Avoiding Delay, Death, and Dirty Air: Framework for Evaluation of Intelligent Vehicle-Highway Systems

Steven Underwood

Frederick Streff

University of Michigan
Ann Arbor

The University of Michigan IVHS

Intelligent Vehicle-Highway Systems
College of Engineering • (313) 764-4332
Professor Kan Chen, 4112 EECS, Ann Arbor, MI 48109-2122, and
University of Michigan Transportation Research Institute • (313) 936-1066
Robert D. Ervin, 2901 Baxter Road, Ann Arbor, MI 48109-2150

IVHS Technical Report: #92-14

Avoiding Delay, Death, and Dirty Air:
Framework for Evaluation of Intelligent Vehicle-Highway Systems

Steven E. Underwood, Ph.D., and Fredrick M. Streff, Ph.D.

Abstract

This paper will discuss a new framework for evaluating Intelligent Vehicle-Highway Systems (IVHS). The framework takes a managerial rather than a methodological perspective, and clarifies three distinct forms of evaluation that may be used to assess the value of IVHS: field test, comparative, and prospective evaluation. *Field test evaluation* seeks to establish the performance (or the effectiveness) of selected systems implemented in field settings. Field test evaluations pose special problems in project planning and management because these evaluations involve many individual studies that require close coordination. Moreover, field test evaluations of IVHS are likely to involve tighter experimental control than traditional traffic evaluations. *Comparative evaluation* seeks to establish the relative merit of competing system architectures. Comparative evaluation efforts may incorporate the results from site-specific testbed field comparisons, but more likely, they will compare systems architectures in simulation testbeds in order to control for external factors that would confound comparisons between site-specific field tests. *Prospective evaluation* applies extrapolative and analytical forecasting methods to assess possible benefits of increasing market penetration. Critical ingredients of prospective evaluations are forecasts (or assumptions) of market penetration and public acceptance for IVHS products and services. An important relationship among the three approaches to evaluation is the sharing of forecasting and impact models. Simulation models may be calibrated in field test settings and then applied to extrapolate benefits and compare the merits of alternative systems architectures.

Avoiding Delay, Death and Dirty Air: Framework for Evaluation of Intelligent Vehicle-Highway Systems

Steven E. Underwood, Ph.D., and Fredrick M. Streff, Ph.D.

The problems of traffic have been with us for ages. Motorists have come to expect the unexpected when it comes to road travel, including congestion due to incidents on our urban highways and surface streets, wasted time due to miscues and navigational errors, and the possibility of accident-related injuries and fatalities. Other difficulties are more regular: recurring congestion at peak periods, exposure to harmful vehicle emissions, the increasing expense of fuel consumption, and general travel-related stress and uncertainty.[1] These are the types of problems that Will Rogers lamented when he quipped "the only way to solve the traffic problems of the country is to pass a law that only paid-for cars are allowed to use the highways." While this may be a humorist's notion of a solution, it is certainly not the ONLY way to solve the traffic problem, and since the advent of intelligent vehicle-highway systems the number of options for dealing with traffic appears to be continually growing.

Intelligent vehicle-highway systems (IVHS) are the application of advanced electronics and communications technology to improve the routing, flow, and general performance of road transportation. IVHS promises to extend the capabilities of drivers and system managers through new information and control alternatives, including motorist information, navigation, route guidance, signal coordination, collision warning, vehicle control, and other similar systems.[2] The problem with IVHS, however, is that the proposed systems are largely untested; we do not really know whether they will deliver benefits as the designers claim and the public expects. The best way to find out is to test them, preferably in limited deployment under controlled conditions, to determine whether specific options are worth further investment. In fact, the need for a systematic evaluation of IVHS has turned the attention of the transportation community to an emerging series of planned operational field tests and system architectures studies. The time has come in the IVHS community to move speculation and entrepreneurial claims aside, and replace them with proven concepts and demonstrated benefits.

The IVHS community faces an uncommon opportunity at this time to assess the impacts and relative merits of alternative system deployments. Rarely do private initiatives address the question of societal benefits. It is equally rare for public evaluation efforts to be initiated with (1) clear program objectives, (2) adequate administrative authority to assure planning and control, and (3) sufficient resources for effective implementation. However, these are precisely the conditions that exist as we embark on the evaluation phase of IVHS development and deployment. The Congressional mandate of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) and the strategic planning effort organized by IVHS AMERICA have provided the structure and impetus needed to bring the wide variety of partners together to establish a set of IVHS community goals, targets, and evaluation procedures. On the public side the U.S. Congress has seen fit to support IVHS development efforts tied to this strategic plan and concrete evaluation plans. Similarly, organizations in the private sector have demonstrated their willingness to establish partnerships with public organizations for joint ventures in IVHS. These appear to be ideal conditions for establishing public-private partnerships for tests and evaluations.

The stage is now set for the initiation of a meaningful effort designed to (1) assess the net benefits of IVHS, (2) evaluate new field deployments, and (3) compare alternative deployment schemes using the consensus of criteria stated in the strategic plan. What is needed at this point is some framework and set of standardized guidelines for evaluation. This paper presents a sketch for the first of these two needs; that is, it describes an initial framework for evaluating IVHS. The paper's intent is to detail components of a comprehensive national effort for IVHS evaluation and to describe relationships between these components. We take a project management and planning approach as opposed to the standard methodological approach to evaluation.[3]

The paper begins with a review of the mandate for evaluation of IVHS in the U.S. and North America. It then moves to a discussion of the special role that field tests will play in the overall evaluation effort, and the importance of formal evaluation. The discussion then turns to delineate the proposed evaluation framework. This framework establishes three distinct forms of evaluation -- field test, comparative, and projective -- and their relationships in the broader evaluation effort. The paper concludes with a review of some planning and implementation issues of particular relevance to IVHS evaluation efforts. The final section provides an outline review of the main points of the paper.

A Mandate for Evaluation

Peter Rossi (1971), a leading authority on evaluation methodology, emphasizes that "evaluation research is more than the application of methods. It is also a political and managerial activity, an input into the complex mosaic from which policy decisions and allocations emerge for the planning, design, implementation, and continuance of programs to better the human condition." With this in mind, it is useful to review the political mandate for evaluation of IVHS in the United States as a basis for formulating a managerial framework. This section reviews the evolution of activities at the federal level leading to requirements for evaluation of IVHS.

Since the genesis of the IVHS movement in the United States, there has been widespread recognition of the vital need for systematic data collection and evaluation of new IVHS deployments in order to improve estimates of the benefits and obstacles associated with their application. Public support of infrastructure requires demonstrated benefits. Yet, little empirical data exists at this point.

The United States General Accounting Office (GAO) conducted an exhaustive review of operational field tests in the U.S. and found positive but limited data on the demonstrated effect of various new and possible future deployments. They applauded the Department of Transportation's recent efforts to design and conduct broad multiple-measure evaluations on Pathfinder and TRAVTEK. They concluded that while IVHS are promising, they require further testing to examine effects of IVHS on a range of potential benefits (General Accounting Office, 1991).

Before passage of the 1991 highway bill the GAO recommended to Congress that "IVHS legislation should require DOT to select, design, and evaluate high-priority operational field tests in accordance with a strategic IVHS research plan." A central theme of the GAO report was that "policy should be aimed at guiding the development of evaluative information that will allow for among other benefits, knowledgeable decisions about the appropriate federal investment in IVHS and how best to target it."

ISTEA picked up on these recommendations in an unprecedented funding of research and development to provide new answers to the congestion and safety questions. In the IVHS arena, the measure focusses on promotion of compatible standards, promotion of widespread use of

IVHS technologies, and the establishment of evaluation guidelines and clearinghouse for IVHS operational field tests.

Given the scope of the federal IVHS program,[4] the challenge for the Federal Highway Administration (FHWA) (the designated lead agency for the Department of Transportation's (DOT) activities in IVHS) is to deliver a program that is approximately 40 times as large as the FY-90 program and 9 times as large as the FY-91 program. The task is to deliver this program in a manner that will leverage federal funds, will be acceptable to Congress, and will be acceptable to the IVHS partners. Accountability for disbursement of these funds is a reasonable concern given the disproportionately small increase in FHWA staff. It appears obvious that the traditionally conservative FHWA procurement and project management mechanisms will have to adjust to this new set of circumstances.

ISTEA emphasizes the need for operational field tests of new technologies. According to the act, the highest priority will be given to projects that (1) contribute to the goals of the strategic plan, (2) minimize the proportion of Federal funding, (3) build on previous work to advance the current state of knowledge, and (4) require a written evaluation consistent with established guidelines. It is incumbent on the Secretary of Transportation to establish the requisite guidelines and requirements of the field and related operational tests. The guideline requirement may be in part credited to the GAO recommendations. ISTEA further stipulates that DOT "maximize the involvement of the United States private sector, colleges and universities, and State and Local governments in all aspects of the program, including design, conduct (including operations and maintenance), evaluation, and financial or in-kind participation."

Recent drafts of the IVHS AMERICA strategic plan[5] provide an indication of what the IVHS community believes to be important goals and what is the appropriate role of the operational field tests and associated evaluation studies. According to the plan, "operational tests are needed to determine whether a promising technology or system is ready for deployment, whether the expected benefits can be achieved at the expected cost, and to assist in the transition to the market place. In order to maximize the cumulative benefits of many such tests, a set of national selection and evaluation criteria and standardized evaluation methodologies will need to be administered with federal leadership. Longer term operational evaluation will be needed to ensure that developed systems are reliable, durable, maintainable, cost effective, and fail safe, and that they perform as intended."

The IVHS AMERICA strategic plan provides few details on the nature of the evaluation guidelines that will be prepared by FHWA as required by ISTEA. It backs away from specific guidelines and instead talks about a more general framework for evaluation. The plan states that it is "essential to provide a structured framework for evaluating the societal benefits and costs of particular IVHS technologies. The use of such a framework will allow consistent and comparable results of alternative approaches for meeting the particular needs of different areas of the country, while still assuring essential uniformity. It will also allow private sector organizations to be more effective in developing products that meet the needs of their customers." The plan recognizes the need for baseline measurements and the systematic collection of information on accident rates, congestion levels, travel speed, delays, and mobility measures. It also specifies organizational roles vis-a-vis operational field tests with U.S. DOT taking responsibility for the development of uniform evaluation guidelines and management of the operational test program, state and local governments taking responsibility for conducting field tests, and academia serving as a "major participant."

Role of Operational Field Tests

Based on the preceding discussion one can see that the operational field tests serve several purposes in support of a national program in IVHS. Four of these purposes are outlined in the following section: (1) technology transfer, (2) proof of concept and product readiness, (3) design improvement through field learning, and (4) systematic evaluation of alternative systems.

First, the operational field tests should facilitate the transfer of technology from other industries to problems in the transportation sector . In many cases this is merely a matter of applying existing technologies to transportation-sector problems. For example, technologies developed in the defense industry may find many applications in IVHS. This type of transfer also typically involves transfer of technology from private-sector laboratories to publicly installed and maintained transportation systems. Education and training of public agency engineers is often a prerequisite to successful public-private transfers of technology.

Technologies, with potential applications to road transportation that exist in the laboratory, may be slow to make it to the field without the benefit of operational field tests. For example, image processing technologies that could be used for traffic sensing may never make it out of the laboratory without operational testing opportunities in the field. Field tests provide the opportunity to collect cost and performance data. Only after a system is tested, and proven to work at a reasonable cost, will it then become a candidate for success in the market place. Perhaps more important, the field test opportunities may serve as magnets for the application of new technologies that would not have otherwise been considered for use in road transportation systems.

Second, operational field tests are useful for proving system concepts and product readiness. It is one thing to devise an architecture and make claims about potential benefits, and it is another to successfully implement the concept in the field. The architecture may appear clever and worthwhile on paper or in the lab, and yet encounter insurmountable obstacles (e.g., unexpected costs, unreliable components, etc.) when it comes to field testing. For example, a city transit authority recently installed navigational devices in a portion of its fleet to test a concept for real-time bus tracking. It turned out that the navigational devices could not withstand the rough conditions produced by regular bus travel and the system was shelved, at least temporarily. Moreover, what appears to be a useful service in concept may turn out to be of little value or even annoying to the end user. This is a concern with most of the voice-based guidance systems because voice-based driver warnings have been known to irritate a large segment of drivers.

Third, technical difficulties should be expected in a first-time implementation of a new concept; operational field tests provide the opportunity to learn about potential obstacles and adjustments to overcome them. If the transceiver for automatic toll collection does not work properly in initial trials, it is probably better to adjust the mounting and redirect the antennae before concluding that the concept was a failure. Obstacles are rarely terminal. Rather, they are more often paths to learning and making improvements in the system. In fact, learning about what is useful and workable in the field should be seen as the primary objective of operational field tests. Therefore, operational field tests should be planned so that opportunities for learning occur within each of the field tests and between successive field tests. Each new field test should draw on the lessons from previous attempts and systematic evaluation.

Finally, individuals and organizations will want to make informed decisions on whether to support, develop, and/or purchase these new systems, and they will require sound data with which to support these decisions. Operational field tests provide the means for systematically evaluating the systems and their components. These evaluations should include assessments of benefits and drawbacks, costs, consumer satisfaction and utilization, willingness-to-pay, and extrapolation to broader levels of implementation into the future.

Role of Formal Evaluation

Formal evaluation is a tool that, if employed appropriately, can assist all parties with interests in the impact of IVHS. Evaluation is not the same as operational field testing, nor is evaluation even the sole purpose of the operational field tests. Instead, formal evaluation is a systematic process for generating information about IVHS for design improvements and decision making. It is a formal learning process that involves the systematic application of research procedures in assessing the concept, design, implementation, and value of a system or technology.

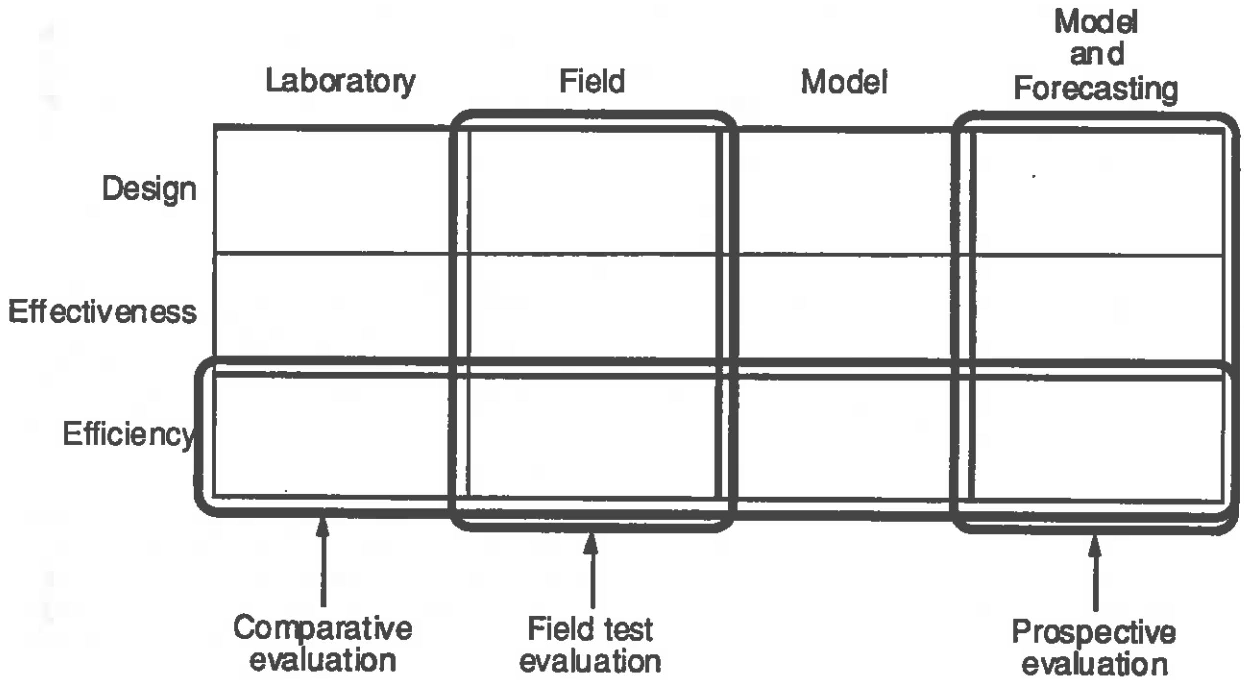
Figure 1 presents the domain of IVHS evaluation activities as a combination of evaluation purpose and research platform. Possible purposes include (1) improving system design, (2) assessing system effectiveness, and (3) comparing systems for the most efficient alternative. Research platforms include (2) laboratory, (2) field, (3) model, and (4) model with forecasted parameters. The combination of purpose and platform results in 12 alternative approaches to evaluation of IVHS, each of which is of value in the appropriate context. However, of the 12 approaches only three groupings will be selected for their relevance to IVHS evaluation. These are

1. *Field test evaluation* which may be conducted for improving system design, assessing system effectiveness, and comparing alternative sytem approaches,

2. *Comparative evaluation* which may be conducted in the laboratory, field, or using models which may forecast into the future, and

3. *Prospective evaluation* which may be conducted for improving system design, assessing

Figure 1. Domain of IVHS Evaluation Alternatives



system effectiveness, and comparing alternative approaches over time.

By taking a systematic approach, formal evaluation increases the impact and efficiency of the learning process. Planned and controlled comparisons assure answers to critical research questions. Generally accepted procedures for formal evaluation enhance the practical significance of the evaluation findings by standardizing the logic of observation and by providing a common framework for analysis and communication of results.

If measures are carefully selected to reflect interests of various stakeholders, and tests are carefully controlled, then stakeholders can make informed assessments of whether their interests will be met by adopting one system over another. For example, auto manufacturers may need to decide whether it is in their strategic interest to invest in the production of optional equipment packages. Communications companies may need to decide whether it is in their interest to install a communications infrastructure. The DOT may need to support decisions on allocating resources to support local and state governments for purchase of the infrastructure. Consumers may want to know if it is worth buying that optional equipment. For that matter, the federal government needs to know whether it is worth spending taxpayer money for the continued development of these systems.

These and other stakeholders are interested in whether they will benefit from these systems and whether it is worth the cost for any marginal improvement. Only through systematic evaluation can the necessary information for making these critical assessments and decisions be gathered and assessed rationally. The sooner the winners and losers can be identified, the better off everyone is.

In general, evaluation should provide a set of systematic, uniform, and formal procedures for collecting, analyzing and interpreting data that will, in the final analysis, assist in communicating what happened in the operational field tests, why things happened the way they did, and what results mean for various interest groups. This information will assist in making decisions about directions to pursue.

Framework for Evaluation

A sound framework for evaluation is essential for planning and coordinating the early phases of IVHS testing and deployment. The framework detailed in this paper describes the basic structure of an effective evaluation and includes such things as the scope of the evaluation, component activities, and relationships between the components. This is not a detailed set of guidelines or a cookbook for IVHS evaluation; rather, it is a framework that distinguishes several distinct forms of evaluation and their relationships.

A good framework will meet the following criteria: (1) it will provide insight into appropriate means for implementing the Congressional mandate with regard to IVHS; (2) it will address the goals and concerns of the DOT-IVHS AMERICA strategic plan, (3) it will address system functionality, relative merit of alternative systems, and projected and aggregated net benefits; (4) it will be general enough to cover range of IVHS technologies being considered in the U.S., and (5) it will provide a useful structure for organizing results and synthesizing information for the purpose of making infomed decisions about IVHS. This section describes a framework that addresses the need for site-specific operational field testing, as well as the needs for comparing the relative merit of various deployments and projecting the aggregated net benefit of IVHS.

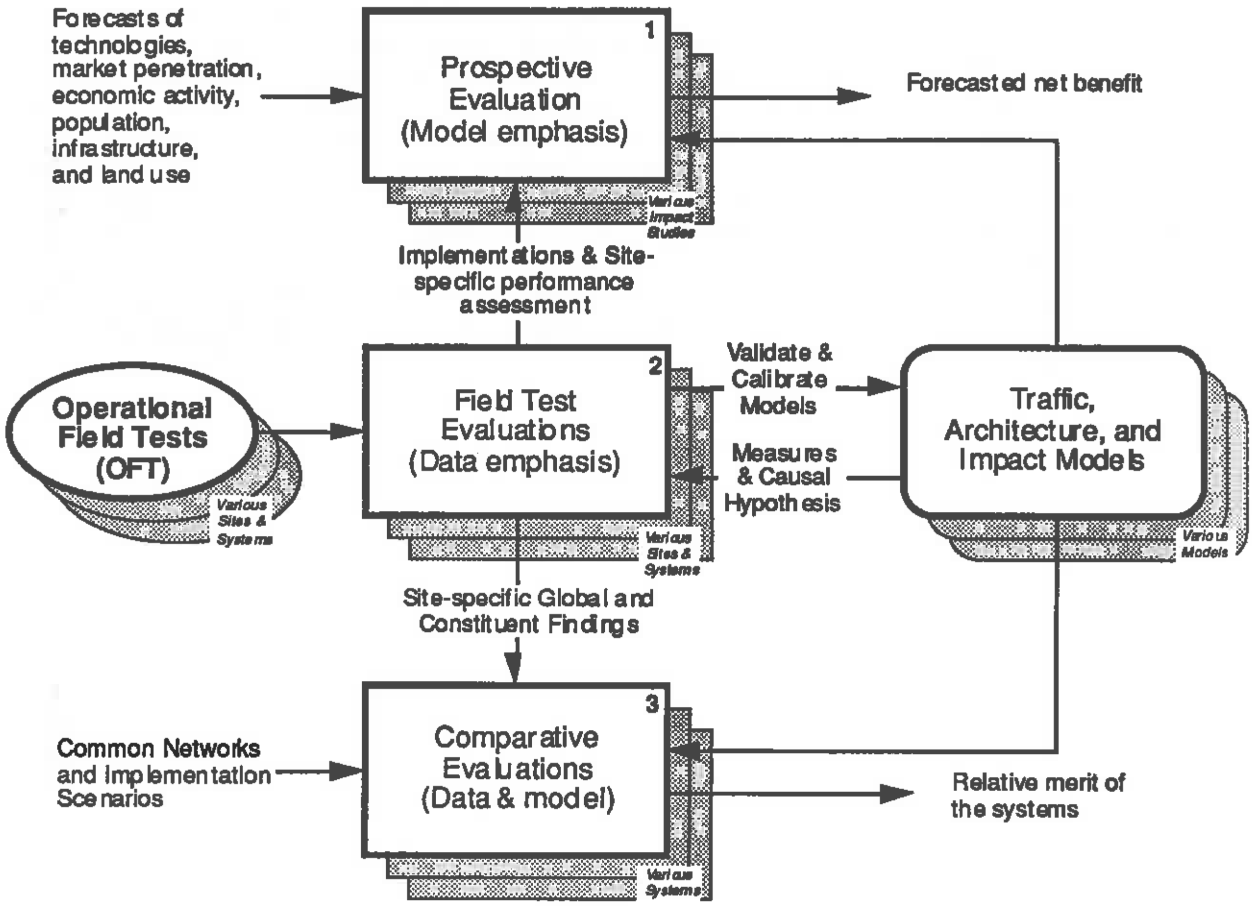
An overview is presented in figure 2 that highlights the major components of the framework and the interrelationships between these components. As the highest level of organization, the range of evaluation efforts can be divided into three categories by purpose: (1) prospective evaluation, (2) field test evaluation, and (3) comparative evaluation. Each category is unique in terms of its overall purpose, data needs, methods, and products. However, they may also be highly interconnected if we wish to take advantage of their complementarity.

Prospective evaluation involves forecasting demographic and market conditions, assessing impacts, and assigning value to the impacts. The emphasis is on determining future net benefits under assumed performance conditions. Field test evaluation involves the empirical assessment of sytem performance under controlled operational conditions. The emphasis is on optimizing performance and determining impacts under realistic field conditions. Comparative evaluation involves controlled deployment of multiple architectures to determine the most cost-effective alternative. In comparative evaluation, the emphasis is placed on controlled cross-system comparison and system cost.

Understanding relationships between the types of evaluation is important. Field test evaluations (box 1) are the centerpiece and can make important contributions to prospective evaluation (box 2) and comparative evaluation (box 3). For example, site-specific benefits established by field test evaluations may serve as anchors for forecasts of net benefits at larger levels of aggregation. Similarly, site-specific field comparisons from selected testbeds may serve as the critical input to comparative evaluation efforts. In some instances, selected systems and subsystems may be implemented and compared on the same stretch of road in the same time frame. Michigan's DIRECT project is an example of system comparisons in the field (Gilbert, et al., 1991). Perhaps the most important relationship among evaluation efforts is the sharing of traffic, architectural, and impact models for various applications in the evaluations.

It is important to recognize the opportunity provided by operational field tests to calibrate and validate selected models that may be used in all three types of evaluation. Traffic simulations are an excellent example of the types of models that may be shared. Once these models are calibrated properly they may then be used for assessing future impacts by changing traffic demand levels and market penetration assumptions. Traffic simulations may also be used in the comparative

Figure 2. Framework for Evaluation of IVHS



evaluations to control for network, traffic, weather, and other environmental conditions that may confound multisite empirical comparisons.

Field Test Evaluations

A second type of evaluation, the type receiving the most attention at this time, is the evaluation of operational field tests. Field test evaluations, represented in box 2 of figure 2, are designed to assess the operational functionality of specific system configurations in the field and to assess the impacts and benefits derived from implementations. The U.S. Department of Transportation is encouraging the development of private-public partnership to mount these ventures. Recently initiated projects, like Pathfinder, TRAVTEK, ADVANCE, FAST-TRAC, and DIRECT have been widely publicized, and the success or failure of these projects will influence to a great extent the public perception of the national IVHS effort. As a result, field test evaluations will play a major role in determining public acceptance of IVHS, as well as providing the means for assessing site-specific benefits of the selected systems.

Field test evaluations are designed to evaluate selected systems (e.g., a specific traffic management system, motorist information system, route guidance system, or vehicle control system) or possibly combinations of systems. TRAVTEK, PATHFINDER, and ADVANCE are examples of an operational field test that assesses a single route-guidance system, and FAST-TRAC is an example of a design that combines an adaptive signal control system and a beacon-based route guidance architecture. The emphasis in the field test evaluations is system performance under realistic field conditions.

System performance and consumer satisfaction are the concerns targeted by the field test evaluations. The structure for this type of inquiry is provided by the logic of experimental and quasi-experimental design. Figure 3 shows that the system under consideration is compared with a control group in a before-and-after evaluation of selected measures of effectiveness. This is the standard experimental design with subjects being randomly assigned to the control and alternative groups. The standard experimental design is appropriate and preferred where random assignment of subjects is feasible. This is the case where individual subjects are the unit of analysis like in studies of human factors and behavioral response to information systems. Quasi-experimentation may be more appropriate in addressing system level measures where random assignment is impossible or infeasible, as in the case of assessing the impact of IVHS on total annual accidents in a region (e.g., see Campbell & Stanley, 1963; Cook & Campbell, 1979). The issue of whether standard experimental designs or quasi-experimental designs are more appropriate is discussed further in the next section.

Methodological standards should facilitate intersystem comparisons and the accumulation of knowledge from successive field trials. It is especially important that research designs, measurement procedures, and instrumentation be standardized to establish some basis of comparison. However, it should be understood that even with the benefits of methodological standardization it may not be possible to make one-to-one comparisons between systems where network, traffic, weather, and other factors serve to confound the implementation environment. Rather, the chief merit of standardization is that it facilitates the accumulation of knowledge over a series of implementations at different sites.

According to the IVHS AMERICA strategic plan, "in order to maximize the cumulative benefits of many such tests, a set of national selection and evaluation criteria and standardized evaluation methodologies will need to be administered with federal leadership." To facilitate the accumulation of knowledge, the system-site characteristics for one operational test should complement those of other tests so as to minimize the duplication of efforts. The primary purpose of the operational field tests is to learn the best way to design and operate these systems in a variety of alternative environments, and duplication of tests would not make progress toward this end. One way to approach this would be to field test the alternative system architectures that emerge from the national architectural design efforts currently underway. Similarly, geographic and network diversity in the operational field test would enhance our understanding about the implementation of these systems in a range of operational environments.

Figure 3. Framework for Field Test Evaluation

	Pretest	Experimental treatment	Posttest
Control group	MOE 1 MOE 2 MOE 3 • • •	None	MOE 1 MOE 2 MOE 3 • • •
Alternative 1	MOE 1 MOE 2 MOE 3 • • •	System 1	MOE 1 MOE 2 MOE 3 • • •

In some cases, especially in those cases where several competing system alternatives offer similar functionality, we might learn more by holding the geographical setting constant to facilitate the comparison of similar and competing systems. This seeming conflict between site diversity for comprehensiveness and site uniformity for comparability between systems may be resolved by establishing national test beds that reflect the desired level of geographic diversity, but can also serve as sites for controlled trials of similar and competing system types.

Test beds where several different types of IVHS could be implemented at the same site would assist in comparing the impacts of different system types, or assessing combinations of system types. The Santa Monica corridor and Southeastern Michigan are examples of sites where different systems and combinations of systems are being implemented. Limited networks that comply with the provisions of the IVHS Corridors program may be good candidates for test bed status. Such test beds would also benefit the evaluation efforts by reducing the need to repeatedly instrument new areas of roadway for standardized measurement of efficiency.

Much of the foregoing discussion implies a summative orientation. That is, the evaluation is designed with the intent of reporting on the final results on the system successes and failures for the benefit of some external audience or decisionmaker. The information is generally intended to support decisions on the following options: continue the system at other sites, continue or increase support at initial site, make modifications to the system and reevaluate, or discontinue the system. So summative evaluation provides objective summative information to assist in these decisions. For reasons of credibility and to avoid conflicts of interest it is wise to involve external evaluators in the summative evaluations. External evaluators would have the knowledge and skills to perform the evaluation, but have little stake in the success of a specific field test or the publics' perception of IVHS. Given this criterion, university centers and national laboratories are good candidates for conducting summative evaluation.

Less has been said about the important role of formative evaluation in operational field tests. In contrast with summative evaluations, formative evaluations are conducted during the initial development of a system with the intent of troubleshooting difficulties and improving overall design. Robert Stakes (1982) summed up the distinction as follows: "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative." Internal personnel are best suited for formative evaluation because they have a stake in the outcome and would presumably prefer to see a successful implementation. As is obvious in the soup analogy, the formative evaluation should precede the summative evaluation. Formative evaluations take many forms, and do not have to correspond with the summative evaluation. In fact, formative evaluations are often less formal, relying on field monitoring and projective assessments, rather than analytical or experimental designs. A model of learning is perhaps a more appropriate framework for formative evaluation of IVHS.

A common pitfall of summative evaluations is to start collecting data too soon after system implementation (that is, prior to the shakedown of the system) thus, not allowing enough time to assure operation. IVHS implementations are complicated and will be prone to lapses and system failure in the initial stages of deployment of perhaps months in duration. By establishing a monitoring and formative testing stage prior to the summative evaluation stage, the evaluation team may avoid costly false starts in data collection. It is suggested that the system development contractor and the responsible public agency build formative evaluation into the implementation plan to assess system reliability and readiness for deployment. The formative evaluation could include many of the measures and instruments planned for the summative evaluation, which would test the summative experiments while monitoring system readiness.

Relative Merit of Field Test Methods

What types of research methodologies are preferable? This depends on many factors including what measures are being assessed, accessibility to subjects, and the ability of the evaluation team to control the treatments and operational environment. It is beyond the scope of this paper to address the relative merits of all methods potentially used in evaluation.[6] However, for the purpose of empirical evaluation one can assert that it is generally better to use randomized experiments over surveys and case studies when practical constraints do not suggest otherwise.[7] This is because experimental approaches provide the necessary information for establishing cause and effect.

Although the following discussion may seem basic to professional researchers, the concepts inherent in experimental and quasi-experimental paradigms are essential for clarifying distinctions between the demonstration and evaluation components of an operational field test, and therefore should be understood by all participants in evaluation design. Where a demonstration is designed to show that a system can be developed and made operational (an important aspect of any operational field test) an evaluation is designed to show that implementation of the system has led to some nonrandom and measurable changes in select target populations. The evaluation component is more ambitious than the demonstration component, in that it requires the evaluator to adhere to certain control principals whenever possible: random assignment of subjects to treatment groups to assure uniformity among the groups, measurement of pre-installation variables to establish before-and-after comparisons, and specification of control groups to demonstrate the results of not implementing the treatment.

Randomization procedures mark the dividing line between experimental designs on the one hand, and quasi-experimental and survey designs on the other, and are of great practical benefit to the evaluator if they can be employed in a pragmatic manner. The rationale for randomized experimentation is provided in the manipulability theory of causation, which states that causal inferences can be drawn from experiments where the value of the dependent variable is controlled through manipulation of an independent variable. This assures that expected pretest differences between the test groups is zero, which makes it easier to say treatment A caused an effect on measure B because it falsifies many competing interpretations that could have plausibly caused an observed relationship even if the treatment had not taken place. This validity of a causal inference is often called *internal validity* because it addressed the internal structure of the experiment. In practice, random assignment provides relief from the anxiety of considering and estimating the value of innumerable alternative causal explanations in order to assure internal validity. In other words, the evaluator is relieved from the burden of holding everything else constant in order to draw valid conclusions without ever being quite certain that complete control is being achieved. This is why random assignment is desirable, and in many situations quite practical.

In those situations where random assignment is not possible or highly impractical, quasi-experiments may be designed to improve internal validity and facilitate causal inferences. Without random assignment the groups being tested may differ in ways that mimic what the treatment might achieve. These are threats to internal validity of the experiments that may be managed through rather direct forms of control.[8] Two approaches are indispensable to the evaluator under these circumstances. First, pretest measures may be taken on the same scale as the posttest in order to establish change due to the manipulation. Second, similar control groups may be used as a no-treatment baseline, which can be compared with the experimental groups to establish change due to the treatment.

Numerous quasi-experimental designs may be devised using the these fundamental building blocks and may be applicable to operational field tests. These will be most appropriate when assessing system-level measures of effectiveness like average queue length at an incident or average monthly accident rates. Quasi-experimental designs of particular merit for these purposes include the interrupted time-series and regression-discontinuity designs (Cook & Campbell, 1979). However, it cannot be overstated that direct controls are not substitutes for, and are generally inferior to, random assignment procedures. In combination with random assignment, direct controls form the basis for the simple experiment.

Experiments and quasi-experiments should be the dominant research designs, but these designs should be used in conjunction with a priori knowledge and theory to build models of the technologies and their impacts on the vehicle-highway system. Such model-based evaluations should be more efficient and yield more information about how to achieve the desired effect of a fully developed intelligent vehicle-highway system. The strategy for integrating models into the evaluation is to construct and validate models of the deployment area, perhaps encompassing a larger geographical area than the operational field tests, and then make experimental runs of the simulation that reflect scenarios of increasing travel demand and levels of market penetration. Models developed for the operational field tests will provide an opportunity to calibrate and validate the simulations that may be used in the parallel extrapolative and comparative evaluation efforts. Data collected for calibration and validation will be invaluable in the effort to develop improved evaluative and real-time dynamic traffic forecasting models. Model development and calibration efforts should also provide insight that may lead to improved implementation of the operational field tests.

Experimental and quasi-experimental strategies have their limitations. In particular, they tend to neglect relevant qualitative contextual evidence and often depend too heavily on a few quantified abstractions, to the neglect of contradictory and supplementary qualitative evidence (Weiss & Rein, 1970). As a supplement to control-based research designs, narrative program histories provide a rich bounty of novel data and provide information on those aspects of the operational field test upon which the quantitative methods are generally mute: the implementation of the systems, possible system side effects, and possible causal processes that produce the observed outcomes (Campbell, 1970; 1979). Although histories and other qualitative approaches have greater flexibility and a dynamic quality, they are not considered very convincing when separated from the hard data produced through the experimental and quasi-experimental approaches.

It is worth noting that surveys have sometimes been overused in transportation evaluation efforts. As a main research tool, a survey can yield much descriptive data (Backstrom & Hursch-Cesar, 1981; Dillman, 1978; O'Muircheartaigh & Payne, 1977). Surveys are especially useful in situations where the objective is to learn about people's attitudes, values, beliefs, experiences, and intentions. For example, we will want to know what the subjects think about the systems being tested, which they prefer, which they would purchase, and at what price. A survey can also be a helpful tool for identifying relationships among variables, which may lead to hypotheses about relationships and more controlled studies. However, because a survey is a correlational method, it cannot be used to establish the causal relationship between a treatment and an outcome that is essential in evaluation work. For example, if a survey reveals that drivers that use motorist information systems enjoy driving more than those who do not, this is no basis for claiming that motorist information caused driving enjoyment. Since correlational methods do not establish temporal precedence, one cannot rule out the alternative possibility that people who enjoy driving tend to use motorist information systems. Furthermore, since correlational methods do not establish *ceteris paribus* ("all else being equal"), one cannot eliminate the possibility that some other factor (e.g., smoking while driving) causes less use of motorist information and less enjoyment with driving.

It is useful to work survey design into the implementation of experimental of quasi-experimental approaches whenever possible. Questionnaires can be used as dependent variables in experiments or quasi-experiments. This may allow the use of random assignment, pretesting, and control groups to establish causality, in which case one may find that motorist information does indeed cause driving pleasure.

Integrating the results of the many individual studies into some projective assessment will also be a challenge. Careful planning at the front-end, perhaps using a matrix technique discussed above, should help when it comes time to draw summary conclusions with persuasive arguments. However, it is important to understand that individual field tests that implement only a single system for evaluation will not produce conclusive evidence of comparative value. In other words, one may find that the evidence supports that the users value the system when compared with not having a system at all; however, one cannot conclude on the basis of this evidence that this system is any better or worse than other systems that were not tested. Furthermore, there will be a question of whether the results may be generalized to other networks or other regions of the country. Field test evaluations that show that the system works and is of value to the travelers and general public must be supplemented with other evaluative efforts designed to compare competing system designs under identified site conditions.

Planning Field Test Evaluations

Planning of the field test evaluation is difficult because of the technical complexity of the systems being deployed, the immense breadth of coverage and impacts experienced by system users, and the number of objectives, methods, subject groups, treatments, and instruments that must be coordinated in a single evaluation plan. The management of the evaluation effort is also likely to be difficult for the very same reasons. This section reviews some planning methods that are likely to be appropriate for many of the operational field tests and how they might be coordinated in a comprehensive evaluation plan.

One difficulty that will be faced in every field test evaluation is coordination among the many constituent studies that contribute to the findings. Operational field tests are large and complex undertakings and the evaluation plans will reflect this complexity. Figure 4 provides some indication of the range of relevant methodologies that may be used to address the many objectives in an operational field test. Along the top of the table are categories of issues or objectives that may be addressed in an evaluation of an operational field test. The first column identifies several types of methods that are more or less appropriate for use in addressing the objectives listed along the top of the table. The dots in the cells of the matrix indicate the relative appropriateness of a method for addressing a category of objectives. Although this table conveys only our personal assessment of the methods, the approach demonstrates that a variety of methods may be usefully employed in IVHS operational field tests. Furthermore, it suggests that one or more approaches may be appropriate for any given set of objectives, and that most any methodology may be adapted to address several sets of objectives. Project planning will involve assessing the appropriateness of the methodology as well as assessing the prospects for consolidating many studies into fewer studies.

Consolidation is the key for improving the efficiency of multimethod studies. The trick is to consolidate constituent designs into broader studies that use similar subjects and more general instrumentation, and still produce the same breadth of results. For example, the subjects and design of an experiment on drivers' behavior may be combined with an assessment of the drivers' preferences for various implementations of the technology by simply tacking on a questionnaire on drivers' preferences to the end of the driver behavior experiment. Similarly, standard questionnaire may be designed to assess subjects' reports of behavior as well as their preferences toward the system. The point is that coordination and consolidation among constituent designs may

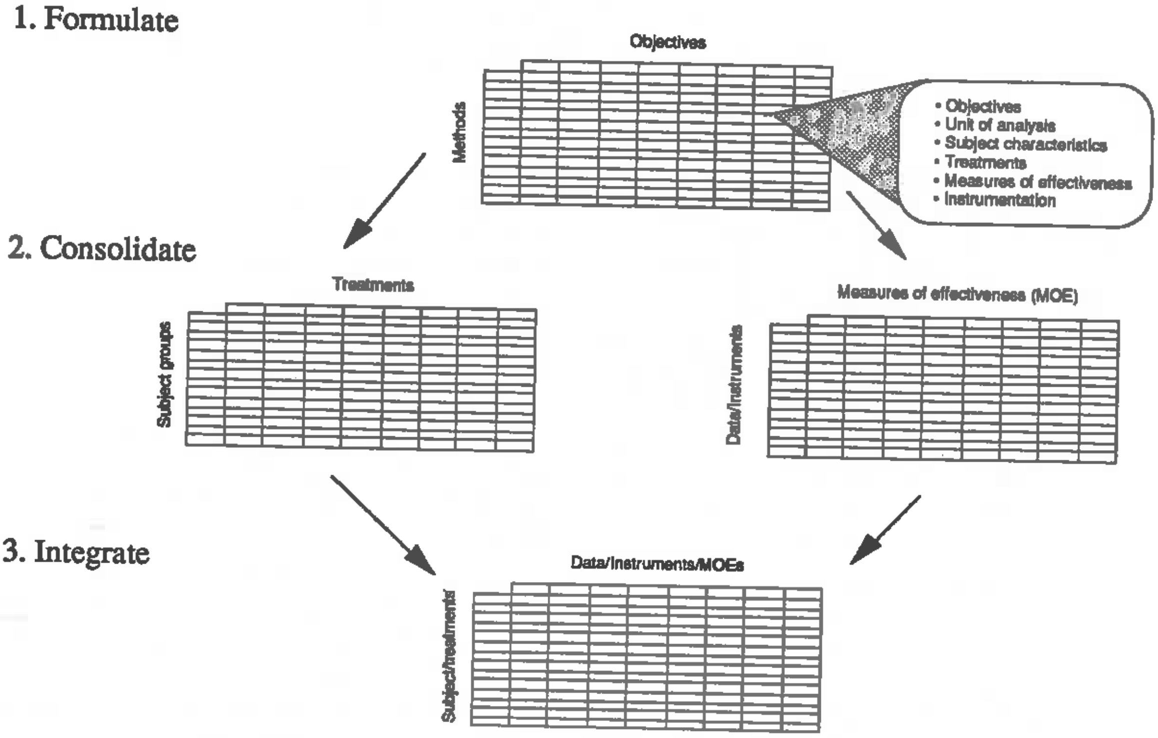
Figure 4. Appropriateness of Methodologies in Evaluating an Operational Field Test

	A System Technical Performance	B Safe and Efficient User Interface	C Indiv. Driver Behavior & Impact	D User Attitudes and Preferences	E Traffic Behavior and Impacts	F Environment and Energy Impacts	G Institutional & Socio-eco-nomic
1 Field experiment	Most	Most	Often	Often	Occasionally	Occasionally	Least
2 Field quasi-experiemt	Often	Often	Most	Often	Most	Most	Least
3 Surveys	Least	Occasionally	Occasionally	Most	Least	Least	Often
4 Subject de-briefing	Occasionally	Occasionally	Occasionally	Most	Least	Least	Occasionally
5 Cost analysis	Most	Least	Often	Least	Most	Most	Least
6 Extrap. impact assess.	Least	Least	Most	Least	Most	Most	Least
7 Narrative case studies	Often	Occasionally	Occasionally	Occasionally	Occasionally	Occasionally	Most

Most appropriate (solid circle) — Often appropriate (top half filled) — Occasionally appropriate (bottom half filled) — Least appropriate (empty circle)

increase the overall efficiency of the effort, and opportunities of consolidation may not be readily apparent without reviewing the larger picture of how the constituents may relate to each other in terms of objectives, methods, subjects, treatments, instruments, and measures of effectiveness. Figure 5 indicates several types of consolidation that may be recognized through a simple matrix approach to evaluation planning.

Figure 5. Research Planning Matrices



Matrix planning is accomplished in three steps over a series of meetings among the evaluation team and the partners of the operational field test. First, the evaluation team considers all of the objectives that must be addressed in the evaluation and conducts a series of brainstorming sessions to generate methods that would address these objectives. This assumes that the objectives have already been generated by the partners, perhaps through facilitation by members of the evaluation team. Each method would be described in terms of objectives, units of analysis, subject characteristics, treatments, measures of effectiveness, and model features. This continues in a series of meetings until the methods have been exhausted. Second, each of the methods is listed in a table that cross-references subject characteristics and treatments. The objective here is to identify like subject-treatment groups that can be consolidated into a single approach. The same is done for measures of effectiveness and instruments in order to reduce the total number of instruments required in the evaluation. Finally, the methods are reconstituted in a more efficient form through the use of a subjects-treatments-instruments-MOE table. This may sound complex and cumbersome, but is really quite systematic and intuitive. The result is a more efficient plan, which consolidates many approaches into fewer approaches.

Field test evaluations should be designed so that evaluators can collect any data they might need in forecasting impacts. This might include special demographic, population growth, urban development, economic development, geographic, market, and impact data that would not otherwise be collected for the immediate system being tested.

Following the generation and consolidation of evaluation approaches, it is helpful to list the steps required for accomplishing each approach, precedent relationships among the steps, resource requirements of the steps, and then generate a temporal plan that optimizes the given resources. Most project managers are familiar with PERT (Program Evaluation and Review Technique) and CPM (Critical Path Method) which are essential in accomplishing this step given the number of activities and resources that must be managed.[9]

Comparative Evaluation

The third form of evaluation in the framework is what we will call comparative evaluation. In comparative evaluation, the various systems and architectures under consideration are compared in terms of their benefits, drawbacks, and costs under controlled implementation conditions. More than one system needs to be implemented under highly controlled conditions to make such comparisons. Edwards et al. (1975) note in their introduction to evaluation that it is "a common administrative fiction, especially in Washigton" that national programs are comparable from place to place and from time to time; "we have frequently encountered the idea that a program is a fixed, unchanging object, observable at various times and places." However, this is not the case with IVHS or any other type of transportation improvement that may be implemented along our roadways. Each test site and time is unique and should be treated that way.

The structure for comparative evaluation is both structural and procedural. The structure for comparitive evaluation is that of controlled experimental comparison as shown in figure 6. The idea is to assess measures of effectiveness for various alternatives under identical conditions. The procedure for comparative evaluation is as follows: (1) select comparison criteria, (2) select appropriate testbed, (3) design the experiments or comparisons, (4) collect the data, and (5) assign values to the outcomes or measures of effectiveness. Cost-effectiveness is the most common comparison criterion. Testbeds can be devised in the laboratory, the field, or through simulation.

The methodological problem faced in comparative evaluation, then, is one of experimental control; how does one control all of the implementation factors to assure identical implementation conditions and internally valid comparisons? For example, how does one control the implementation of competing route guidance architectures in order to compare the time savings attributable to each? If we want to assure a completely valid comparison we would want to use a purely experimental design with random assignment of subjects to the architectures where environmental conditions like network geometry, traffic on the network, and weather would be held constant. In other words, the randomly assigned drivers would be linked or "yoked," and "race" each other from the same starting point to the same ending point at the same time to determine which experiences the least travel time.

In the route guidance example, the yoked design requires that all of the competing systems be implemented at the same site. Although a yoked comparison will assure a measure of internal validity, one can argue that the single-site requirement is quite restrictive. In practice, systems will be implemented at different sites or under different conditions and still compared. Furthermore, many of the evaluation measures, including most of the technical and operational variables, are not as likely to be confounded by site-specific variation. So, when making these system-to-system comparisons, why not hold all of the conditions as identical as possible, acknowledging that there will be some differences, and make as good a comparison as possible? There is some appeal to this approach if one can control for network, traffic, and weather conditions, possibly through statistical manipulation. However, we will not be able to rule out alternative explanations in most system performance comparisons if the subjects are not yoked, and system successes and failures can be just as easily attributed to differences in the implementations or sites as they are to differences in the systems themselves. Ambiguous empirically

Figure 6. Framework for Comparative Evaluation

	Pretest	Experimental treatment	Posttest
Control group	MOE 1 MOE 2 MOE 3 • • •	None	MOE 1 MOE 2 MOE 3 • • •
Alternative 1	MOE 1 MOE 2 MOE 3 • • •	System 1	MOE 1 MOE 2 MOE 3 • • •
Alternative 2	MOE 1 MOE 2 MOE 3 • • •	System 2	MOE 1 MOE 2 MOE 3 • • •
Alternative 3 • • •	MOE 1 MOE 2 MOE 3 • • •	System 3	MOE 1 MOE 2 MOE 3 • • •
Alternative n	MOE 1 MOE 2 MOE 3 • • •	System n	MOE 1 MOE 2 MOE 3 • • •

based comparisons are likely to occur, and may even be quite persuasive among the unwitting consumers, especially if more valid alternative evaluation designs are not implemented.

This further supports the testbed idea, in this case for the purpose of making yoked comparisons between competing architectures that provide similar functionality. This does not imply that the multiple-site, operational, field-test evaluation of IVHS should be abandoned or replaced by a single-site evaluation scheme; implementations of operational field tests at multiple sites provides important and cumulative information about system performance under differing circumstances and conditions. Rather, in those cases where a direct comparison of architectures is essential, those parts of the evaluation that are conducted primarily for making the intersystem comparison should follow a yoked design on the same test site. In most cases this is a small portion of the operational field-test-evaluation effort and the benefits derived from controlling for selected comparisons should far outweigh the incremental cost. So, for example, if it is deemed important to compare the map format and the arrow format route guidance systems for driver time savings and other performance measures, then this would call for a yoked design to be implemented at a limited site that supports both systems.

Another way to compare architectures is to model their operation and impacts in a simulated traffic environment that provides complete control over the environmental variables. Traffic simulations now exist that represent individual vehicles and the logic that guides these vehicles as they travel through fairly realistic networks (e.g., Van Aerde & Yagar, 1988; McGurrin & Wang, 1991). The vehicles then move simultaneously through the network, from their origins to their destinations, according to specified departure rates and routing logic. These simulations can model intelligent vehicles that follow route guidance instructions, intelligent vehicles that respond to motorist information, or background vehicles that rely on conventional guidance approaches such as maps or experience. Moreover, the simulated vehicles can actually obtain guidance or control instructions through direct linkage with the system architecture. The new traffic simulators could be used as testbeds where the operator has complete control over network geometry, road and intersection capacity, signal location and timing, demand between selected origins and destinations, background traffic, incidents, and the percentage of vehicles with various forms of route guidance. This provides the means of testing the architectures in a variety of controlled traffic environments, including the possibility of direct face-offs between competing architectures. Other simulations have the capacity to model intervehicle relationships and vehicle control systems; these could be used in the same manner.

However, simulated comparisons are not without their limitations. The chief criticism is the simulation's lack of realism; by using simulation to improve the internal validity one also compromises external validity. As a result, even though one may demonstrate that the simulation with motorist information consistently produces better results than the simulation without motorist information, it may become difficult to persuade audiences that this demonstration has any relevance to the "real world."

Valuation methods are again relevant in comparing architectures. Cost-effectiveness analysis holds sytem performance measures constant across the alternatives and compares the alternatives on the basis of cost. This is the most common approach to comparing systems. However, an alternative approach would be to have the R&D team design to a set cost level and then evaluate the alternatives in terms of performance. In this later case the performance measures would be quantified and synthesized in accordance with accepted multicriteria value-assessment procedures.

Prospective evaluation

Most models of IVHS implementation require the participation, support, and cooperation of both private and public sector organizations. The need for cooperative deployment produces what has been widely described as a "chicken-and-egg" problem: the public sector has little incentive to provide infrastructure if vehicles are not equipped, and private organizations have little incentive to provide vehicles if there is no infrastructure support. In order to attain cooperative deployment there must be some understanding or agreement about the partners' respective roles, parallel responsibilities and actions, and the implicit mutual desirability of the goal. An initial step toward the attainment of such an understanding is to determine the net value of system deployment and the allocation of value to the parties. An initial assessment of future benefits will tell the parties whether it is worth pursuing this at all.

Prospective evaluation (represented by box 1 in figure 1) is the process whereby one assesses the overall net benefits that may result from future system deployment, and the allocation of benefits and costs to various stakeholders over time. The output provides some sense of future benefits, drawbacks,and costs over time. Prospective evaluation combines demographic forecasting, technological forecasting, social impact assessment, and value assessment in determining future net benefits under assumed or projected IVHS performance conditions. The general approach and range of methods are common to those used in technology assessment and impact analysis.[10]

Figure 7. Prospective evaluation

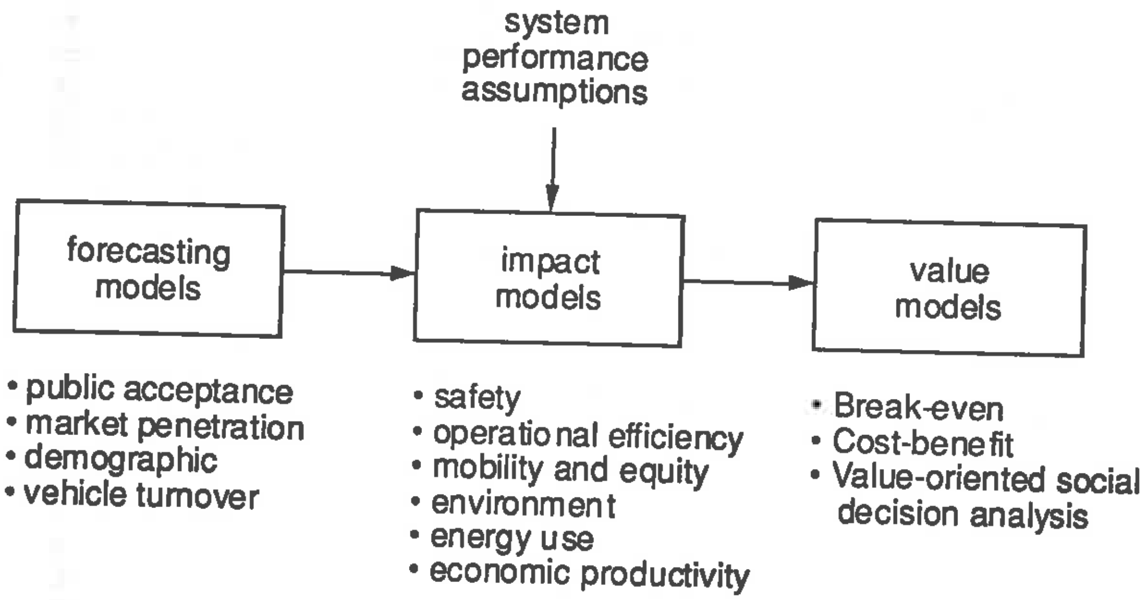


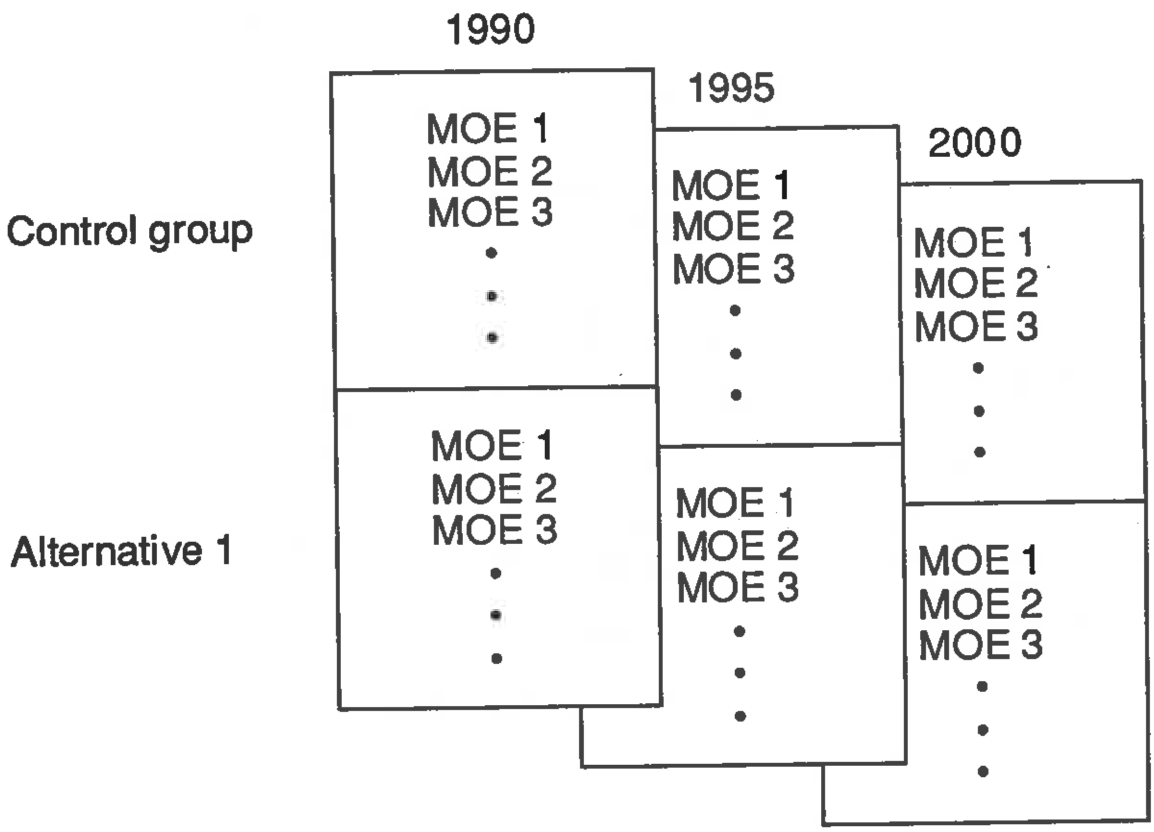
Figure 7 shows the logical sequence of activities in prospective evaluation. The first step is to forecast the societal context for the delivery of selected IVHS. The primary objective is to determine the market penetration for the systems. This may require assessing demographic shifts for selected market segments, the acceptance and installation of public infrastructure, and vehicle ownership, use, and turnover for the time period of the assessment.[11] If the societal context is not forecasted, then these key determining variables must take on assumed value in the impact assessment stage. A range of forecasting methods may be used depending on the existence of models, data availability, and a host of practical constraints. [12]

The structure for comparative evaluation draws from the logic of a control-group time-series design. For the purposes of prospective evaluation, alternative predictive models of the system would be compared at various times in the future.

The second step is to assess the impacts that result from implementation of the systems at the forecasted levels of market penetration, assuming a utilization rate and a level of system performance. Alternative scenarios may be assessed to determine the sensitivity of impacts to projections or assumptions regarding market penetration, utilization, and system performance. Impacts may include measures pertaining to safety, operational efficiency, mobility and equity, the environment, energy use, and economic productivity. Impacts may be aggregated by geographic or jursidictional boundaries.

The third step is to assess the value of the impacts. This last step is optional and involves aggregating the impacts and synthesizing the impact measures using some analytical formula for

Figure 8. Framework for Prospective Evaluation



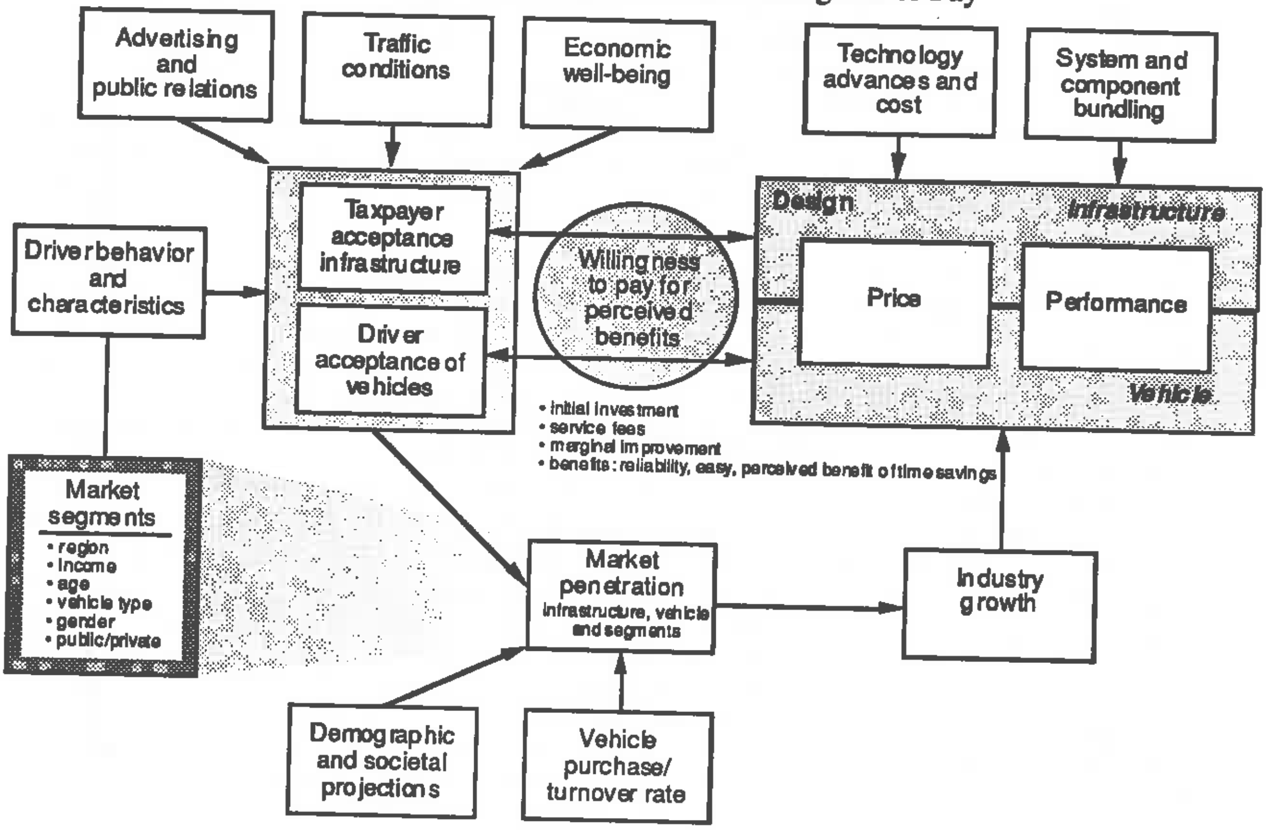
stakeholder value. A common method for assigning value is to determine monetary values based on willingness to pay. This is the approach of financial analysis and cost-benefit analysis. An alternative approach is to develop utility-of-value functions for society as a whole or for selected stakeholders.

Prospective evaluation relies on IVHS deployment and market scenarios, which describe uncertain future states. Focus groups and expert panels may help in producing realistic estimates of cost, design, performance, price, market, and public deployment assumptions. By providing a range of scenarios one can indicate the sensitivity of the assessment to these assumed values.

Estimates of market penetration may be obtained from holistic expert judgment techniques (e.g., Underwood, et al., 1991) or through more complex analytical approaches as depicted in figure 9.

A consumer model of willingness to pay is depicted in figure 9. This shows the critical relationships among product design, willingness to pay, and market penetration. The figure shows how a product (which, in the case of IVHS may combine vehicle-based systems with infrastructure-based systems) is influenced by technology advances as well as system and component bundling. Each product has two attributes: design and price. Products will be evaluated by the public who may be (1) taxpayers who would foot the bill for the infrastructure, or (2) vehicle owners/users who would very likely pay for the vehicle-based components, although infrastructure-based services may be covered by the users too.

Figure 9. Consumer Choice Model of Willingness to Pay



Based on their perception of likely benefits, the public would decide what they would be willing to pay for products and services. This is influenced by their economic status, driving patterns, traffic frequently encountered, and advertising and public relations. Willingness to pay could be segmented by region, income, age, vehicle type, gender, etc. Estimates of market penetration could then be estimated through demographic projections of these market segments and factoring in frequency of purchase and vehicle turnover rates. It would be useful to establish estimates for these variables and their relationships prior to projecting net benefits. However, in the absence of empirically derived estimates, meaningful forecasts may still be generated through the development of alternative implementation and market penetration scenarios. The existing Delphi projections (Underwood, et al., 1991) could assist in generating meaningful scenarios.

Willingness to pay is also a centerpiece of the economists' approach to evaluation. In this case, a formal framework for prospective evaluation is provided in the normative economic theory of consumer surplus, consumer demand, and equilibrium pricing. The foundations go back to Dupuit (1844) who in his classic paper on *consumers surplus* can be said to have laid the groundwork that, in modified forms, still informs the contemporary work on economic evaluation. If one postulates that social value is measured by willingness to pay, then in a perfectly competitive economy, market prices will reflect this willingness to pay and social cost as well. However, in cases like road transportation where the economic incentives deviate from the perfectly competitive ideal, then the theory of shadow pricing can help the evaluator adjust for the deviations. The theory is well documented in the literature on normative economics (Mishan, 1981) and provides the justification for the application of methods of economic valuation embodied in cost-benefit analysis (Mishan, 1976; Sassone & Schaffer, 1978).[13]

Cost-benefit analysis involves accounting for all the impacts, both positive and negtive, and assigning a monetary value to them. Despite its theoretical merits, cost-benefit analysis may not be appropriate for prospective evaluation of IVHS. Although cost-benefit analysis provides a calculus for synthesizing pieces into a projective measure (i.e., net present value [NPV]), it also suffers from economic myopia and undue rigidity. The criterion of positive NPV provides little insight into the nature of the benefits and costs and their allocation across stakeholders. It also assumes that meaningful answers can be formulated by valuing the good and bad effects in monetary terms. Furthermore, cost-benefit analysis assumes that market, or market-like values are meaningful quantities.

Two additional issues in comparative evaluation are (1) what criteria do you use to compare the systems, and (2) whose values do you base the comparison on? A criterion is an explicit attribute or characteristic used for the purpose of comparative evaluation. In a projective approach to evaluation one would use a single criterion of value, say the net present value of a project. A more analytical approach would use multiple criteria, perhaps later synthesized into a single criteria. Which attributes are used in an analytical approach would depend on the objectives, the system, and the constraints of the evaluation design. They might include such measures as person hours of travel, volume/capacity ratio, operating cost per passenger trip, average cost of in-vehicle unit, total infrastructure cost, tons of emissions, and accident rate.[14] Multicriteria approaches provide the advantage of multiple attributes and a formal logic for mathematical synthesis.[15] Evaluation problems are decomposed into attributes, utilities, utility weights, and probabilities. Each of these components is evaluated for an individual or group and then synthesized into a single value for the outcome of an alternative. Usually a single decision maker is assumed for the purpose of analysis. However, alternative stakeholder groups may have different values and expectations, which may be reflected in multiple evaluations. For example, automotive manufacturers may value an architecture differently than public sector representatives due to differences in vehicle and infrastructure costs. Value Oriented Social Decision Analysis (VOSDA) is an approach that seeks to jointly optimize the value of multiple parties (Chen, et al., 1979).

In general, cost-benefit analysis is good for evaluating actions where the impacts are shallow and broad. That is, there are few second or later-order impacts and the impacts are relatively unsubstantial at the individual level. By contrast, analytical impact assessment techniques are better suited than cost-benefit analysis for addressing "deeper" impacts like safety improvements with a smaller incidence.

What about negative impacts? Drawbacks are just as important as benefits in a balanced assessment. For example, it may turn out that the design of certain control systems intended for improving safety may actually reduce some forms of safety because drivers become distracted by in-vehicle displays. In another example, it may turn out that in order to avoid traffic, drivers will take longer, circuitous routes that may reduce travel time but add to vehicle emissions. Because of the longer distance traveled, such examples highlight the importance of assessing drawbacks in a complete accounting of net benefits. Many IVHS critics suspect that by using technologies to improve conditions for an individual driver, the secondary effect will be to increase demand over time until a new equilibrium is reached with more traffic; the tertiary effect will be being increased emissions and throughput. These types of second and third order effects must also be traced.

Several steps will facilitate uniform and consistent synthesis of forecasts and impacts. First, forecasting methods and databases should be standardized in order to promote internal consistency in the market estimates. In some cases existing forecasts, (e.g., existing population forecasts for Standard Metropolitan Statistical Areas [SMSA]), may be adopted with minimal modification. However, in other cases (e.g., forecasts of market penetration) specific forecasting efforts will have to be undertaken. Projections should use identical base years and time increments, and they should be checked for consistency over common time horizons. Ideally, the forecasts will address the smallest geographical units first (e.g., specific urban or regional areas), and then be aggregated over regional levels, and then at the national level.

Second, standardized geographical, jurisdictional, and market segmentation procedures should be followed; these segments should reflect the heterogeneity of the populations. Regional and local deployment scenarios will be critical for generating forecasts that are relevant to local conditions. For example, it is unlikely that IVHS deployment scenarios in Montana will match deployment scenarios for Southern California in either form or scale in the near future, and these regional differences must be taken into account in projecting market penetration and impacts. Consumer descriptors like gender, age, and income may also be considered as part of an overall market assessment.

Potential mismatches between likely investment sources and actual beneficiaries can be identified at this stage and addressed as potential barriers to deployment. It will be useful to subdivide potential users into specific market segments that will either use or be impacted by the system for more detailed tracking and analysis. It will also be useful to identify key stakeholders and the expected impacts on specific groups such as disabled, elderly, poor urban, and rural Americans.

Traffic simulation is probably the best way of determining the impact of market penetration on time savings, vehicle miles traveled, and derived measures like safety and emissions. Estimates can be made for a range of network types under different scenarios of travel demand and market penetration. These can then be aggregated to assess the net impact and benefit at selected points in the future.

It is important that the models used in the prospective evaluation be based on data generated from empirical measurements reflecting current conditions as accurately as possible. For example, any simulation of the impact of motorist information or vehicle control systems should be calibrated and validated using data on existing traffic networks and traffic conditions. The evaluation of the operational field-test deployments provides a significant opportunity to obtain this type of data. Steps should be taken in the operational field tests to collect data that will be needed in the prospective evaluations.

Although it is likely that forecasts will rely extensively on conventional forecasting techniques, new methods for the development of scenarios may be useful in looking at the role of institutions, political decision making, and other less predictable influences on the adoption of net technologies and systems. New approaches are also needed to synthesize the disparate forecasts pertaining to technologies, system design, system performance, market penetration, public acceptance, etc. One solution is to hold a series of *policy exercises* which would enable experts, stakeholders, and policy makers to participate in a process to construct well-integrated and contextually rich future histories on the deployment and adoption of IVHS technologies (Underwood, 1988).

Such scenario techniques may incorporate backcasting, and can be designed to include consideration of discontinuities and institutional processes. Policy exercises have already been used to address long-run and large-scale environmental issues, such as the impact of climate change on the Great Lakes Basin, and have been found to enhance communication among interested parties and to help integrate forecasts of the many factors that can impact the environment, including policy shifts.

Issues in IVHS Evaluation

It is hoped that this framework will provide a new and broader perspective to the many seemingly isolated evaluation studies that will take place as part of the overall national IVHS effort. Furthermore, it is hoped that the framework will lead to further discussion on the effective planning and control of the numerous evaluation efforts that are to be coordinated nationwide. In addition to the framework we would like to touch on a couple of issues related to planning and implementing evaluations of IVHS. We are highlighting these issues because they are likely to influence the planning and conduct of most evaluation activities.

First, the planning and implementation of nearly all of the evaluation efforts, and especially the evaluation of the operational field tests, will involve diverse groups of people representing a wide range of values, motives, expectations, knowledge, and skills. Participants in these efforts will carry with them their own personal perspectives on the situation, as well as the perspectives of their respective organizations. One prevalent distinction may be made on the basis of public-sector and private-sector organizational cultures; but other factors may influence a person's perspective, including disciplinary training, and personal and organizational stakes in the project. Another important distinction is between technology-oriented participants, who view field tests as primarily efforts to measure the performance of equipment against selected technical measures, and people-oriented participants who view the field tests as primarily efforts to assess the impacts of the systems on people. Although both views are important, there are likely to be disputes regarding their relative emphasis.

These differences in viewpoint are likely to result in some conflicts, due in part to miscommunications and in part to real differences of opinion among the participants. Therefore, in each step of an evaluation effort, ranging from specification of project objectives to the final presentation of results, it will be important to address these differences directly through procedures designed to facilitate open discussion, mutual education, joint problem solving, and the integration of perspectives. Open discussion of issues and mutual education go a long way toward dispelling communication-based conflicts. Real conflicts may be diffused by logrolling or trading on the issues, combining complementary capabilities, working for a common goals, and exploiting economies of scale. What is most important is for the project team to make efforts to ensure that partners (1) listen to critics as well as supporters, (2) explicitly identify the interests of the various parties and stakeholders that are involved in the project, and (3) think creatively about how to meet the interests of as many parties as possible.

Second, although guidelines will be recommended for the design of field-test evaluations, each field test will be unique, and the overall evaluation effort will benefit from tailored evaluation plans for each of the tests. The guidelines required by ISTEA will go a long way toward standardizing aspects of the operational field tests, as well as assuring uniformity and accumulation of knowledge. The guidelines may include both core elements and options. The core will include those aspects that are mandatory for all qualifying operational field tests. The options will suggest additional tests that may be relevant. As long as the core guidelines are adhered to, the evaluation team should have the freedom to adapt the plan, primarily through adding evaluation activities, so that it meets the interests of partners.

One way to tailor the guidelines to the needs of the individual operational field tests is to employ a sophisticated contingency of decision tree approach. The contingency logic may be computer coded in the form of an *artificial-intelligence* application package for operational field tests design. Artifical intelligence packages for research design and analysis are becoming commonplace in the behavioral research community and could be of significant benefit in this context (McAuliffe & Hamel, 1991; Nachtsheim, 1987).[16]

Finally, the management of expectations will be critical to the perceived success of all of the evaluation efforts described above. Whenever there are parallel efforts to aggregate net benefits and to test field implementations there is the potential for discord among the results of the approaches. While specific field-test evaluations are likely to generate overly conservative estimates of net benefits, the aggregated prospective evaluations are likely to generate seemingly optimistic assessments of net benefits. This is due in part to inherent biases in the methods. Quasi-experimental designs of the field-test evaluations may ignore many of the impacts that result from the implementation. Projective methods are notoriously optimistic about market penetration of new technologies. There is a distinct possibility in the case of IVHS that near-term and optimistic prospective evaluations will set the public up for disappointment when inherently conservative field-test results roll in. This can be remedied in part by linkage between the results of the operational field tests and the prospective evaluations, where the field-test results will feed into the projective studies.

It is also common for the public and for clients to be unaware of potential limitations of evaluation plans and designs, and, consequently, to have unrealistically high expectations about what can result. Again, this gets back to the issue of expectations and the need to manage them effectively for appropriate response to the evaluation results.

Summary and Prescriptions for Evaluating IVHS

The following section is a summary of the previous sections with a focus on prescriptions for effective IVHS evaluation. This may be used for quick reference and as an analytical table of contents to the document. As with the main body of the text this section discusses the general requirements for an evaluation framework and then discusses each of the types of evaluation in turn: prospective evaluation, field test evaluation, and comparative evaluation.

Overall Framework

- An effective overall evaluation framework will
  - provide insight into appropriate means for implementing the Congressional mandate with regard to IVHS
  - address the goals and concerns of the DOT-IVHS AMERICA strategic plan
  - address system functionality, relative merit of alternative systems, and projected and aggregated net benefits
  - be general enough to cover range of IVHS technologies being considered in the U.S.
  - provide a useful mechanism for organizing results and synthesizing information relevant to the evaluation efforts.
- An evaluation framework should contain elements examining
  - prospective evaluation
  - field test evaluation
  - comparative evaluation
- Components of evaluations must be considered as complementary; that is, elements from each type of evaluation generate information enhancing the value of other evaluation types.
- Evaluation results must be used to improve the calibration and validation of traffic, architectural, and impact models.

Field Test Evaluations

- Field test evaluations are designed to assess the operational functionality of specific system configurations in the field and to assess the impacts and benefits derived from implementations.
- It is especially important that field-test evaluation research designs, measurement procedures, and instrumentation be standardized to establish a uniform basis of comparison between field tests.
- Even with highly standardized methods, one-to-one comparisons of different systems operating at different sites may not be meaningful. Thus, the chief merit of standardization is the cumulation of knowledge over a series of system-site implementations.
- To further facilitate cumulation of knowledge, system-site characteristics for one operational test should complement those of other tests to minimize duplication of effort.
- There is value in developing a select set of test-bed sites that reflect the desired level of site diversity for comprehensiveness and site uniformity for comparability between systems.
- Summative field-test evaluations are designed with the intent of reporting results for the benefit of an external audience or decisionmaker.
- To increase the credibility of results and to avoid conflicts of interest, it is best to employ evaluators external to the field-test implementation for summative evaluations.
- Formative evaluations are conducted during the initial development of a system with the intent of troubleshooting difficulties and improving overall design.
- Internal field test personnel are best suited for formative evaluation because they have a stake in the outcome.
- Formative evaluations should precede summative evaluations.
- The key to effective and efficient field-test evaluations is to consolidate evaluation components into broader studies that use similar subjects and more general instrumentation.
- Apply matrix planning procedures to maximize efficiency of evaluation consolidation.
- To establish cause and effect relationships it is best to use randomized experimental designs over survey and case-study methods.

Comparative Evaluations

- In comparative evaluation, the various systems and architectures under consideration are compared in terms of their benefits, drawbacks, and costs under controlled conditions.
- More than one system needs to be implemented under highly controlled conditions to make meaningful comparisons.
- When examining intersystem effects, yoked designs at a single test site are most appropriate.
- Sufficient experimental control can be achieved for examining intersystem differences using simulated traffic environments.
- There is often a potential tradeoff between internal and external validity in comparative evaluations. While we must strive to maximize external validity, internal validity is essential to a meaningful evaluation.

Prospective Evaluation

- Prospective evaluation is the process whereby overall net benefits, and the allocation of benefits and costs to IVHS stakeholders are assessed.

- Prospective evaluation involves forecasting, quantifying, and aggregating various forms of societal, environmental, and economic impacts derived from IVHS, including mechanisms for determining the incidence of IVHS benefits, drawbacks, and costs.
- Analytic approaches to prospective evaluation include queuing models, network simulations, and emissions models that break impacts into components that are measured and aggregated into a synthesized measure.
- Analytic approaches are concerned with
  - forecasting factors that may influence future IVHS deployment, such as public acceptance, market penetration, technology development, economic growth, lifestyle trends, infrastructure deployment, changing land use patterns
  - assessing the range of impacts that result from projected levels of IVHS deployment, such as traffic safety, operational efficiency, mobility and societal equity, traveler comfort and convenience, the environment, energy use, economic productivity, and the IVHS industry.
- Projective approaches include trend analysis and other more qualitative methods that examine impacts as a whole that responds to environmental changes.
- Forecasting methods and databases should be standardized to promote internal consistency in market estimates.
- Standardized geographical and jurisdictional segmentation procedures should be used to reflect population heterogeneity.

Notes

[1] See Deborah Gordon's *Steering a New Course: Transportation, Energy, and the Environment* for a current and comprehensive account of transportation ills and possible remedies ranging from incentives to double mass-transit ridership to intelligent vehicle highway systems.

[2] There is by now a large number of introductory reviews of IVHS in the U.S. (e.g., Automotive Engineering, 1991; Jurgen, 1991; Plous, 1992). More technical surveys are also available (e.g., Walker, 1990; Saxton, 1991)

[3] Wherever methodological procedures and issues may be of further inerest, the text provides extensive references to guide the reader. Standard evaluation methods are well documented in a number of professional and scientific texts (e.g., Rossi & Freeman, 1982)

[4] The scope of ISTEA's support of IVHS can be summarized with the following arithmetic. The IVHS component of ISTEA was established and funded at roughly $660 million over six years. $501 million is provided for an IVHS corridors program ($71 million for FY`92, and $430 million for FY`93 to FY`97). $158 million is provided for other IVHS research and development ($23 million for FY`92, and $135 million for FY`93 to FY`97). To this one must add appropriations of $119.8 million (FY`92) for Congested Corridors and $20 million for IVHS research (FY`92) to get a complete picture of IVHS funding for FY-92. The bottom line is that a total of $233.8 million has been appropriated for FY`92, $143.0 million has been appropriated for FY`93. Of this, nearly $110 million of the FY`92 appropriation has been earmarked by Congress.

[5] IVHS AMERICA is an incorporated, for-profit, educational and scientific association that has been chartered as a formal advisory committee to the U.S. DOT. As a chartered advisory committee IVHS AMERICA will play a central role in the design of a national program of IVHS research, development, and deployment. The IVHS AMERICA strategic plan is important at this juncture because it is a critical input to, and the foundation for, the congressionally mandated DOT strategic plan, which is due at the end of the year.

According to current drafts of the IVHS AMERICA strategic plan, the mission of the IVHS community in the U.S. is "first, to improve surface transportation by deploying IVHS technology broadly throughout the nation and in cooperation with Mexico and Canada, throughout North America, and second, to develop a U.S.-based IVHS industry to provide technology in the U.S. and abroad." Toward this end the plan calls for near-term and appropriate deployment of existing technology, investment in infrastructure and research and development, and formation of new public/private/academic partnerships.

The specific goals of the IVHS community, as enumerated in the IVHS AMERICA Strategic Plan, are (1) to improve safety , (2) to increase capacity and operational efficiency, (3) to enhance mobility, convenience and comfort, (4) to reduce negative environmental and energy impacts, (5) to improve the productivity of individuals, organizations, and the economy as a whole, (6) to develop a viable and profitable U.S.-based IVHS industry, (7) to redirect and educate transportation professionals, and (8) to develop a new institutional structure for technology development and deployment. Specific objectives, many of which provide quantified targets, are delineated in the strategic plan for each of these goals. These goals and objectives should provide guidance for the measures that are relevant for any operational field test.

[6] Several books outline appropriate experimental design strategies and the most effective design of evaluation tools such as surveys and interviews (e.g., Charles River Associates, 1972; Billheimer & Trexler, 1980).

[7] Numerous textbooks and handbooks are available to explain the experimental design process in considerable detail (e.g., Keppel, 1982; Montgomery, 1984; Hicks, 1982; Box, Hunter & Hunter, 1978; Brown & Melamed, 1990).

[8] Common threats to internal validity are listed and analyzed by Cook and Campbell (1979). The classic list includes such considerations as history, maturation, testing, instrumentation, statistical regression, selection, mortality, interactions with selection, ambiguity about the direction of causal inference, diffusion or imitation of treatments, compensatory equalization of treatments, compensatory rivalry by responsendt receiving less desirable treatments, and resentful demoralization of respondents receiving less desirable treatments.

[9] See Moder, et al. *Project Management with CPM, PERT, and Precedence Diagramming* for details on these approaches.

[10] For a review of methods used in technology assessment and impact analysis see Porter, et al. (1980).

[11] See Button, et al. *Car Ownership Modeling and Forecasting* and de Jong *Some Joint Models of Car Ownership and Use* for details on forecasting car ownership.

[12] See Millet and Honton (1991) for a review of forecasting methods.

[13] The application of cost-benefit approaches to transportation projects is described in several useful handbooks (e.g., Lee, 1987; Campbell & Humphrey, 1988).

[14] For a review of attributes commonly used in road transportation evaluation see *Measures of Effectiveness for Multimodal Urban Traffic Management* by Abrams and Di Renzo.

[15]Procedures for application of multicriteria evaluation methods in the context of evaluation research are decribed in Pitz and McKillip (1984) and Edwards & Newman (1982). Reviews of mathematical principles are available in Keeney & Raiffa (1976), Zeleny (1982), and von Winterfeldt & Edwards (1986).

[16] Examples of rule-based research-design software packages are Designer Research by Idea-Works, Experimental Design by Statistical Programs, and Statistical Consultant by Software Labs. All three of these programs are available for IBM PCs and compatibles. Statistical Consultant is based on *A Guide for Selecting Statistical Techniques for Analyzing Social Science Data* ; which was written for the Institute for Social Research (ISR) at the University of Michigan (Andrews, Klem, et al, 1981).

References

Abrams, C., and Di Renzo, J. (1979). *Measures of effectiveness for multimodal urban traffic management*. Department of Transportation Report FHWA-RD-79-113, Washington, D.C., December.

Andrews, F.M., Klem, L., Davidson, T.N., O'Malley, P.M., and Rodgers, W,L. (1981). *A guide for selecting statistical techniques for analyzing social science data, 2nd ed.* Institute for Social Research, University of Michigan, Ann Arbor, MI.

Automotive Engineering. (1991). IVHS. Automtove Engineering, 99(3), 13-20.

Backstrom, C.H., and Hursch-Cesar, G. (1981). *Survey research*. New York: Macmillan.

Billheimer, J. & Trexler, R. (1975). *Evaluation hamdbook for transportation impact assessment*. U.S. Department of Transportation Report UMTA-IT-06-0203-81-1, Washington, D.C., December.

Box, G.E.P., Hunter, W.G., & Hunter, J.S. (1978). *Statistics for experimenters*. New York: Wiley.

Brown, S.R., & Melamed, L. (1990). *Experimental design and analysis*. Newbury Park, CA: Sage.

Button, K.J., Pearman, A.D., & Fowkes, A.S. (1982). *Car ownership modelling and forecasting*. Aldershot, Gower.

Campbell, B., Humphrey, T.F. (1988). *Methods of cost-effectiveness analysis for highway projects*. Washington,D.C.: National Cooperative Highway Research Program (NCHRP), Synthesis No. 142, December.

Campbell, D.T. (1970). Considering the case against experimental evaluations of social innovations. Administrative Science Quarterly, 15, 110-113.

Campbell, D.T. (1979). Assessing the impact of planned social change. Evaluation and Program Planning, 2, 67-90.

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Charles River Associates. (1972). *Measurement of the effects of transportation changes*. Report CRA-166-2, Cambridge, MA.

Chen, K., Mathes, J.C., et al. (1979). *Value oriented social decision analysis: enhancing mutual undersatnding to resolve public policy issues*. IEEE Transactions on Systems, Man, and Cybernetics, SMC -9, 576-580.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin Co.

De Jong, G.C. (1989). *Some joint models of car ownership and use*. Ph.D. Thesis, Faculteit der Economishe, Universiteit van Amsterdam, Netherlands.

Dillman, D. (1978). *Mail and telephone surveys*. New York: Wiley.

Dupuit, J. (1844). *On the measurement of the utility of public works*. Transactions in International Economic Papers, 2, London: Macmillan (1952).

Edwards, W., & Newman, J.R. (1982). *Multiattribute evaluation*. Beverly Hills: Sage.

Edwards, W., Guttentag, M., & Snapper, K. (1975). A decision theoretic approach to evaluation research. In E.L. Struening, and M. Guttentag (Eds.), handbook of evaluation research, vol. 1 Beverly Hills, CA: Sage.

General Accounting Office. (1991). *Smart highways: An assessment of their potential to improve travel.* REMDP-91-18, Washington, D.C.: United States General Accounting Office.

Gilbert, K., Underwood, S.E., & DeFrain, L. (1991). *DIRECT: Comparison of alternative driver information systems.* (pp. 397-406). SAE No. 912783, Vehicle Navigation and Information Systems Conference Proceedings, Society of Automotive Engineers, P-253, Dearborn, MI, October 20-23.

Gordon, D. (1991). *Steering a new course: Transportation, energy, and the environment.* Washington D.C.: Island Press

Hicks, C.R. (1982). *Fundamental concepts in the design of experiments.* New York: Holt, Rinehart and Winston.

Keeney, R., & Raiffa, H. (1976). *Decisions with multiple objectives: preferences and value tradeoffs.* New York: Wiley.

Keppel, G. (1982). *Design and analysis: a researchers handbook.* Englewood Cliffs: Prentice-Hall.

Jurgen, R.K. (1991). *Smart cars and smart highways go global.* IEEE Spectrum, May, pp. 26-36.

Lee, D.B. (1987). *Benefit-cost evaluation - staff study.* U.S. DOT, Transportation Systems Center.

McAuliffe, D., & Hamel, C. (1991). *Experimental design.* Computer, March, 86-92.

McGurrin, M., & Wang, P.T.R. (1991). *An object oriented traffic simulation with IVHS applications.* (pp. 551-562). Society of Automove Engineers, Vehicle Navigation and Information Systems Conference Proceedings, P-.253, Dearborn, MI, October 20-23.

Miles, M.B., & Huberman, A.M. (1984). *Qualitative data analysis: A sourcebook of new methods.* Beverly Hills, CA: Sage.

Millett, S.M., & Honton, E.J. (1991). *A managers guide to technology forecasting and strategy analysis methods.* Columbus: Battelle Press.

Mishan, E.J. (1981). *Introduction to normative economics.* New York: Oxford University Press.

Mishan, E.J. (1976). *Cost-benefit analysis.* New York: Praeger.

Mobility 2000 Working Group. (1990). *Proceedings of a national workshop on IVHS.* Dallas, TX: Texas Transportation Institute.

Mobility 2000 Working Group. (1990). *Intelligent vehicle-highway systems: operational benefits.* Dallas, TX: Texas Transportation Institute.

Moder, J.J., Phillips, C.R., & Davis, E.W. (1983). *Project management with CPM, PERT, and precedence diagramming, 3rd Ed.* New York: Van Nostrand Reinhold.

Montgomery, D.C. (1984). *Design and analysis of experiments.* New York: John Wiley & Sons.

Nachtsheim, C. (1987). Tools for computer-aided design of experiments. Journal of Quality Technology, 19(3), 132-160.

Nunnally, J.C. *Psychometric theory.* New York: McGraw-Hill.

O'Muircheartaigh, C.A., & Payne, C. (1977). *The analysis of survey data. Vols. 1 and 2.* New York: Wiley.

Patton, M.Q. (1980). *Qualitative evaluation methods.* Beverly Hills, CA: Sage.

Pitz, G.F., & McKillip, J. (1984). *Decision analysis for program evaluators*. Beverly Hills: Sage.

Plous, F.K. (1992). Smart roads, smart cars. Planning, 58(1), 12-15.

Porter, A.L., Rossini, F.A., Carpenter, S.R., and Roper, A.T. (1980). *A guidebook for technology assessment and impact analysis*. New York: North-Holland.

Rossi, P. (1971). Evaluating educational programs. In F.G. Caro (Ed.), Readings in Evaluation Research (pp. 97-99). New York: Russell Sage Foundation.

Sage, A.P. (1977). *Methodology for large scale systems*. New York: McGraw-Hill.

Sassone, P.G., & Schaffer, W.A. (1978). *Cost-benefit analysis: a handbook*. New York: Academic Press.

Saxton, L. (Ed.). (1991). IEEE Transactions on Vehicular Technology, Special Issue on Intelligent Vehicle-Highway Systems, 40(1).

Stakes, R.E. (1982). How sharp should the evaluator's teeth be? Evaluation News, 3, 89-80.

Underwood, S.E., Chen, K., & Ervin, R. (1991). *The future of intelligent vehicle-highway systems*. Transportation Research Record, 1305.

Underwood, S.E. (1988). The policy exercise: cooperative learning for long-run policy assessment, In Crookall, et al. (Eds.), Simulation-gaming in education and training. Oxford: Pergamon Press.

Van Aerde, M., & Yagar, S. (1988). Dynamic integrated freeway/traffic signal networks: A routing-based modeling approach. Transportation Research-A, 22A(6), 445-453.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research. Cambridge*: Cambridge University Press.

Weiss, C.H., & Rein, M. (1970). The evaluation of broad-aim programs: Experimental design, its difficulties, and an alternative. Administrative Science Quarterly, 15, 97-109.

Zeleny, M. (1982). *Multiple criterion decision making*. Englewood Cliffs, NJ: Prentice-Hall.