In terms of being qualified for HOV lane use and choosing not to use them, the most common reasons were "all traffic moves fast enough" and "slower than regular lanes." Less than 10 percent listed HOV lane safety as a reason for not using HOV lanes. Although this is a comparatively low figure, it shows that the HOV lane safety issue is still a fairly serious concern among some travelers (Table 1). Finally, the frequency of HOV lane usage on commuters' past five commutes (i.e., after the lane conversion) is given in Table 1. This shows some tendency toward regular mode switching (i.e., values in the range of 1 to 4 are more than 10 percent of the total) or not using the HOV lane when they are qualified to do so.

The survey also showed that the commuters in this corridor actively seek alternative route and departure times to shorten their commutes, with more than 30 percent indicating that they changed their route or departure times at least once in the past five commutes in an attempt to avoid traffic congestion. In terms of long-term changes, Table 1 shows that more than 35 percent of the commuters changed their usual departure times for work after the HOV lane conversion and more than 21 percent changed their usual routes. However, only 2.1 percent attributed these changes to the HOV lane conversion. Thus, the effect of the HOV lane conversion on route and departure time choice does not seem to be perceived by commuters to be significant. If this is the case, it indicates that the HOV lanes are not having a large impact on the well-being commuters derive from existing route and departure time choices [see the reports by Mannering and Hamed (*6*) and Small (*7*) for discussions of the impacts of HOV lanes on commuter welfare]. This important matter will be explored in forthcoming sections.

The HOV lane conversion appears to have had little impact on commuters' mode choice. Nine SOV users became carpool or vanpool users, and one became a bus user. However, five carpool or vanpool users became SOV users and two bus users became SOV users. Statistically, the HOV lane conversion had no significant impact on mode choice. Some caution should be exercised in interpreting these findings because the survey approach disproportionately samples SOV users (i.e., because it is based only on the vehicles observed to be traveling in the study area).

Commuters' opinions toward HOV lanes and HOV lane conversions also revealed some interesting findings. First, 47 percent believe that HOV lanes do not help save time for all commuters, and 41 percent do (Table 1). This suggests some lingering doubts as to the effectiveness of HOV lanes. This doubt is underscored by the 69 percent of respondents who believe that HOV lanes are not being adequately used. In fairness, it is possible that some of this negativity is an outgrowth of the fact that Seattle's HOV lane system is not yet complete. This could be a major source of HOV lane underutilization and a subsequent reduction in perceived effectiveness.

There also seems to be general public support for HOV lanes. Table 1 shows that only 36 percent of the survey's respondents believe that HOV lanes should be converted to general-purpose lanes. In terms of lane conversion, public opinion is negative, with 45 percent disagreeing (39 percent agreeing) that regular lanes should be converted. However, it is important to note that although

TABLE 1 Sample Summary Statistics (Averages Unless Otherwise Noted)

Sex (% male/female)	62.3/37.7
Age in years	42.9
Annual household income in thousands of dollars	75.6
Level of education	16.0
Household size	2.9
Number of household members older than 15 years old	2.2
Number of household members working outside of home	1.9
Number of household vehicles	2.4
Work schedule of commuters (% fixed/flexible)	48.4/51.6
Percent of usual travel mode in area highways between 6-9 am and 3-6 PM (SOV/ carpool or vanpool/ bus/ other)	77.3/17.1/5.3/0.3
Percent having ever used HOV lanes in the Seattle area between 6-9 am and 3-6 PM	70.2
Percent of usual travel mode when using HOV lanes in the Seattle area between 6-9 am and 3-6 PM (SOV/ carpool or vanpool/ bus/ motorcycle)	2.7/85.2/10.3/1.8
Percent sometimes qualifying for HOV lane use but not using them	37.9
Percent of reason for not using the HOV lane when qualified (slower than regular lanes/ too much trouble to change lanes/ HOV lanes are not safe/ all traffic moves fast enough/ forget to use HOV lane/ other)	24.4/7.9/9.4/38.6/7.9/11.8
Percent of commuters who used HOV lanes on I-90 during past five commutes (not at all/ 1-4 times/ every day)	75.6/10.8/13.6
Percent of commuters who changed usual departure time to work after new HOV lanes added	35.2
Percent of commuters who changed usual departure time to work because of HOV lanes	2.1
Percent of commuters who changed usual route to work after new HOV lanes added	21.3
Percent of commuters who changed usual route to work because of HOV lanes	2.1
HOV lanes help save all commuters time (disagree strongly or disagree/ neutral/ agree strongly or agree)	47/11/42
Existing HOV lanes are being adequately used (disagree strongly or disagree/ neutral/ agree strongly or agree)	69/13/18
HOV lanes should be open to all traffic (disagree strongly or disagree/ neutral/ agree strongly or agree)	57/7/36
Converting some regular highway lanes to HOV lanes is a good idea (disagree strongly or disagree/ neutral/ agree strongly or agree)	45/16/39

negative, there does not seem to be a strong public resentment toward lane conversion. This is certainly a shift from the lane conversion resistance observed in the 1970s.

Finally, the comments gathered at the end of the survey provided some interesting information, because it allowed respondents to air their frustrations or opinions about WSDOT's HOV lane and lane conversion policies. A large proportion of the respondents (about 51 percent) provided no comments. For the 49 percent of the respondents who did provide comments, the authors carefully screened the comments and classified them as negative (anti-HOV lane), positive (pro-HOV lane), and neutral. Nearly 50 percent had negative comments, 37 percent had positive comments or positive comments with criticism, and 13 percent were neutral. The relatively high percentage of negative comments shows some persistent dissatisfaction with HOV improvements or projects, but the overall proportion (slightly less than 25 percent of the entire sample) is relatively small.

Although the statistics discussed earlier provide some information as to the public's acceptance of HOV lanes in general and lane conversions in particular, a true multivariate analysis will allow determination of the characteristics of individuals who have predisposed positive or negative opinions toward HOV lanes or HOV lane conversions, or both. This type of information is critical, because it will permit state agencies to effectively market HOV lane projects by targeting specific commuter market segments. It will also allow agencies to forecast probable acceptance of HOV facilities on specific corridors once the socioeconomic and commute characteristics of the corridor are known.

Two types of multivariate analyses will be conducted in this paper. First, a model is developed to determine the probability that a survey respondent offers a negative, neutral, or positive comment, given that the respondent provides a comment. Second, a model is developed for each of the questions relating to commuters' opinions toward HOV facilities. These models will enable us determination of commuters' likelihood of disagreeing, being neutral, or agreeing with specific HOV-related statements. A description of these models and the model estimation results are provided in the following sections.

MODELING NATURE OF COMMUTERS' COMMENTS

For the 159 individuals who provided comments on their completed survey forms (roughly half of the 322 respondents), comments with regard to HOV lanes and HOV lane conversions were classified as being negative, neutral, or positive. Given the discrete nature of the three possible choice alternatives, a multinomial probabilistic choice model is a natural selection. In developing such a model it is assumed that a respondent will make the comment that provides the most satisfaction. Therefore, the probability of individual *n* making comment type *i* from the set of comment alternatives *I* is

$$P_n(i) = P(U_{in} \geq U_{In}) \ \forall \ I \qquad (1)$$

where *P* denotes probability and U_{in} is the satisfaction provided by comment type *i* to individual *n*. To estimate this probability, the satisfaction function (or in economic terms, the utility function) must be specified. This is usually done in a linear form such that

$$U_{in} = \beta \mathbf{X}_{in} + \epsilon_{in} \qquad (2)$$

where

$\mathbf{X}_{in}$ = a vector of measurable characteristics that define utility (e.g., age, gender, current mode of travel, departure time changes, and so on),
β = a vector of estimable parameters, and
ϵ_{in} = an error term that accounts for unobserved factors influencing an individual's utility of making comment type *i*.

The term $\beta\mathbf{X}_{in}$ in this equation is said to be the observable portion of utility because the vector $\mathbf{X}_{in}$ contains measurable characteristic variables (e.g., age of individual *n*), and ϵ_{in} is the unobserved portion.

Given Equations 1 and 2, the following can be written:

$$P_n(i) = P(\beta\mathbf{X}_{in} + \epsilon_{in} \geq \beta\mathbf{X}_{In} + \epsilon_{In}) \ \forall \ I \qquad (3)$$

With Equation 3 an estimable discrete choice model can be derived by assuming a distributional form for the error term. A natural choice would be to assume that this error term is normally distributed. If this is done a probit model results. However, probit models are computationally difficult to estimate. A more common approach is to assume that the ϵ_{in} terms are generalized extreme value (GEV) distributed. The GEV assumption produces a closed-form model that can be readily estimated by standard maximum likelihood methods. It can be shown (*8*) that the GEV assumption results in the multinomial logit model

$$P_n(i) = \exp[\beta\mathbf{X}_{in}]/\Sigma_I \exp[\beta\mathbf{X}_{In}] \qquad (4)$$

where all variables are as defined previously and the vector β is estimable by standard maximum likelihood methods.

Multinomial logit model coefficient estimates for the three types of comments are presented in Table 2 (note that only 155 of the 159 did not have missing data in either the dependent or independent variables; this explains the 155 observations given at the bottom of Table 2). Turning to specific estimation results, the model shows that individuals under the age of 21 are more likely to give a negative comment (i.e., negative coefficients for neutral and positive comment utilities). Also, since the coefficient for the positive comment alternative is more negative (i.e., larger absolute value) than the negative coefficient for the neutral comment (-1.367 versus -0.732), this age group is more likely to be neutral than positive. Although these coefficients are not highly significant (*t*-statistics just greater than 1), they do suggest the presence of anti-HOV sentiments among the young.

Next, the coefficients for the higher education dummies indicate that individuals with postgraduate work are more likely to give positive comments (i.e., a positive coefficient in the positive utility function) and less likely to be neutral (a negative coefficient in the neutral utility function). This shows that postgraduate education polarizes opinions with a greater likelihood of being positive.

Surprisingly, the same result was found with regard to the number of household vehicles per person. That is, a high number of vehicles per person made the individual more likely to make a positive comment and less likely to make a neutral comment. This finding appears to be an artifact of the sample that consists of affluent suburbanites with high vehicle ownership levels (Table 1).

Next we find that individuals with fixed work hours are less likely to make a neutral comment. The absence of work departure time flexibility seems to have polarized this population segment into making either a positive or a negative comment.

Individuals that indicated that SOV was their usual mode of travel were less likely to give a neutral comment and, as expected, less likely to give a positive comment. The tendency toward negative comments from SOV users is not surprising given their frustration in seeing what many of them consider to be underused HOV lanes during congested periods.

Individuals who were observed to change their departure times after the HOV lane conversion were much more likely to give a negative response and much less likely to give a positive response (as indicated by the highly significant negative coefficient in the positive alternative). As Table 1 shows, nearly 38 percent of commuters changed their usual departure times between September 1993 and June 1994, but only 2.1 percent listed the HOV lane conversion as the reason for this change. The most common reason for the change was an increase in traffic congestion (42 percent), which may have been due in some part to the reduction in SOV capacity due to the loss of a lane [change in work hours was the next most common reason (16 percent)]. It appears that these departure time change dummy variables are capturing the frustration of commuters in having to change their usual departure times, which has been shown to cause a significant loss in commuter welfare (*6*).

Finally, the route change dummy coefficient indicates that commuters who changed their usual routes after the HOV lane conver-

TABLE 2 Multinomial Logit Estimation Results for Comments About HOV

Variable*	Estimated coefficients (t-statistics)
Constant [0]	2.280 (2.906)
Constant [P]	-1.291 (-1.727)
Younger age dummy (1 if age is less than 21, 0 otherwise) [0]	-0.732 (-1.061)
Younger age dummy (1 if age is less than 21, 0 otherwise) [P]	-1.367 (-1.093)
Higher education dummy (1 if post graduate, 0 otherwise) [0]	-0.693 (-1.581)
Higher education dummy (1 if post graduate, 0 otherwise) [P]	0.790 (1.472)
Number of adults in a household (greater than 15 years old)[0]	-0.328 (-1.471)
Number of household vehicles per person [0]	-0.527 (-1.037)
Number of household vehicles per person [P]	0.811 (1.949)
Fixed-work dummy (1 if work-schedule is fixed, 0 otherwise) [0]	-0.458 (-1.235)
SOV dummy (1 if SOV is a usual mode in area highways between 6-9 AM and 3-6 PM, 0 otherwise) [0]	-1.416 (-3.099)
SOV dummy (1 if SOV is a usual mode in area highways between 6-9 AM and 3-6 PM, 0 otherwise) [P]	-0.787 (-1.290)
Departure time change dummy (1 if changed usual departure time to work after new HOV lanes, 0 otherwise) [0]	-0.665 (-1.580)
Departure time change dummy (1 if changed usual departure time to work after new HOV lanes, 0 otherwise) [P]	-1.689 (-2.265)
Route change dummy (1 if changed usual route to work after new HOV lanes, 0 otherwise) [0]	-0.774 (-1.434)
Log-likelihood at zero	-170.28
Log-likelihood at convergence	-136.18
Number of observations	155

* Numbers in brackets indicate variables defined for: [N] Negative opinion, [0] Neutral opinion, [P] Positive opinion alternatives.

sion were less likely to give a neutral comment. As was the case with departure time, the most common reason for the 21 percent of respondents who changed their routes was increasing traffic congestion (about 46 percent), with only 2.1 percent citing HOV lanes as the cause of the change. The polarization of the route change response (i.e., respondents are more likely to make positive or negative responses) seems to indicate that some commuters are happy with their new routes (the result of congestion being to force them to find a possibly better route in terms of travel time) and others are less pleased. The more consistent negative response of departure time changers suggests that departure time has a greater impact on commuter utility than the route choice. This is consistent with the earlier findings of Mannering and Hamed (*6*).

MODELING COMMUTER OPINIONS

The questions relating to commuter opinions have responses that range from disagree strongly to agree strongly. This type of data is referred to as ordered (because there is a consistent transition from disagreeing to agreeing) and can be translated into an integer form for the purposes of model estimation. In their case the statistical analysis showed that the data can best be grouped into three categories: (a) disagree (which includes strongly disagree and disagree), (b) neutral, and (c) agree (which includes agree and strongly agree). This grouping suggests that respondents did not adequately distinguish between strongly and simply agreeing and disagreeing. This reordering will have no effect on the substantive findings of the forthcoming statistical analysis.

Translating these three choices into integer form provides the following: 1 as disagree, 2 as neutral, and 3 as agree With this ordering, an ordered probability model can be derived (*9*). Such models begin by defining an unobserved variable, *z*, that is used as a basis for modeling the ordinal ranking of the data. This unobserved variable is specified as

$$z = \beta \mathbf{X} + \epsilon \qquad (5)$$

where

$\mathbf{X}$ = a vector of characteristics determining individuals' choice of ranking category,
β = a vector of estimable parameters, and
ϵ = a random disturbance.

By using this equation observed ordinal rankings, y (ranging from 1 to 3 here), are defined as

$$y = 1 \text{ if } z \leq \mu_1 \qquad (6)$$

$$y = 2 \text{ if } \mu_1 < z \leq \mu_2$$

$$y = 3 \text{ if } z > \mu_2$$

where the μ's are estimable parameters that define y, which corresponds to integer rankings. Note that without the loss of generality, μ_1 can be constrained to be zero so that only the threshold μ_2 needs to be estimated.

If the disturbance term in Equation 6 is assumed to be standard normal (with the equal to mean zero and variance equal to 1) an ordered probit model results, and if the disturbance is assumed to be standard logistic, an ordered logit model results. Unlike the case of the discrete choice model presented in the previous section, the ordered logit model does not have a significant computational advantage over the ordered probit. The choice of one model over the other is often made purely on theoretical grounds, and because of the widespread use of the normal distribution in statistics, a standard normal distribution of the error term is assumed and a series of ordered probit models is estimated. Ordered probit models of attitude statements made in the survey are discussed below.

HOV Lanes Help Save All Commuters Time

As shown in Table 1, more people believe that HOV lanes do not save all commuters time relative to those that do. It is important to understand this skepticism with regard to the value of HOV lanes. Model estimation results for this statement are presented in Table 3. The estimation results show that respondents less than 21 years old are less likely to agree with this statement. This is consistent with the tendency of this group to provide negative comments, as shown in the preceding section.

Higher-income households were also less likely to agree with this statement (i.e., a negative coefficient). This may be because higher-income households are more dependent on automobiles than their lower-income counterparts. The greater the number of adults in the household, the less likely the respondent is to agree with this statement. This suggests some lingering skepticism among larger, older households as to the effectiveness of HOV lanes.

Respondents with fixed work hours were less likely to agree that HOV lanes save all travelers time. This group of travelers has limited ability to adjust departure times to avoid congestion, and in the absence of what they believe are reasonable modal alternatives, they may harbor bitter feelings toward losing a lane of capacity to HOVs.

People who are regular SOV users do not tend to believe that HOV lanes save all travelers time, and people who are regular HOV users tend to believe that HOV lanes save all travelers time (as indicated by the negative and positive coefficients, respectively). This sort of modal bias is an expected result.

Finally, the 2.1 percent of respondents who indicated that they changed their usual departure times because of the presence of HOV lanes were less likely to agree with the statement that HOV lanes save all commuters time. Correct or not, these respondents seem to be blaming their forced departure time changes on the presence of HOV lanes.

Existing HOV Lanes Are Being Adequately Used

Table 1 shows that nearly 70 percent of respondents do not believe that HOV lanes are being adequately used. From a policy perspec-

TABLE 3 Ordered Probit Estimation Results for Opinion of HOV Lanes Saving all commuters time

Variable*	Estimated coefficients (t-statistics)
Constant	0.900 (3.913)
Younger age dummy (1 if age is less than 21, 0 otherwise)	-0.276 (-1.156)
Higher income dummy (1 if annual household income is greater than $75K, 0 otherwise)	-0.238 (-1.646)
Number of adults in a household (greater than 15 years old)	-0.147 (-2.075)
Fixed-work dummy (1 if work-schedule is fixed, 0 otherwise)	-0.319 (-2.160)
SOV dummy (1 if SOV is a usual mode in area highways between 6-9 AM and 3-6 PM, 0 otherwise)	-0.684 (-3.254)
HOV use dummy (1 if used HOV lanes on I-90 during past five commutes, 0 otherwise)	0.347 (1.627)
Departure time change due to HOV lanes dummy (1 if changed usual departure time to work due to presence of HOV lanes, 0 otherwise)	-0.943 (-1.353)
Threshold μ_2	0.312 (6.260)
Log-likelihood at zero	-302.30
Log-likelihood at convergence	-280.17
Number of observations	313

* Dependent variables: 1 is base (disagree), 2 is neutral, 3 is agree

tive, such a belief is clearly a matter of concern with regard to future expansions of HOV systems. Ordered probit estimation results of opinions on this statement are presented in Table 4. The results show that men are more likely to agree with this statement, although the level of statistical significance ($t = 1.261$) is not very high. This finding may be an outgrowth of the demographic characteristics of the sample. A sample drawn from lane conversions in other corridors would provide evidence to either support or refute this finding.

Older respondents (older than 50 years) and respondents from higher-income households (greater than $75,000) were less likely to believe that HOV lanes are being adequately used. Again, this could be the result of their greater dependence on SOV travel and their concern over the loss of roadway capacity caused by HOV lanes.

Respondents who were more highly educated were more likely to agree with this statement. This is consistent with the earlier finding that such respondents were more likely to make a positive comment on the survey.

Both the higher the number of adults in the household and having a fixed work schedule reduced the likelihood of believing that HOV lanes are being adequately used. This finding shows skepticism among people with these characteristics, as was the case with their believing that HOV lanes saved all commuters time.

Finally, as expected, respondents who currently list HOV modes as their usual modes of travel are more likely to believe that HOV lanes are being adequately used. This result is consistent with earlier findings.

HOV Lanes Should Be Opened to All Traffic

The statement that HOV lanes should be opened to all traffic is interesting because it asks consumers to pass judgment on a national transportation policy. As shown in Table 1, 36 percent of respondents agreed with this statement. Although this is not a majority, it is nonetheless a disturbingly high figure. The ordered probit estimation results presented in Table 5 provide some insight into the characteristics of respondents who are likely to agree or disagree with this statement.

Many of the results are consistent with earlier findings that isolated characteristics of respondents made them likely to have opinions that favor or oppose HOV lanes. For example, older respondents, respondents from higher-income households, respondents with fixed work start times, regular SOV users, and individuals who attribute departure time changes to the presence of HOV lanes are all more likely to favor opening HOV lanes to all traffic. These consistent findings clearly isolate the characteristics of individuals who are likely to oppose HOV policies.

Table 5 shows that regular HOV users and households with a large number of children were factors that increased the likelihood of disagreeing with this statement. The presence of a large number of children increases the likelihood of qualifying for HOV lane usage (i.e., transporting children) and thus results in a more favorable attitude toward future HOV lane use. Finally, it is important to note that the negative coefficient of the constant term indicates a general disposition of the public to oppose opening HOV lanes to all traffic.

Converting Some Regular Highway Lanes to HOV Lanes Is a Good Idea

The statement that converting some regular highway lanes to HOV lanes is a good idea had a response showing 45 percent disagreeing, 16 percent neutral, and 39 percent agreeing (Table 1). It is clear that opposition toward lane conversion exists, but it is by no means overwhelming. Ordered probit estimation results for this statement are presented in Table 6.

The model results presented in Table 6 closely parallel the findings of earlier models. Regular SOV users and respondents who attribute departure time changes to the presence of HOV lanes are

TABLE 4 Ordered Probit Estimation Results for Opinion of Existing HOV Lanes Being Adequately Used

Variable*	Estimated coefficients (t-statistics)
Constant	-0.583 (-3.146)
Gender dummy (1 if male, 0 if female)	0.204 (1.261)
Older age dummy (1 if age is greater than 50, 0 otherwise)	-0.449 (-2.202)
Higher income dummy (1 if annual household income is greater than $75K, 0 otherwise)	-0.181 (-1.165)
High education dummy (1 if post graduate, 0 otherwise)	0.358 (2.169)
Number of adults in a household (greater than 15 years)	-0.216 (-2.683)
Fixed-work dummy (1 if work-schedule is fixed, 0 otherwise)	-0.150 (-0.954)
HOV use dummy (1 if used HOV lanes on I-90 during past five commutes, 0 otherwise)	0.996 (5.628)
Threshold μ_2	0.487 (6.993)
Log-likelihood at zero	-294.77
Log-likelihood at convergence	-239.57
Number of observations	314

* Dependent variables: 1 is base (disagree), 2 is neutral, 3 is agree

TABLE 5 Ordered Probit Estimation Results for Opinion That HOV Lanes Should Be Opened to All Traffic

Variable*	Estimated coefficients (t-statistics)
Constant	-0.740 (-2.369)
Older age dummy (1 if age is greater than 50, 0 otherwise)	0.290 (1.600)
Higher income dummy (1 if annual household income is greater than $75K, 0 otherwise)	0.216 (1.479)
Number of children 0-15 years	-0.156 (-1.953)
Fixed-work dummy (1 if work-schedule is fixed, 0 otherwise)	0.310 (2.025)
SOV dummy (1 if SOV is a usual mode in area highways between 6-9 AM and 3-6 PM, 0 otherwise)	0.900 (3.819)
HOV use dummy (1 if used HOV lanes on I-90 during past five commutes, 0 otherwise)	-0.375 (-1.593)
Departure time change due to HOV lanes dummy (1 if changed usual departure time to work due to presence of HOV lanes, 0 otherwise)	1.568 (2.158)
Threshold μ_2	0.246 (5.334)
Log-likelihood at zero	-285.44
Log-likelihood at convergence	-254.08
Number of observations	311

* Dependent variables: 1 is base (disagree), 2 is neutral, 3 is agree

likely to oppose lane conversion, whereas regular HOV users and households with a large number of children are likely to favor lane conversions.

IMPLICATIONS OF FINDINGS AND CONCLUSIONS

The findings of the survey of commuters using the I-90 HOV lane conversion corridor show that the lane conversion was not overwhelmingly accepted by the public. In fact, more respondents oppose lane conversions than favor them. Still, the percentage of people who oppose lane conversions is just slightly greater than the percentage who favor them, suggesting that the long-held resistance of the public to lane conversions may be waning. However, ordered probit model results show that lane conversion resistance is higher among the young (commuters less than 21 years old), higher-income households, SOV users, and individuals who changed their departure times as a result of the presence of HOV lanes. Given the size of some of these population groups (e.g., more than 77 percent are usual SOV users), it is clear that considerable marketing is needed before a significant majority of

TABLE 6 Ordered Probit Estimation Results for Opinion About Converting Some Regular Highway Lanes to HOV Lanes

Variable*	Estimated coefficients (t-statistics)
Constant	0.259 (0.987)
Number of children 0-15 years	0.152 (2.026)
SOV dummy (1 if SOV is a usual mode in area highways between 6-9 AM and 3-6 PM, 0 otherwise)	-0.702 (-3.510)
HOV use dummy (1 if used HOV lanes on I-90 during past five commutes, 0 otherwise)	0.458 (2.276)
Departure time change due to HOV lanes dummy (1 if changed usual departure time to work due to presence of HOV lanes, 0 otherwise)	-1.031 (-1.370)
Threshold μ_2	0.445 (7.640)
Log-likelihood at zero	-322.94
Log-likelihood at convergence	-297.30
Number of observations	313

* Dependent variables: 1 is base (disagree), 2 is neutral, 3 is agree

the public comes to accept HOV lane conversions as a tolerable transportation policy.

With regard to HOV lanes in general, the public is not completely dissatisfied. Only 36 percent of the commuting public believe that HOV lanes should be opened to all traffic. On the down side 47 percent do not believe that HOV lanes save all commuters time, and more than 69 percent believe that HOV lanes are not being adequately used. Ordered probit models show that individuals most likely to have a negative bias toward HOV lanes are young (less than 21 years old), are from higher-income households, have a large number of adults in their households, indicate SOVs as the usual mode of travel, and have fixed work hours. Apparently, individuals who fit this mold have yet to be convinced of the purported virtues of HOV lanes.

In terms of the types of comments individuals made on their survey forms, slightly more than 50 percent made no comment at all. Of those who did comment, the majority responded negatively to the lane conversion or HOV facilities in general, or both. Multinomial logit estimation results show, as was the case with the opinion models discussed earlier, that commuters who were likely to make negative comments were younger (less than 21 years old), were regular SOV users, and had fixed work schedules. One different finding is that individuals were more likely to make negative comments if they changed their usual departure times after the lane conversion, regardless of the reason (previous results on HOV opinions show this to be important only if respondents attribute the departure time change to HOV lanes). It appears that many respondents are venting their frustrations about having to change their usual departure times.

In summary, from a public opinion point of view, the I-90 lane conversion in the Seattle area can be classified as a qualified success. Although a slight majority of commuters oppose the conversion, public opinion for and against is surprisingly close. It appears that with effective marketing and careful implementation, lane conversions can be successfully made. However, it is important to recognize that significant opposition may arise from the young, higher-income households with a high number of adults, commuters with fixed work times, regular SOV users, and commuters who will be forced to make departure time changes. Commuters who fit this mold should be dealt with through informational campaigns and other strategies in an effort to reduce their opposition.

REFERENCES

1. Batz, T. *High Occupancy Vehicle Treatments, Impacts and Parameters,* Vol. I. Procedures and conclusions.;ro Report 86-017-7767. New Jersey Department of Transportation, Trenton, Aug. 1986.
2. Billheimer, J., R. Bullemer, and C. Fratessa. *The Santa Monica Freeway Diamond Lanes.* Report DOT-TSC-UMTA-77-44. Transportation Systems Center, Cambridge, Mass. 1977.
3. Simkowitz, H. *Southeast Expressway High Occupancy Vehicle Lane Evaluation.* Report DOT-TSC-UMTA-78-25. Urban Mass Transit Administration, Washington D.C., 1978.
4. Auslam, M. *HOV Lane Conversions in California (1989–1994).* California Department of Transportation, June 1994.
5. Gard, J., P. Jovanis, R. Kitamura, and V. Narasayya. *Public Attitudes Toward Conversion of Mixed-Use Freeway Lanes to HOV Lanes.* Report UCD-ITS-RR-94-6. Institute of Transportation Studies, University of California, Davis, 1994.
6. Mannering, F., and M. Hamed. Commuter Welfare Approach to High Occupancy Vehicle Lane Evaluation: An Exploratory Analysis. *Transportation Research,* Vol. 24A, No. 5, 1990, pp. 371–379.
7. Small, K. Bus Priority and Congestion Pricing on Urban Expressways. In *Research in Transportation Economics,* Vol. I (T. Keeler, ed.), JAI Press, 1983, pp. 27–74.
8. McFadden, D. Econometric Models of Probabilistic Choice. In *Structural Analysis of Discrete Data with Econometric Applications* (C. Manski and D. McFadden, eds.), MIT Press, Cambridge, Mass., 1981, pp. 198–272.
9. Greene, W. *Econometric Analysis.* Macmillan Publishing, New York, 1993.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.

Safety Evaluation at Three-Leg, Unsignalized Intersections by Traffic Conflict Technique

NABEEL K. SALMAN AND KHOLOUD J. AL-MAITA

The FHWA procedure detailed in *Traffic Conflict Techniques for Safety and Operations* was used to collect data on conflicts at 18 three-leg, unsignalized intersections. The objectives were to identify the average and abnormal conflict rates at three-leg, unsignalized intersections in the city of Amman, Jordan, and to relate traffic conflicts to traffic volumes and accidents. The average and abnormal limits can be used to evaluate comparable three-leg, unsignalized intersections by the traffic conflict technique. A risk index was developed on the basis of various conflict types and related accidents to priority rank the intersections for adopting countermeasures. The study found that the predominant conflict type was the left-turn, same-direction conflict. Multiple regression technique was applied between different conflict types and combinations of traffic flows generating those conflicts at various intersection categories. Most of the predicted relationships indicated that traffic conflicts are dependent on different combinations of the generating flows. In addition, multiple regression technique was applied to relate accidents to traffic conflicts. The results revealed that accidents and conflicts are related by a linear relationship.

Traffic accidents are the most direct measure of safety for a highway location. Because attempts to estimate the relative safety of a highway location are usually hampered by the problems associated with the unreliability of accident records and the time required to wait for adequate sample sizes, the Traffic Conflict Technique was developed as a substitute measure. An earlier attempt to estimate the relative safety of a highway by observing erratic driving, unsafe maneuvers, and near misses at problem locations was first formalized elsewhere (*1*). The first formalized procedures for identifying and recording near-miss events between vehicles at intersections were developed (*2*) later. Additional research in 1979 through 1985 provided standard definitions and refined data collection procedures and applied this technique to estimating the number of predicted accidents at an intersection (*3*). Parker and Zegeer (*4*) defined traffic conflict as an event involving two or more road users in which the action of one user causes the other user to make an evasive maneuver to avoid a collision. The evasive maneuvers are identified by brake lights or lane changes.

OBJECTIVES

The specific research objectives of the present study were as follows:

1. Determination of the predominant conflict type among all conflict types at three-leg, unsignalized intersections.
2. Identification of the average and abnormally high values for traffic conflicts at studied intersections.
3. Determination of the most hazardous intersection among all studied intersections.
4. Development of prediction models that relate traffic conflicts to traffic volumes and total accidents.

DATA COLLECTION

The data collection consisted of three stages. The first stage involved selecting intersections for the study, the second stage consisted of collecting accident data from the traffic police administration department, and the third stage consisted of making field measurements of both traffic conflicts and traffic volumes.

Extensive field work was conducted at 18 unsignalized, three-leg intersections located in Amman, Jordan, during the summer of 1992. The guidelines that were considered in selecting the intersections were the availability of accident records for each intersection, minimal pedestrian traffic, no unusual sight distance restriction, no turn restriction or one-way street, no appreciable grade, and no parking restriction (*4*). A brief general description of the characteristics of each individual intersection site is presented in Table 1.

Accident reports on all accidents that occurred at the 18 intersections over the 3-year (1989 to 1991) period were reviewed. The total number of accidents for each intersection was summarized. Accidents on wet roads, accidents involving pedestrians, nighttime accidents, and weekend accidents were not used in the analysis.

The recommended procedures detailed in the *Traffic Conflict Technique for Safety and Operations Engineers Guide* (*4*) and the *Traffic Conflict Technique for Safety and Operations Observer Manual* (*5*) were followed in conducting the study.

Two observers were used to collect the data at each intersection. All observers were taken to the field to observe traffic conflicts and to classify and distinguish different types of conflicts. Figure 1 shows a typical three-leg, unsignalized intersection with three approaches and two observer locations.

Nine types of conflicts that occurred at a three-leg, unsignalized intersection were studied. Figure 2 shows the four types of conflicts that may occur exclusively from the south approach. They were (a) opposing left turn, (b) right turn, same direction, (c) right turn from right, and (d) left turn from right.

Figure 3 shows the three conflict types that may occur exclusively from the north approach: (a) left turn, same direction, (b) right turn from left, and (c) left turn from left. Finally, Figure 4 shows the two conflict types that may occur from either direction: (a) slow vehicle, same direction, and (b) lane change. The lane

---

N. K. Salman, Department of Civil Engineering, University of Jordan, Amman, Jordan. K. J. Al-Maita, Civil Engineering Department, Mu'tah University, P. O. Box 6551, Postal Code 131/14, Zarka, Jordan.

TABLE 1 General Description of Three-Leg, Unsignalized Intersections Studied in Amman, Jordan, 1992

	Major Road [a]		Minor Road [a]			
Intersection Number	Name	No. Of Lanes	Name	No. Of Lanes	Control Device	Speed Limit km/hr
1	Om-Omara	2LU	Al-Rawdah	2LU	Stop	40
2	Ibrahim Al-Kattan	2LU	Barakat Al-Zoaby	2LU	Stop	40
3	King Faisal	2LU	Shatt Al-Arab	2LU	Stop	50
4	Al-Shaima'a Al-Sadieh	2LU	Amrawah	2LD	Stop	40
5	Sakher Al-Ahmasy	4LD	Al-Shaima'a Al-Sadieh	2LU	Stop	40
6	Isam Al-Ajlony	2LU	Hamameh	2LU	Stop	50
7	Al-Rawdah	4LD	Abd. Elhafith Abu Al-Feelat	4LD	Stop	50
8	Abdalleh Ghoshih	4LD	Barakat Al-Zoaby	2LU	Stop	40
9	Sa'ad Bin Aby Waqas	4LU	Shatt Al-Arab	2LD	Stop	50
10	Isam Al-Ajlony	2LU	Abu Sofian	2LU	Stop	50
11	Sakher Al-Ahmasy	4LD	Om Al-Fadel	4LD	Stop	50
12	Al-Razi	2LU	Aqraba	2LD	Stop	40
13	Mohammad Ali Janah	4LU	Al-Qahera	4LD	Stop	50
14	Al-Rawdah	4LD	Al-Wefaq	2LD	Stop	40
15	Al-Razi	2LU	Jamal Al-Deen Al-Afagany	2LD	Stop	50
16	Abdallah Goshih	4LD	Al-Bader	4LU	Stop	40
17	Sakher Al-Ahmasy	4LD	Jothama Al-Kenany	2LU	Stop	40
18	Princes Basma	4LD	Fawzi Al-Qawqjy	2LU	Stop	40

[a] See Figure 1.

2LU = Two-Lanes Undivided;
4LU = Four-Lanes Undivided;
2LD = Two-Lanes Divided;
4LD = Four-Lanes Divided

change type of conflict can occur only when two or more lanes are present in one direction.

In all of the foregoing conflict situations, when the second vehicle makes an evasive maneuver it may place another road user (a third vehicle) in danger of a collision. Figure 5 shows an example of secondary conflict. Only one secondary conflict for any initial conflict was counted even if a whole line of cars stopped, because the first vehicle was in a conflict situation. The event would be recorded as one conflict (according to its category) and one secondary conflict.

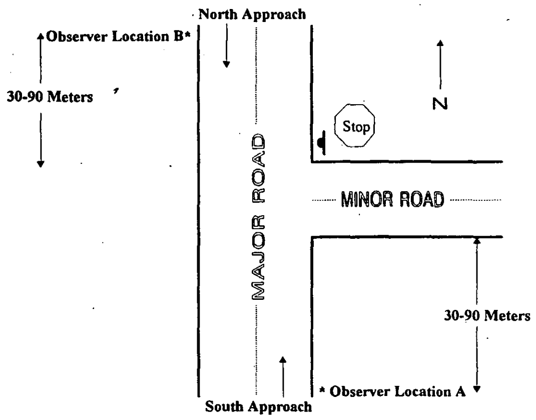


FIGURE 1 Typical three-leg, unsignalized intersection with approximate observer locations (*4*).

The daily counts were conducted from 7:00 a.m. to 6:00 p.m. during typical weekdays over the summer months (July and August) of 1992. Two observers collected a full day's data at each intersection, where one observer counted traffic conflicts and the other one recorded traffic volumes from one approach and documented the data in 20-min increments. Then the two observers alternated to the other approach of the intersection and counted the traffic conflict and volumes on that approach. The locations of the observers were in the range of 30 to 90 m from the intersection (Figure 1), depending on the available space for observation purposes. In addition, a roadway inventory (i.e., street alignment and width, number of lanes, curb radii, pavement markings and condition, channelization, parking, sight distance obstructions, and signs) was carried out by the observers during the same survey day.

DATA ANALYSIS

The conflict data collected in the field were compiled and presented in a format suitable for analysis. The conflict counts in the field were used to produce certain sums and rates, which were needed for analysis. FHWA procedures for summarizing the data (*4*) were followed to determine the hourly conflict frequency and daily conflict counts and to calculate the conflict rate per 1,000 vehicles. The data summarized for each intersection are presented in Table 2.

Conflict Frequency Distribution

To determine the type of conflict that was the most predominant among all other conflict types, the conflict frequency distribution

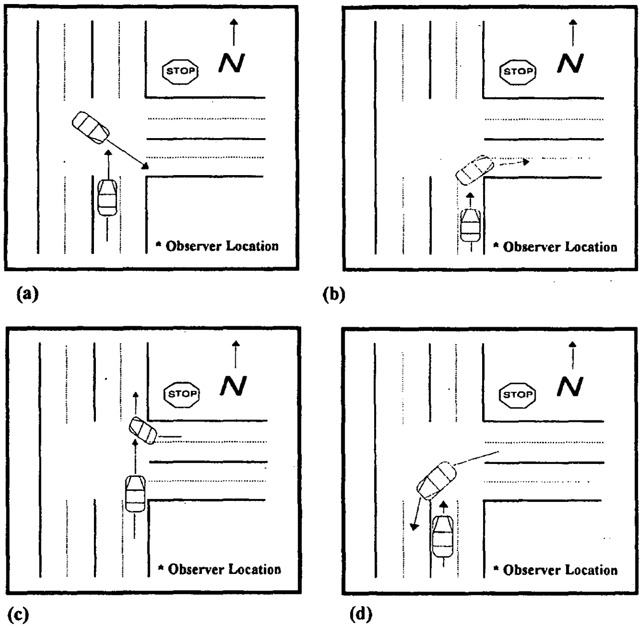


FIGURE 2 Traffic conflicts from south approach: (*a*) opposing left turn conflict; (*b*) right turn, same-direction conflict; (*c*) right turn from right conflict; (*d*) left turn from right conflict.

was determined. Each conflict type was combined into one group and the total number of conflicts for all conflict types was computed, and then the percentage of each conflict type from the total number of conflicts was also calculated. Finally, conflict types were ranked in descending order according to these percentages. The rank orders are presented in Figure 6 along with the relative cumulative frequency distribution of conflicts. The results indicate that the left-turn, same-direction conflict type is first in the rank order. This may be for many reasons such as the absence of a left-turn storage lane, improper channelization, poor road marking, poor sight distance for left-turning vehicles with respect to oncoming through traffic, and the large volume of both left-turning and opposing through vehicles at most of the studied intersections. The rank of right turn from left conflict type was last in the rank order. This could be because the occurrence of these conflicts is rare. This conflict occurs when the first vehicle on the left-hand side makes a right turn across the center of the main street roadway into an opposing lane, thus forcing the driver in the second vehicle to take an evasive action.

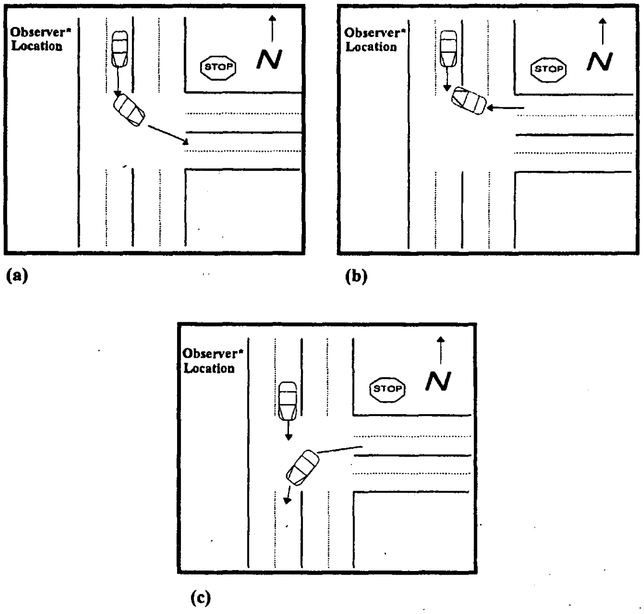


FIGURE 3 Traffic conflicts from north approach: (*a*) left turn, same-direction conflict; (*b*) right turn from Left Conflict; (*c*) left turn from left conflict.

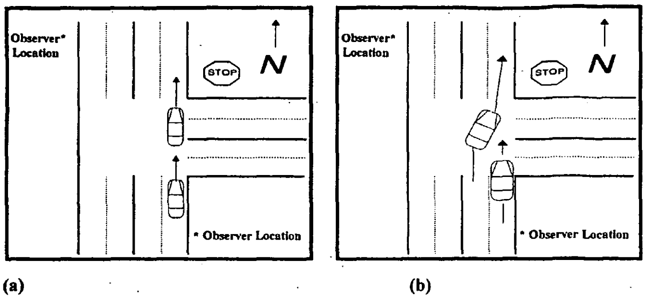


FIGURE 4 Traffic conflicts from either approach: (*a*) slow-vehicle, same-direction conflict; (*b*) lane change conflict.

Safety and Operational Analysis

To identify the potential hazards and operational problems, the limits of safety problems were determined, and the identification of problems and the selection of countermeasures were then carried out. The magnitude of the problem was determined by developing average and abnormal daily conflict counts in the study area by using the procedures detailed by Parker and Zegeer (*4*) and Glauz et al. (*6*). The results are presented in Table 3. These results did not demonstrate similarities to the previous results obtained by Crowe (*7*), who identified the average and abnormal conflict counts at three-leg, unsignalized intersections from data collected in Houston, Texas. The mean conflict count for the most frequently occurring conflict type in the present study (left turn, same direction) was approximately 5 times higher than the value in previous results. Generally, these results indicated that daily conflict counts were much higher in the area around Amman, Jordan, than daily conflict counts in the area around Houston, Texas.

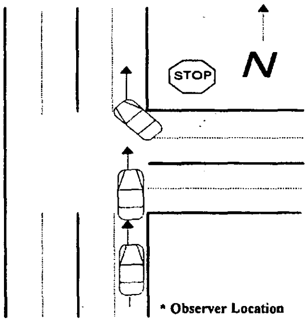


FIGURE 5 Right turn from right secondary conflict.

TABLE 2 Daily Conflict Counts at Studied Intersections

			Conflict Type									
Intersection No.[a]	No. of Accidents	Two Way Approach[b]	LT. Same Direction	RT. Same Direction	Slow Vehicle	Lane Change	Opposing LT	LT from right	RT from right	LT from left	RT from left	All Same Direction[c]
1	3	3900.0	65.5	22	134.5	0.0	98	80.0	23	73.5	22	222.0
2	4	5446.5	83.5	97	104.0	0.0	48	156.5	35	251.5	46	284.5
3	5	7762.5	138.0	120	173.5	0.0	43	164.0	46	140.5	0	431.5
4	3	7433.0	65.0	115	168.0	0.0	80	97.0	37	77.0	28	348.0
5	4	7568.5	89.0	168	113.0	182.0	102	201.0	36	160.0	0	552.0
6	9	8269.0	292.0	71	406.0	0.0	174	304.0	307	152.0	115	769.0
7	10	8403.5	300.5	36	156.0	452.0	383	62.0	32	219.0	0	945.0
8	5	8345.5	236.0	33	185.5	10.0	143	50.0	43	89.0	0	464.5
9	9	8980.5	177.5	264	135.0	102.5	134	234.0	18	334.0	0	679.0
10	4	9415.0	25.5	0	65.5	0.0	6	179.0	0	175.5	0	91.0
11	10	11232.0	660.5	104	259.0	239.5	439	261.0	117	721.5	0	1263.0
12	9	11498.0	640.0	283	292.0	0.0	350	236.0	108	299.5	168	1215.0
13	9	11953.5	843.0	45	180.5	115.0	473	297.0	39	216.0	0	1183.5
14	7	12302.0	632.0	0	214.0	25.0	458	150.0	120	118.5	0	871.0
15	9	10149.0	561.5	34	114.5	0.0	166	57.0	219	538.0	80	710.0
16	8	12921.0	193.0	65	313.5	0.0	120	290.0	42	213.0	0	571.5
17	7	17924.5	315.5	165	316.0	79.0	279	252.0	187	194.0	0	875.5
18	6	19194.0	494.0	0	371.0	106.0	303	128.0	53	103.0	0	971.0

[a] See Table(1).

[b] 11-Hour Vehicle Volumes.

[c] Includes LT Same-Direction, RT Same-Direction, Slow Vehicle, and Lane Change Conflicts.

LT = Left Turn
RT = Right Turn

The average and abnormally high percentile in Table 3 can be used to evaluate individual intersections after conflict data have been collected. For example, Figure 7 shows that intersection 13 exceeds the 95th percentile limit and intersection 11 exceeds the 90th percentile limit, which was abnormally high for the left-turn, same-direction conflict type. The high number of vehicles turning left and the absence of a left-turn storage lane at intersection 13 were the main contributing factors for the high occurrence of this type of conflict. Similarly, the high number of opposing through vehicles with the absence of a left-turn storage lane at intersection 11 were the main contributing factors for the high occurrence of this type of conflict. General countermeasures could be recommended by installing traffic signals if warranted by the *Manual on Uniform Traffic Control Devices for Streets and Highways* (*8*) and by considering the need for installing a traffic signal with separate left-turn phasing as well as to provide left-turn lanes or to prohibit left turns on the approach. For a further explanation of the possible causes of the abnormally high numbers of conflicts and general countermeasures for each conflict type, refer to the *Engineers Guide* (*4*).

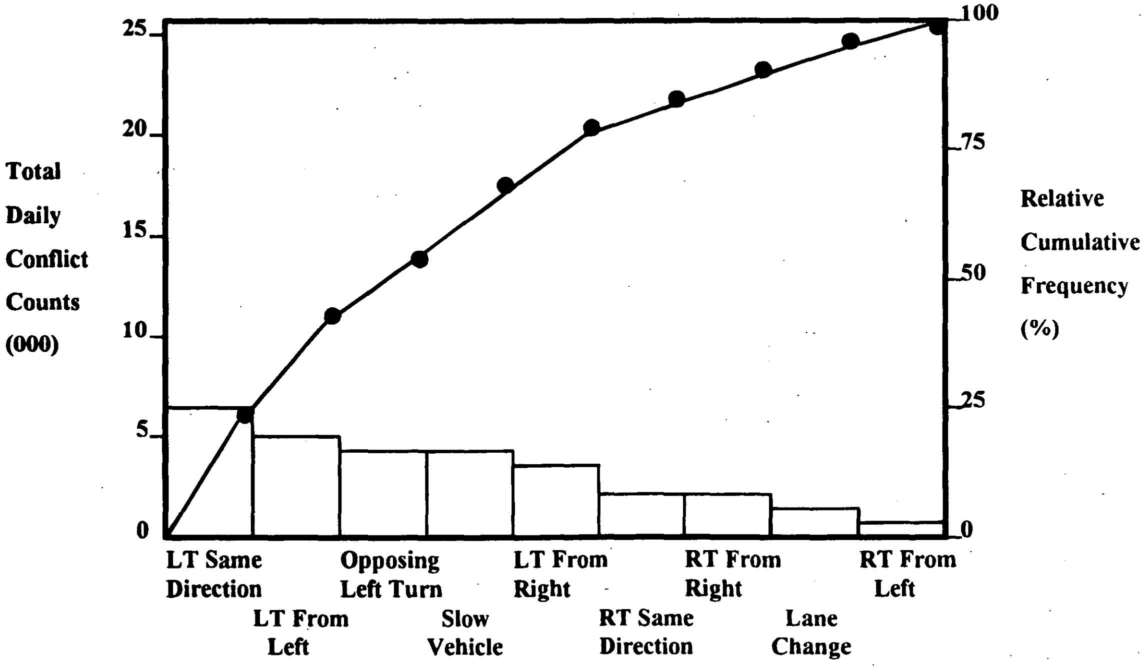


FIGURE 6 Cumulative frequency distribution of conflicts of three-leg, unsignalized intersection.

TABLE 3 Mean and Abnormal Daily Conflict Counts for Three-Leg, Unsignalized Intersections

Conflict Types		Mean Conflict Count	Abnormally High Conflict Count 90th- Percentile	Abnormally High Conflict Count 95th- Percentile
LT Same Direction	C	322.44	656.58	814.31
	S & C	430.92	918.92	1159.89
RT Same Direction	C	90.11	200.71	257.84
	S & C	101.83	230.09	297.23
Slow Vehicle	C	205.67	337.53	389.61
	S & C	221.92	366.60	424.03
Lane Change	C	131.15[a]	305.52	400.29
	S & C	154.05[a]	367.60	488.51
Opposing LT	C	211.06	416.46	510.70
	S & C	245.11	497.61	616.51
LT From Right	C	177.72	293.29	339.13
	S & C	210.83	357.42	416.76
RT From Right	C	81.22	187.56	244.44
	S & C	93.28	218.59	287.41
LT From Left	C	226.42	448.29	550.42
	S & C	283.47	612.89	778.12
RT From Left	C	25.50	-----	-----
	S & C	27.92	-----	-----
All Same Direction	C	691.08	1159.31	1347.28
	S & C	837.42	1455.51	1710.60

LT	Left Turn
RT	Right Turn
C	Primary Conflict
S & C	Secondary and Primary Conflict

[a]Includes only 4-lane roadway

---- Indicates that this type of conflict is so rare that any number observed at an intersection should be considered abnormal.

Priority Ranking of Intersections

A priority ranking of the 18 three-leg, unsignalized intersections was conducted to determine the most hazardous intersections among all of the intersections studied. A priority ranking of the intersections was developed on the basis of an assessment of the risk level (risk index) for each intersection. Some conflict types were better risk indicators than others. Muhlard (*9*) defined risk as the probability of injury-producing accidents. She developed three levels of risk, as follows:

1. Conflict types that would lead to a right-angle or head-on collision indicate a very high potential of injury accidents.
2. Conflict types that would lead to a rear-end collision, with or without a turning vehicle, show a medium potential of injury accidents.
3. Other types of conflicts were normally less dangerous and mostly reflect additional driving difficulties because they may require too much of the driver's attention.

Conflict types were ranked according to the three levels. Opposing left turn, left turn from left, and right turn from left have the highest level of risk, whereas left turn, same direction; right turn, same direction; lane change; right turn from left; and left turn from right have the lowest level of risk. For the purpose of calculating the risk index the following formula proposed by Taylor and Thompson (*10*) was used:

$$R.I_{ij} = K_i \cdot (I.V.)_{ij} \qquad (1)$$

$$R.I_j = \sum_{i=1}^{n} R.I_{ij} \qquad (2)$$

where

$R.I_{ij}$ = risk index for conflict type i at intersection j,
K_i = relative weight for conflict type i,
$= W_i / \sum_{i=1}^{n} W_i$,
$(I.V.)_{ij}$ = indicator value for conflict type i at intersection j,
n = number of conflict types,
W_i = weighting factor for conflict type i, and
$R.I_j$ = total risk index for intersection j

The weighting factor was determined by assigning a value of 3 to the conflict type with the highest rank, a value of 2 to conflict type

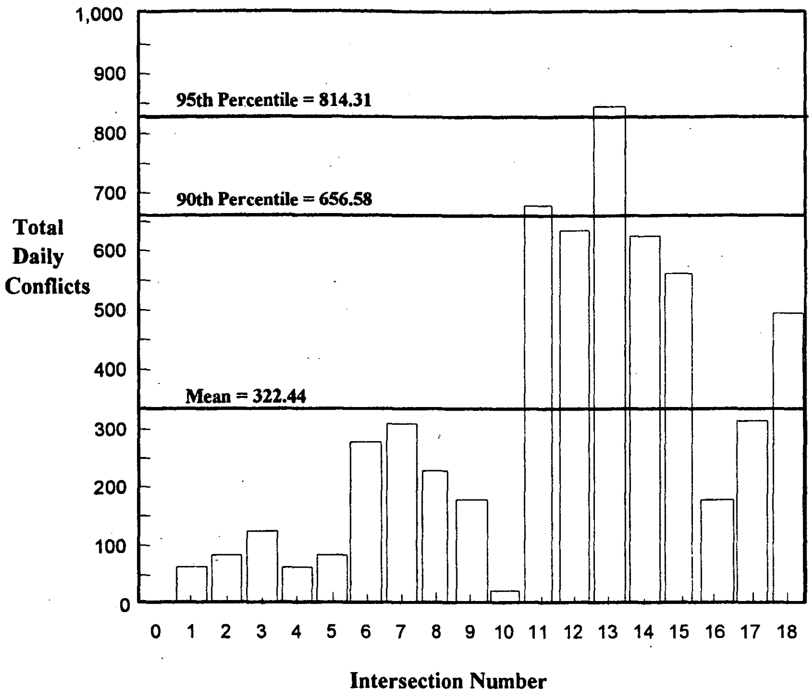


FIGURE 7 Mean and percentiles of left-turn, same-direction traffic conflicts at studied intersections.

with the next highest rank, and 1 to the conflict type with the lowest rank, and then the relative weight for each type of conflict was calculated. Additionally, the primary conflict rate per 1,000 vehicles for each conflict type and intersection was used as an indicator value for that conflict type.

After estimation of the risk index for each type of conflict, the risk index was calculated for each intersection. Finally, a priority ranking of the studied intersections was developed according to the risk index. Table 4 shows the risk index and rank order for each intersection. The results indicated that intersections 11, 6, and 7 are the most hazardous from a safety standpoint and have the greatest priority for countermeasures among all of the intersections studied.

Linear Regression Analysis

The linear regression analysis technique was used to model the relationships between traffic conflicts and volumes, accidents and traffic volumes, and accidents and traffic conflicts. Although previous studies had shown that traffic volume is not a good predictor of conflict, the study tried to verify if such a relationship may exist.

In the traffic conflict–traffic volume model prediction the total traffic conflicts could not be compared with the total entry volumes at the studied intersections. This could be explained by the fact that the FHWA procedure does not include a procedure for collecting data for the minor intersected street (the third approach in this case); therefore, the total entry volumes could not be determined. The researchers, however, believe that the lack of a procedure for collecting data for the minor intersected street was a weak point of the FHWA procedure. Since, as has been shown earlier, there were difficulties in investigating relationships between total traffic conflicts and total entry volumes, it was considered justified to see if relationships existed for the conflicts involving the separate maneuvers with different combinations of the flows generating these conflicts. The hourly conflicts were compared with the sums and square root of the product of the concerned flows generating these conflicts for various intersection categories to calibrate the models. The sum and square root of the product of traffic volumes were considered, as suggested by Spicer et al. (*11*). Intersection categories included the two-lane intersection roadway, four-lane intersection roadway, and all intersections. The following conflict types with their generating flows (Figure 1) were investigated.

TABLE 4 Risk Index and Priority Ranking of Studied Intersections

Intersection Number[a]	Risk Index	Priority Ranking
1	37.86	7
2	35.74	8
3	22.52	14
4	20.36	16
5	28.49	11
6	53.06	2
7	51.11	3
8	24.24	12
9	31.48	10
10	14.36	18
11	54.81	1
12	48.16	5
13	49.29	4
14	39.77	6
15	31.41	9
16	21.66	15
17	24.10	13
18	20.22	17

[a] See Table 1.

Left-Turn, Same-Direction Conflicts

A left-turn, same-direction conflict was generated from conflicting traffic coming from the north approach along the major roadway with traffic turning left into the minor roadway. The results of these investigations are presented in Table 5. Table 5 shows the coefficient of determination (R^2) for the predicted relationships between conflicts as a dependent variable and each traffic flow combination as an independent variable.

Table 5 shows that for a two-lane intersection roadway category, 74 percent of the total variation of this type of conflict could be explained by the summation of the two generating flows, whereas 81 percent of the total variation of this type of conflict could be explained by the square root of the product of the two generating flows. The variation of this type of conflict is more affected by the variation in the square root of the product of the two generating flows than by the summation of these flows. This finding supports the results found by Spicer et al. (*11*). This could be explained by the fact that the product of two flows will take into account the effect of one of the flows being small or zero. The sum, however, would not show this, especially when one flow is high. Table 5 also shows that the variation of this type of conflict at the two-lane intersection roadway category is more affected by the variation of traffic flow than the four-lane intersection roadway category. This could be explained by the fact that at the two-lane intersection roadway category the left-turn and through traffic flows use the same lane, whereas at the four-lane intersection roadway category the left-turn traffic flow mostly uses the inside lane and the through traffic flow mostly uses the outer lane of the street. Therefore, at the two-lane intersection roadway category this type of conflict is more affected by the variation in traffic flows than the four-lane intersection roadway category.

Right-Turn, Same-Direction Conflicts

A right-turn, same-direction conflict is generated from conflicting traffic that was coming from the south approach along the major roadway with traffic turning right into the minor roadway. The results of the regression analysis are presented in Table 6.

Table 6 shows that weak relationships exist between right-turn, same-direction conflicts and the summation of the generating flows for different intersection categories. For these relationships the coefficients of determination (R^2) were 0.23, 0.04, and 0.06 for two-lane, four-lane, and all intersection categories, respectively, whereas linear relationships were found between this type of conflict and the square root of the product of the generating flows at different intersection categories.

Right-turn, same-direction conflicts were less affected by traffic flows than left-turn, same-direction conflicts. This could be explained by the fact that the right-turn vehicles use the outer lane of the roadway before reaching the intersection area, in addition to the existence of proper channelization with large curb radii of some studied intersections. All of these factors assist in reducing the effects of traffic flow on right-turn, same-direction conflicts.

Opposing Left-Turn Conflicts

An opposing left-turn conflict was generated from conflicting traffic that was coming from the south approach along the major roadway with opposing traffic turning left into the minor roadway. The results of the regression analysis are presented in Table 7.

Table 7 shows that for a two-lane intersection roadway category 63 percent of the total variation of this type of conflict could be explained by the summation of the two generating flows, whereas 71 percent of the total variation of this type of conflict could be explained by the square root of the product of the two generating flows. This could be explained, as shown previously, by the fact that the product of two flows will take into account the effect of one of the flows being small or zero. The summation, however, would not show this, especially when one flow is high.

Table 7 also shows that the variation of the opposing left-turn conflicts at the two-lane intersection roadway category was affected by the variation in traffic flows more than four-lane and all intersection roadway categories were. This could be explained by the fact that at the two-lane intersection roadway category the absence of a left-turn storage lane forces the left-turning vehicles to implement their maneuvers without stopping to give right-of-way to the opposing through vehicles, whereas at the four-lane intersection roadway category the presence of an additional lane could help to reduce this type of conflict.

All Same-Direction Conflicts

An all same-direction conflict type was generated from the conflicting traffic of through, left-turn, and right-turn flows along or from the major roadway. The results of these investigations are presented in Table 8.

TABLE 5 Linear Regression Models Between Left-Turn, Same-Direction Conflicts and Different Combinations of Traffic Flows

Intersection Category	Regression Model	R-Square
2-Lane Intersection	C = -24.74 + 0.1216VT	0.74[a]
	C = -19.26 + 0.2459SROP	0.81[a]
4-Lane Intersection	C = -4.64 + 0.0648VT	0.61[a]
	C = -8.80 + 0.1625SROP	0.76[a]
All Intersections	C = -10.32 + 0.0782VT	0.64[a]
	C = -9.66 + 0.1671SROP	0.68[a]

C = Total hourly left-turn, same direction conflicts
VT = The summation of the two hourly flows generating these conflicts ($V_1 + V_2$)
SROP = The square root of product of the two hourly flows generating these conflicts $\left(\sqrt{V_1 * V_2}\right)$

[a] Significant at $\alpha = 0.001$

TABLE 6 Linear Regression Models Between Right-Turn, Same-Direction Conflicts and Different Combinations of Traffic Flows

Intersection Category	Regression Model	R-Square
2-Lane Intersection	C = -3.35 + 0.0354VT	0.23 [a]
	C = -7.49 + 0.1221SROP	0.56 [a]
4-Lane Intersection	C = 4.07 + 0.0082VT	0.04
	C = -6.15 + 0.0815SROP	0.53 [a]
All Intersections	C = 3.6 + 0.0111VT	0.06
	C = -3.71 + 0.0719SROP	0.40 [a]

C = Total hourly right-turn, same direction conflicts
VT = The summation of the two hourly flows generating these conflicts ($V_1 + V_2$)
SROP = The square root of product of the two hourly flows generating these conflicts ($\sqrt{V_1 * V_2}$)

[a] Significant at $\alpha = 0.001$

Table 8 shows that linear relationships existed between all same-direction hourly conflicts and hourly mainline traffic volumes for the three intersection categories. These relationships were statistically significant at the 0.05 level of significance.

Table 8 shows that 75 percent of the total variations of all same-direction conflicts could be explained by the mainline traffic volumes for the two-lane intersection roadway category. On the other hand, only 43 percent of the total variation of these conflicts could be explained by the mainline traffic volumes for the four-lane intersection roadway category. This could be explained by the fact that the two-lane intersection roadway category the total number of vehicles per one direction of movement (i.e., through, right turn, and left turn) will use the same lane, which increases the opportunities for conflicts between these vehicles. Therefore, any variations in the number of these vehicles will affect the number of these conflicts. On the other hand, at the four-lane intersection roadway category the total number of vehicles per one direction will be divided into two lanes, the through flow will use one of the two lanes, and the turning flows will use the other lane, which reduces the opportunities for conflicts between these traffic flows. Therefore, the variation of all same-direction conflicts would be explained more by the variation in traffic volume at the two-lane intersection roadway category than by that at the four-lane intersection roadway category.

Finally, in accident–traffic conflict model prediction, the total number of accidents per intersection was compared with the mean of the hourly conflict counts per intersection.

The following model was developed:

$$Y = 2.233 + 0.0349X \qquad (3)$$

TABLE 7 Linear Regression Models Between Opposing Left-Turn Conflicts and Different Combinations of Traffic Flows

Intersection Category	Regression Model	R-Square
2-Lane Intersection	C = -5.74 + 0.0488VT	0.63 [a]
	C = -4.05 + 0.1028SROP	0.71 [a]
4-Lane Intersection	C = 6.49 + 0.0322VT	0.30 [a]
	C = -6.49 + 0.1309SROP	0.61 [a]
All Intersections	C = -0.753 + 0.0406VT	0.48 [a]
	C = -2.81 + 0.1041SROP	0.66 [a]

C = Total hourly opposing left-turn conflicts
VT = The summation of the two hourly flows generating these conflicts ($V_1 + V_2$)
SROP = The square root of product of the two hourly flows generating these conflicts ($\sqrt{V_1 * V_2}$)

[a] Significant at $\alpha = 0.001$

where

Y = is the total number of accidents for the previous 3 years, and
X = is the mean hourly conflict counts.

The model was linear and statistically significant at the 0.001 level of significance.

CONCLUSIONS

The following conclusions were reached on the basis of the results of the analysis:

1. The left-turn, same-direction conflict type was the most predominant type compared with the other types of conflicts.
2. The average and abnormal limits for traffic conflicts at a three-leg, unsignalized intersection were determined.
3. A risk index was developed on the basis of subjective weighting of three levels of severity of the accidents produced from different types of conflicts. This index was used to provide a priority ranking of the studied intersections.
4. The relationships between left-turn, same-direction conflicts and the combinations of flows generating those conflicts for various intersection categories were linear. However, a higher correlation coefficient was evident between this type of conflict and the two combinations of the generating flows at the two-lane intersection roadway category rather than at the other intersection categories. Similarly, a higher correlation coefficient was evident between this type of conflict and the square root of the products of the two generating flows. Similarly, the opposing left-turn conflict yields the same results.
5. Weak relationships existed between right-turn, same-direction conflicts and the summation of the two generating flows for various intersection categories. The relationships between this

TABLE 8 Linear Regression Models Between All Same-Direction Conflicts and Mainline Traffic Volumes

Intersection Category	Regression Model	R-Square
2-Lane Intersection	C = -32.78 + 0.1178MLV	0.75 [a]
4-Lane Intersection	C = 13.97 + 0.0603MLV	0.43 [a]
All Intersections	C = -0.4236 + 0.0709MLV	0.51 [a]

C = All same direction hourly conflicts
MLV = Hourly mainline traffic volume
[a] Significant at $\alpha = 0.001$

type of conflict and the square root of the product of the two generating flows for various intersection categories were linear; however, a slightly higher correlation coefficient was obtained between this type of conflict and the square root of the products of the two generating flows at the two-lane intersection roadway category rather than at the other intersection categories.

6. The relationships between all same-direction conflicts and the mainline traffic volumes for various intersection categories were significantly linear; a higher correlation coefficient was evident between these conflicts and mainline traffic volumes at the two-lane intersection roadway category rather than at the other intersection categories.

7. The relationship between accidents and total conflicts was linear and statistically significant. Generally, as the number of traffic conflicts increased, the number of accidents also increased.

RECOMMENDATIONS

1. The average and abnormal conflict count limitations are recommended to be used as a guideline when evaluating other three-leg, unsignalized intersections in the city of Amman.

2. The traffic conflict technique is recommended for use in other cities in Jordan and on different types of intersections so that those cities can develop their own average conflict rates and provide the needed information to analyze intersections for their safety.

3. Volume counting along with the traffic conflict technique is recommended to be used to support the information needed to analyze intersections for their safety.

4. It is recommended that the risk index be used to provide a priority ranking of intersections for the implementation of countermeasures. For future studies, however when estimating the risk index, the results of the study recommend that the weighting factors should be determined according to the costs of the accidents produced in Jordan for each type of conflict.

REFERENCES

1. Williams, M. J. Validity of the Traffic Conflicts Technique. *Accident Analysis and Prevention,* Vol. 13, 1981, pp. 133–145.
2. Zegeer, C. V., and R. C. Deen. Traffic Conflict as a Diagnostic Tool in Highway Safety. In *Transportation Research Record 667,* TRB, National Research Council, Washington, D.C., 1978, pp. 48–55.
3. Glauz, W. D., and D. J. Migletz. *NCHRP Report 219: Application of Traffic Conflict Analysis at Intersection.* TRB, National Research Council, Washington, D.C., 1980.
4. Parker, M. R., and C. V. Zegeer. *Traffic Conflict Technique for Safety and Operation Engineers Guide.* Report FHWA-IP-88-026. FHWA, U.S. Department of Transportation, 1989.
5. Parker, M. R., and C. V. Zegeer. *Traffic Conflict Technique for Safety and Operations Observers Manual.* Report FHWA-IP-88-027. FHWA, U.S. Department of Transportation, 1989.
6. Glauz, W. D., K. M. Bauer, and D. J. Migletz. Expected Traffic Conflict Rates and the Use in Predicting Accidents. In *Transportation Research Record 1026,* TRB, National Research Council, Washington, D.C., 1985, pp. 1–12.
7. Crowe, E. C. Traffic Conflict Values for Three-Leg, Unsignalized Intersections. In *Transportation Research Record 1287,* TRB, National Research Council, Washington, D.C., 1990, pp. 185–194.
8. *Manual on Uniform Traffic Control Devices for Streets and Highways.* FHWA, U.S. Department of Transportation, 1978.
9. Muhlard, N. Systematic Behavioral Observation: The SBOT Experimented in the Philippines. *Proc., Second International Committee on Traffic Conflicts Technique Workshop,* No. 2, Munich, Germany, November 23 to 24, 1989.
10. Taylor, J. I., and H. I. Thompson. *Identification of Hazardous Locations.* Report FHWA-RD-76-44. FHWA, U.S. Department of Transportation, 1989.
11. Spicer, B. R., A. H. Wheeler, and S. J. Older. *Variation in Vehicle Conflicts at T-Junction and Comparison with Record Collisions.* Report SR 545. Transport and Road Research Laboratory, Crowthorne, Berkshire, United Kingdom, 1980.

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

Overreporting and Measured Effectiveness of Seat Belts in Motor Vehicle Crashes in Utah

J. MICHAEL DEAN, JAMES C. READING, AND PATRICIA J. NECHODOM

Motor vehicle crashes are a leading cause of injury and mortality in the United States. The effectiveness of seat belt use has been difficult to study because of the unavailability of population-based crash data bases that include all noninjured occupants. The 1991 Utah Crash Outcome Data Evaluation System data base was developed to determine seat belt effectiveness. It includes occupants of all police-reported motor vehicle crashes. Seat belt effectiveness may be overestimated, however, because of self-reporting of seat belt use when crash occupants are questioned by police. The effect of misclassification of seat belt use on the calculated odds ratio associated with seat belt use was studied by using logistic regression models of four levels of injury. The odds ratio associated with seat belt use for any degree of injury was 0.448 [95 percent confidence interval (CI) 0.425 − 0.473]; the odds ratios associated with seat belt use for injuries requiring outpatient emergency care, hospitalization, or fatalities were 0.476 (95 percent CI 0.449 − 0.504), 0.203 (95 percent CI 0.170 − 0.241), and 0.148 (95 percent CI 0.097 − 0.226), respectively. Adjustment of the fraction of correct classification of seat belt use among reported belt users decreased the protective effect associated with seat belt use for all four levels of injury. This is consistent with overestimation of seat belt effectiveness associated with non-differential misclassification. Based on the assumption that the 1991 observational use rate applies to the 1991 crash population, odds ratios were corrected for seat belt self reporting bias. The corrected odds ratio associated with protection from any degree of injury was 0.723 (95 percent CI 0.685 − 0.763); the corrected odds ratios associated with seat belt use and injuries requiring outpatient emergency care, hospitalization, or fatalities were 0.747 (95 percent CI 0.705 − 0.791), 0.505 (95 percent CI 0.421 − 0.606), and 0.455 (95 percent CI 0.296 − 0.697), respectively. The study results confirm the protective effect of seat belts in motor vehicle crashes and provide a methodology for correcting seat belt effectiveness estimates.

Motor vehicle crashes are a major cause of death and morbidity in the United States, accounting for an estimated cost of $137 billion in 1990. The use of safety restraint devices, including seat belts, shoulder straps, child safety seats, and airbags, has been associated with reduced mortality and morbidity. The total effect of restraint use may be underestimated, however, because previous studies have used patients in trauma centers (1–7). In past studies, victims of motor vehicle crashes were not considered unless they require treatment at the facility where the studies were based. Such studies may underestimate seat belt effectiveness, because they did not include persons who were uninjured.

The Utah Crash Outcome Data Evaluation System (CODES) project was designed to establish a population-based crash data base that includes all victims of reportable motor vehicle crashes in Utah. The data base includes all police crash reports, emergency medical service (EMS) run reports, and files (computerized discharge summaries of medical records) from hospitals and clinics. These hospitals and clinics include outpatient, emergency department, inpatient, and rehabilitation facilities. The 1991 data sources have been linked to a single crash data base using probabilistic linkage, with an estimated linkage efficiency of greater than 80 percent.

Seat belt use information is obtained from police crash reports, and in most instances this is reported by the crash occupant. Because the police officer usually does not witness the accident, uninjured occupants and occupants with minor injuries may be outside the vehicle when the officer is obtaining information concerning seat belt use. The police officer obtains information about seat belt use by asking the occupants. Because seat belt use is mandatory in Utah, crash occupants may report that they used seat belts to avoid a citation and fine. However, seat belt use by more severely injured or killed occupants can be directly assessed by police officers, particularly if extrication is required. Overreporting of seat belt use by uninjured occupants might be expected to overestimate the effectiveness of seat belts.

In Utah, the statewide rate of seat belt use was 46.9 percent as measured by direct observational studies in 1991. In the 1991 CODES data base, however, the reported use rate was 74 percent, suggesting that significant numbers of crash occupants overreport seat belt use to avoid a citation and fine. This is consistent with findings of other investigators, who note significant problems with self-reported seat belt use rates (*8–10*). It is clear that mandatory restraint laws increase self-reported use (*11,12*), but a disparity remains between self-reported rates and direct observational studies (*8,9*). Thus, estimates of seat belt efficacy may be biased.

Seat belt overreporting is a differential misclassification problem, because misclassification is much less likely if the occupant is killed or incapacitated. In addition, occupants who wear seat belts are extremely unlikely to deny seat belt use, because they will incur a citation and fine. Methods have been developed for correcting bias related to exposure variable (*13–15*), confounding covariate (*16–21*), and outcome (*22,23*) misclassification. Corrections of parameter estimates from logistic regression models have been developed (*17,24*) but relate to nondifferential misclassification.

The purpose of this paper is to evaluate the effect of seat belt overreporting (information biased differential misclassification) on the odds ratio and confidence limits relating seat belt use and injury. By using independent observational studies from the same time period, corrected odds ratios may be obtained to provide more reasonable estimates of seat belt effectiveness.

Primary Children's Medical Center, 100 North Medical Drive, Salt Lake City, Utah 84113.

MATERIALS AND METHODS

Utah 1991 CODES Data Base

The 1991 motor vehicle crash records were obtained from the Utah Department of Transportation, Division of Traffic and Safety. This file includes all crashes reported on public roads in Utah, including all injury accidents and accidents believed associated with property damage over $750. Crash data are obtained by local law enforcement officers who investigate each accident, including all crashes with estimated property damage of $750 or greater. The Utah crash reports include all occupants in a vehicle, including uninjured occupants. In 1991, there were 47,443 accidents involving 74,595 vehicles containing 103,812 occupants. For 7,983 occupants, information concerning restraint use was not available, and these cases were excluded from this study. For this study, only drivers of passenger vehicles and light trucks were included (N = 66,035).

Crash records include extensive information about the circumstances of the accident, including road and weather conditions, lighting conditions, vehicle descriptions, damage descriptions, and driver intentions. Vehicle occupant information includes age, gender, position within the vehicle, use of safety devices, and injury codes.

EMS data records for 1991 were obtained from the Bureau of Emergency Medical Services, Utah Department of Health. Most of these were submitted electronically by pre-hospital providers (>80 percent). All records submitted in hard copy form were reviewed by Utah CODES clerical staff, and incidents involving motor vehicle accidents were added to the computer file. Files containing summary discharge information were obtained from all hospitals in Utah. These files were complete for outpatient emergency department care, inpatient acute care, and rehabilitation care. EMS and hospital files were used to determine whether a crash occupant required on-scene or hospital-based medical care related to the crash.

Linkage of CODES Data Base Files

The CODES data base was constructed by linking the crash file to EMS and hospital files using probabilistic linkage. This methodology is described in detail elsewhere (*25–27*). Successful linkage is related to whether an individual was actually injured, errors in the data sources, and the effectiveness of the algorithm used to achieve the linkage. On the basis of detailed study of the linkage between the crash records and EMS run reports, it is believed that at least 80 percent of individuals who actually required an EMS response, outpatient, or inpatient medical care have been linked in this data base (unpublished observation).

Injury Stratification

Each driver record was classified according to treatment level as follows:

Level 0—No injury indicated on crash report nor linkage to medical care.

Level 1—Injury indicated on crash report but no linkage to an EMS, emergency department, or inpatient record.

Level 2—Injury indicated on crash report and transported by an EMS agency or treated in an emergency department but not hospitalized or killed.

Level 3—Hospitalized but did not die within 30 days of crash.

Level 4—Died within 30 days of crash.

Subsequently, four dependent variables A, B, C, and D were constructed to include populations with increasing levels of injury. Injury variable A was coded as 1 for all drivers with any treatment level above Level 0. Injury variable B was coded as 1 for all drivers with treatment Levels 2, 3, or 4. Injury level C was coded as 1 for all drivers with treatment Levels 3 or 4. Injury level D was coded as 1 for all drivers with treatment Level 4. Variable A is useful for detecting the occurrence of any injury, variables B and C provide information about increasingly severe injuries, and variable D provides information about fatalities. This stratification approach was developed by the seven states participating in CODES projects and staff from NHTSA.

Statistical Analyses

The data were analyzed using a logistic regression model with injury (stratified into Levels A through D as discussed) as the dependent variable. This logistic regression model used was developed by seven states involved in CODES projects and staff at NHTSA. To eliminate differences related to seating position, only drivers were used for this study. The independent variables were seat belt use, whether the vehicle rolled over, whether the crash was a single vehicle, fixed object or multiple vehicle collision, rural versus urban, age, gender, and posted speed limit. A driver was considered belted if the police crash report explicitly reported the use of seat belts or the combination of lap belt and shoulder straps; drivers were considered unbelted if the police crash report explicitly reported nonuse of any of these devices. Drivers for whom belt use was unknown or for whom airbags deployed (N = 43) were excluded from all analyses. In each model there were 43,017 drivers for whom all variables were available.

Stochastic Simulations of Misclassification

As already described, killed or incapacitated individuals (as assessed by the investigating police officer) were assumed to be correctly classified for seat belt use. In addition, reported nonusers were assumed to be correctly classified. All other reported users were considered susceptible to self-reporting bias.

By using a uniform random number generator, fractions f of the potentially misclassified drivers were changed to nonuser status, and the logistic regression model was recomputed with the same variables as described. The step was repeated numerous times for values of f of 0.05, 0.10, 0.15, and 0.25 to assess the variability between simulations for injury level B only. Subsequently, the models were recomputed for all fractions of f between 0 and 1.00 for injury levels A, B, C, and D.

Based on the assumption that the observational rate of 46.7 percent applies to the crash population, the fraction $f = 0.34$ would yield this total rate of use. The logistic models were run with $f = 0.34$ for all 4 injury levels, A through D.

Computer Software

Probabilistic record linkage was performed using Automatcher 2.0 (Matchware, Inc., Silver Spring, Maryland). Data base manipula-

TABLE 1 Seat Belt Use and Sustained of Drivers in 1991 CODES

Seat Belt Use	Total Drivers	Level of Injury None	A	B	C	D
No	19,285	14,888	4,397	3,621	484	86
Yes	46,750 (71%)	40,678 (73%)	6,072 (58%)	5,184 (59%)	288 (37%)	34 (28%)
	66,035	55,566	10,469	8,805	772	120

tions were done with Foxpro 2.0 (Microsoft Corp., Redmond Washington). Logistic regression was done with the logistic procedure (PROC LOGISTIC), and random numbers were generated with the RANUNI function in the SAS statistics system (SAS Institute, Inc., Cary, North Carolina).

RESULTS

The reported use of seat belts by drivers in the Utah CODES data set is shown in Table 1. The table also shows the numbers of drivers, belted and unbelted, who sustained injury at Levels A through D. The overall reported use rate among drivers was 71 percent, which contrasts with the directly measured observational rate of 46.9 percent. There were 55,566 drivers who sustained no injury, and 10,469 drivers sustained some level of injury (injury Level A). As the injury level increases from A through D, the reported seat belt use rate drops from 73 percent (uninjured) to 28 percent (fatalities).

The odds ratio associated with seat belt use varied by injury level. Table 2 shows the full model for injury Level A; the seat belt odds ratio is 0.448 (95 percent CI 0.425 − 0.473). The odds ratios associated with belt use for injury levels B, C, and D were 0.476 (95 percent CI 0.449 − 0.504), 0.203 (95 percent CI 0.170 − 0.241), and 0.148 (95 percent CI 0.097 − 0.226), respectively. Thus, as the level of injury increases, stronger protective effects are seen from the use of seat belts.

Logistic regression was performed with fractions $f = 0.05$, 0.10, 0.15, and 0.25 multiple times, using B level injury as the dependent variable. Figure 1 shows a plot of the resulting odds ratios. It can be seen that, as increasing fractions of reported users are randomized to nonuser status, the odds ratio increases, indicating an attenuation of effect. This confirms the expectation that overreporting exaggerates the effectiveness of seat belts. The figure also demonstrates that repetitive stochastic simulations result in similar odds ratios. For this reason, in subsequent results, single simulations at various fractions f for each injury level are reported.

Logistic regression was then performed with fractions from 0.05 to 1.00 for injury Levels A through D. Figures 2 through 5 show the resulting plots of odds ratios against fractions f. For all injury levels, overreporting clearly causes an exaggeration of seat belt efficacy, though the curves behave somewhat differently. Figure 6 shows all four curves on the same graph.

Note for each injury level that the odds ratio increases rapidly as f approaches 1.00. This is to be expected because of the assumption that all incapacitated or killed individuals were correctly classified with respect to seat belt use. As f approaches 1.00, seat belt use preferentially occurs in incapacitated and killed individuals.

By using observational data to adjust use rates, 15,776 (34.3 percent) reported users must be assumed to be nonusers (again, it is assumed that killed or incapacitated users are correctly classified). Logistic regression models were recomputed for each injury level after randomly switching this number of users to nonusers. The corrected odds ratio associated with protection from any degree of injury was 0.723 (95 percent CI 0.685 − 0.763); the corrected odds ratios associated with seat belt use and injuries requiring outpatient emergency care, hospitalization, or fatalities were 0.747 (95 percent CI 0.705 − 0.791), 0.505 (95 percent CI 0.421 − 0.606), and 0.455 (95 percent CI 0.296 − 0.697), respectively.

DISCUSSION OF RESULTS

In this study, it has been demonstrated that (a) seat belt use reduces injury from motor vehicle crashes with an odds ratio of approximately 0.723, and (b) misclassification of seat belt use exaggerates the apparent effectiveness of seat belt use in preventing injury and

TABLE 2 Result of Unadjusted Logistic Regression Model: Level A is dependent Variable

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	Odds Ratio	95% Confidence Interval
Model Intercept	-1.6009	0.0571	786.43	0.20	0.18 - 0.23
Seat Belt Use	-0.8022	0.0278	830.52	0.45	0.42 - 0.47
Rollover	1.6686	0.0573	848.84	5.31	4.74 - 5.94
SVFO	0.3799	0.0843	20.30	1.46	1.24 - 1.73
SVO	-0.0944	0.0445	4.49	0.91	0.83 - 0.99
MVH	1.162	0.0958	147.01	3.20	2.65 - 3.86
Rural	-0.1946	0.0331	34.66	0.82	0.77 - 0.88
Age	0.00198	0.0008	5.91	1.002	1.000 - 1.004
Male	-0.5061	0.0267	359.41	0.60	0.57 - 0.64
Speed Limit	0.0179	0.0012	218.12	1.018	1.016 - 1.020

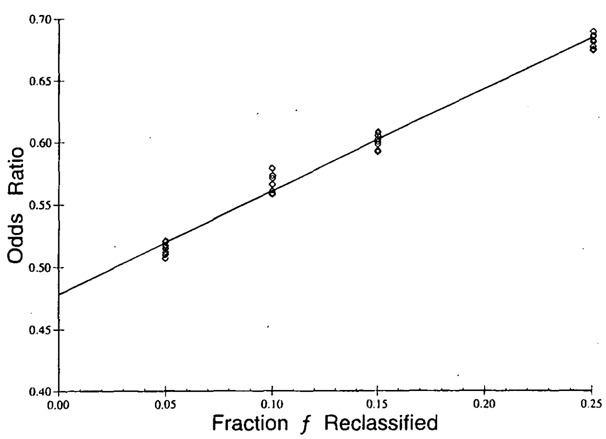


FIGURE 1 Odds ratio for injury Level B associated with seat belt use, function of fraction *f* of drivers randomly changed from user to nonuser status. Each point represents separate simulation, multiple simulations conducted for $f = 0.05, 0.10, 0.15$, and 0.25.

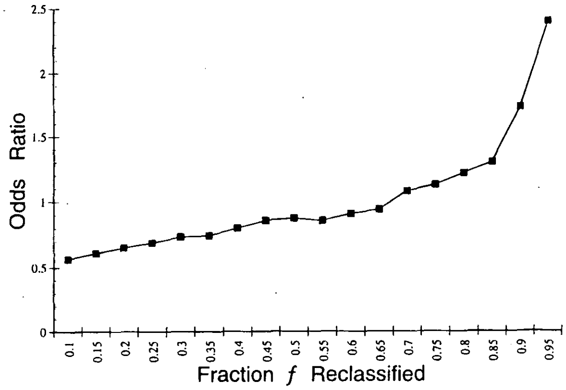


FIGURE 3 Odds ratio for injury Level B associated with seat belt use, function of fraction *f* of drivers randomly changed from user to nonuser status. Each point represents separate simulation.

mortality (to an odds ratio of 0.45 for injury of any type). Statewide observational use data are used to adjust the odds ratios to obtain more reasonable estimates of seat belt efficacy for each level of injury from minor to fatal.

Seat belt use has been known to reduce injury in motor vehicle crashes, but few previous studies have used population-based data (*28*). This is likely the first report of seat belt effectiveness based on a comprehensive, population-based, statewide crash data base that includes all drivers, injured and uninjured. The odds ratio obtained from this population-based study is consistent with previous estimates of seat belt effect (*29*).

The major findings of this study relate to the effect of differential misclassification of seat belt use on the association between seat belt use and noninjury. The observational seat belt use rate in Utah was 46.9 percent, and the expected true value of *f* was 0.343 in 1991. Assuming different probabilities *f* of accurate classification, Figures 1 through 4 demonstrate the exaggeration of the protective effect of seat belts with overreporting bias. It is believed that the parameter estimate obtained at $f = 0.343$ represents a reasonable estimator of the true odds ratio associated with belt use.

It is assumed that police reports accurately classify seat belt use by killed or incapacitated drivers, and adjustments of seat belt classification was restricted to other drivers. Killed or incapacitated drivers are more likely to remain in the vehicle until being extricated by external observers (police, fire fighters, or bystanders), and these external observers are likely to validate the investigating police officer's assessment of belt use. The reported seat belt use rates for killed and incapacitated occupants were 28 and 39 percent, respectively. These rates are well below the statewide observational rate of 46.9 percent. This suggests that there is not a large amount of misclassification in this subset of drivers or the reported use rate would be considerably higher. In contrast, police reports indicated that 73 percent of the uninjured drivers used seat belts, well above the observational rate. In these instances, the drivers are likely to

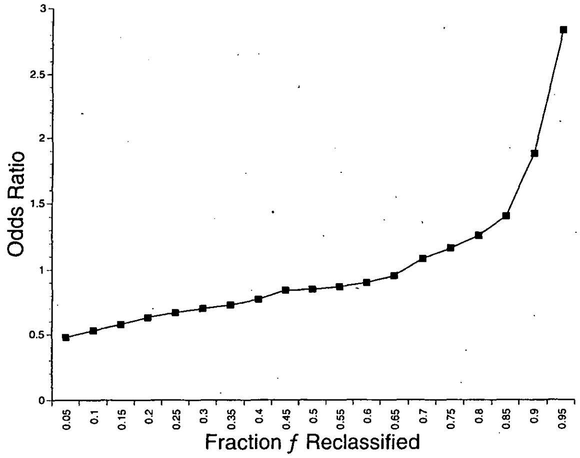


FIGURE 2 Odds ratio for injury Level A associated with seat belt use, function of fraction *f* of drivers randomly changed from user to nonuser status. Each point represents separate simulation.

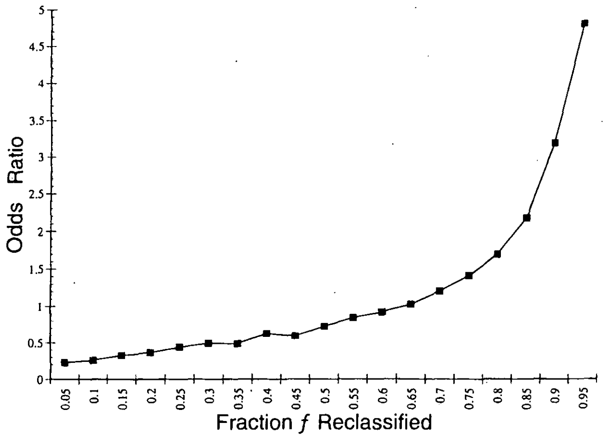


FIGURE 4 Odds ratio for injury Level C associated with seat belt use, function of fraction *f* of drivers randomly changed from user to nonuser status. Each point represents separate simulation.

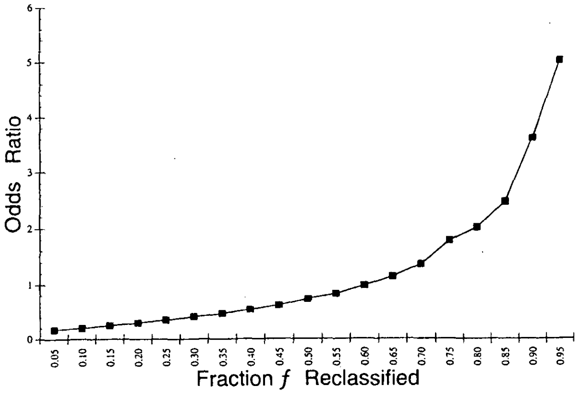


FIGURE 5 Odds ratio for injury Level D associated with seat belt use, function of fraction *f* of drivers randomly changed from user to nonuser status. Each point represents separate simulation.

self-report belt use to the police officer, as such drivers will have already left their vehicle before the arrival of the police officer. These reported use rates (lower than observational) suggest that the police assessment of belt use may be reasonably accurate for killed or incapacitated occupants, but significantly overestimates belt use by uninjured occupants.

To assess this assumption, the simulations were calculated with the assumption that only the dead drivers and nonusers were correctly classified. The odds ratios were lower, not higher, for injury Levels A through C for values for fraction *f* from 0.05 through 0.40 (the range tested). For fatalities (Level D), odds ratios were higher by no more than 0.06 (value at 0.40 reclassification of drivers). At the reclassification rate of 0.343, assuming incapacitated drivers were as likely to be incorrectly classified as less severely injured drivers, odds ratios were 0.654, 0.687, 0.460, and 0.512 for injury Levels A through D, respectively. Thus, the assumption of correct classification of incapacitated drivers renders a more conservative estimate of seat belt efficacy.

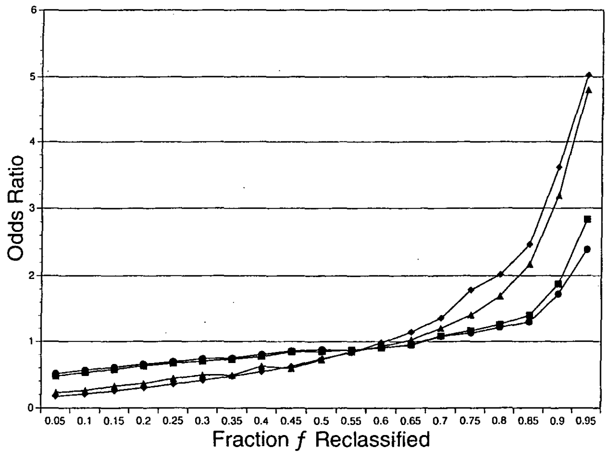


FIGURE 6 Odds ratio for all four injury levels A (square), B (circles), C (triangles), and D (diamonds) associated with seat belt use, function of fraction *f* of drivers randomly changed from user to nonuser status. Each point represents separate simulation.

Using a single statewide assumption for seat belt misclassification may be criticized, because it is known that males less often use belts than females, usage rates are lower in rural areas, and usage rates are higher on freeways and highways. However, the low variability seen in Figure 1 for repetitive stochastic simulations suggests that the parameter estimates are stable and are much more affected by the relatively large number of drivers who are assumed misclassified. It is not believed that a more sophisticated simulation, stratified for gender, roadways, or rural location, would yield significantly different results from those observed in Figures 2 through 6.

In summary, seat belts significantly reduce injury from motor vehicle crashes. The magnitude of the effect of seat belts on the rate of injury is exaggerated by overreporting seat belt use. Stochastic simulations were used to evaluate the bias associated with misclassification of seat belt use. By using independent observational use data, the original odds ratios were corrected to provide a more reasonable estimate of seat belt efficacy in Utah.

ACKNOWLEDGMENTS

The authors thank the Utah Crash Outcome Data Evaluation System Advisory Committee members for their assistance in obtaining the crash related data files, helpful comments, and suggestions. The authors gratefully acknowledge J. Lynn Lyon for his critical review of this manuscript. This research was supported by Cooperative Agreement DTNH22-92-Y-57329 from NHTSA.

REFERENCES

1. Bueno, M. M., N. Redeker, and E. M. Norman. Analysis of Motor Vehicle Crash Data in an Urban Trauma Center: Implications for Nursing Practice and Research. *Heart and Lung*, Vol. 21, 1992, pp. 558–567.
2. Hooker, E. A., D. F. Danzl, D. M. Thomas, F. Miller, and W. Zupances. Economic Impact of Motor Vehicle Restraints in Kentucky: a Trauma Center's Experience. *Journal of Kentucky Medical Association,* 1990, Vol. 88, pp. 59–61.
3. Kaplan, B. H., and R. A. Cowley. Seatbelt Effectiveness and Cost of Noncompliance among Drivers Admitted to a Trauma Center. *American Journal of Emergency Medicine,* Vol. 9, 1991, pp. 4–10.
4. Orsay, E. M., T. L. Turnbull, M. Dunne, J. A. Barrett, P. Langenberg, and C. P. Orsay. Prospective Study of the Effect of Safety Belts on Morbidity and Health Care Costs in Motor-Vehicle Accidents. *Journal of The American Medical Association,* Vol. 260, 1988, pp. 3598–3603.
5. Orsay, E. M., M. Dunne, T. L. Turnbull, J. A. Barrett, and P. Langenberg, C. P. Orsay. Prospective Study of the Effect of Safety Belts in Motor Vehicle Crashes. *Annals of Emergency Medicine,* Vol. 19, 1990, pp. 258–261.
6. Redelmeier, D. A., and P. J. Blair. Survivors of Motor Vehicle Trauma: an Analysis of Seat Belt Use and Health Care Utilization. *Journal of General Internal Medicine,* Vol. 8, 1993, pp. 199–203.
7. Rutledge, R., A. Lalor, D. Oller, A. Hansen, M. Thomason, W. Meredith, M. B. Foil, and C. Baker. The Cost of Not Wearing Seat Belts. A Comparison of Outcome in 3396 Patients. *Annals of Surgery,* Vol. 217, 1993, pp. 122–127.
8. Comparison of Observed and Self-Reported Seat Belt Use Rates—United States. *Morbidity and Mortality Weekly Report,* Vol. 37, 1988; pp. 349–351.
9. Streff, F. M., and A. C. Wagenaar. Are There Really Shortcuts? Estimating Seat Belt Use with Self-Report Measures. *Accident Analysis and Prevention*, Vol. 21, 1989, pp. 509–516.
10. Robertson, L. S. The Validity of Self-Reported Behavioral Risk Factors: Seatbelt and Alcohol Use. *Journal of Trauma,* Vol. 32, 1992, pp. 58–59.
11. Escobedo, L. G., T. L. Chorba, P. L. Remington, R. F. Anda, L. Sanderson, and A. A. Zaidi. The Influence of Safety Belt Laws on Self-Reported Safety Belt Use in the United States. *Accident Analysis and Prevention,* Vol. 24, 1992, pp. 643–653.

12. Fielding, J. E., K. K. Knight, and R. Z. Goetzel. The Impact of Legislation on Self-Reported Safety Belt Use in a Working Population. *Journal of Occupational Medicine,* Vol. 34, 1992, pp. 715–717.
13. Birkett, N. J. Effect of Nondifferential Misclassification on Estimates of Odds Ratios with Multiple Levels of Exposure. *American Journal of Epidemiology,* Vol. 136, 1992, pp. 356–362.
14. Brenner, H, and M. Blettner. Misclassification Bias Arising from Random Error in Exposure Measurement: Implications for Dual Measurement Strategies. *American Journal of Epidemiology,* Vol. 138, 1993, pp. 453–461.
15. Flegal, K. M., C. Brownie, and J. D. Haas. The Effects of Exposure Misclassification on Estimates of Relative Risk. *American Journal of Epidemiology,* Vol. 123, 1986, pp. 736–751.
16. Greenland, S. The Effect of Misclassification in the Presence of Covariates. *American Journal of Epidemiology,* Vol. 112, 1980, pp. 564–569.
17. Rosner, B., D. Spiegelman, and W. C. Willett. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Measurement Error: the Case of Multiple Covariates Measured with Error. *American Journal of Epidemiology,* Vol. 132, 1990, pp. 734–745.
18. Cox, B., and J. M. Elwood: The Effect on the Stratum-Specific Odds Ratios of Nondifferential Misclassification of a Dichotomous Covariate. *American Journal of Epidemiology,* Vol. 133, 1991, pp. 202–207.
19. Savitz, D. A., and A. E. Baron. Estimating and Correcting for Confounder Misclassification *American Journal of Epidemiology,* Vol. 129, 1989, pp. 1062–1071.
20. Vach, W., and M. Blettner. Biased Estimation of the Odds Ratio in Case-Control Studies Due to the Use of Ad Hoc Methods of Correcting for Missing Values of Confounding Variables. *American Journal of Epidemiology,* Vol. 134, 1991, pp. 895–907.
21. Brenner, H. Bias Due to Non-Differential Misclassification of Polytomous Confounders. *Journal of Clinical Epidemiology,* Vol. 46, 1993, pp. 57–63.
22. Copeland, K. T., H. Checkoway, A. J. McMichael, and R. H. Holbrook. Bias Due to Misclassification in the Estimation of Relative Risk. *American Journal of Epidemiology,* Vol. 105, 1977, pp. 488–495.
23. Hsieh, C. C. The Effect of Non-Differential Outcome Misclassification on Estimates of the Attributable and Prevented Fraction. *Statistics in Medicine,* Vol. 10, 1991, pp. 361–373.
24. Rosner, B., D. Spiegelman, and W. C. Willett. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Random Within-Person Measurement Error. *American Journal of Epidemiology,* Vol. 136, 1992, pp. 1400–1413.
25. Jaro, M. A. Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida. *Journal of the American Statistical Association,* Vol. 84, 1989, pp. 414–420.
26. Fellegi, I. P., and A. B. Sunter. A Theory for Record Linkage. *Journal of the American Statistical Association,* Vol. 64, 1969, pp. 1183–1210.
27. Newcombe, H. B., and J. M. Kennedy. Record Linkage. *Communication of the Association for Computing Machinery,* Vol. 5, 1962, pp. 563–566.
28. Campbell, H. J. Safety Belt Injury Reduction Related to Crash Severity and Front Seated Position. *Journal of Trauma,* Vol. 27, 1987, pp. 733–739.
29. Evaluation of the Effectiveness of Occupant Protection. National Highway Traffic Safety Administration Interim Report, June 1992.

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