REPORT OF THE AD HOC SYSTEMS STUDY GROUP

SETH BONDER
Assistant Professor
Industrial Engineering

DONALD E. CLEVELAND
Associate Professor
Civil Engineering

DEAN WILSON
Associate Professor
Industrial Engineering

THE UNIVERSITY OF MICHIGAN

Contributing Authors

JOHN M. ARMSTRONG
DARYL BEM
ALLEN COOK
JOHN DUNCAN
ROBERT FARRELL
BRUCE GILLOGLY
KERRY KILPATRICK
WAYNE VAN WAGONER

1967

Published by
HIGHWAY SAFETY RESEARCH INSTITUTE
Institute of Science and Technology
The University of Michigan

FOREWORD

This report describes the work of an ad hoc systems study group sponsored by the Highway Safety Research Institute of The University of Michigan. The objective of this three-month study was to develop a highway systems model to be used as a management guide for the allocation of research resources.

The contents of this report, including the conclusions and recommendations, represent views of the systems study group and should not be considered as having HSRI approval, either expressed or implied, until reviewed, evaluated, and subsequently endorsed by the Institute.

We would like to acknowledge the contributions of Mrs. Toni Kennedy and Mr. Robert Farrell, who edited the text of the report, and Mrs. Elaine Buckle who patiently typed and proofread the drafts.

CONTENTS

1 DEVELOPMENT OF THE HIGHWAY SYSTEMS MODEL

SETH BONDER

INTRODUCTION

BACKGROUND

In recent months the subject of automobile safety has figured prominently in articles, features, and editorials in the popular press, and has produced at least one best-selling book. It has been the topic of extensive government hearings which resulted in the passage of new legislation aimed at improving safety in our vehicles and on our highways. There is little doubt that this concern on the part of the public, industry, and government is well-founded. During 1964 automobile accidents took 47,000 lives--an increase of one third over the annual traffic death figure a decade earlier (Johnson, 1965). During 1965 the traffic death figure rose to approximately 49,000, and it is estimated that 50,000 will die on our roads and streets in 1966. These increasing numbers cannot be rationalized by the fact that we drive more each year, for the annual fatality rate--the number of deaths per one hundred million miles of travel--is also rising. In 1961 the fatality rate was 5.2; by 1964 it had risen 10% to 5.7. Campbell (1965) notes that "...traffic accidents are the leading cause of death among all persons from age 5 to 31."

The number of traffic fatalities alone is reason enough for public concern; however, the economic cost of accidents and injuries is also increasingly a matter for national concern. The National Safety Council has estimated that the total cost of motor vehicle accidents was 8.3 billion dollars in 1964. Other estimates (Bureau of Public Roads, 1965) placed the cost at 10-15 billion dollars annually. Although we are a nation of large resources they are not unlimited, and their waste must somehow be controlled.

This concern for highway safety is not new. Over 40 years ago President Coolidge said, "With deplorable and continuing increase in highway mortality and injury, the time is highly appropriate for a comprehensive study of the causes, that we may have proper understanding of conditions and so may intelligently provide remedies" (Campbell, 1965). However, a resulting research study conducted in

1926, and other safety research conducted during that decade, provided few solutions. Although this failure can probably be traced to many factors, it was perhaps mainly a result of the deficiencies of the existing mechanistic point of view. Most accidents were attributed to the careless, irresponsible, unqualified, and inept driver (Michaels, 1965), an attitude that led, quite naturally, to the belief that a program of social controls (traffic laws, driver education, licensing, etc.) was the necessary and sufficient solution. Recent knowledge (Michaels, 1965), could readily have predicted the inadequacy of this approach.

The issue of highway safety lay relatively dormant until 1949, when an advisory group to President Truman's Committee for Traffic Safety outlined a five-year program of specific safety research projects. This five-year plan was never implemented as planned, and, in fact, several still relevant projects remain undone. In 1958 the President's Committee for Traffic Safety sponsored a meeting of outstanding scientists and professional traffic safety experts at Williamsburg, Virginia. The purpose of this conference was to develop fresh approaches to the problem of safety research. Although many of the excellent research recommendations generated at that time have yet to be pursued, the Williamsburg Conference provided much of the initial stimulus for recent safety research efforts.

Since 1960, commitment of scientific skills and funds to highway safety research has been increasing at an accelerating rate. The majority of these efforts (some of which are described in the succeeding chapters of this report) have been component studies devoted to determining specific causes of accidents and to improving the vehicle, the road, traffic controls, driver education and selection, and legal and enforcement activities. However, this extensive allocation of research resources has not yet provided any meaningful solutions to the highway safety problem (Arthur D. Little, 1966).

This continued failure to achieve applicable results is surprising, if not perplexing. In contrast to the comparatively unsophisticated research efforts of the middle twenties, many of the recent individual studies have taken advantage of advances in scientific principles and methods such as statistical inference, time series analysis, and computer simulation.[1] One might conclude from this that highway transportation phenomena are characterized by a degree of variability that precludes scientific description and useful prediction. However, recent literature would suggest rather that the restricted view of uncoordinated, individual studies is at fault (Arthur D. Little, 1966; Michaels, 1965; Bureau of Public Roads, 1965; Campbell, 1965; Whitton, 1965). Research predicated upon identifying, isolating, and then modifying or removing some one significant cause of accidents has proven ineffective. The recent Arthur D. Little report (1966) notes that "...cause is not a meaningful concept in most traffic accidents since they usually result from the combination of a number of human, vehicle, and environment factors." The Bureau of Public Roads emphasizes this point in a report to a Senate subcommittee:

> The best analyses show that most traffic accidents happen to normal people driving in a normal way on normal roads in all kinds and makes of cars. There isn't a special type of person or road or car that accounts for most accidents. It isn't the drunk or the young; it isn't the convertible or the ill-maintained car; and it isn't the road with curves or grades that can account for significant numbers of accidents (Bureau of Public Roads, 1965).

Although they have provided some measure of basic knowledge, recent researches directed at component design improvement have also been ineffective in reducing accidents. Whitton, in a letter to U. S. Senator Abraham Ribicoff (1965), explains:

> Historically, we have sought the solution to safety through improvement in the vehicle or the highway or the driver separately. The significant increase in accident rates that has occurred over the past four years demonstrates that such classical safety techniques, although necessary, are not sufficient to reverse the rising tide of accidents. Sufficient evidence has now accumulated to convince us that attempts to improve each part in isolation cannot alone succeed in reducing accidents. Thus, it seems inescapable to me that we are coming to an end of an era in safety and, in order to make significant progress, a new approach is required.

---

[1] A discussion of scientific methodology is given in a later section of this chapter.

The failure of scientific research to provide proven solutions to the safety problem has resulted in continued reliance upon unsupported judgment as a basis for countermeasures. Historically, the "common sense" approach has led highway safety one step forward and two giant steps backward. For example, Garrett(1961) has pointed out that

> ...Until the last decade it was widely accepted as a "fact" that when an automobile accident occurred, occupants who were thrown from the car were safer than those who remained inside. Indeed, the connotation of the phrase "thrown clear" was that the ejectee who survived this experience would otherwise have been killed. Although such accidents have in fact been documented, a 1954 automotive crash injury research (A.C.I.R.) report showed that they were the exception rather than the rule...that door opening was both a frequent and a hazardous event: In injury-producing automobile accidents, about 34% of the cars had one or more front doors opened, and, contrary to general opinion, occupants who were hurled through these doors were often "thrown clear" to eternity--not to safety....

A recent Bureau of Public Roads report (1965) notes that

> ...The arbitrary installation of traffic signals on rural primary highways, or of median barriers on multi-lane highways...installed on the basis of judgment, frequently have increased, rather than decreased, accidents....

Stonex (1963) has shown that the standard guardrail end, still being built into some new highways, has proved unnecessarily dangerous.

This reliance on "trial and error" in developing and evaluating safety programs contrasts sharply with procedures employed by the military services and public health areas. The Department of Defense has recognized the futility and cost (in time and resources) of trial and error development of weapon systems, and has developed planning procedures to assess the system prior to its operational use (Bonder, 1966). Similar procedures were employed to evaluate the Salk and Sabin vaccines for their efficacy, cost, and safety prior to their utilization.

There is growing recognition among professional safety researchers and government officials of the need for parallel safety evaluation procedures. Whitton (1965) states this requirement clearly in his letter to U. S. Senator Abraham Ribicoff: "Objective and systematic methods must be employed for knowing what will produce safety and for determining how to design and build safety into highway transport." That these methods are currently nonexistent* is evident from the recent literature search by Arthur D. Little (1966).

In December 1965, with growing public demand for improved automotive safety and professional recognition of the need for a more effective approach to highway safety research, the automotive industry gave The University of Michigan a 10-million-dollar gift to establish a Highway Safety Research Institute (HSRI). An ad hoc advisory committee from the faculty of the University defined the purpose of the new Institute as follows: "The mission of the Highway Safety Research Institute shall be to create, collect, collate, analyze, and disseminate knowledge of the various factors influencing safety on the highways" (Highway Safety Research Institute , 1966). It was appropriately suggested, in view of the failure of component efforts to achieve measurable success to date, that the Institute should pursue a systems approach to the study of highway safety.

---

*Concurrent with the conduct of this study, Booz-Allen Research Associates has been engaged in developing a transportation systems model study under a contract from the U. S. Bureau of Public Roads.

THE HSRI SYSTEMS PROJECT

While the term "systems approach" is very much in vogue, it is quite appropriate to ask what constitutes a systems approach to highway safety, and how one should be organized and conducted. To answer these questions, and to provide HSRI management with some guidelines for allocating research resources, a systems project comprised of faculty and graduate students at The University of Michigan was funded by the Highway Safety Research Institute. The purpose of this report is to describe the early work of this group and to set forth some tentative recommendations.

In order to establish a reasonable objective for its own efforts, the systems group had to assume the functions or role of HSRI in the safety research community. To understand the specific role presumed as a basis for the work reported herein, and to provide the lay reader with sufficient background to understand succeeding chapters of this report, a brief sideroad into scientific methodology and related topics is presented. The reader familiar with modeling concepts, operations research, systems analysis, etc., may bypass this discussion.

Scientific Method: The Use of Models. A detailed discussion of scientific principles is obviously beyond the scope of one section, or indeed, one volume.[2] A brief introduction is perhaps best accomplished by focusing on the general activity of modeling.

A model is a representation of reality which is used for prediction. Its purpose is to make it possible to determine how changes in one or more aspects of the modeled entity affect other aspects, or the whole. Use of the model permits the investigator to make such predictions by manipulating the model rather than the modeled entity, thus eliminating costly (in time and dollars) experimentation with the real system.

A schematic of the modeling procedure is given in Fig. 1-1. This figure illustrates two methods of moving from past and present systems to future systems. One is the path of "expert judgment,"

---

[2] The interested reader is referred to the classic text, An Introduction to Logic and Scientific Method, by M. R. Cohen and E. Nagle (Harcourt, Brace, and Company, 1934).

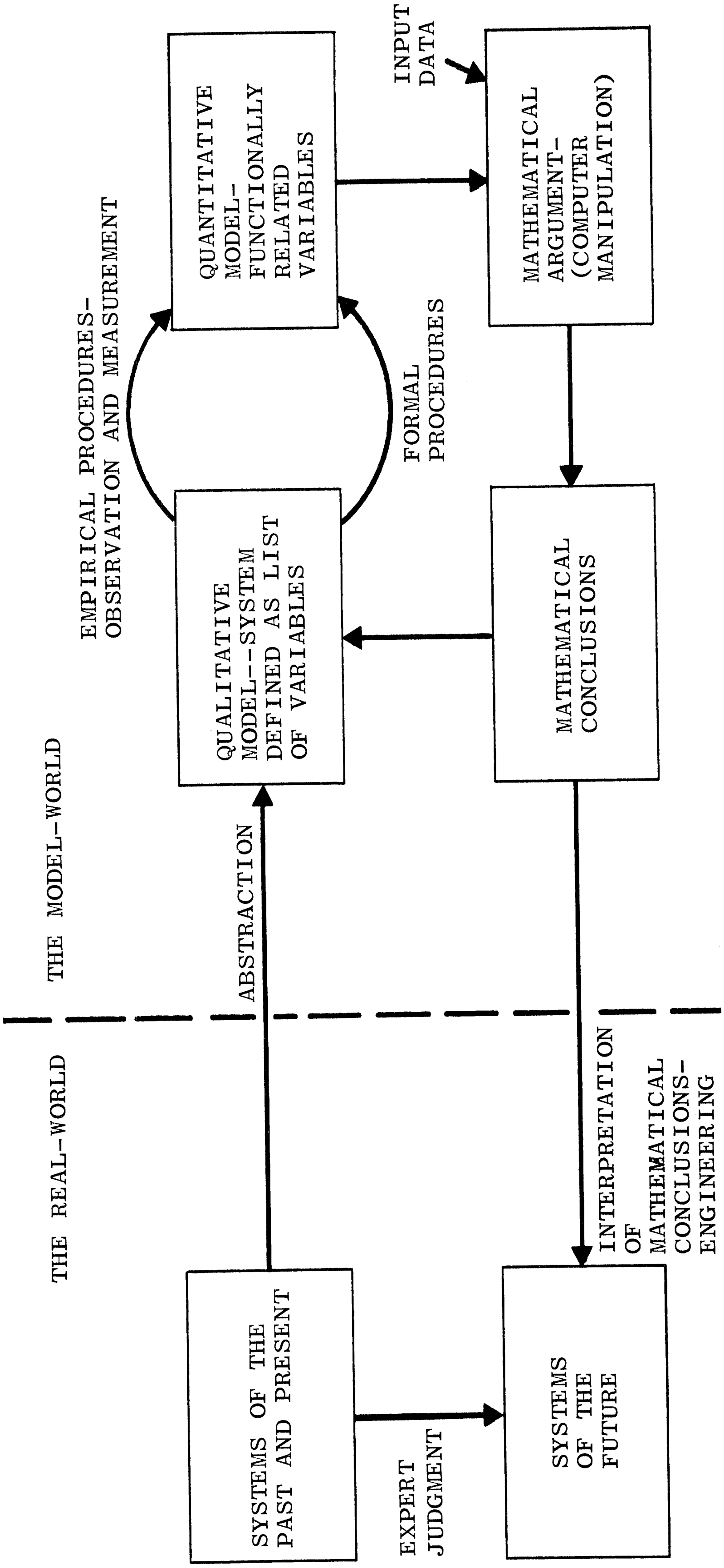


FIGURE 1-1. THE MODEL CONCEPT (Adapted from Thrall, Coombs, and Davis; 1954)

which at times proves adequate when dealing with relatively simple phenomena and in times of stable technology. However, this approach has all the drawbacks exemplified by the failure of "common sense" countermeasures described in the previous section.

The other is the modeling approach, which provides a way to deal with complex systems and rapidly developing technology when the extensive technical and operational experience upon which expert judgment is based may be lacking.

The modeling process begins by defining the system, usually by selecting those variables that reflect relevant aspects of the system (Ashby, 1958). The development of relationships among system variables is usually accomplished by both empirical (real-world) procedures and the procedures of formal logic. As a coherent body, the latter is usually referred to as rationalism, and is a method by which explicit propositions about the phenomenon under study are formulated and alternative hypotheses are developed for eventual comparison with empirical observations of the phenomenon. Its procedure is straightforward. Premises are stated, a logical structure developed, and conclusions deduced from the structure. If the conclusions (hypotheses) have been obtained by consistently applying formal rules of logic to the assumptions or premises underlying the model, the model is said to be valid or internally consistent. Nothing is implied regarding the empirical truth of the conclusions of a valid model. The conclusions are a direct result of the premises and no more. If the premises refer to the future, then the conclusions or hypotheses may be taken as predictions. Figure 1-1 illustrates this procedure as well as indicating the requirement to provide input data to the model and the importance of the computer for manipulative purposes.

Hypotheses generated by rational methods must be formulated so that they may be subjected to experimental confirmation. This is accomplished via the feedback loop from the mathematical conclusions to the beginning of the modeling process (see Fig.1-1). Before any confidence can be placed in conclusions derived from the model, they must be tested against experimental data, a process called model verification. It should be noted that the function of experiment is eliminative; that is, experimental data are used to eliminate some or all of the alternative hypotheses. Experiment does not demonstrate the truth of a hypothesis but merely identifies hypotheses that should be rejected, thereby narrowing the field so that true ones may be found. If the data agree with the model conclusions, this lends confidence to the use of the model, and to its conclusions as predictions of future events. Those hypotheses not rejected are submitted to further study by using data from the verification procedure to enrich the logical structure. Herein, then, lies the relationship between formal and experimental (i.e., empirical) methods. The process is a regenerative one leading to identification of fruitful hypotheses (predictions) which are continually tested for compatibility with future observations.

Since the modeling process is a regenerative one requiring constant revision and enrichment, where does one begin? This is a matter of choice based upon the experience of the researcher. If his knowledge of the phenomenon is adequate to formulate reasonable premises, then he will begin by formalizing, for this will yield the insight necessary for constructing reasonable experiments. If adequate knowledge does not exist, then the investigator must observe the phenomenon to detect meaningful regularities in the data.

Models have been characterized by many different schemes (Churchman, Ackoff, Arnoff, 1959); however, for our brief introduction to the subject, we need only distinguish between the deterministic and the probabilistic (stochastic) models, and then classify those models that have been used in analyzing highway safety problems.

(1) The deterministic model is a model in which there are no chance elements, so that the output of the model is uniquely determined by the inputs. The term (expected value) is often used to describe a deterministic model in which the inputs are expected values of random variables.

(2) The stochastic model is a model that contains random or chance elements so that the output is not uniquely determined but rather is described by a probability law.

Earlier we defined a model as a representation of reality; however, this abstraction takes many forms, as illustrated in Fig. 1-2, which depicts a continuum bounded at one end by methods using exclusively analytic mathematical descriptions and on the other by experimenting with the real world. Experimentation with the real world is included as a model and thus, by definition, is not completely realistic since experimental controls necessary to proper scientific study will introduce artificialities.

In many complex systems it is not possible to use completely analytic mathematical descriptions of a complex system because the tools are simply not available. On the other hand, direct manipulation of the real world for experimental purposes may be impractical (e.g., if the system is only in the paper design stage, or if cost or risk to human life is prohibitive.)

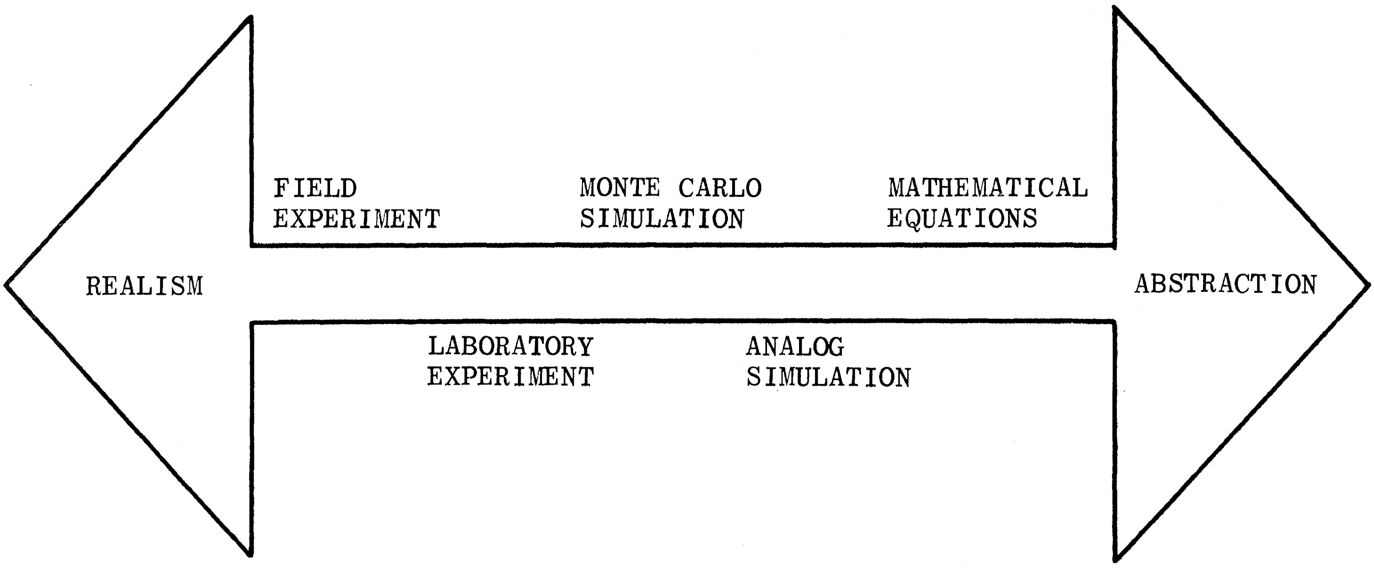


FIGURE 1-2. SPECTRUM OF MODELS

Systems Approaches. With this background on the concept of modeling, we can consider next the topic of systems and systems approaches. Many definitions of the term system have been offered. Simply, we may consider that a system is a complex of functional components, principally men and machines, which has been organized to perform specified tasks in different social, industrial, military, economic, and physical environments. Ideally, the system performs these tasks by utilizing its resources to satisfy some measureable criteria. Ashby (1958, p. 40) defines a system as "...not a thing, but a list of variables," which implies, quite correctly, that (a) the boundaries of the system are established by the analyst, and (b) the "systems approach" is just a variation on the technique of modeling.

In practice, this variation is one of scope or breadth; i.e., in the systems approach one attempts to model the entire system, with all its interacting functional components or subsystems, into what might be called a composite model, as opposed to the component model with its implied limitations in scope. The composite or systems model was developed to assist decision-makers to understand situations too complex to be described by component modeling alone. For example, among highway safety researches, if one finds a study in which the road is assumed a constant and driver behavior is the principal variable, one can also find a study in which driver behavior is assumed constant and the road is the principal variable. Neither study yields jointly acceptable conclusions because, within the overall complex of highway safety research, both the highway and driver behavior are variables.

The composition of models is a chaining process. Each submodel is linked to the others by a series of input-output links (i.e., the output of each submodel is the input to the next). In this process it is often necessary to expand the set of inputs and outputs that would be used in each submodel by itself. For instance, if we are to combine a submodel of driver behavior at stop lights with a submodel of automobile design, the dependence of driver behavior on such design parameters as windshield area must be made explicit.

Large and complex systems are not themselves particularly new phenomena. However, the need to understand their operation, and thus, the need to develop effective methods for analyzing these systems, is relatively recent, dating approximately to the early days of World War II. It was during this period that the first "operational research" group (better known today as operations research, or OR), was organized by Professor Blackett in response to the urgent need to solve new and complex military operational problems (Blackett, 1962).

Blackett's "circus" (the term was coined to describe the interdisciplinary nature of his staff, considered a novelty at that time) was organized within the United Kingdom Aircraft Command Research Group. Its success in applying the scientific method to problems of the British war effort led to the creation of similar groups in the United States. Scientific staff was drawn from Princeton, Columbia, the M.I.T. Radiation Laboratory, and the Naval Ordinance Laboratory, to work with operational units of the military services. Accounts of some of their early activities are recorded by Trefethen (1954).

The basic concepts of operations research were, of course, not new, drawing as they did upon well established principles of both science and management. However, the synthesis of an interdisciplinary set of scientific methods, techniques, and tools to solve a particular class of problems was an innovation. Current systems approaches represent a development and outgrowth of these initial OR efforts.

After the war a number of organizations (the RAND Corporation and the Weapon Systems Evaluation Group, to name just two) were formed to render continuing scientific support to the military services; on a much smaller scale, other operations research groups were organized to deal with industrial problems. These organizations directed their early efforts to formulating mathematical models of some of the OR procedures which had proved successful during the war. As operations research developed in the mid-fifties "...honors went to practitioners who used or improved mathematical techniques like linear programming or queueing theory and found new applications for them" (Quade, 1964).

The optimization model (Fig. 1-3), often considered synonymous with operations research methodology, is described by Churchman, Ackoff, and Arnoff (1959):

> This model expresses the effectiveness of the systems under study as a function of a set of variables at least one of which is subject to control. The general form of the operations research model is
>
> $$E = f(x_i, y_j)$$
>
> where E represents the effectiveness of the systems, x_i the variables of the system which are subject to control, and y_j those variables which are not subject to control. The restrictions on values of the variables may be expressed in a supplementary set of equations and/or inequations.

In other words, an "optimum" solution may be determined mathematically if formal statements of the relationships between variables are available.

Although the operations research paradigm proved sucessful in solving a wide range of military and industrial problems, it had its limitations and, by current standards, is concerned with relatively component problems. The chief difficulty was in handling problems in which a single measure of effectiveness or criterion function was not available.

The development of optimization techniques for operations analysis of existing weapon and industrial systems continued throughout the 1950's. However, led by the work of The RAND Corporation for the Air Force, emphasis gradually shifted to problems of broader scope and complexity, such as the development of weapon systems, force structures, and disarmament. The analysis of weapon systems for possible future warfare presented a new kind of problem, one quite different from those treated by operations analysis in World War II. The key word was planning, with its implied estimates of future needs. As one military man has pointed out, "...reading a crystal ball is a talent claimed by few, but

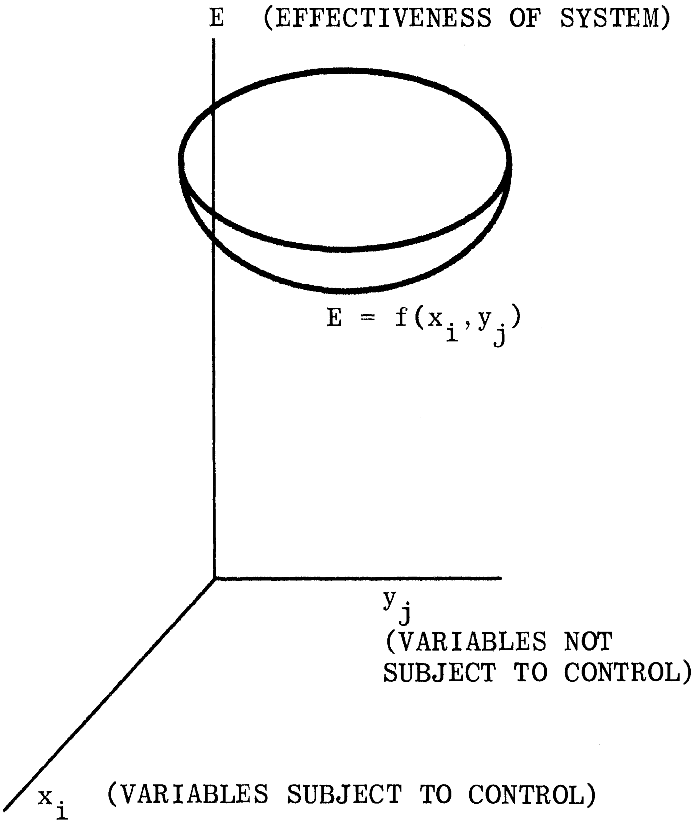


FIGURE 1-3. THE OPERATIONS RESEARCH MODEL

needed by all who work in the planning field." The process of planning for future systems is, to a large extent, synthesis, in that the environment has to be forecast, the alternatives designed, and the operational laws invented.

The approach used to solve these problems emphasized a comprehensive treatment by an integrated, simultaneous examination of all major relevant problems, i.e., the "systems approach." This approach to large-scale systems problems has been given many names ranging from the original operations research to systems engineering, systems analysis, systems research, and cost-effectiveness analysis. Efforts to distinguish clearly among these have met with little acceptance; however, at the risk of renewing the semantic controversy, the following definitions appear to summarize the essential differences:

1. Operations Research: Analysis to increase the efficiency with which existing man-machine systems are operated.
2. Systems Engineering: Detailed engineering design involved in realizing a future system in terms of specific hardware.
3. Systems Analysis[3]: Systematic approach to the comparison of alternative systems for carrying out some specific task or tasks. If differences in cost are considered, it is referred to as a "cost-effectiveness" analysis.
4. Systems Research: Development of procedures for the planning of new industrial or military systems to better perform existing operations or implement operations never before performed.

In essence, then, operations research analyzes existing systems, systems engineering designs specific future systems, systems analysis compares the effectiveness of alternative systems, and systems research develops system planning procedures.

---

[3] Quade (1964) notes that this term is often used in the commercial world to describe the analysis of business office systems.

Having discussed some basic modeling concepts and distinguished among various "systems approaches," we can turn now to some statements of objectives for both the Highway Safety Research Institute and the HSRI systems project. As a necessary basis for formulating its own objectives, the systems project had to assume a set of reasonable objectives for the Highway Safety Research Institute. It must be emphasized that while these objectives, set forth in the following section, represent the current views of the systems project group, they are not to be interpreted as having official HSRI endorsement at this time.

OBJECTIVES FOR THE HIGHWAY SAFETY RESEARCH INSTITUTE

In the belief that the component-study approach is inadequate to deal with the complex of interrelated elements contributing to highway safety, the Highway Safety Research Institute is appropriately committed to developing a broad and comprehensive approach to the achievement of highway safety. It is assumed that the long-range objective of the HSRI is to develop and utilize methods for evaluating the broad effects of specific resource allocations to improve highway systems. Specifically, it is assumed that the Institute will devote its efforts to developing (through systems research) a comprehensive highway systems model that may be employed for operations and systems analysis of highway systems. This model, which would be available to industrial and government management, would be used to evaluate (a) the effects of making design changes in the existing system, (b) the effects of instituting changes in the operating policy of an existing system, (c) the feasibility and cost-effectiveness of proposed systems, and (d) the comparison of alternative systems. The model would, in addition, serve the scientific community as a tool for identifying areas where research is most needed and might yield the most significant advances.

The second objective of HSRI follows quite naturally from the first, although the two may at times appear to be in conflict. This second objective is to make efficient use of the University's research capabilities in order to develop the basic knowledge necessary for a comprehensive model of highway transportation systems.

Assuming these two objectives for HSRI--to develop a highway systems analysis model and to stimulate and sponsor basic research on highway system phenomena--the primary objective of the systems project may be formalized as follows: to provide a mechanism for the rational selection of research projects leading to the development of a comprehensive highway system model.

Rational selection requires consideration of both priority and integrability. It is perhaps more important to determine priority of research, but it is also usually more difficult to obtain the information necessary to establish these priorities. Campbell (1965, p. 350) suggests the following approach to establishing research priorities:

> A basis for isolated projects is the accident facts as gathered by any jurisdiction. Accident Facts, published by the National Safety Council, is an excellent source for the selection of projects. One can see at a glance where the tough but fertile problem areas are by looking at rates and absolute numbers in the analyses of circumstances, location and direction of movement, and improper driving actions. For example, excessive speeds, passing on the left, crossing movement, failure to yield, and crossing the streets by pedestrians are some of the most impressive areas. And, finally, one should select those problem areas where the accidents tend to increase geometrically each year.

Unfortunately, despite the apparent rationality of selecting projects with high marginal payoff for safety, this approach is misleading. It must be remembered that this simplified view of "cause of accident" has led us nowhere in the past, and that recent literature clearly indicates that accidents are products of various combinations of human, vehicle, and environment factors. Campbell's suggested approach would almost inevitably result in "Type III" statistical error--the right answer to the wrong question. Thus, given the hypothesis that a multiplicity of factors contribute to accidents, a rational selection of project priorities in terms of high payoff would depend upon an understanding of the effects of interactions among these factors. We have come full circle, since this understanding can only be obtained by developing a comprehensive systems model--precisely the objective of the Highway Safety Research Institute. For these reasons, the systems project's recommendations for research project priorities have been based upon the identification of serious gaps in knowledge, rather than upon anticipated payoff. The identification of these areas proceeds directly from the development of the integrability criterion.

A major effort of the project was the development of a systems framework to ensure the integrability of component research efforts. In essence, this framework will also form the initial structure of the model to be used for systems analyses of highway transportation systems. In formulating this systems analysis model the overall research task is structured so that initial integration of component research efforts is possible. Without such an integrating structure a research program can be little more than a series of component studies which must somehow be organized and interrelated <u>after</u> they are completed. That this is almost impossible is evidenced by the plethora of existing component studies which have yet to be integrated to produce effective accident countermeasurers.

In addition to providing the mechanism for integrating component research studies, the systems model will provide the following:

1. An explicit statement of research needs and a structure for the long-term research program, leading to a usable model for systems analysis. It will specify the measures which are necessary and the relationship which must be developed.

2. A basis for the interdisciplinary research; i.e., it will identify the common ground upon which the various disciplines in the long-term project converge. The model, as it is continually enriched, will identify the research goals within a framework that ensures they can be clearly understood by everyone concerned with the Institute's research program. Thus, in a sense, the model will provide a statement of the overall strategy.

In the remaining sections of this chapter, the initial qualitative highway model is described, the function and technical content of each of the component research areas examined are summarized, and guidelines for allocating research resources are presented. Succeeding chapters treat the various component research areas in greater depth, and include more detailed recommendations.

A QUALITATIVE HIGHWAY SYSTEM MODEL

It should be clear from the preceding discussion that the primary objective of the highway systems project is to develop a research framework within which component studies may be performed and then integrated. It is intended that this integrated structure, or systems model, be used to perform systems analyses on proposed highway systems. That is, the systems model is to be a mechanism for predicting how alternative designs and operational policies will affect the costs (safety and monetary) and service provided by the proposed systems. In this section we shall present such a systems model by (a) defining the different classes or levels of factors considered in the model and (b) describing the sets of hierarchial relationships by which these factors are interrelated in appropriate submodels.

Before embarking on this itinerary, we pause to establish our basic vocabulary. Earlier we defined a system in general as a complex of functional components, principally men and machines, organized to perform specified tasks in different environments. Here we will be specifically concerned with the highway transportation complex organized to provide safe transportation service for society. This system is comprised of elements or <u>subsystems</u> such as the vehicle, the highway, the human operator, the casualty recovery facilities, and others. These subsystems are, in turn, made up of <u>components</u>. For example, the vehicle subsystem is the sum of its components--the engine, transmission, suspension, etc.--and their attendent interactions.

The inclusion and designation of particular elements as subsystems or components of the highway transportation system is neither unique nor static. Initial specifications are based on expected usefulness for systems analysis. As more is learned about the system, and as experience with system modeling grows, additions, deletions, and redesignation of elements will, in all likelihood occur.

SYSTEM FACTORS

The service and safety of highway systems in operating environments is dependent on, and in fact constrained by, (a) limitations in driver capabilities and hardware technology for design[4] of subsystems, and (b) control decisions which structure or plan the operational use of design capabilities. Accordingly, estimates or predictions of the service and safety provided by a highway system should be related to design and control factors, and, conversely, subsystem design decisions should be based on operational requirements (demand, and minimum acceptable safety standards). This interaction among design, control, and operational factors is shown in Fig. 1-4. As this figure indicates, the system's design and operational requirements must be considered concurrently when planning changes in an existing system or conceptualizing proposed new highway systems.

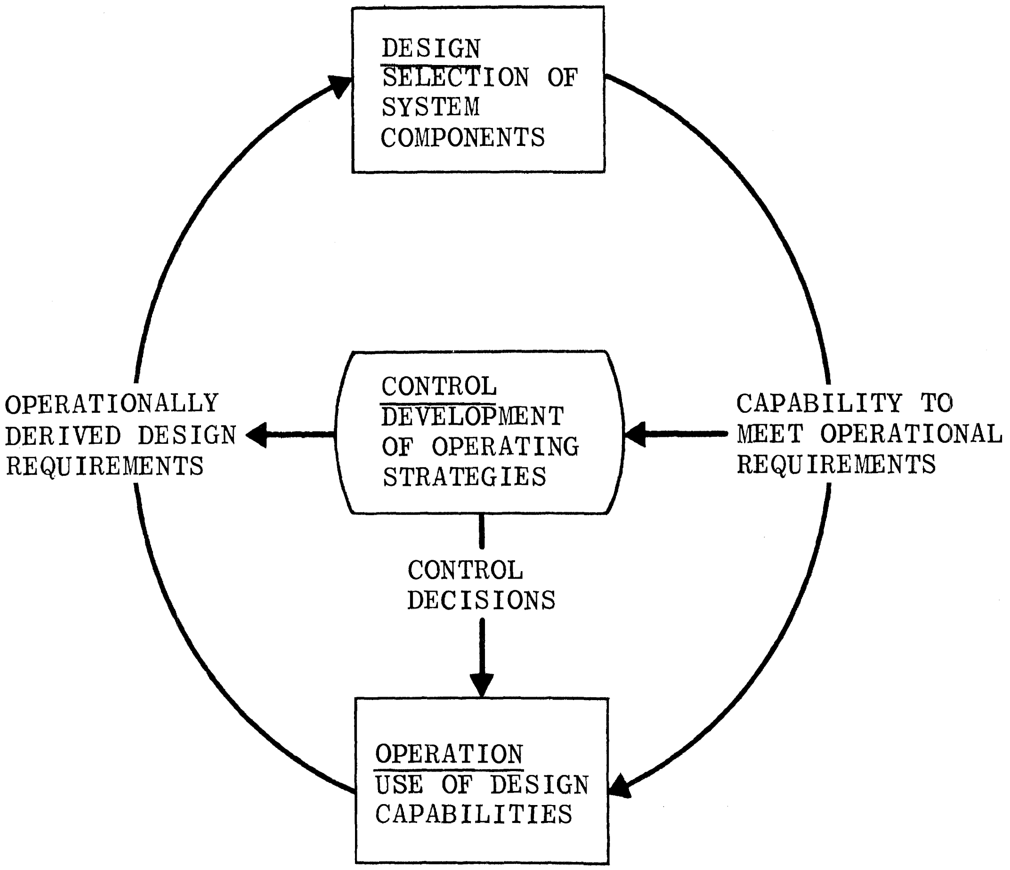


FIGURE 1-4. INTERACTION AMONG DESIGN, CONTROL, AND OPERATION

---

[4] Design is taken to be the selection of subsystem component hardware such as engine for the vehicle, materials for the highway, and medical equipment for a recovery system.

In accordance with this philosophy, we define the following levels or classes of factors used in conceptualizing related design, control, and operations models. The chosen sets of factors are neither unique nor complete at this stage of the study; however, they do facilitate an initial organization of the system's complexity.

System Effectiveness Measures (e_i): These are measures of the system's worth in terms of the quality of service and level of safety provided. In contrast with the typical operations research approach, which employs one measure of effectiveness, multiple measures (each of which provides decision-makers with specific information) are utilized in a systems study. Some representative effectiveness measures are given in Table 1-I. A more complete list is given in Appendix 1-A.

TABLE 1-I. SYSTEM EFFECTIVENESS MEASURES

Service	Safety
No. of Passenger-Hours	Frequency of Accidents (by type)
No. of Ton-Hours	Frequency of Accidents/Mile of Road
Average Concentration	Frequency of Accidents/ Mile of Vehicle Travel
Average Travel Time	
	Frequency of Injury/Passenger Mile

Subsystem Performance Measures (y_j): These are measures of the performance of subsystems such as the vehicle, the highway, the maintenance subsystem, and the casualty recovery subsystem. The subsystem performance measures contain both static and dynamic elements. The term "performance capabilities" is used to describe static (absolute) measures (independent of time) which are numbers that would be used to specify capability requirements of a subsystem. Dynamic measures, called "performance functions," are explicit functions of time (state of the subsystem) as viewed in an operational context. The distinctions between performance capabilities and performance functions are more explicitly treated in the chapters devoted to specific areas of study.

It is emphasized that these are not measures of the benefit or worth of the system, but rather of the performance which ultimately affects the system's worth. Table 1-II lists a few representative subsystem performance measures; a more complete list is given in Appendix 1-B. Since some of these measures are dependent upon a number of subsystems, it is convenient to group them under conceptual headings.

TABLE 1-II. SUBSYSTEM PERFORMANCE MEASURES

Mobility

Speed
Acceleration
Stopping Distance
Road Passing Distance
Headway Maintenance
Minimum Turning Radius

Recovery

Recovery Time
Case Fatality Ratio

Survival

Vehicle Reliability
Road Serviceability
Required Maintenance (vehicle, road)
Pr(Transmitted Forces = j | Impact Vector X)

Medical

Pr(Injury Level i | Transmitted Force = j)
Pr(Survival | Injury i and No Treatment)

Examination of these factors indicates that they are in a sense conceptual rather than physical. That is, they describe performance measures for many kinds of transportation vehicles other than cars; for example, helicopters, or zero-ground-pressure vehicles. In fact, these are just the factors that one would use as input to a systems analysis of alternative transportation systems.

Control Strategies (s_k): These are not measurable variables, but are decisions or plans to adjust the current state of any subsystem or combinations of subsystems. These plans are implemented through the use of the subsystem performance functions (y_j) noted above. It is convenient to consider the use of these capabilities as operational tactics. For example, the decision of a human driver to pass in a particular traffic situation might be the strategy, and the requisite use of the steering and acceleration capabilities would constitute the necessary operational tactics.

In the existing highway system, the strategist whose actions must be described for various operational situations is the human operator. If the human operator is supplanted (as would be the case on an "automatic" highway) control strategies might well be preprogrammed. Some representative strategies are shown in Table 1-III. A more detailed listing is given in Chapter 6.

TABLE 1-III. CONTROL STRATEGIES

Maintain Current State
Pass
Avoid
Maintain Safe Speed
Maintain Comfortable Speed

Hardware Performance Measures (x_i): These are measures of (a) subsystem hardware performance capabilities (x_1, ..., x_r) and (b) component hardware performance capabilities (x_{r+1}, ..., x_n). The purpose of the dichotomy of factors within a class of measures will become evident when the submodels and their functions are described. As the name implies, the component hardware measures describe performance of such components as the engine of the vehicle subsystem. In contrast, the subsystem hardware measures describe characteristics of the subsystem as a whole and are, in general, dependent upon the component measures.

Human motor, sensory, and information-processing performance capabilities such as reaction time and acceleration tolerance are included in this class of component and subsystem hardware performance measures. This is consistent with the view that one of the human's roles as an operator is that of enhancing or deterring performance of subsystems. This function, in essence as another component, should not be confused with his decision-making role as a control strategist.

Representative hardware performance measures are shown in Table 1-IV. Although many of these factors are identified in succeeding chapters, a comparative list for all subsystems has not been developed to date.

TABLE 1-IV. HARDWARE PERFORMANCE MEASURES

Component	Subsystem
Engine Horsepower	Vehicle Gross Weight
Pavement Stiffness	Coefficient of Friction
Human Vibration Tolerance	Human Steering Capability
Transmission Efficiency	Load Capacity

An examination of these factors will indicate that they are generally conceptual rather than physical descriptors, in that they describe performance capabilities of the population of automobiles and are not restricted to a particular car. For example, engine horsepower is a measure common to all automobiles; however, the magnitude of this measure varies with the physical characteristics of the specific engine and car under consideration, i.e., gas turbine, reciprocating piston, etc.

Physical Hardware Characteristics (z_j): These are the physical characteristics of subsystem components such as brake lining thickness, wheel diameter, highway material, the color of the road signs, cylinder length, and the distance from brake pedal to accelerator.

Environmental Parameters (θ_k): These are measures that describe the environment over which it is assumed we have little or no control. We distinguish between two such groups of parameters. The first (θ_k; $1 \leqq k \leqq r$) describes the natural and man-made environment such as climate and noise. The second group (θ_k; $r < k \leqq n$) consists of measures describing the dynamic state of the vehicle-operator system, e.g., "lane density is 250 vehicles/mile" or "a vehicle is approaching at 60 mph." Although in many studies road characteristics are included as part of the environment, they are not so included here, since the road is under our design control.

Identifying and assigning particular factors to the appropriate class of measures is, like the selection of relevant subsystems and components, neither unique nor static. Although most factors can readily be recognized as belonging to a particular class, others leave room for considerable doubt. For example, is "roadway illumination" an environmental parameter or a component hardware performance measure? The resolution of such questions will obviously be a function of the analysts' experience, but in all cases should be based on the primary role the factor in question plays in the overall analysis.

QUALITATIVE SYSTEM RELATIONSHIPS

Having defined the many systems factors, we are now in a position to describe a symbolic paradigm which, it is hypothesized, defines and depicts the interrelationships among all factors of the highway system model. The relationships are given by

$$e_\ell = F_\ell(Y,\ S,\ t) \qquad [1]$$

$$s_k = \psi_k(Y,\ X_h,\ \theta_d) \qquad [2]$$

$$y_j = f_j(X,\ \theta) \qquad [3]$$

$$x_i = g_i(X) \qquad 1 \leq i \leq r \qquad [4]$$

$$x_i = h_i(Z,\ \theta_s) \qquad r < i \leq n \qquad [5]$$

where $S = (s_1,\ \ldots,\ s_k,\ \ldots,\ s_n)$

$Y = (y_1,\ \ldots,\ y_j,\ \ldots,\ y_n)$

$X = (x_1,\ \ldots,\ x_i,\ \ldots,\ x_n)$

$Z = (z_1,\ \ldots,\ z_j,\ \ldots,\ z_n)$

$\theta = (\theta_1,\ \ldots,\ \theta_k,\ \ldots,\ \theta_n)$

$$\theta_d = (\theta_{r+1}, \ldots, \theta_n)$$

$$\theta_s = (\theta_1, \ldots, \theta_n)$$

and X_h is the set of component and subsystem hardware performance measures that describe human motor, sensory, and information processing capabilities only. Qualitatively, the set of equations [1] represent the system operations model, [2] the control model, and [3-5] the design model.

Equations [1] qualitatively represent an operations model which is used to integrate the many subsystem performance capabilities (y_j) and the control strategies (s_k) to assess the worth of the system in terms of the service provided and safety afforded. The symbol (t) is included in the formulation to indicate that all the activities in the operations model are time dependent. The model is used to evaluate the system by examining the dynamic changes in subsystem performance measures and their utilization to fulfill an operating demand.

Equations [2] represent the control model which, for the case of a human controller, describes the decisions or control strategies made by the operator when confronted with particular operational situations. The equation indicates that these strategies (passing, pulling off the road, etc.) are partially related to the driver's knowledge or awareness of (a) some of the subsystem performance capabilities, (b) his own capabilities, and (c) the particular dynamic environment in which he is driving. For example, one might hypothesize that a driver's decision to pass is related to his knowledge of (a) the vehicle's acceleration capability, (b) his ability to control the vehicle during the passing maneuver, and (c) the conditions in the opposing lane at that moment in time.

Equations [3] are used to estimate subsystem performance measures when the system is operated in a particular environment specified by the parameter set θ. In all likelihood, more than one such relationship will be required for some performance capabilities, depending on the specific nature of the x_j and θ_k. Consider, for example, vehicle speed capability. Maximum speed over straight and level highway would be limited by the power package capabilities, maximum speed over rough and bumpy roads might be limited by the driver's tolerance to vertical acceleration loads, and maximum speed during limited visibility conditions would be governed by the driver's sensory capabilities under such conditions. Obviously, these different aspects cannot be included in any one relationship. Rather, each must be investigated separately and included in the model. The reader should recognize another aspect of Eq. [3]. It was noted that the y_j are descriptive of many modes of transportation; however, the x_i describe the hardware performance capabilities of specific vehicles. Accordingly, the relationships f_j would need to be developed separately for each mode of transportation before alternative modes (cars, helicopters, etc.) could be systematically analyzed as competitors to fulfill an operational demand.

Equations [4] represent an interaction model which serves two major functions:

(1) It is used to account for the fact that changes in one component of the system will affect the component characteristics of other parts of the system. For example, changes in the strength characteristics of the highway may well affect the coefficient of friction between the tires and the highway surface.

(2) It is used to resolve a problem unique to systems planning, that is, that prior to actual design of the subsystems, their hardware performance characteristics will not be known. As a simple example, the gross weight, center of gravity, and moments of inertia of an automobile are not physically measurable until the car is actually designed. Accordingly, some method of predicting these measures as a function of engine, transmission, suspension and other component characteristics is needed.

Finally, Eq. [5] is a statement of engineering relationships that exist between hardware component performance capabilities and the physical hardware characteristics and natural environmental parameters. For example, the engine brake horsepower is functionally related to the area and length of the cylinder, the cylinder pressure, the number of cylinders and engine speed; and the braking reaction time of the human might be related to the physical distance from accelerator to brake pedal and to ambient temperature. This level of detail may appear to be unnecessarily microscopic; however, as can be seen in succeeding chapters of this report, it is precisely at this level that the majority of the fundamental component research efforts have been and will continue to be directed.

SYSTEM SUBMODEL AND PROJECT STUDIES

Equations [1] through [5] constitute a symbolic description of a highway system model. Figure 1-5, a schematic representation of the model, provides an overview of its possible utilization for assessing the service, safety, and costs of alternative highway systems. In addition to depicting the relation between the design, control, and operations models, the figure indicates the inputs and submodels deemed necessary for reasonable initial evaluations of alternative or modified highway systems. The purpose of this section is to summarize how the model can be used for such an evaluation, and to describe briefly the function of the submodels, indicating the existing knowledge (or lack of knowledge) in the various study areas.

Inputs to the various models are:

(a) Hardware performance measures (x_i) and/or physical hardware characteristics (z_j) of the highway systems to be studied. These may be characteristics of proposed systems or modifications to existing ones.

(b) The geographic area in which the proposed system is to be evaluated.

(c) Operating policies of some of the subsystems, such as recovery and maintenance.

(d) Social and legal factors which, for the initial modeling, are assumed to affect only vehicle controller behavior[5]. It is recognized that later models must include their effect on design characteristics of the vehicle and other subsystems and their effect on various aspects of the operations model.

The procedure for utilizing the model is relatively straightforward and essentially follows a path through the system equations from [1] through [5] . The hardware performance measures (x_i: $r<i\leqq n$) which can be independently specified are used by the highway hardware interaction model to compute the remaining dependent hard-

---

[5] Social factors are implicit in the demand model.

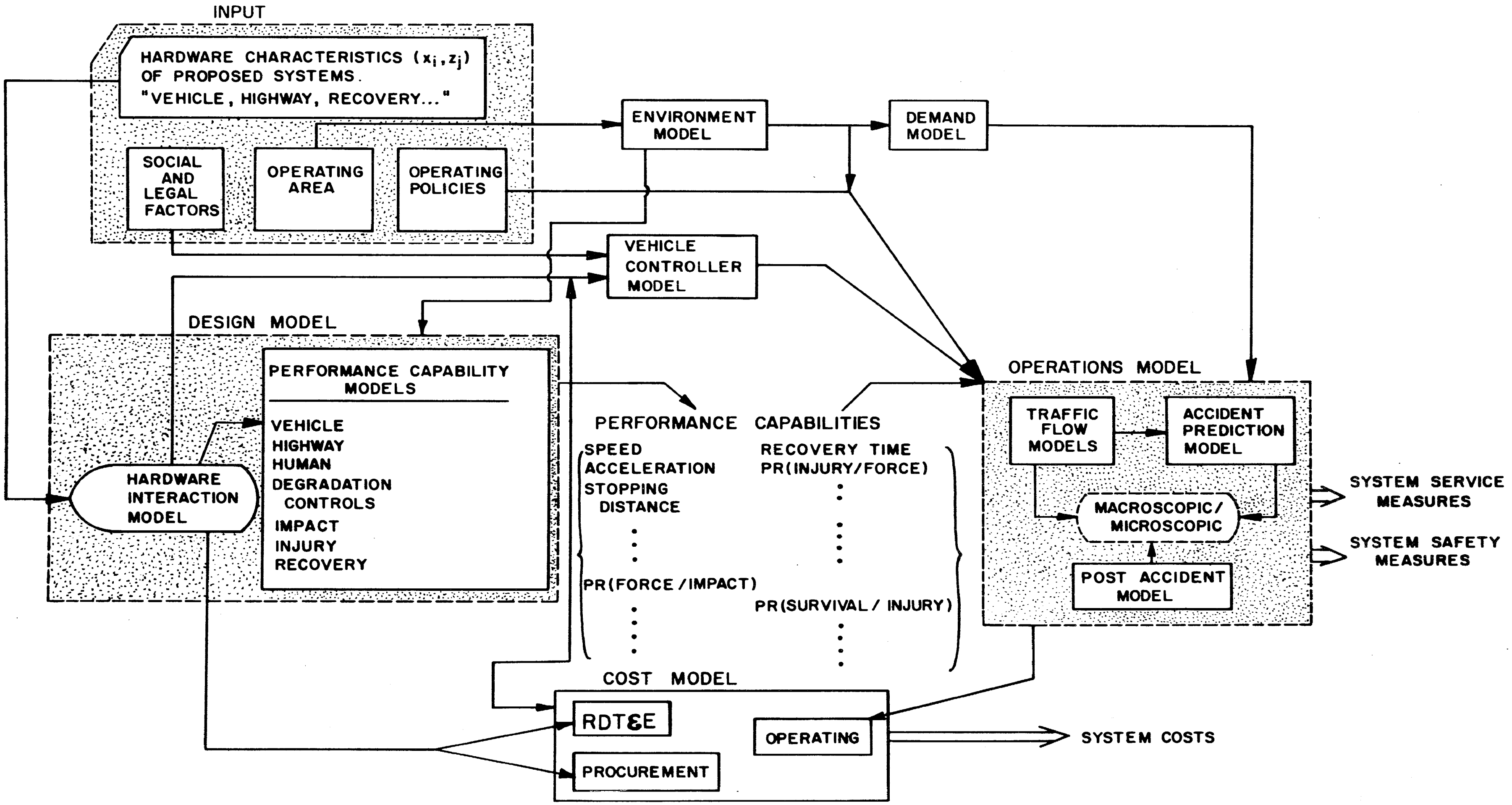


FIGURE 1-5. SCHEMATIC OF THE HIGHWAY SYSTEMS MODEL

ware performance measures (x_i: $1 \leq i \leq r$). The total set of hardware performance measures, along with appropriate environmental parameters, are used as input to the performance capability submodels to estimate performance capabilities and performance functions for the various highway subsystems. Using the design model outputs, the control model basically determines how the human controller makes use of the hardware and performance capabilities under the restraints of social and legal factors. Performance estimates from the design model, control strategies from the control model, estimates of traffic demands from the demand model, environmental parameters from the environment model, and operating policy inputs are all used in the central operations model to estimate the service and safety aspects of the system. Using both design and operational model inputs, the cost model estimates the total system costs which consider the research and development of subsystems, their procurement, and the costs associated with system operations. The resultant predictions of system effectiveness and total costs may then be evaluated by highway system planners. If the results are not acceptable, appropriate changes in the system's design (hardware characteristics) and/or operating policies can be made and the analysis repeated.

A brief description of the functions of the submodels follows; information deficiencies identified by the study are also noted. Areas examined in some depth are described in succeeding chapters.

In the following submodel descriptions an asterisk is used to denote those cases where little or no study effort was allocated because of limitations in time.

HARDWARE INTERACTIONS*

As noted previously, this model describes the important hardware interaction phenomena and serves two major functions. It is used to explain the mechanism by which changes in one component of the system affect the component characteristics of other parts of the system, and it is used to estimate some of the hardware performance measures that cannot be known prior to actual design of subsystems. These estimates are obtained through the use of equation set [4]. In mathematical terms we have a set of r equations in n unknowns ($r \leq n$) which has an infinite number of solutions. When the component hardware measures are specified as inputs, however, the system reduces to r equations in r unknowns, and yields a unique solution for the total set of system hardware measures. Thus, values for the hardware measures $[x_1$ through $x_n]$ are made available for use as inputs to the performance capability models. No attempt was made to develop equations of the interaction model, since a composite set of hardware performance measures for all subsystems still needs to be developed. The literature indicates that many such relationships in two dimensions $x_i = f(x_j)$ are available; however, their use in this model is questionable since many of the interactions are multidimensional in nature.

VEHICLE CAPABILITIES

The purpose of this model is to predict the performance capabilities and dynamics of vehicle subsystems. Study focused on the dynamic performance of vehicles, a key factor in accident analysis and prediction: traffic accidents or collisions can occur only when the trajectory of a vehicle intersects with one or more trajectories of other vehicles and/or pedestrians, or crosses a point occupied by an immobile object. Hence, the output capabilities of this model are used as inputs to the accident prediction model. Study indicated that the prime shortcoming in the area of modeling vehicle dynamics is the lack of a single composite model which relates vehicle component inputs and environmental conditions to vehicle trajectories. Some approximate models were found to describe dynamic performance for special circumstances only. There is a need to extend present models to consider simultaneously various combinations of the following control actions--braking, accelerating, steering, and vertical tracking. These should include the weight transfer among the wheels. The models should explicitly account for forces that will not pass through the center of mass of the vehicle.

THE HIGHWAY SUBSYSTEM

The highway subsystem includes the road, roadside, obstructions, and significant elements only visible from the highway. Concern is with the geometry and the materials making up the significant elements of the subsystem. Traffic informational and control devices are treated. This subsystem interacts with most other subsystems since the scene of highway operations is always on a highway and the state of the highway system is of significance to all moving vehicles. For example, the highway interacts with the environment subsystem to produce a surface condition, and with the vehicle subsystem to provide a vehicle operating speed.

In the chapter dealing with the highway subsystem the relationships between the physical hardware characteristic variables and the component hardware performance capabilities were found to be generally well understood. On the other hand the subsystem hardware performance characteristics are not known as explicit functions of the components. Knowledge appears to be particularly lacking in the relationship between the coefficient of friction between the tire and the road and the energy absorption characteristics of roadside obstacles. There appears to be a lack of definition and knowledge of subsystem performance measures to be used in the development of strategies and effectiveness measures. Work is needed in some of these measures in the areas of street lighting, traffic control and informational devices, and highway geometry. These of course must be developed jointly with measures from other subsystems.

HUMAN CAPABILITIES

It was noted earlier in this chapter that the human has a dichotomized role in the highway system in that he acts both as a component of the system and as its controller or decision-maker. As a system component his capabilities are not focused on any one subsystem but rather cut across many of them[6]. Within this role he collects information through his sensory system and processes it through his intellectual system, producing a motor response which either changes the dynamic variables of the vehicle subsystem or

---

[6] Human factors enter into the total system at nearly every point, and they do not in themselves coalesce into a meaningful conceptual unit of analysis that can be treated independently within the systems model. This is not to imply that different basic psychological principles are necessarily required to predict and explain the different human behaviors that operate within each system component. At the present stage of development, however, it does appear that different conceptual treatments of human behavior will be differentially fruitful at different points of the analysis

enables him to take in new information. Performance capabilities do not describe what the human does in a particular set of circumstances--those are the control strategies--but rather what he is capable of doing. These capabilities have been grouped into a number of general categories--sensory, motor, and information processing. The literature of engineering psychology indicates human performance capabilities are reasonably well documented in terms of the variables to which they are related, and that research in the area of highway safety can profit from the general functional relations developed in this field.

DEGRADATION

The term degradation is used to refer to the phenomenon of loss in component performance with age and usage. The function of the degradation model is to determine the relation between degradation of specific components and the resultant change in subsystem performance capabilities. Changes in the latter are then used as direct inputs to the accident prediction model, and as indirect inputs via the vehicle controller model (i.e., through the hypothesis that the human control strategies are modified by changes in subsystem performance capabilities). Control of degradation process may be accomplished by improving component designs, by screening components prior to selection, by limiting the load on components, and by optimal policies of inspection, maintenance, and replacement. Each of these degradation control methods requires a monetary investment accounted for by the system cost model.

The literature indicates little or no knowledge regarding either degradation processes themselves or the means of controlling degradation. Initial efforts of research in this area should be directed toward determining what subsystem performance capabilities are affected by degradation processes, as well as determining appropriate

---

in order that the human factors may be characterized either in terms that are commensurate with the other input and output variables of the subsystem under consideration or in terms that lend themselves to a specific set of analytic techniques. The outline given in Appendix 1-C lists the various roles played by human factors in the total system and indicates those that are considered at greater length in subsequent chapters of this report. The role of the human as the operator of the vehicle is selected for the most detailed analysis in terms of the systems model (Chapter 6).

measures of degradation or degradation process parameters. A major theoretical problem is the description of degradation in the light of the dependencies that exist among the components and subsystems. Secondary research efforts should be concerned with the various means of controlling degradation via inspection and maintenance activities.

IMPACT AND INJURY SUBMODELS

These two areas are considered together, partly because they received minimal study, but also because sequentially they form the main connecting link between the accident prediction model and the recovery subsystem model. If the accident prediction model can specify frequency of accidents in terms of their impact direction and force level, the role of the impact model is to determine the attenuation of those forces eventually transmitted to the driver and passengers of the vehicle. The specific measure hypothesized as relevant to this phenomenon is the probability of transmitting a particular force, given an accident described by a specific direction and level of impact force. The model should include what are commonly called "first-" and "second-collision" effects. Discussions with other safety research personnel suggest that this information cannot be generated from existing knowledge in the impact research area.

Given an estimate of the impact forces transmitted to the occupants, the function of an injury model is to determine the type and level of injury sustained. The measure considered descriptive of this phenomenon is the probability of sustaining some particular level of injury given exposure to a particular level of force. Although needed, however, no taxonomy of injuries relevant to the purpose of the highway systems model appears to exist. Research is needed to relate force levels to injury levels, and to develop useful measures of the injury level. It is suggested in Chapter 7 that injury level might well be measured by probability of survival as a function of time in the absence of medical treatment.

RECOVERY SUBMODEL

The casualty recovery subsystem is concerned with other subsystems whose design and operation are directly related to the location, treatment, and transportation of traffic-accident injury victims. The three aspects of casualty recovery (location, treatment, and transportation) are subsumed under the hypothesized major measure of such a system--the probability of survival given an injury of a particular level. The literature indicates that little is known about the performance of current recovery systems, and that no efforts have been devoted to the design and efficient operation of future systems. Initial research should be concerned with the effect of system response time and medical treatment level on such measures of fatality as the case fatality ratio. As noted above, measures describing injury levels are needed as well as measures describing treatment levels.

VEHICLE CONTROLLER MODEL

The purpose of this model is to produce output describing the behavior of the vehicle controller in various dynamic situations. This behavior has been termed control strategy. Control strategies are modified to correspond to the real system performance capabilities which may be limited by the human performance capabilities or subsystem performance capabilities. As shown in Fig.1-5, the control strategies may also be modified by social and legal factors. Past and present research in the area has been directed toward discovery of the variables which affect the strategies but little or no effort to development of functional relations to predict them. Research is required to determine the form of the control strategies and to determine the influence of different methods of presenting data to the human control system.

ENVIRONMENT MODEL

By environment we refer to those factors to which a system generally must adapt, but which are independent of the system and generally not subject to practical modification. The function of the environment model is to predict the frequency of occurrence of relevant aspects of the environment. Research is needed to develop ways to apply available meteorological data to a highway system visibility model for input to the vehicle controller model. An adjunct to this is the development of a predictive model which will give the probability of occurrence of meteorological conditions. A descriptive model of wind force, with emphasis on spatial and time characteristics, is needed as an input to the vehicle dynamics model.

DEMAND MODEL

The traffic-generation or demand model describes the set of trips undertaken in a given highway transportation system. This description provides the input for the operational model. The demand is affected by service and safety levels, by the provision of alternative modes of transport, and by general sociological factors describing the pattern of life around the highway system. In the present system structure the dependencies on service and safety are not explicitly modeled. The dependencies on sociological variables appear only in determining the pattern of traffic generation in time and space. Other effects are considered to be second-order in nature.

Considerable work is being done with demand and traffic assignment. There remain, however, no quantitative methods to define "unreasonable" delay or restriction of movement. Despite having established definite levels of capacity for given modes of transport and highway types, understanding of capacity is incomplete. Expressions are needed to relate congestion to demand and, hence, effectiveness and safety. Current demand models suffer from the absence of such expressions.

SYSTEM COST MODEL

The purpose of this model is to predict total systems cost,[7] including the research and development cost of equipment, the procurement of these assets, and the operating costs of materials, personnel, facilities, training, etc. The overall cost model has been constructed so that the cost of research and development and of procurement are estimated from the hardware performance measures and/or subsystem performance capabilities described earlier. The costs of operating, maintenance, inspection, etc., are computed directly from the predicted cost of fuel, maintenance times, etc., derived from knowledge of the hardware characteristics and traffic flow

---

[7] No monetary costs are assigned to the items specifically measured by the service and safety outputs of the model. That is, no attempt is made to assign a monetary value to injury and death toll.

measures. The literature indicates that research and development, procurement, and operating costs have been developed for other systems--particularly defense systems. Some of these can be directly applied to the highway transportation system once the appropriate data are acquired. A number of operating cost models of the highway transportation systems were found. No work has been performed which accounts for the important interactions and dependencies of costs in the research and development, procurement, and operating areas.

TRAFFIC FLOW MODELS

The traffic model accepts (1) the demand output, (2) the vehicle controller description or strategies, (3) the highway network description, and (4) the environment description, and produces measures of the flow of vehicles, and thus passengers and cargo, through the system. The output of the model provides the data from which the system service measures are calculated. Outputs are also used as direct input to the accident prediction model. The literature contains four different types of traffic flow models which were not in any way integrated, although integration could be easily performed. The queueing model and car-following models are structures that are analogous (but by no means isomorphic) to the real-world process. It is recommended that initial research be devoted to extending queueing and car-following theories to accept augmented sets of inputs, and, in the case of queueing theory, to produce statements about the physical state of a queue as well as the number of vehicles in it. Secondary research efforts should be directed toward unifying the present knowledge incorporated in capacity and flow-density relation theories.

ACCIDENT PREDICTION MODEL

The purpose of this model is to predict the location, type, and frequency of accidents as a function of the operational situation. Specifically, the model accepts as input the character of traffic, the performance capability of the vehicle subsystem, the vehicle controller's strategies, the highway network, and the environmental characteristics. The literature search failed to reveal any models in which the actual events preceding and leading up to an accident are taken into account. This represents the most serious deterrent to the evaluation of highway transporation systems, since the output from the accident prediction model, together with the output from the post-accident event model (see below), produces the measures of safety which are required for systems analysis.

POST-ACCIDENT MODEL

The purpose of this model is to integrate the information produced by the accident model, the impact model, the injury model, and the recovery model to produce eventual measures of the safety of the system. One possible means of performing this integration is suggested in Appendix 1-D.

GUIDELINES FOR RESEARCH ALLOCATION

In the introduction to this chapter it was recommended that research resources be directed to those systems model areas characterized by inadequate knowledge rather than to areas presumed to offer a high marginal payoff in safety. Areas in which knowledge was found to be deficient were indicated in preceding sections of this chapter. The chapters which follow present a more detailed discussion of the state of knowledge in the areas examined by this study, along with suggestions for particular researches.

In concluding this chapter, we present some overall guidelines for the selection of systems model research directed toward the assumed long-range objectives of the Highway Safety Research Institute--to develop and then to utilize scientific methods to assess the effects of various resource allocations in efforts to improve highway systems.

1. Research concerned with developing a transfer function should be analytically oriented, since the structure is to be analogous to the real process. In many of the areas examined during this study, the research produced correlational models with little or no regard to the physics of the process being modeled. Research should aim toward a level of conceptualization beyond raw empiricism if the advantages of systems analysis are to be realized. Implicit in this guideline is the recommendation that the variables of the research studies be conceptualized in a way that does not limit the findings to specific hardware components; i.e., the variables should be behavioral in nature (Ashby, 1958; p.1).

2. Research studies should be compatible with the system structure presented herein (or one modified and enriched by further work of a systems group). This does not imply that particular input and output measures are binding but rather that the proposed research effort be identified with one of the levels (equations) of the systems model. Proposed inputs to the research effort and stipulated outputs should be readily identifiable with appropriate systems factors (s_k, y_j, x_i, etc.). Since the nature of the systems model described in this report is descriptive, rather than prescriptive , this implies that research should be concerned with extending the state of knowledge in areas related to highway systems, rather than with examining remedial measures. The failure of past attempts to find remedies on an individual or "patchwork" basis has proved the need to work out approaches within a larger system structure.

3. Research in those areas having essential voids in knowledge should take precedence over research to improve existing model structures in other areas. Our literature review has indicated seriously inadequate knowledge in (a) accident prediction methods, (b) control strategy descriptions, (c) information describing the sequence of events from impact to injury sustainment, (d) degradation descriptions, and (e) casualty recovery models. The precedence criterion stated above, however, should only be applied to items a, b, and c since items d and e are not necessarily requisite to the operation of the initial systems analysis model. Analysis can begin by assuming no degradation, degradation of a constant average value, etc., and a binary, treatable-fatal, recovery system performance.

APPENDIX 1-A
SYSTEM EFFECTIVENESS MEASURES

Service Measures

Number of Passenger-Hours
Number of Ton-Hours
Average Concentration (vehicles/lane mile)
Average Travel Time (hours/mile)
Trip Travel Time (specified origin, route, destination)
    Average Time
    Normative Time
Average Speed (miles/hour)
Flow (vehicles/hour)
Average Number of Stops (number/mile)
Trip Number of Stops
Average Delay (hours/average free-velocity trip time)
Trip Delay (hours/free-velocity trip time)
Rate of Change in Demand (trips/day)
Capacity vs. Flow

Safety Measures

Frequency of Accidents (by type)
Frequency of Accidents per Mile of Road
Frequency of Accidents per Mile of Vehicle Travel
Frequency of Accidents per Hour of Vehicle Travel
Frequency of Injury (by level)--Passenger and Pedestrians
Frequency of Injury per Passenger Mile
Frequency of Injury per Passenger Hour
Frequency of Injury per Passenger Trip
Frequency of Injury per Vehicle Trip

APPENDIX 1-B

SUBSYSTEM PERFORMANCE MEASURES

Speed

Acceleration Capability

Minimal Stopping Distance

Panic Stopping Distance

Range (given a vehicle velocity)

Extended Trip Range (time)

Average Nonextended Trip Range (time)

Sight Distance

Headway Maintenance

Minimum Turning Radius

Brake Fade

Brake Pedal Sensitivity

Deceleration and Acceleration Weight Transfer

Turning Weight Transfer

Steady State Weight Transfer

Vertical Disturbance

Acceleration Distance (from zero velocity)

Passing Distance (acceleration distance from nonzero velocity)

Accelerator Sensitivity

Static Steering Sensitivity

Dynamic Steering Sensitivity

Vehicle Slip Angle

Turning Characteristic Index

Stability Margin

Pr(fire |accident)

Pr(destruction |accident)

Pr(injury level i |accident)

Pr(injury level i |transmitted forces = j)

Pr(transmitted forces = j |impact vector x)

Pr(impact vector x |accident)

Pr(survival |injury level i and no treatment)

Pr(survival |accident)

Recovery Time

Case Fatality Ratio

Road Serviceability

Required Maintenance Effort (vehicle, road)

Internal Comfort Index

APPENDIX 1-C

ROLE OF THE HUMAN IN THE HIGHWAY SYSTEM

I. The human as a user of the highway system (see Chapter 8).

II. The human as a vehicle operator (see Chapter 6).

   A. His performance capabilities
   B. His control strategies

III. The human as a pedestrian

IV. The human as the agent responsible for subsystem selection and maintenance

   A. The Vehicle

      1. His selection of type and model and frequency of replacement
      2. His selection of optional safety equipment
      3. His use of safety equipment in the vehicle
      4. His maintenance and inspection of the vehicle

   B. The Vehicle Operator

      1. His decision to take driver training or not
      2. His maintenance of his medical fitness for driving
      3. His responsibility for maintaining a valid license
      4. His maintenance of his driving capabilities in the short run (e.g., alcohol consumption, fatigue levels, etc.)

   C. Miscellaneous

      1. His role as a taxpayer and voter in supporting highway construction, mass transit programs, controlling legislation over drivers and automobile manufacturers, etc.
      2. His role as an insurance buyer
      3. His consumption of parking and storage facilities
      4. His role as a Good Samaritan or reporter of accidents (see Chapter 4).

V. The human as a contributing factor to accidents (see Chapter 11).

VI. The human as an accident victim (see Chapter 4).

APPENDIX 1-D

MEASURES FOR POST-ACCIDENT EVENT MODEL

Let

A = accident event
I_ℓ = ℓth level of injury
t_k = kth level of forces transmitted to occupants
x_i = ith level of impact force
θ_j = jth impact angle off the longitudinal axis of vehicle
S = survival event

Then

$$\Pr(I_\ell|A) = \sum_k \Pr(I_\ell|A \cdot t_k)\ \Pr(t_k|A) \tag{1}$$

where

$$\Pr(t_k|A) = \sum_i \sum_j \Pr(t_k|A \cdot x_i \cdot \theta_j)\ \Pr(x_i\theta_j|A) \tag{2}$$

Substituting Eq. 2 into Eq. 1:

$$\Pr(I_\ell|A) = \sum_k \Pr(I_\ell|A \cdot t_k)\ \left[\sum_i \sum_j \Pr(t_k|A \cdot x_i \cdot \theta_j)\ \Pr(x_i \cdot \theta_j|A)\right] \tag{3}$$

$$= \sum_k \Pr(I_\ell|A \cdot t_k)\ \left[\sum_i \sum_j \Pr(t_k|A \cdot x_i \cdot \theta_j)\ \Pr(x_i|\theta \cdot A)\ \Pr(\theta_j|A)\right] \tag{4}$$

Since

$$\Pr(x_i \cdot \theta_j|A) = \Pr(x_i|\theta_j \cdot A)\ \Pr(\theta_j|A)$$

The first term of Eq. 3 is obtained from an injury submodel, the second term from an impact submodel, and the third term from the accident prediction model. Recovery aspects are simply included as

$$\Pr(S|A) = \sum_\ell \Pr(S|I_\ell \cdot A)\ \Pr(I_\ell \cdot A) \tag{5}$$

REFERENCES

Arthur D. Little, Inc., An Analysis of the State of the Art of Traffic Safety. 1966.

Ashby, W R., An Introduction to Cybernetics. Chapman & Hall: London, 1958.

Blackett, P. M. S., Studies of War: Nuclear and Conventional. Hill & Wang: New York, 1962.

Bonder, S , "Introduction to Military Planning and Modeling Concepts." Topics in Military Operations Research, Engineering Summer Conference, The University of Michigan: Ann Arbor, 1966.

Bureau of Public Roads, Engineered Safety Through Research and Development (report to a Senate subcommittee on executive reorganization). May 1965.

Bureau of Public Roads, U. S. Department of Commerce, A National Program of Research and Development for Highway Transportation. July 19, 1965.

Campbell, M. E., "Highway Traffic Safety--Is It Possible?" Traffic Quarterly, Vol. 19, No. 3, 1965, p. 333.

Churchman, C. W., Ackoff, R. L , & Arnoff, E L., Introduction to Operations Research. John Wiley, 1959.

Garrett, J W., "An Evaluation of Door Lock Effectiveness: Pre-1956 vs. Post-1955 Automobiles." Automotive Crash Injury Research, Cornell University, July 1961.

Highway Safety Research Institute, Institute of Science and Technology, The University of Michigan, Final Report of the Faculty Program Advisory Committee. June 1966.

Johnson, W G., "Systems Development." Presentation to the Board of Directors, National Safety Council, October 26, 1965.

Michaels, R. M., Traffic Safety: A Challenge to Systems Engineering. Office of Research and Development, Bureau of Public Roads, U. S. Department of Commerce; address to the Washington Society of Engineers, March 3, 1965.

Quade, E. S., Analysis for Military Decisions. Report prepared for USAF project RAND, The RAND Corporation: Santa Monica, November 1964.

Stone, K. A., "Relation of Cross-Section Design and Highway Safety." Traffic Safety Research Review, Vol. 7, No. 4, pp. 2-14, 1963.

Trefethen, F. N., "A History of Operations Research." *Operations Research for Management*, McCloskey and F N. Trefethen (ed.) The Johns Hopkins Press: Baltimore, 1954.

Whitton, R. "Research in Highway Safety for the Next Decade by the Bureau of Public Roads." Remarks addressed to the American Trial Lawyers Association, Law Day Conference, May 1, 1965.

2 VEHICLE DYNAMICS

John Duncan
Bruce Gillogly

INTRODUCTION

For this project, studies in the general area of vehicle description and performance were limited to a consideration of a specific subarea, vehicle dynamics. The general purpose is to investigate the area of vehicle dynamics in relation to a possible future systems analysis of the highway transportation network. If the analysis and prediction of traffic accidents is to be considered in terms of intersecting vehicle (or vehicle and pedestrian) trajectories, or of a vehicle trajectory intersecting a stationary object, it is necessary to understand the dynamic characteristics and capabilities of automobiles and other vehicles.

The first step toward such an understanding is the determination of a set of parameters which adequately describe the dynamic performance of a vehicle. These measures serve as inputs to the operating models, particularly to the accident-prediction model.

This section hypothesizes vehicle performance measures for four basic vehicle subsystems or component groups, (1) braking, (2) accelerating, (3) steering, and (4) suspension.

The component groups may be described by the functional relationships among control settings, vehicle state (velocity, engine rpm, etc.), and environmental factors. But the specification of entire function descriptions is unduly complicated: each function must instead be described by a small set of parameter values. Some such parameter descriptions are:

The function may be considered determined by the output value at some specified input: this output value is the parametric descriptor.

The function may be determined by some small finite set of output values at specified inputs.

The function may be approximated by a linear function: the parameters (slope, matrix, intercept) of the linear function are considered parameters of the actual function.

The function may be considered as the composition of two or more separate simple (often linear) functions or as the sum or product of simple functions of subsets of the input.

In addition to the performance measures for each group, the operational function, the inputs and outputs, and the degradation processes are also listed since they give additional information about the subsystem being investigated.

For each of the four groups listed, a literature search was conducted to uncover any work done in these areas that would provide useful analytic relations for the performance measures chosen or would indicate an approach to the development of such an analytic model. The results of this literature search are summarized at the end of each section under the headings of Existing Models and Model Deficiencies. Under Existing Models are listed the basic characteristics of those models that relate to the stated performance measures, and the assumptions used to develop these models. Under Model Deficiencies are statements of the shortcomings of these models.

BRAKING

OPERATIONAL FUNCTIONS

Hydraulic, mechanical, electrical, or pneumatic braking systems transfer the control signals or forces originating with the operator into reaction forces at the base of the wheel which cause vehicle deceleration.

INPUTS/OUTPUTS

The inputs of a braking system are the operator's control signals or forces. These may be measured either in terms of his applied force or of the pedal or other control position. The system outputs are either deceleration vs. time or vehicle position vs. time. If deceleration occurs over a curved path, it may be necessary to describe the output in two dimensions.

COMPONENTS (Hydraulic Braking System)[1]

1. Brake pedal and mechanical linkage between pedal and master cylinder

---

[1] Nonhydraulic braking systems are not analyzed here.

2. Brake fluid reservoir and master cylinder (pump if power brakes)
3. Hydraulic lines
4. Wheel cylinders
5. Brake shoes and brake drum
6. Tires
7. Roadway surface

DEGRADATION PROCESSES

1. Shear or jamming of mechanical linkages
2. Loss of hydraulic pressure resulting from fluid losses
3. Wear and deterioration of moving surfaces, brake drum-shoe, hydraulic pistons and cylinders, tires and road surface, pivots and pins of mechanical linkage
4. Breakdown of lubricants
5. Loss of tire pressure
6. Contamination of brake drum-shoe interface and tire-road interface caused by a film of foreign material resulting in lower coefficients of friction

PERFORMANCE MEASURES

1. Minimal Stopping Distance: The function parameter is the stopping distance for some initial velocity using the control tactic or input which results in the shortest possible stop.

2. Panic Stopping Distance: The parameter is the stopping distance for some initial velocity when the operator applies a large but constant brake pedal force.

3. Brake Fade: The percentage increase in panic stopping distance for some initial velocity after the vehicle has been subjected to a fixed series of accelerations and decelerations. This parameter is assumed to be composable (by multiplication) with either 1 or 2.

4. Brake Pedal Sensitivity: The average rate change of steady-state deceleration with respect to the applied force of the operator (parameters established by linear approximation).

EXISTING MODELS

1. Stopping Distance. The literature search on the subject of braking dynamics has uncovered no models for the hydraulic-mechanical linkage between the brake pedal and brake drum; however, simple models were found for the dynamics of a one-wheel vehicle. These models start with a braking torque at the brake drum and describe the braking dynamics for this simple vehicle in terms of nonlinear differential equations. One such model by Chandler (1960-61) accounts for slip between the wheel and road. The model requires the assumptions of:

1. A single pneumatic tire
2. A piece-wise linear approximation to the coefficient of friction
3. No drag forces
4. A constant wheel load

The model is

$$M\dot{v} + Mgn(s) = 0 \tag{1}$$

$$I\dot{w} + MR'\dot{v} = -T_b \tag{2}$$

$$s = 1 - Rw/v \tag{3}$$

where M = total mass supported by the wheel
I = moment of inertia of the wheel
R' = the standing height of the axle
w = angular velocity of the wheel
v = instantaneous linear velocity of the moving wheel
T_b = constant braking torque
f = reaction torque due to forces between road and tire
g = acceleration due to gravity
$n(s)$ = frictional coefficient between road and tire at slip s
R = rolling radius of the tire
s = proportional slip between road and tire
t = time
$\cdot$ denotes time differentiation

The first equation states that the only external factor causing linear deceleration is the road reaction force, $Mgn(s)$. The torque equation assumes a step braking torque at the brake drum and ignores the operator response time, the response of the mechanical-hydraulic linkage, and any hydraulic pressure losses. Solutions to these equations for v(t) exist (provided n(s) is piece-wise continuous) and may be found numerically for any vehicle given the initial conditions for position and velocity.

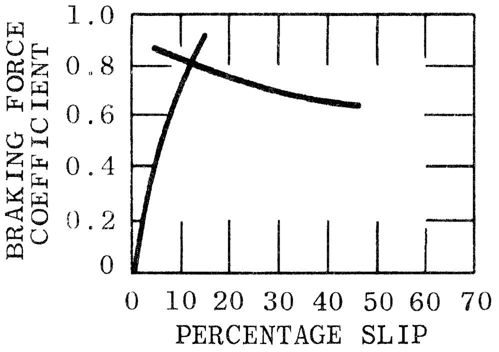
Chandler makes some progress toward an analytic solution using a piece-wise linear approximation to the wheel-road coefficient of friction. He solves for the distance from brake application in terms of the slip, s. Since the final value of slip is not generally known, the expression for stopping distance cannot be evaluated directly. No computational comparison is made between this method and conventional numerical methods for solving differential equations.

Solutions to these equations obtained by using coefficients of friction similar to those in Fig. 2-1 and typical parameters for passenger cars show that the wheel tends to lock rapidly for moderate braking torques (see Fig. 2-2). This means that Eq. 1 could be used to find approximate stopping distances for a panic stop. These equations also show that for a constant applied torque there will be an optimal value for braking torque (i.e., the torque resulting in the shortest braking distance). This torque decelerates the vehicle in such a way that the average road reaction force is maximum (for the coefficient of friction shown in Fig. 2-1 wheel slip would be maintained at a value of 0.1 for as long as possible).

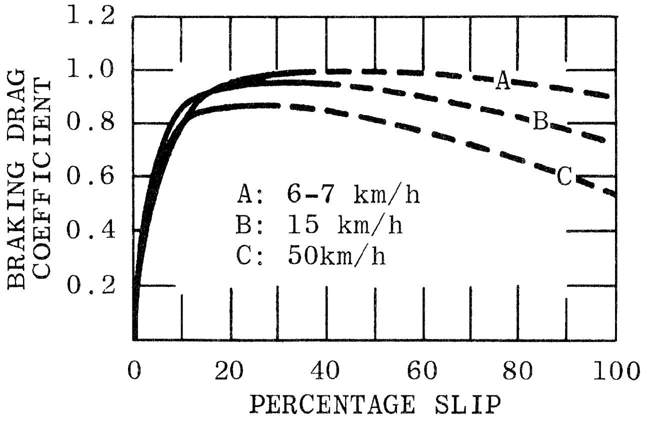
Chandler's model assumes the retarding forces come entirely from the brakes (see Eq. 1); he notes that Mgn(s) is the only retarding force present. But during an actual stop, retarding forces come from engine drag, wind drag, and rolling drag in addition to those caused by the brakes. Newcomb (1964) shows that the total of these other retarding forces has the form:

$$f_d = a + bv + cv^2 \tag{4}$$

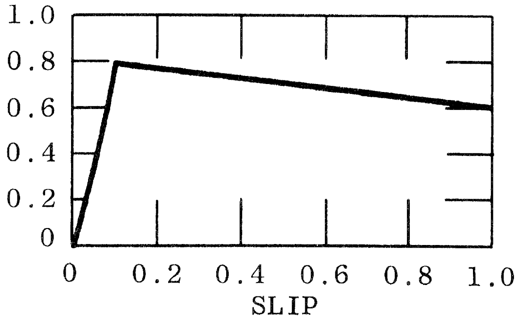
where a, b, and c are constants and v is the velocity of the vehicle. More realistic stopping distances can be computed by including this term in Eq. 1. In fact, Newcomb states from his numerical solutions that for a vehicle decelerating at 0.1 g, brakes account for only 30% of the deceleration, while the engine drag accounts for 30%, and the



a. Experimentally Determined Force-Slip Function

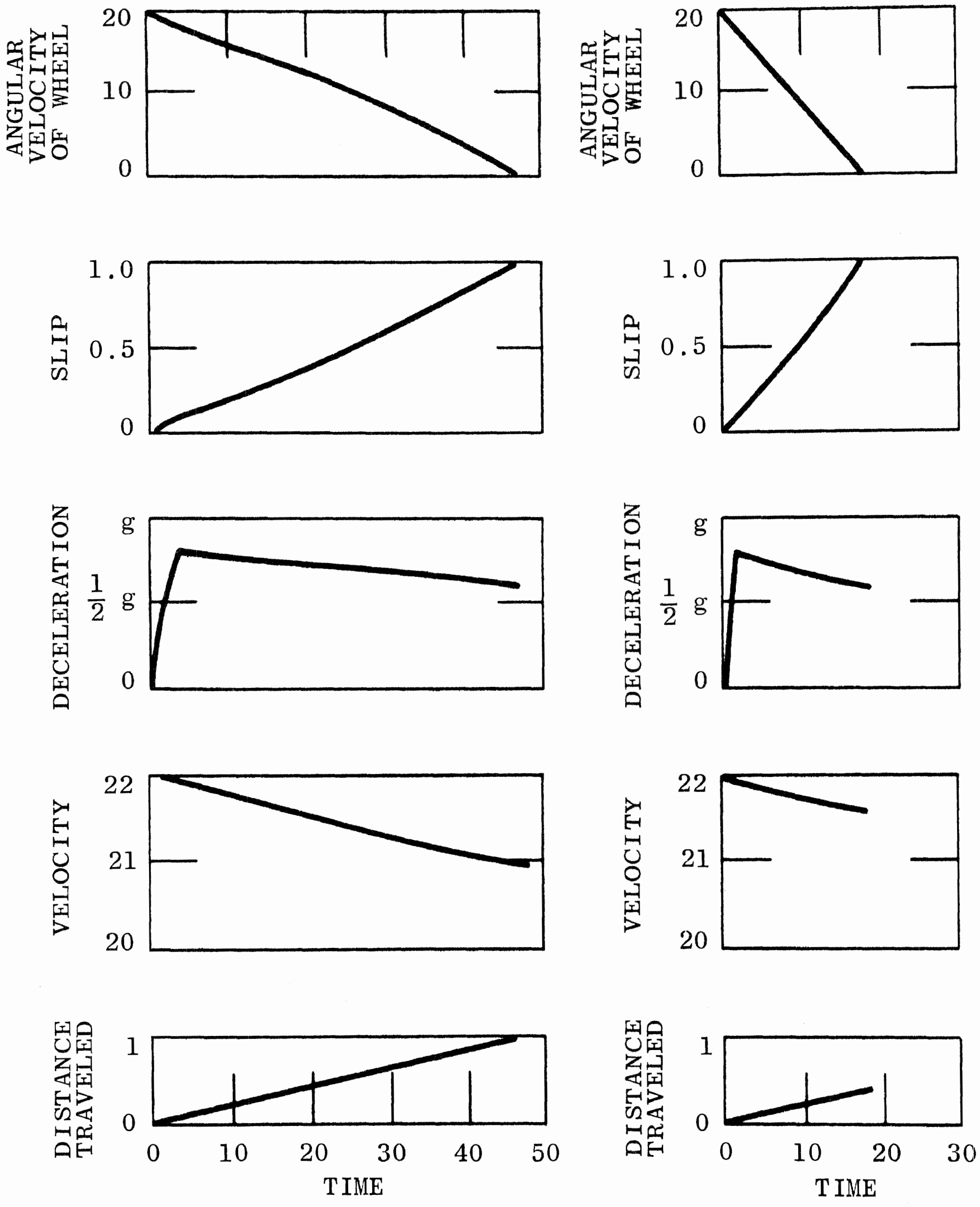


b. Experimentally Determined Force-Slip Function



c. Simplified Force-Slip Function

FIGURE 2-1. BRAKING FORCE SLIP FUNCTIONS



a. Brake-Shoe Torque Equivalent to 1.2 g Deceleration.

b. Brake-Shoe Torque Equivalent to 3 g Deceleration.

FIGURE 2-2. GRAPHS FOR A LOADED WHEEL FROM 15 MI/HR rolling drag and wind forces account for 40%. However, for a deceleration rate of 0.5 g, which is more nearly a panic stop, these percentages are 86%, 6%, and 8% respectively. From these examples it may be concluded that panic stopping distances may be in error approximately 10% if the additional drag forces are not taken into account.

2. Brake Fade. Newcomb (1960) discusses the relation between the work done by the braking system during a stop and the final temperature of the brake shoe-drum. He relates the work of a series of stops to final temperature by considering the geometry and specific heat of the drum-shoe and cooling for an equivalent infinite slab. Using his model to determine final or steady-state temperatures of the drum and tire after a series of stops, it should be possible to find the maximum braking torque. The model is

$$T_{b_{max}} = Rn(v, T)F_{h_{max}}$$

where R = effective radius of the brake shoe

$n(v, T)$ = the coefficient of friction between the drum and shoe for slip velocity, v, and temperature, T

$F_{h_{max}}$ = maximum forces of the hydraulic system

A simplified model such as this could be used with Chandler's equations to find the panic stopping distance vs. brake drum temperature or for any series of stops.

MODEL DEFICIENCIES

The models uncovered thus far treat the kinematics of simple vehicles only; such models do not account for:

1. Weight transfers between wheels
2. The dynamic characteristics of the linkages which transfer brake pedal forces to brake drum torques
3. Centrifugal forces, wind gusts, and any other external forces which might act on the vehicle
4. Forces not applied through the center of mass of the vehicle

For accurate prediction of vehicle trajectories, a knowledge of weight transfers is essential since the frictional forces at each wheel are proportional to instantaneous loads. Hence when braking results in an equal wheel loading, the wheels will lock at different times, causing the vehicle to turn. In addition, the total deceleration reaction time, i.e., the sum for all the wheels, is not the same as that for a one-wheel vehicle with the same mass. Hence, the stopping distances for the two models are different.

The lack of hydraulic system models presents a serious problem in the area of reliability since degradation of this system component is thought to affect stopping distances greatly.

ACCELERATION

OPERATION FUNCTION

Vehicle acceleration systems supply the energy required to accelerate the vehicle or maintain its velocity against drag forces.

INPUT/OUTPUT

The input to the acceleration system is either the force applied to the accelerator pedal, or the position of the accelerator pedal. The output of the acceleration system is the vehicle acceleration vs. time or the vehicle position vs. time. It could also be the acceleration force vs. time.

COMPONENTS

1. Engine
2. Transmission
3. Drive train
4. Tires

DEGRADATION PROCESSES

1. Shear or jamming of mechanical (solid or fluid) linkages
2. Breakdown of lubricant
3. Loss of tire pressure
4. Wear and deterioration of moving surfaces
5. Chemical degradation of fuel
6. Chemical or mechanical degradation of ignition mechanism (spark plugs)
7. Contamination of the tire and road-surface interface.

PERFORMANCE MEASURES

1. Acceleration Distance: The distance required to reach a given velocity from a standing start with fixed throttle position.

2. Passing Distance: The distance required to reach a given velocity from some initial velocity with fixed throttle position.

3. Sensitivity: Average rate of change of acceleration with respect to applied force.

EXISTING MODELS

A search of the available literature on vehicle dynamics failed to disclose any useful models for vehicle acceleration. But it can be seen from the braking model developed by Chandler (1960-61) that the difference between accelerating and braking as far as the equations of motion are concerned is only a change in sign of the term b_o. Of course this model only takes into account the effect of the change in coefficient of friction with tire slip and is subject to all the shortcomings that are mentioned in the braking section. But when this model is augmented as suggested in the braking section, it offers the possibility of becoming a useful model.

MODEL DEFICIENCIES

This model has many deficiencies, and, in addition to those mentioned in the braking section, the following are of importance for its use as an acceleration model. The Chandler model assumes a constant braking torque and, if used as an acceleration model, would imply a constant accelerating torque when in actuality the accelerating torque varies with engine speed and transmission ratio. It is possible to use a piece-wise linear approximation of torque during acceleration, but this would require a very complex procedure for determining the output of the model in a continuous manner.

An additional problem is that no available models give the relationship between the hardware characteristics of an engine and the torque output of the engine. Many data are available on specific engines and such a model would not be difficult to construct, but none exist presently. It is obvious that while the Chandler model cannot be simply modified to apply for acceleration, it is all that is presently available and can be used until a more applicable model is developed.

STEERING

OPERATION FUNCTION

The operator applies force to a steering control device, which is linked to the wheels. The reaction of the pavement with the tires creates lateral forces which turn the vehicle axis horizontally.

COMPONENTS

1. Steering wheel
2. Steering gearbox
3. Steering linkages
4. Tires

DEGRADATION PROCESSES

1. Shear or jamming of mechanical linkage
2. Breakdown of lubricant
3. Loss of tire pressure
4. Wear and deterioration of moving surfaces
5. Contamination of tire-roadway interface

PERFORMANCE MEASURES

1. Static steering sensitivity's the force required to turn the wheels and is given by

$$\frac{\partial B}{\partial F}\left(\frac{\text{change in position}}{\text{change in force}}\right)$$

2. Dynamic steering sensitivity: same as above, but higher speed.

3. Vehicle slip angle: the angle between vehicle path and vehicle axis for a certain radius turn at a certain velocity

4. Curvature response: this index describes in part the dynamic characteristics of a vehicle while turning (especially its stability).

$$\frac{\partial d}{\partial v}R = \text{constant}$$

where d = front wheel angle
v = velocity
R = turn radius of curvature

5. Stability margin is given by

$$1 - \frac{V^2}{ugR(\ell_f \text{ or } \ell_r)}$$

where V = vehicle velocity

u = coefficient of friction

R = radius of turn

ℓ_f or ℓ_r = front or rear wheel thrust factor

EXISTING MODELS

The analysis of vehicle steering behavior has progressed steadily since the initial work of Olley (1957) and Segel (1956-57). Recent work has been done by Radt and Milliken (1960), Goland and Jindra (1961), Ellis (1963), and Hoffmann (1964). Of these, the work of Hoffmann is of particular interest, since his work is part of a larger program studying driver-vehicle-highway interactions being undertaken by Melbourne University.

Hoffmann starts with the basic equations of motion and applies them first to a simple linear model that does not allow skidding and then to a more sophisticated nonlinear model that takes tire skidding and breakaway into consideration. Of interest here is the second or nonlinear model since it is a more realistic representation of the actual behavior of rubber-wheeled vehicles.

The equations of motion are:

$$\frac{MV^2}{2R} = F_f + F_r \quad (5)$$

$$aF_f = bF_r \quad (6)$$

where M = total mass of the vehicle

V = velocity of the vehicle

R = radius of the turn

F_f = side force at front tires

F_r = side force at rear tires

a = distance from front wheels to vehicle center of gravity

b = distance from rear wheels to vehicle center of gravity

Solving these equations for the steady-state case, Hoffman develops the following relations between front wheel angle, vehicle slip angle, and radius of curvature:

$$d = \left(\frac{a + b}{R}\right) + \frac{uM}{2(a + b)}\left[\frac{\ell_r a}{k_r C_r} \log \left(1 - \frac{V^2}{\iota_r rgR}\right) - \frac{\ell_f b}{k_f C_f} \log \left(1 - \frac{V^2}{\ell_f ugR}\right)\right] \tag{7}$$

$$B = \frac{b}{R} + \frac{\ell_r uaM}{2k_r C_r (a + b)} \log \left(1 - \frac{V^2}{\ell_r ugR}\right) \tag{8}$$

where d = front wheel angle

u = coefficient of friction

ℓ_r, ℓ_f = rear or front wheel thrust factor

k_r, k_f = rear or front tire elastic constant

C_r, C_f = rear or front wheel cornering stiffness

g = acceleration due to gravity

B = vehicle slip angle

These two equations (7 and 8) are the source of three performance measures of the steering subsystem. First, the equations themselves give the input/output relationship between front wheel angle or steering wheel position as the input and vehicle slip angle and radius of curvature as the output. Second, the partial derivative of Eq. 7 with respect to the velocity gives the following expression for the curvature response of the vehicle:

$$\frac{\partial d}{\partial V} = \frac{uM}{2(a + b)}\left[\frac{2\ell_f bV}{k_f C_f\left(\ell_f ugR - V^2\right)} - \frac{2\ell_r aV}{k_r C_r\left(\ell_r ugR - V^2\right)}\right] \tag{9}$$

Third, it can be seen from Eq. 7 that a zero for either of the log terms will indicate that the vehicle has gone into a skid. Hence, the value of the smaller of these two terms is the vehicle steering stability margin:

$$S_m = 1 - \frac{V^2}{ugR(\ell_f \text{ or } \ell_r)} \tag{10}$$

Additional information is available from these performance measures as they are applied to vehicles. Any one of these three will give an indication of the state of the vehicle in terms of its "oversteer" or "understeer" characteristics. In addition, the first set of equations allows a determination of the vehicle position or attitude with respect to the actual path of the vehicle. The second measure is of value in determining whether a vehicle will undergo a sudden change in steering behavior as its velocity changes, and the third provides an indication of the velocity at which a skid will occur.

MODEL DEFICIENCIES

The model developed by Hoffmann considers only the effects of mass, inertia, and tires. It does not include the effects of weight transfers which can be caused by a number of factors, including the independent front suspension, rear roll steer, body roll, wind gusts and loads, uneven road surface, and nonsymmetrically loaded vehicles. The model also assumes that there is no loss of contact between the tire and the road and that the coefficient of friction is a constant. The basic equations do not take into account the forces produced by grade or superelevation, or the forces caused by wind loads.

It is apparent that the factors not included are of considerable importance when skidding is imminent. At this point weight transfer causes unequal wheel loading and the wheels do not break away in the uniform manner predicted by the model. Further, the changes in coefficient of friction and wheel loads could combine in such a way that the vehicle would react in a manner opposite to that indicated by the model. This is particularly likely on steep grades where the weight transfer would be very large, causing the vehicle to break away rear-end first rather than front-end first as the model would predict.

OPERATION FUNCTION

Vehicle suspension systems absorb road shock, distribute dynamic reaction forces so that wheel loading is approximately uniform, and damp undesirable motion.

SUSPENSION

INPUTS/OUTPUTS

The input to a suspension system is the distribution of road elevations at each wheel and the forces determined by steering, braking, and accelerating activities. The road input can be conceptually measured in terms of a vertical distance vs. position on the road. The output of a suspension system is the wheel loading, the vehicle motion in the vertical direction, and the angular motion about the center of gravity. This output may be measured in terms of the vertical position of the vehicle's center of gravity, and pitch, roll, and yaw angles, vs. time.

COMPONENTS

1. Wheels
2. Springs or torsion bars
3. Shock absorbers
4. Vehicle body

DEGRADATION PROCESSES

1. Shear of springs
2. Tire blow-outs
3. Fatigue of springs resulting in variations in spring constants
4. Wear of shock absorbers resulting in lower damping coefficients
5. Deterioration of the road resulting in a change of its power spectrum (cracks, ruts, patches, deformation of foundation)

PERFORMANCE MEASURES

1. Deceleration and acceleration weight transfer: Maximum percentage of weight transfer from rear to front wheels during panic stop or maximum acceleration.

2. Turning weight transfer: Maximum percentage of weight transfer from one side to the other during a constant-radius, constant-velocity turn.

3. Steady-state weight transfer: The peak or RMS (root-mean-square average) weight transfer about the pitch or roll axis resulting from the surface irregularities of some "standard" road surface.

4. Vertical disturbance: The peak or RMS vertical movement of the center of gravity resulting from the surface irregularities of a "standard" road.

EXISTING MODELS

The total vehicle-driver suspension system can be modelled in terms of spring-mass systems to determine motion relative to vehicle trajectory. Roughly the system consists of four wheels, the vehicle body, and the driver. A comprehensive model might allow 6 degrees of freedom for the body (3 in translation and 3 in rotation); 2 for each wheel, assuming them independently sprung (1 for vertical translation and 1 for rotation about the vehicle roll axis); and 1 for the vertical driver motion. However, as shown by M. Mitschke (1961), a simpler system gives results which agree well with experimental data. This model, shown schematically in Fig. 2-3, will allow us to relate the listed performance measures to vehicle design and "standard roads." Writing the usual force and torque equations leads to a system of linear differential equations where the dependent variables are the vertical positions of the front wheels (assumed to be tied together), the rear wheels (also assumed tied), the body, the driver, and the pitch angle of the body. The nonhomogeneous part of these equations, or "the forcing functions," are the road surface perturbations in the vertical direction and the reaction torques of the road from acceleration and deceleration. Through the Laplace or Fourier transformations, the system of differential equations becomes a system of algebraic relations which can be solved for the independent variables in terms of the applied forcing functions. For example,

$$\theta(w) = H_1(w)D(w) + H_2(w)T_a(w) \qquad (11)$$

where θ is the pitch angle of the vehicle

$H_1(w)$ and $H_2(w)$ are transfer functions for the differential equations

$D(w)$ is the Fourier transform of the road surface

$T_a(w)$ is the effective reaction torque resulting from changes in vehicle velocity

w is the Fourier transform variable

Similar equations can be written for vertical motions. For any given road configuration $D(w)$ and vehicle acceleration $T_a(w)$, the pitch motion $\theta(t)$ or other vertical motions can be found numerically, if not analytically, using such relations.

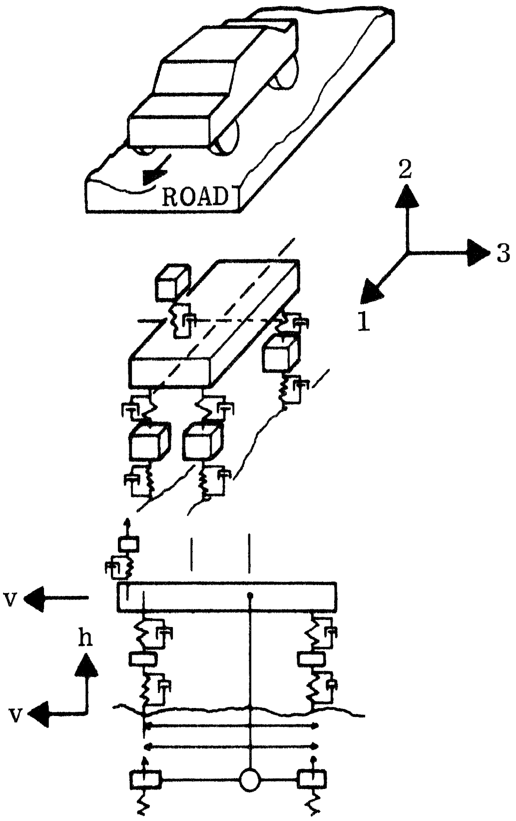


FIGURE 2-3. EQUIVALENT VIBRATION DIAGRAM OF A VEHICLE

From the computed motions, the front- and rear-wheel loading and the accelerations of the occupant can be computed.

Weight Transfers for Panic Stops and Maximum Acceleration. During accelerations the road forces do not act through the center of gravity and result in pitch motions or dynamic weight transfers. If during panic braking sufficient weight is transferred to the front wheel, the car will be unstable (i.e., it will spin out of the driver's control). According to Lister (no date given) this phenomenon occurs in a large number of accidents. Spin during panic braking is a highly irregular phenomenon as illustrated by Fig. 2-4. The tendency for rear brake lock depends upon the geometry of the vehicle, the road-wheel coefficient of friction, and the front-wheel weight transfer; therefore acceleration weight transfer (AWT), as a single performance measure, will not completely predict braking spins. It will, however, indicate such tendencies.

To calculate AWT from the diagram, one must find the acceleration reaction torque, T_a. The equations of motion previously described do not include the linear vehicle motion which is coupled to this torque. In any case, a first approximation would be

$$T_a(t) = Mh\ddot{X}(t)$$

Where h is the static height of the vehicle's center of gravity. Assuming a constant deceleration and a smooth road, one could find

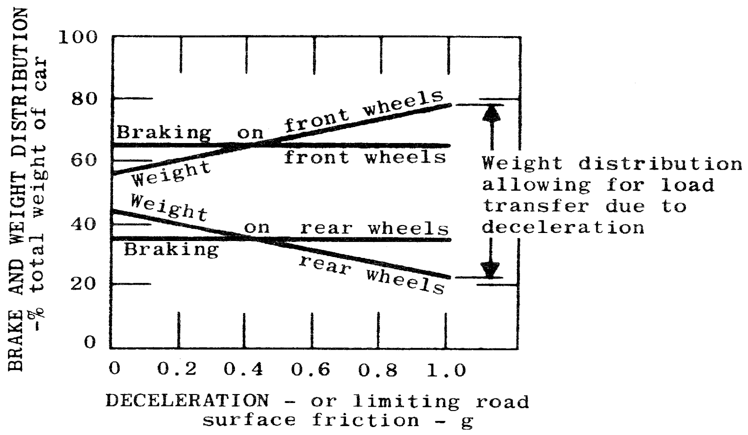
$$\theta(t) = MhA \int_o^t H_2(t - r)\ dr$$

where A is the amplitude of deceleration and $H_2(t)$ is the impulse torque response (i.e., the inverse Laplace transform of $H_2(s)$). From $\theta(\infty)$ the compression of the front wheel springs and the force or weight transfer to the front wheels are calculated.

Random Roads. The calculation of steady-state weight transfers and vertical disturbances from road irregularities requires an approach slightly different from AWT computations. This is because the forcing functions of the road are random. The approach of some workers (Bussman, 1964) has been to assume vertical perturbations for the road which are wide-sense stationary stochastic processes. This

Accidents where Loss of Control Occurred
(% of all accidents in which a car was involved).

Road conditions	Dry Roads No.	Dry Roads %	Wet Roads No.	Wet Roads %	Dry Roads No.	Dry Roads %	All Conditions No.	All Conditions %
Roads restricted to 30 mph ("slow")	116	10	22	17	5	63	53	13
Loads restricted to 40 mph or no limit ("fast")	87	21	113	35	28	68	331	29
London-Birmingham Motorway	66	64	45	69	20	100	131	70



Passenger car, calculated brake and weight distribution (height of cg. 23.5 in.: wheelbase, 108.5 in.)

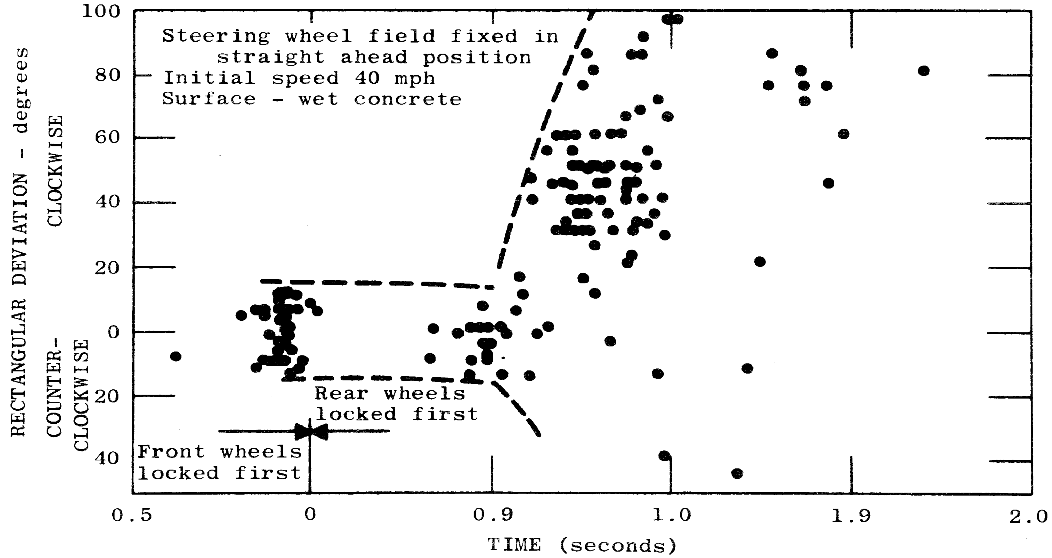


FIGURE 2-4. DEVIATION OF A CAR WHEN ALL FOUR WHEELS DID NOT LOCK AT THE SAME INSTANT (Time reckoned from instant rear wheel locked to instant when second front wheel locked)

assumption allows computation of the autocorrelation function for $\theta(t)$ using

$$R_\theta(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(\tau - t)h_1(u)h_1(u + t)\, du\, dt$$

where R(t) is the autocorrelation function for the road. By the assumption of stationarity, the RMS pitch motions are found from R(0). In a similar fashion the RMS vehicle weight transfer and the vertical motions or accelerations of the driver or vehicle can be computed.

Computations along these lines have been made by Mitschke. The power spectra used in these computations were measured by Walls et al. (undated).

Peak Motions and Acceleration. To find peak values for accelerations and motion requires assumption of a Gaussian stochastic process. With this assumption, the vehicle motions are also Gaussian and hence a distribution function for the difference in initial and final vehicle positions for the time interval, say t_1 to t_2, can be computed. From the distribution function the probabilities of peak motion can be found.

MODEL DEFICIENCIES

A. Few measurements of road power spectra are available.

B. Extreme road surface irregularities such as "chuck holes" cause some vehicles to respond according to nonlinear equations because the wheel stops limit vertical wheel travel. Linear models will not predict these events.

C. The reviewed models do not account for air lift.

D. The models do not include linear vehicle motion and therefore weight transfer computations require an assumed accelerating reaction torque.

E. No models were found for vehicle roll motion although models similar to those used here for pitch motion will probably work.

G. Most existing models do not treat the cross correlation between inputs from different wheels. This not only complicates the model, but presents the additional difficulty of measuring the cross correlation functions for roads.

CONCLUSION

Vehicle dynamics are related to traffic safety in our study through the use of subsystem performance measures. Prediction of vehicle dynamics permits estimation of these measures which in turn allows prediction of collision frequencies. The listed performance measures are important descriptors of vehicle capability; however, their importance to accident prediction and analysis remains to be established.

Once a comprehensive, useful set of performance measures is established, then models relating these performance measures to hardware design and operation must be derived and evaluated if the effects of design and operational modifications are to be evaluated in terms of collision frequency. For the components listed, no such models were found for the following performance measures.

1. Brake pedal sensitivity
2. Static steering sensitivity
3. Dynamic steering sensitivity
4. Turning weight transfer
5. Steady-state weight transfer (pitch)
6. Acceleration sensitivity

Even for those models which have been postulated, it was possible to find little empirical verification.

The greatest shortcoming in the area of modeling vehicle dynamics is the lack of a single composite model which relates the vehicle inputs (steering, braking, and acceleration controls and environmental forces), design parameters, and environmental conditions to vehicle trajectories. Such a model could be useful for:

1. Simulating precollision events
2. Reducing the cost of obtaining data on vehicle dynamics
3. Relating performance measures to hardware design

Those approximate models found for describing subsystem performance hardly suffice for trajectory simulation since they predict vehicle dynamics only for special circumstances. These models most likely could not be used for predicting vehicle response to multiple inputs to the vehicle since the equations of motion are nonlinear.

Therefore, a summary of this section can be expressed in the following two recommendations:

1. Find a comprehensive set of performance measures for vehicle dynamics description which can serve as inputs to the accident prediction model.

2. Develop a single model for vehicle dynamics that includes all types of inputs (control, environment, and design).

REFERENCES

Bussman, D., Vibration of a Multi-Wheeled Vehicle, Ohio State University, Technical Report RF 573-64-1, 30 August 1964.

Chandler, K., "Theoretical Studies in Braking," Proc. Inst. Mech. Eng., 1960-61, No. 4, pp. 147-161.

Ellis, J., "Understeer and Oversteer," Automobile Engineer, May 1963.

Goland, M., and Jindra, F., "Car Handling Characteristics," Automobile Engineer, August 1961.

Hoffmann, E., "Notes on Vehicle Stability," Australian Road Research, March 1964.

Lister, R., "Retention of Directional Control When Braking."

Mitschke, M., "Influence of Road Vehicle Dimensions on the Amplitude of Body Motions and Dynamic Wheel Loads," SAE 310C, January 9-13, 1961.

Newcomb, T., "Thermal Aspects of Vehicle Braking," Automobile Engineer, July 1960, Vol. 50, p. 288.

Newcomb, T., "Work Done by Brakes," Automobile Engineer, March 1964.

Olley, M., "Road Manners of the Modern Car," Inst. Auto. Eng., 1957.

Radt, H. S., Jr., and Milliken, W. F., Jr., "Exactly What Happens When an Automobile Skids?" J. Soc. Auto. Eng., December 1960.

Segel, L., "The Response of the Automobile to Steering Control," Proc. Automotive Div., Inst. Mech. Eng., No. 7, 1956-57.

Walls, J., Hocbolt, J., and Press, H., "Some Measurements and Power Spectra of Runway Roughness," National Advisory Committee for Aeronautics, Technical Note 3305.

3 THE HIGHWAY

DONALD E. CLEVELAND
WAYNE VAN WAGONER

INTRODUCTION

In the system structure of highway operations, the environment within which the vehicle operates has been subdivided into four separate areas, highway geometry, highway materials, traffic control devices, and the dynamic environment. Highway Geometry concerns the geometry of the environment, i.e., the spatial relationships that describe the relative location and elevation of the road and adjacent surfaces, obstacles, and other elements significant in the geometry of motion and of visibility from vehicles. Highway Materials concerns the properties of materials in the environment, particularly their strength and visibility properties; within this section the microscopic geometry significant in vehicular control and deformation studies is also treated. Traffic Control Devices concerns the message characteristics of traffic control devices. Finally, those aspects of the environment that are dynamic in nature are treated separately in Chapter 9.

HIGHWAY GEOMETRY

Before an analysis of auto behavior can be made, the geometrics of the highway and the surrounding environment must be defined. Thus, an effort has been made to define the highway in purely geometric terms. No consideration has been given to the vehicle or materials involved. Also no use of current design standards or trends are considered since these considerations are dependent on the vehicle and materials as well as geometrics. However, a list of current design standards is included at the end of each section for convenience.

It is readily apparent that there is a great variation in the geometry of different highways. As the service level of a highway rises, geometric standards become more stringent. Thus, the general types of highways can generally be distinguished by their general geometric characteristics. Also, between highways of the same class, each facility has its own exclusive geometry. Thus each highway can be uniquely defined. This will allow the vehicle to be placed in any type of existing or predicted environment with highway geometrics strictly defined.

The form of the geometric definitions must be such that when combined with material, driver, and vehicle inputs, the resulting model is meaningful. The geometric definitions must be such that when a vehicle is placed on this geometry, its movement will be measured by an interaction of geometric and vehicle characteristics.

It is believed that the best way to define highway geometrics is first to locate the centerline with respect to a basic x, y, z coordinate system. Thus the beginning of a highway might be considered the origin (0,0,0). This base coordinate system should be oriented with the x-axis in a west-east direction and the y-axis in a south-north direction. This convention will facilitate the location of coordinate points since tangents are given in bearing notation.

After the location of a point on the centerline has been defined in the basic coordinate system, a secondary coordinate system (u, v, w) is defined at that point. The u-axis of this secondary system should be tangent to the plane of the centerline, and contained in the x-y plane. This allows the w-axis to be perpendicular to the horizontal x-y plane, and parallel to the z-axis at all points. Thus, cross-sectional elements can be defined in a continuous way along the highway centerline.

Even though the two coordinate systems can be related to each other at any point, it is felt that a more satisfactory approach would be to keep them as independent systems. Since the cross-sectional elements are defined by a secondary coordinate system for each point of the centerline, a total description of any portion of the highway can be obtained by examining the elements of the secondary coordinate system between (x_1, y_1, z_1) and (x_2, y_2, z_2). This will yield a total description of the highway geometrics.

I. GEOMETRIC VARIABLES

A. Longitudinal Element Definitions

1. Tangent
    a. In plan: straight section of highway
    b. In profile: section of highway with constant grade

2. Grade--longitudinal slope of highway expressed as a percent of grade

   Percent of grade (G) = $\frac{\text{feet of change in elevation}}{\text{100 feet}} \times 100$

   (Sign convention: + = rising slope in direction of study)

3. Circular Curve

   a. Degree of curve--central angle subtended by an arc 100 feet in length (D)

      Possible range:

      $0 \leqq D$

      Design practices:[1]

Speed (mph)	D_{Max}
30	25.0°
40	13.5°
50	8.5°
60	5.5°
70	4.0°

   b. Change in direction--angle between extended incoming tangent and outgoing tangent in direction of curve (I)

      Possible range:

      $0 \leqq I$

   c. Length* = $\frac{I}{D} \times 100$

4. Transition--curve linking a tangent and a circular curve, which obeys the following mathematical requirement: The radius of the spiral at any point is inversely proportional to its length.

   From this mathematical requirement the following statement can be made: The degree of curve of the spiral increases at a uniform rate from zero at its connection with the tangent to the degree D of the circular curve to which it is connected.

   The constant rate of increase in D per station along the spiral is represented by k.

   $k = \frac{100D}{L_s}$ where L_s = length of spiral in ft.

   a. Length--distance needed to go from $\theta = 0$ to $\theta = D$

---

[1] American Association of State Highway Officials, A Policy on Geometric Design of Rural Highways (1965), p. 170, Table III-9.

*All lengths are projections on the x-y plane.

b. Design practices[2]

Speed (mph)	Min. Length of Spiral (ft)
30	100
40	125
50	150
60	175
65	190
70	200
75	220
80	240

5. Vertical Curve--curve linking two grade tangents; generally a portion of a parabola. The slope at the ends of the vertical curve are identical to the connecting tangents.

   a. d^2y/dx^2 = constant

   b. Possible range $g_1 = g_2$, $g_1 = \infty$, $g_2 = -\infty$

      $0 \leqq dy/dx$ where g is grade

   c. Existing range

      $$L = \sqrt{\frac{2V}{R}}$$

      L = length in feet of parabola

      V = speed in mph

      $$A = R = \frac{G_2 - G_1}{L}$$ = rate of change of grade per 100 ft. station.

B. Definition of Cross-Sectional Elements

1. Rate of Superelevation--Cross-sectional change in elevation per foot of horizontal width of pavement.

2. Pavement--That portion of the highway provided with a wearing surface for vehicle travel.

3. Roadway--That portion of a road which is improved, designed, as ordinarily intended for vehicular use.[3]

4. Shoulder--The portion of the roadway contiguous with the pavement for accommodation of stopped vehicles, for emergency use, and for lateral support of base and surface courses.[4]

5. Lane--Pavement width allocated to one vehicle.

---

[2] Ibid, p. 176, Table III-11.

[3] Highway Capacity Manual (1965), Highway Research Board, Special Report 87, p. 9.

[4] American Association of State Highway Officials, A Policy on Geometric Design of Rural Highways, 1965, p. 234.

6. Right of Way--A general term denoting land, property, or interest therein, usually in a strip, acquired for or devoted to a highway.[5]
7. Median--The portion of a divided highway separating the pavement for traffic in opposite directions.[6]
8. Drainage Ditch--A portion of the roadside shaped to carry off water.
9. Side Slope--A portion of the roadside connecting the shoulder edge with drainage ditch edge or natural topography.
10. Lateral Obstruction--Discontinuity or a break in a smooth surface.
11. Crown--The actual perpendicular distance from the centerline of the pavement to a line connecting the outer edges of the pavement.
12. Roadside--A general term denoting the area adjoining the outer edge of the roadway.[7]

II. HIGHWAY GEOMETRY DESCRIBED BY CONIC SECTIONS

The geometric characteristics of the horizontal and vertical alignment of the roadway centerline can readily be defined by the plan and profile in the basic coordinate system. These characteristics include circular curves, spirals, tangents, and vertical curves. The remaining elements are of a cross-sectional nature. Therefore, they are defined in the secondary coordinate system.

In evaluating the cross-sectional elements, it would be a great convenience to be able to define all these elements in a standard way. It is felt that this can be accomplished by considering all cross-sectional elements as portions of conic surfaces. By adopting this convention, the general conic equation:

$$Ay^2 + Byz + Cz^2 + Dy + Ez + F = 0$$

applies to all cross-sectional elements. Each portion of the cross-section can be defined by selecting five points within the section and solving the resulting simultaneous equations to obtain the defining equation. Of course, the number of equations obtained for any given cross-section will be a function of the number of discontinuities in that cross-section. In other words, the discontinuity points would act as the upper and lower boundary points for each defining conic section.

The major drawback of this method is that five points are needed to define the equation of the cross-sectional element in question. If it were assumed, for example, that all cross-sectional elements were defined as a portion of a parabola, the number of points necessary for definition would be reduced. The general conic equation gives more accurate results than the parabola. Also obtaining five reference points in any particular section should pose no difficulty.

---

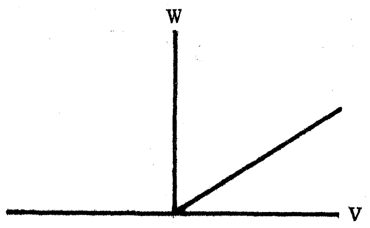
[5] American Association of State Highway Officials, <u>A Policy on Geometric Design of Rural Highways</u> (1954), p. 639.

[6] Ibid, p. 633.

[7] Ibid, p. 634.

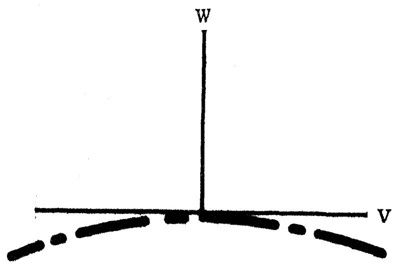
Following is a list of the cross-section structural elements with possible conic defining equations.

1. Superelevation



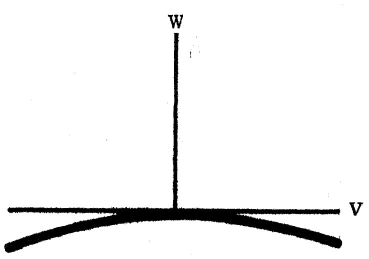
$$Av^2 + Bwr + Cw^2 + Dv + Ew + F = 0$$

2. Shoulder



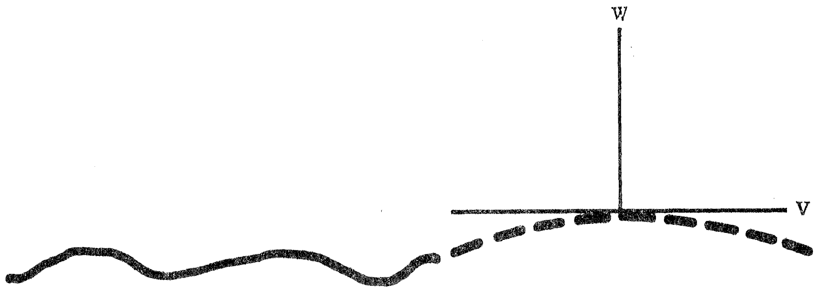
$$Av^2 + Bwr + Cw^2 + Dv + Ew + F = 0$$

3. Crown



$$Av^2 + Bwv + Cw^2 + Dv + Ew + F = 0$$

4. Outward Slope



$$Av^2 + Bwv + Cw^2 + Dv + Ew + F = 0$$

5. Lateral Obstructions

a. Can be defined by conic equation:

$$A'v^2 + Bwv' + Cw^2 + Dv' + Ew + F = 0$$

b. Conic equation will define the shape, size and volume when taken over a distance du.

(1) A sign, for example, is considered as very thin (less than some finite thickness E) and would be reported as only at one point unless it was not parallel to the v' axis.

(2) A tree, on the other hand, when taken with respect to several small differential elements in the u' direction, will produce the nature of the solid in size, shape, and volume.

It should be apparent that a complete geometrical cross-section definition can be made in terms of the summation of consecutive conic equations from discontinuity to discontinuity. When these cross-sectional elements are summed along the centerline, the three-dimensional geometric characteristics of the highway are wholly defined by conic surfaces.

III. PERFORMANCE CHARACTERISTICS OF INDIVIDUAL ELEMENTS

A. Cross-Sectional Elevation

1. Superelevation
2. Crown of Pavement
3. Cross-Sectional Change in Shoulder Elevation
4. Cross-Sectional Change in Side Slope Elevation
5. Cross-Sectional Change in Roadside Elevation

B. Cross-Sectional Width
    1. Width of Pavement*
    2. Width of Lane*
    3. Width of Shoulder*
    4. Width of Roadway*
    5. Right-of-Way Width*
    6. Median Width*
    7. Side Slope Width*
    8. Drainage Ditch Width*

C. Lateral Obstructions (discontinuities)
    1. Regulatory, Traffic Control, and Highway Information
        a. Small signs (supported on one post)
        b. Large signs (supported on one or more posts)
        c. Signals
        d. Channelization islands--devices for separating linear flows of traffic
        e. Billboards
    2. Highway Service and Non-Serving Structures
        a. Bridges
        b. Bridge abutments
        c. Bridge piers
        d. Guard rail
        e. Light poles
        f. Drainage structures (head walls, culverts, erosion control devices)
        g. Utility poles and lines
        h. Other transportation type structures (railroad crossings at grade, airport runway lights)
    3. Landscaping
        a. Trees
        b. Bushes
        c. Rocks and boulders

D. Longitudinal Elements: Intersections of joining or crossing streams of traffic
    1. Number of Roads Intersecting at One Point
    2. Angle of Attack

---

*Possible range in non-negative.

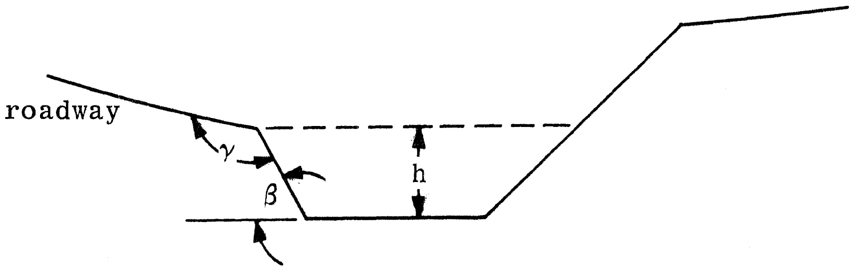
IV. GEOMETRIC PERFORMANCE VARIABLES FOR A SECTION OF HIGHWAY

A. Longitudinal Elements
   1. Tangent: Horizontal distribution of tangent azimuth
   2. Grade
      a. Total + and - grade
      b. Percent of total highway in grade (>.5% or <-.5%)
      c. Distribution of grade
   3. Circular Curve
      a. Total change in direction/unit length
      b. Number of reversals or changes from right to left or left to right/unit length
      c. Total number of curves/unit length
      d. Percent of section that is curved
      e. Distribution of distances between curve centers
      f. Distribution of central angles of curve
      g. Distribution of degree of curve
   4. Transition Curves
      a. Percent of circular curves with transition curves
      b. Distribution by maximum D (change in D) per 100 foot station
      c. Percent of section length in transition curve
      d. Distribution of k
   5. Vertical Curve
      a. Total number of vertical curves
      b. Total number of crests (+ to -) and sags (- to +) curves
      c. Percent of length in vertical curvature
      d. Distribution of distances between centers of vertical curves
      e. Distribution of curves by change in grade per 100-foot station

B. Width Elements
   1. Pavement Width
      a. Total change in width
      b. Total number of changes in width per unit length
      c. Distribution of widths
   2. Width of Shoulder
      a. Total change in width
      b. Total number of changes in width per unit length
      c. Distribution of widths

3. Median Width
    a. Total change in width
    b. Total number of changes in width per unit length
    c. Distribution of widths
4. Width of Roadway
    a. Total change in width
    b. Total number of changes in width per unit length
    c. Distribution of widths
5. Width of Lane
    a. Total change in width
    b. Total increase in width/unit length
    c. Total increase in width/unit length
    d. Total number of changes in width per unit length
    e. Distribution of lane widths
6. Right of Way Widths
    a. Total change in width
    b. Distribution of widths

C. Cross-Sectional Elements
1. Pavement Elevation
    a. Percent of full superelevated length
    b. Percent of superelevated transition length
    c. Changes in crown/unit length
    d. Average crown/unit length of tangent
    e. Distribution of maximum superelevation
2. Shoulder cross-sectional elevations
    a. Joint distribution of cross-sectional lateral changes in elevation and angle of connection between tangent to pavement edge and shoulder.
    b. Longitudinal distribution of angles of connection
3. Side Slopes Cross-Sectional Elevations
    a. Joint distribution of cross-sectional changes in elevation, and angle of connection with the outside edge of the shoulder
    b. Longitudinal distribution of angles of connection
4. Drainage Ditch Cross-Sectional Elevations
    a. Percent of pavement that is paralleled by drainage ditches
    b. Distribution of angles γ and β



c. Distribution of h, the vertical difference in elevation from the bottom of the ditch to the lowest lip or a distance "b," whichever is smaller

5. Elevation From Roadway Outward to the Edge of Right of way or a Distance "a," Whichever is Maximum
    a. Joint distribution of width and slope between inner and outer limits to yield average slope
    b. Distribution of the connection angle between inner edge of roadside and side slope or drainage ditch
    c. Distribution of variances from the average slope
6. Median Cross-Sectional Elevation
    a. Percent of section paralleled by a median
    b. Distribution of connection angles between the outside edge of the shoulder and the median
7. Curb
    a. Percent of section paralleled by curbing
    b. Joint distribution of curb elevation and lateral width of rise

D. Lateral Obstructions
1. Number of lateral obstructions/unit length
2. Joint distribution of longitudinal distances, and lateral distances between obstruction
3. Lateral distribution of distances from the highway centerline to lateral obstruction
4. Distribution of widths of lateral obstructions perpendicular to the centerline
5. Distributions of vertical distance from roadway to overhead obstructions

E. Intersections
1. Number of intersecting streams per unit length
2. Distribution of types of intersecting roads
3. Distribution of distance between the centers of intersections
4. Distribution of the smallest angle between intersecting stream, and highway under consideration

PERFORMANCE CHARACTERISTICS

1. Circular Curve--ability of a vehicle to change direction.
2. Sight Distance--ability of a driver to see objects ahead in the roadway. Sight distance is not unique to highway geometrics, but is dominated by it.
3. Design Speed--speed at which vehicle will track highway centerline.
4. Speed and Acceleration--ability of a vehicle to accelerate and maintain speed. Speed and acceleration are not unique to highway geometrics.

There are several vehicle performance characteristics which are directly affected by highway geometric inputs. These vehicle performance characteristics, denoted as "y's" are listed below.

y_1	Speed
y_2	Acceleration
y_3	Turning radius
y_4	Gradability
y_5	Endurance
y_6	Headway maintenance
y_7	Sight distance (passability)
y_8	Sight distance

The above vehicle performance characteristics are in part functions of the highway geometric characteristics mentioned earlier in this chapter. For convenience these geometric characteristics will be listed below and will be denoted by "x's".

x_1	Degree of curve
x_2	Transition curve
x_3	Vertical curve
x_4	Superelevation
x_5	Shoulder width
x_6	Pavement width
x_7	Lane width
x_8	Distance of lateral obstruction from pavement edge
x_9	Frequency of lateral obstruction
x_{10}	Intersections
x_{11}	Accessibility
x_{12}	Median width
x_{13}	Length of acceleration lane
x_{14}	Cross walk
x_{15}	Sight distance
x_{16}	Railroad crossings

The functional relation which exists between the vehicle performance characteristics and highway geometric characteristics follow.

$$y_1 = f(x_1, 000, x_{16}, \theta)$$

Where θ is all other nongeometric variables.

$$y_2 = f(x_3, x_{11}, x_{13})$$
$$y_3 = f(x_1, x_5, x_6, x_{10}, x_{12})$$

y_4 = f(nongeometric variables such as power, initial speed, resistances, etc.)

$y_5 = f(x_1, x_3, x_{10}, x_{12}, x_{14}, x_{16})$

$y_6 = f(x_1, x_6, x_{10}, x_{11}, x_{13}, x_{14}, x_{15}, x_{16})$

$y_7 = f(x_1, x_2, x_3, x_4, x_6, x_8, x_9, x_{11}, x_{15})$

$y_8 = f(x_1, x_2, x_3, x_4, x_6, x_8, x_9, x_{15})$

The vehicle performance characteristics can be evaluated if two facts are known about the geometric functions involved. First, if the geometric characteristic is known and defined in terms of the vehicle characteristic; secondly, if the relative weight of each of the geometric characteristics is known.

In the first case, most of the geometric characteristics have been defined in terms of vehicle performance characteristics. The relative weight of each of the geometric characteristics however has not been defined. This area must be pursued if accurate functional relationships are to be achieved.

Existing relationships between highway geometric and vehicle performance characteristics are given below:

(1) Stopping Sight Distance[8]

$$D_s - d_r + d_b$$

where dr = distance traveled during perception-reaction time

d_b = braking distance (locked wheels)

$$d_r = 1.47\ Vt$$

$$d_b = \frac{V^2}{30(f + G)}$$

$$D_s = 1.47Vt + \frac{V^2}{30(f + G)}$$

where V = speed in miles per hour

d = distance in feet

f = coefficient of friction

G = percent grade divided by 100

t = perception-reaction time

(2) Passing Sight Distance[9]

$$D = d_1 + d_2 + d_3 + d_4$$

---

[8] American Association of State Highway Officials, A Policy on Geometric Design of Rural Highways (1965), p. 136.

[9] Ibid, p. 144.

where d_1 = perception, reaction, and acceleration from trailing speed to encroachment on left passing lane.

$$d_1 = 1.47\, t_1 \left(V - m + \frac{at_1}{2}\right)$$

t_1 = time of preliminary delay

a = average acceleration rate

V = average speed of passing vehicle

m = difference in speed of passed vehicle and passing vehicle

d_2 = $1.47Vt_2$ - (distance in left lane)

t_2 = time vehicle passing occupies left lane

d_3 = clearance length between the opposing and passing vehicle 110 ft to 300 ft

d_4 = distance travelled by opposing vehicle

$d_4 = (2/3)d_2$ (assume opposing vehicle at same speed as passing vehicle)

(3) Horizontal Curve[10]

V	D_{max}	Superelevation
30	24.8°	.10
40	13.4°	.10
50	8.3°	.10
60	5.5°	.10
65	4.7°	.10
70	3.9°	.10
75	3.2°	.10
80	2.8°	.10

$$D = \frac{85{,}900\ (e + f)}{V^2}$$

or

$$V = \sqrt{\frac{85{,}900\ (e + f)}{D}}$$

(4) Transition Spirals[11]

V	Min. Length of Spiral
30	100'
40	125'
50	150'
60	175'

---

[10] Ibid, p. 158.

[11] Ibid, Table III 11, p. 176.

(4) Transition Spirals (Continued)

V	Min. Length of Spiral
65	190'
70	200'
75	220'
80	240'

In general, $L = 100 \text{ ft} + (V-30)\,2.5$

$$V = \frac{L - 100}{2.5} + 30$$

(5) Max. Grade to Design Speed[12]

$$HP = \frac{(R_r)V}{375} + \frac{(R_a)V}{375} + \frac{(R_g)V}{375}$$

where HP = horsepower required

R_r = rolling resistance

R_a = air resistance

R_g = resistance of grade

(a) Rolling Resistance:

$$R_r = 17.9\,W + (1.39V - 10.2) \text{ (in pounds)}$$

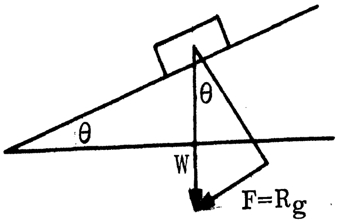
where V = speed (mph)

W = gross vehicle weight (tons)

(b) Air Resistance:

$$R_a = CAV^2 \text{ (in pounds)}$$

(c) Grade Resistance:



---

[12] Hay, William W., An Introduction to Transportation Engineering (John Wiley & Sons, Inc., 1961), p. 177.

$$R_g = W \sin\theta \approx W\theta$$

pounds/% grade for 1 ton

$$R_g = 2{,}000(.01) = 20 \text{ pounds/ton/\% grade}$$

$$R_g = 20 \text{ pounds } (W) \text{ \% grade (in pounds)}$$

Thus:

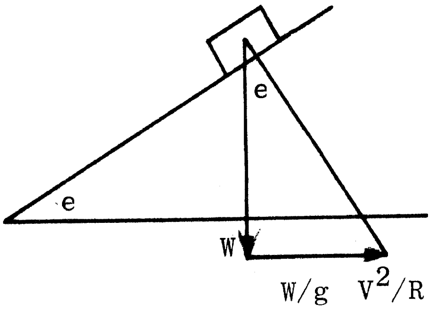
V = f(power and resistance)

$$V = \frac{375 \text{ HP}}{R_r + R_a + R_g} \quad \text{(in miles per hour)}$$

$$V = \frac{375 \text{ HP}}{17.9W + (1.39V - 10.2) + CAV^2 + 20(W)G}$$

(6) Superelevation

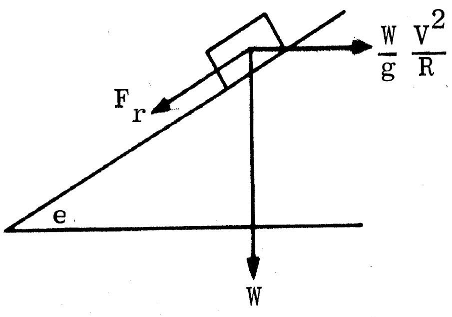
(a) Case I: no side friction (equilibrium speed) $W_p = F_p$

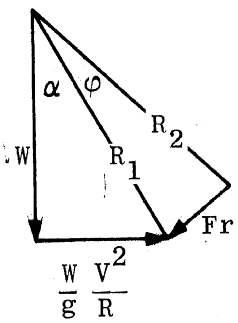


for small angles $e = \frac{v^2}{gR}$ where v is fps

thus $v_{fps} = \sqrt{egR}$ $V_{mph} = \sqrt{15Re}$

(b) Case II: side friction and





$$\tan\ (\alpha + \varphi) = \frac{w}{g}\,\frac{v^2}{R_1}\left(\frac{1}{w}\right) = v^2/gR$$

for small angles

$$e + f = \frac{v^2}{gR}$$

$$v = \sqrt{gR(e + f)}$$

where v = speed (fps)

g = gravity

R = radius of curve in feet

e = superelevation

f = side friction

(7) Length of Vertical Curve

$$L = \sqrt{\frac{2V}{R}}$$

where L = length in feet

V = speed in mph

$A = \frac{G_2 - G_1}{L}$ = rate of change of grade per 100-ft station

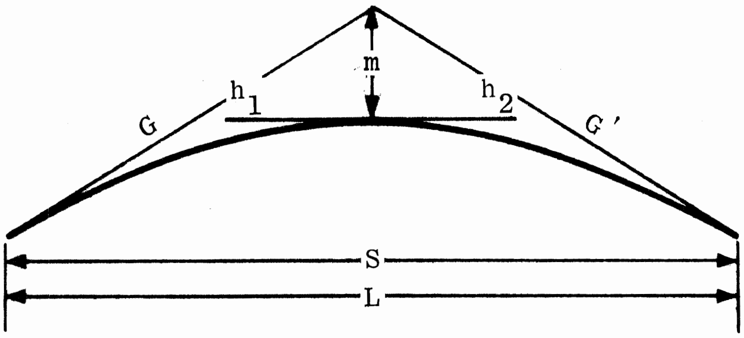
Thus:

$$V = \frac{L^2A}{2}$$

(8) Sight Distance on Vertical Curves[13]

(a) Crest Curve:

1. Sight Distance $\leqq$ Curve Length

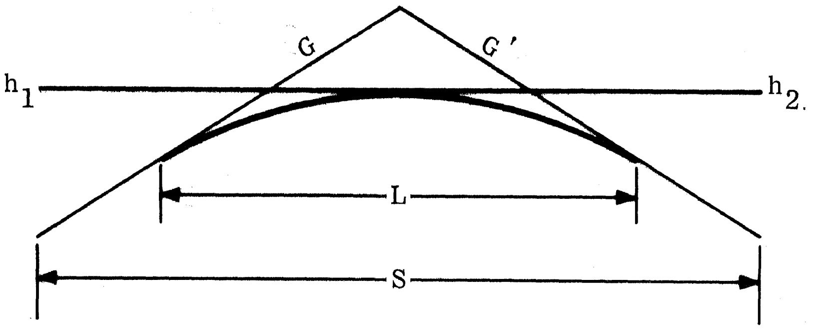


---

[13] American Association of State Highway Officials, A Policy on Geometric Design of Rural Highways (1965), pp. 204-209.

$$L = \frac{(G - G')S^2}{100(\sqrt{2h_1} + \sqrt{2h_2})^2}$$

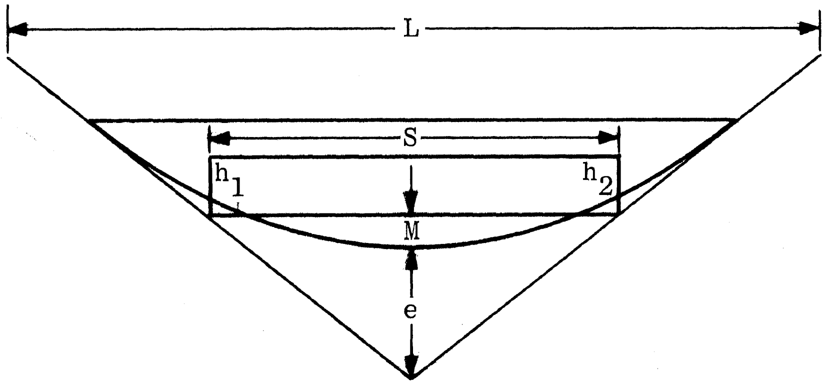
2. $S > L$



$$L = 2S - 200\left[\frac{(\sqrt{h_1} + \sqrt{h_2})^2}{G - G'}\right]$$

(b) Sight Distance at Underpass (where the restriction falls at the PI of the vertical curve)

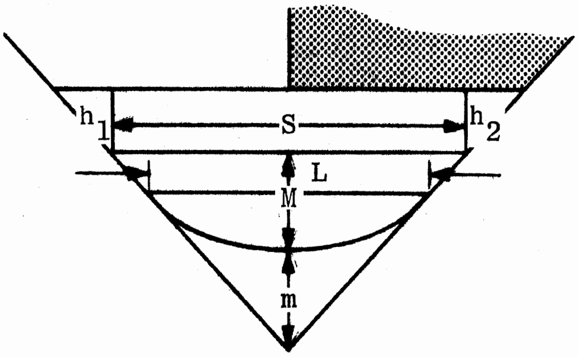
1. $S > L$



$$S = \frac{L}{2} + \frac{LM}{2m}$$

$$L = \frac{2(G - G')S - 100}{(G - G')} = 2S - \frac{100}{(G - G')}$$

2. $S \leqq L$

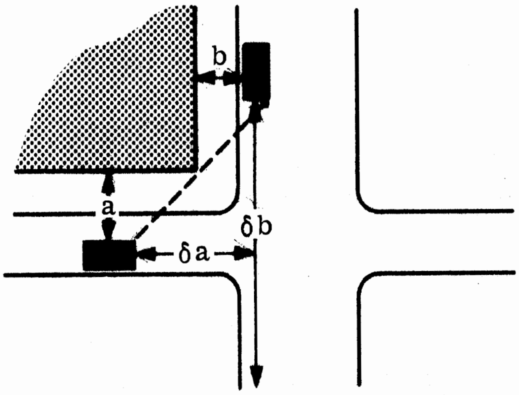


$$L = \frac{S^2(G - G')}{8m}$$

$$S = \frac{L}{2} + \frac{4(C - h_1 + h_2)}{G - G'}$$

$$V = \frac{S^4(G - G')^2}{64m^2} (R/2)$$

(9) Intersection Sight Distance[14] (Case II: Enabling Vehicle to Stop)



---

[14] Ibid, pp. 392-395.

$$d_b = \frac{ad_a}{da - b} = SSD_{V_b} = 1.4V_b t + V_b^2/30(f \pm g)$$

$$V_b = \sqrt{(d_b - 1.4V_b t)30(f \pm g)}$$

Of the highway geometric variables mentioned earlier in this chapter, most have been studied extensively in terms of vehicle-highway geometry interaction. There are some elements, however, that will need a great deal of work. (Vertical cross-sectional elements which need further work include side slopes and roadside slopes. The problem of interaction between vehicle, side area, and materials on the roadside is very important to vehicle escape.)

Lateral obstructions also need extensive study. For example, sign supports, poles, and bridges need to be analyzed in connection with vehicle impact. Likewise, guard rails and drainage structures need more testing and improved design standards.

The environmental geometry described in this chapter has been defined in such a way that it will lend itself to interact with vehicle and material submodels.

HIGHWAY MATERIALS

A performance measure for a highway system component usually describes the ability of that component to perform a function. Frequently it is necessary to describe component performance in terms of several such measures. Before any performance measure of a system is selected and evaluated it is necessary to determine the precise function of the component in question.

The main function of materials in a highway system are:

1. To support a load
2. To reflect or emit light and provide contrast
3. To provide frictional forces
4. To provide barriers control vehicular energy dissipation (desirable or undesirable)

Attention has been given to the first two items above, as described below.

LOAD BEARING (Item 1)

The function of load bearing is to provide a reaction force to moving or stationary objects. The quality of this function or service for a moving vehicle is related to the amount of dynamic disturbance caused by surface inequalities and material compression which result in the actual trajectory of the vehicle departing from its desired trajectory.

In order to describe this performance quality we must select means which can be at least empirically related to disturbances in vehicle trajectory. Two possible performance means are:

1. Deflection for some fixed load per unit area
2. Spectrum for surface irregularities

Since most models for the suspension system of vehicles are described in terms of linear differential equations, it seems reasonable to describe the surface irregularities in terms of Wiener transform which can be used in turn to determine the total time average kinetic energy in the system. Since surface irregularities are a stochastic process over a distance parameter each random surface can have a different surface character. However, if it is assumed that the surface irregularities in a wide sense are stationary over the distance parameter, it is possible to estimate the power spectrum from one rather long surface sample from the ensemble of possible samples. Let the estimate of the autocorrelation function be R(t), then an estimate for the power spectrum is

$$G(W) = \int_{-}^{+} R(t)e^{iwt}\, dt$$

and the energy spectrum for the suspension system with transfer function H(w) is

$$E(w) = H^2(w)G(w)$$

From this equation we can determine either the average kinetic energy in the suspension system of the vehicle or the energy in any frequency interval.

LIGHT REFLECTION AND EMISSION (Item 2)

Directional Reflectance. Directional reflectance is used to describe the optical energy transfer at a plane surface for incremental solid angles of incidence where the directions of incidence are arbitrary. The distribution function depends upon the optical wavelength λ, the incidence direction azimuth and zenith angles (x, y) and the azimuth and zenith angles (x', y') of the reflection direction.

$$r = r(\lambda, x, y, x', y')$$

From this function one can determine the power or energy in the reflected radiation field given the distribution of the incident field. The above distribution has not been measured for many materials although it provides an almost complete characterization of the reflecting properties of materials within an incoherent non-polarized electromagnetic field.

If this function $r(\lambda, x, y, x', y')$ is known one can determine any reflectance performance measure. Other more approximate reflection performance measures are:

Total diffuse reflectance

Specular reflectance: occurs in specular reflection where the directions of incidence and reflection form a plane which includes the perpendicular to the surface of the material and which intersects the angle between the two directions (also a function of λ).

Contrast. The main role of reflected light in the driver's seeing operations is both the detection and discrimination of objects. The detection of an object is usually related to the total amount of energy reflected by this object. The discrimination operation depends upon the geometry of the object and the spectral distribution of light.

A performance measure for contrast can be:

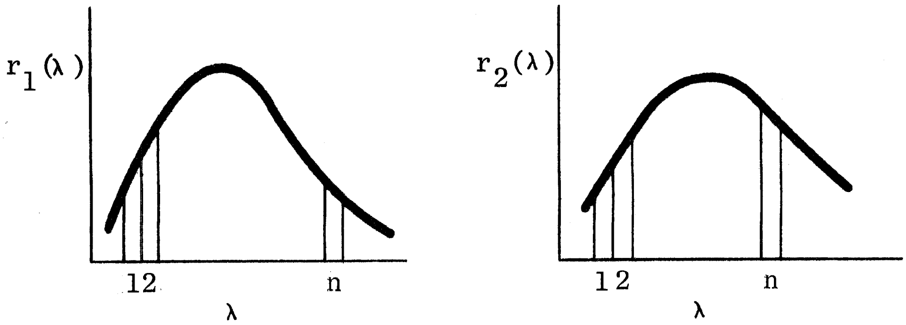
$$C = \frac{L_1 - L_0}{L_1}$$

or

$$C = \frac{L_1 - L_0}{L_1 + L_0} \Big/ 2$$

L_1 and L_2 being the luminance of the two materials.

Another way to measure contrast can be as follows. Consider the r functions of materials 1 and 2 for the same angular characteristics ($x_1 = x_2 y_1 = y_2$, etc.)



Dividing the visible spectrum in intervals or colors one obtains a dimension vector R_n whose coordinates are the $r(\lambda i)$ and whose contrast can be measured by

$$C = R_n(1) - R_n(2)$$

HIGHWAY INFORMATION AND CONTROL DEVICES

An important input to the interaction of vehicle, highway, and materials is the use of traffic control and information. Many of these control and information inputs are of a geometric nature as far as support and physical presence is concerned. The geometric characteristics of these devices have been considered under Highway Geometry.

This section is divided into four general categories: signs, markings, signals, and islands.

Signs are of three general types: regulatory signs, warning signs, and guide signs. These types are defined as follows:[15]

> Regulatory signs, giving the highway user notice of traffic laws or regulations, that apply at a given place or on a given highway, disregard of which is punishable as an infraction, violation, or misdemeanor.

---

[15] National Joint Committee on Uniform Traffic Control Devices Manual on Uniform Traffic Control Devices for Streets and Highways, U. S. Department of Commerce (1961), p. 10.

> Warning signs, calling attention to conditions in and adjacent to a highway or street that are potentially hazardous to traffic operations.
>
> Guide signs, showing route designations, destinations, directions, distances, points of interest, and other geographical or cultural information.

Further, the classification of regulatory signs is as follows:[16]

1. Right-of-way series:
   - (a) Stop sign
   - (b) Yield sign
2. Speed series
3. Movement series:
   - (a) Turning
   - (b) Alignment
   - (c) Exclusion
   - (d) One way
4. Parking series
5. Pedestrian series
6. Miscellaneous series

For descriptive purposes, signs will be described in the following terms:

1. Small signs (signs supported on one post)
2. Large signs (signs supported on more than one post)
3. Lighted signs
   - a. Small
   - b. Large
4. Billboards

The interaction model of signs can best be expressed in the following performance characteristics.

1. Number of signs/unit length
2. Longitudinal distribution of distances between signs
3. Distribution of sign size
4. Distribution of sign shape
5. Lateral distribution of signs
6. Distribution of methods of information transfer
7. Distribution of type of letters, pictures, symbols, and letter spacing.

Markings can be generally divided into three groups: pavement markings, hazard markers, and object markings. These groups can be further divided into several subcategories as follows:[17]

---

[16] Ibid, p. 26.

[17] Ibid, p. 114.

1. Pavement Markings
    a. Center lines
    b. Lane lines
    c. No-passing-zone markings
    d. Pavement edge lines
    e. Paved-shoulder markings
    f. Pavement-width transitions
    g. Channelizing lines
    h. Approaches to obstructions
    i. Turn markings
    j. Stop lines
    k. Crosswalk lines
    l. Approaches to railroad crossings
    m. Parking space limits
    n. Word and symbol markings
    o. Lane-use control markings
2. Curb Markings for Parking Restrictions
3. Object Markings
    a. Objects within the roadway
    b. Objects adjacent to the roadway
4. Reflector Markers
    a. Hazard markers
    b. Delineators

As with signs, pavement marking performance variables can be described as follows:

1. Percent of pavement with each type of marking described above.
2. Distributions of pavement markings described above.

Signals can be described as including "all power-operated traffic control devices, except flashers, signs, and markings, by which traffic is warned or directed to take some specific action."[18]

Traffic signals can generally be classified in four ways:[19]

1. Traffic Control Signals (Stop-and-Go):
    (a) Pretimed signals
    (b) Traffic-actuated signals
        (1) Full traffic-actuated signals
        (2) Semi-traffic-actuated signals
        (3) Traffic-adjusted signals
2. Pedestrian Signals
3. Special Traffic Signals
    (a) Flashing beacons
    (b) Lane-direction-control signals
    (c) Traffic signals at drawbridges
4. Train-Approach Signals and Gates

---

[18] Ibid, p. 155.

[19] Ibid, p. 157.

Performance characteristics of traffic signals can be defined as follows:

1. Number of signals/unit length
2. Longitudinal distribution of distances between signals
3. Distribution of signal type
4. Joint distribution of lateral distance and vertical distance to signals

Finally, traffic control islands can be defined as "a defined area between traffic lanes for control of vehicle movements or for pedestrian refuge. Within an intersection area, a median or an outer separation is considered to be an island. An island may be designated by paint, raised bars, mushroom buttons, curbs, guideposts, pavement edge, or other devices."[20]

Islands can be further subdivided into:[21]

1. Pedestrian refuge islands
2. Traffic divisional islands
3. Traffic channelizing islands

Performance characteristics for traffic islands can be defined as follows:

1. Percent of length with traffic control islands
2. Distribution of type of islands
3. Joint distribution of height and width of different types of islands.

RECOMMENDED RESEARCH IN HIGHWAYS

A complete review of available literature relating the highway subsystem to the other subsystems was not undertaken and should be done at an early date to extend the systems approach. Appendix 3-A presents a review of literature completed; the views of the investigators are presented below.

Within the area of the highway environment, the physical, geometric, and traffic information and control device subsystems, the recommended research program should be directed towards the interactions with other subsystems. Within the area itself meaningful research in materials appears possible with particular emphasis on

---

[20] Ibid, p. 237.

[21] Ibid, p. 237.

better understanding of the significant properties of pavements contributing to the tire-road subsystem coefficient of friction, and investigations of improved energy-absorbing and deflecting subsystems for roadside obstacles. Street-lighting subsystems need research.

Traffic informational and control devices require research associated with their visibility.

At the subsystem level the geometric environment must be better described and it is recommended that the performance measures identified in the chapter be further explored.

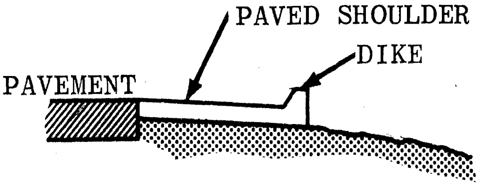
APPENDIX 3-A

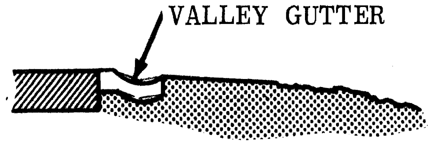
SELECTED BIBLIOGRAPHY AND SUMMARIES

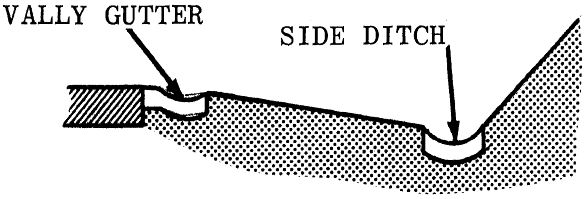
Ackroyd, Richard, "Highway Shoulders as Related to Surface Drainage," HRB Bulletin 151, Symposium on Highway Shoulders, 1957.

Surface drainage affects the stability and durability of shoulders. The author briefly described four situations:

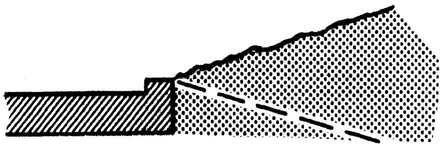
(1) When the roadway is of a substantial grade, longitudinal flow of water will erode the shoulder, usually at the pavement edge, making a small channel at first, then overflowing the shoulder slope and resulting in gullies. The three designs sketched below are frequently used to handle this problem; all of them are intended to carry the runoff longitudinally to an inlet. Yet there is still a need for more efficient and economic design.







(2) The poor design practice shown below permits water (including drainage from the shoulder) and debris to be retained on the pavement where it interferes with traffic. In winter, the sheet flow of ice across the pavement slab is a potential danger to traffic.



(3) Where roads are frequently overflowed by flood water, some kind of protection to the shoulder is needed. Pipes, sod, or paving on the shoulder may be an economical protection depending on the difference in elevation of water surfaces on both sides of the road. Research was being undertaken (1957) to evaluate this relationship.

(4) Another common cause of shoulder failure is poor and inadequate design of culverts. In this case water builds up at the entrance to culverts and overtops the shoulder, causing erosion or slide.

Brant, Frank H., "Highway Shoulder Design from the Roadside Development Viewpoint," HRB Bulletin 151, Symposium on Highway Shoulders. 1957.

The author states the reasons and advantages of turf stabilized shoulders when permeability is not required and economy is the governing factor. The following points are emphasized.

(1) Turf can be easily grown on stabilized compacted soil-aggregate materials. Top-soil is not necessary. Turf reduces the effect of weathering by wind or water. It also provides some shearing resistance for occasional traffic.

(2) The maintenance cost is lowest for turf stabilized shoulders. Some of the problems yet to be solved are also pointed out.

(a) How steep can a turf stabilized shoulder be? Is the present recommended 1 inch/foot too steep?

(b) Should shoulders have a degree of stability considered necessary 24 hours a day, 365 days a year?

(c) What is the acceptable degree of shoulder rutting?

(d) What width of the shoulder should be stable? Should the entire width be stable at all times, or only a portion of it as required to withstand more frequent traffic swerving off the pavement edge?

Burg, Albert, and Slade F. Hulbert, "Predicting the Effectiveness of Highway Signs," HRB Bulletin 324, Free Operations, 1962.

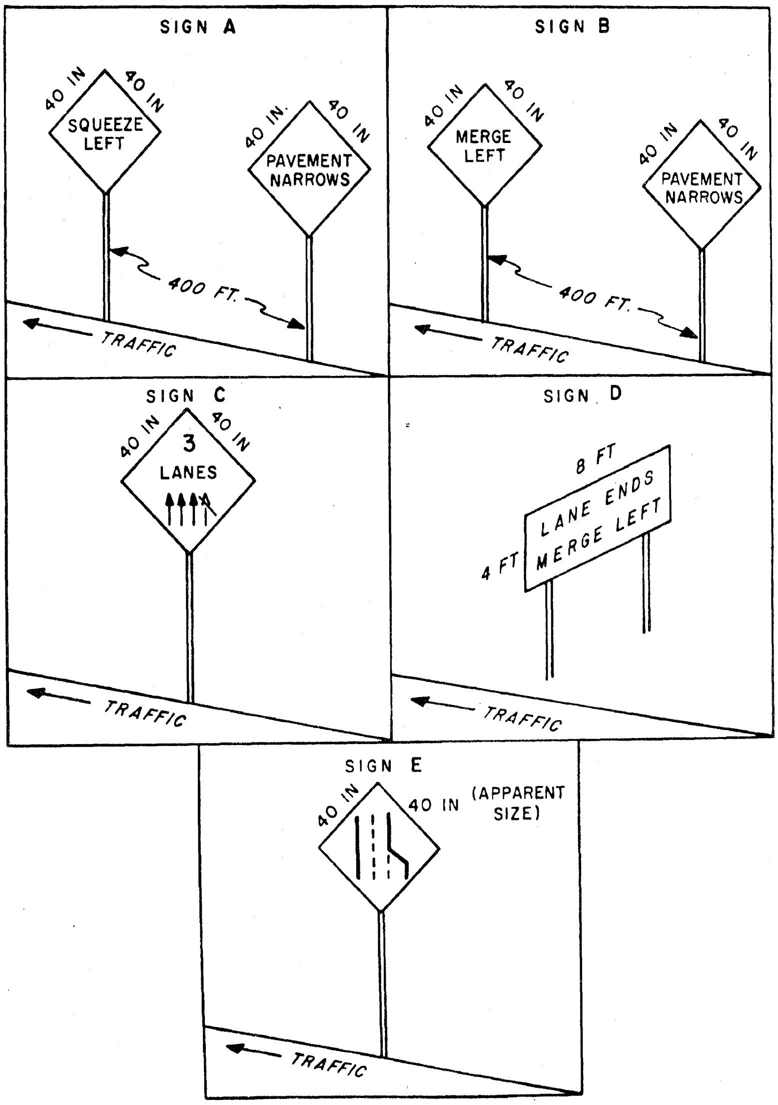
The effectiveness and adequacy of a newly installed sign is difficult to evaluate until after a period of several months or years when complaints and accident experience have been accumulated. A "short-cut" method is therefore necessary to evaluate and choose among several alternatives of sign designs and their locations on the highway prior to actually placing them on the highway.

Four sign configurations were first designed. (See illustration.) A fifth sign was added later for limited evaluation only.

The signs were installed in a standard manner. Motion pictures were taken from the driver's position in a moving vehicle. Each sign was installed and filmed in turn. The resulting films were then projected on a screen to observers.

Three groups of observers were selected. The first group was made up of 560 students in Los Angeles. The second group was composed of 273 attendees at the Western Safety Congress in Los Angeles. A third group of 47 persons was tested in Washington, D.C. in an effort to detect any regional differences in responses to signs. Each group was further divided into subgroups consisting of drivers of both sexes, with a considerable range in both age and driving experience.

Before the film was shown, the experimenter stated to the groups that they were about to see some short films of traffic signs, following each of which they would be asked to answer some questions.


SIGN A
40 IN
40 IN
SQUEEZE LEFT
40 IN.
40 IN
PAVEMENT NARROWS
400 FT.
TRAFFIC
SIGN B
40 IN
40 IN
MERGE LEFT
40 IN
40 IN
PAVEMENT NARROWS
400 FT.
TRAFFIC
SIGN C
40 IN
40 IN
3 LANES
TRAFFIC
SIGN D
8 FT
4 FT
LANE ENDS MERGE LEFT
TRAFFIC
SIGN E
40 IN
40 IN
(APPARENT SIZE)
TRAFFIC

For every group, the order of presentation of the signs was selected at random. After the first film had been shown, the observers were asked to record their first impressions. This procedure was repeated for each film. After the last film, they were asked to indicate their personal preferences among the signs.

Because of limited time on some occasions, only part of the series of films was shown.

The observers' first impressions, their clarily ratings, and their preferences were analyzed. A consistent result was obtained: "In every group Sign D was preferred over the other three signs by a wide margin, Sign A was always last choice, and Signs B and C fell somewhere in-between."

A marked difference of response was observed between the student group and the group of attendees at the safety meeting. Because of the difference in sizes of the two groups (560 students and 273 attendees) no firm conclusion could be drawn.

The Washington group differed from the Los Angeles groups in response to Sign A, ranking it third while the Los Angeles groups ranked it first.

Further studies seem needed in at least two areas.

(1) How do the factors (familiarity, education) influence preferences? To what degree can the true effectiveness of a sign be determined from the subjective opinions of observers?

(2) The method described in this paper requires choosing among several alternatives. Can data collected from such an experiment be generalized to some basic principles of sign design?

Byington, Stanley R., "Interstate System Accident Research," Public Roads, Vol. 32 (11), December 1963, p. 256-266.

The study conducted in 1960 was an introductory analysis using available data from 16 states. It involved a total of 1,000 miles of existing highways, 1,130 miles of Interstate highways, and 114 miles of control highways. Control highways were those routes having design and traffic characteristics similar to the obliterated highways. These three types of highways were divided into homogeneous sections for study. The data collected included highway design features, traffic characteristics, and accident records.

Two types of analysis were used. In the group analysis, comparisons of accident data were made between highway sections having similar traffic volumes. For example, accidents on existing highway sections having traffic volumes between 4,000 and 8,000 vehicles per day were compared with those on Interstate sections, regardless of location, which carried similar traffic volumes. In the individual analysis, comparisons were made only on accident data collected for existing highways and the nearby Interstate study sections. For example, accidents on existing highway sections were compared with those on parallel Interstate sections regardless of the design characteristics of the two types of highways. Within these two types of analysis, three comparisons were made on accident rates:

(1) existing highway sections "before," compared with Interstate System sections;

(2) "before" compared with "after" for existing highway sections;

(3) existing highway sections "before" compared with "after" for Interstate plus existing highway sections.

Accident rates in the comparisons were expressed as the number of accidents per 100 million vehicle-miles. The results showed that accident rates of Interstate highways were 43% and 48% below those of the existing highways "before" and "after" respectively.

In calculating the accident, injury, and fatality rates, two methods were used. In the summation method, the mean accident, injury, and fatality rates were computed by dividing total accidents, injuries, and fatalities occurring on each type of highway by the total vehicle-miles traveled on the respective highways. In the average method, "the smallest data breakdown possible was employed as a single study section for one year." A Chi Square test showed that the accident, injury, and fatality rates for the Interstate highway were significantly lower than for existing highways "before," both in rural and urban areas. For the "before" and after" of existing highway sections only the fatality rates were significantly different. Injury rates declined on existing highways after the construction of Interstate routes, except in urban areas having populations from 5,000 to 50,000. It was noted that the greater the population density, the greater the benefits in accident reduction. Injury rate reductions were greatest in the large metropolitan areas and in rural areas. In rural areas, fatality rates dropped from 11.3 to 3.3. In urban areas the net reduction was 1.2 fatalities per 100 million vehicle-miles.

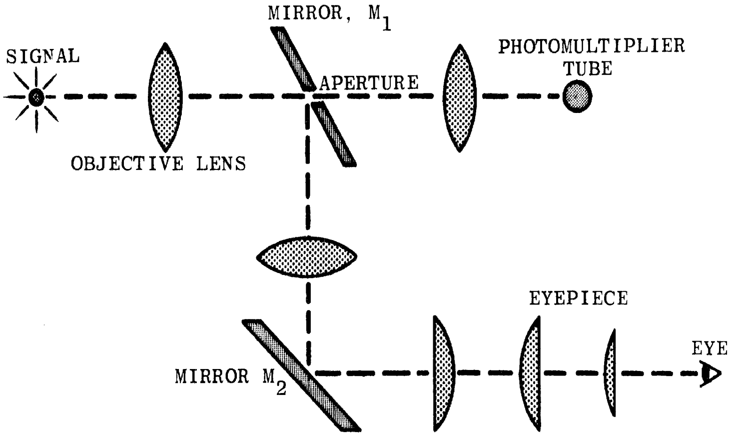
Findings on the controlled-access study indicated that partial access control on urban highways contributed little to safety. This meant that full control of access should be used wherever possible to minimize accidents. Fatality rates on highways with full control of access decreased by 2/3 in rural areas and 1/2 in urban areas as compared to those having no control of access. These data point to the conclusion that control of access is the most important factor on safety. However, other elements of roadway design, such as wide medians, flat curves, etc., are also important. For example, in rural areas the Interstate highways had lower accidents and injury rates than existing highways both "before" and "after," and regardless of the degree of access control.

Tests of Independence and Linearity of Regression for regression coefficients, accident rates, and ADT were made for each type of highway. The results showed that accident rates were dependent on the average daily volumes for all types of highways (except 3-lane existing "before" highways).

Rear-end and same-direction sideswipe collisions decreased by about 50% on Interstate highways compared to existing highways "before" and "after." Fixed-object collisions and noncollision accidents remained about the same. Angle collisions on existing highways were increased by more than 50% after the Interstate highways were opened to traffic, but head-on, sideswipe and noncollision accidents reduced by about one half.

A Spectra-Pritchard photometer was used for the measurement as illustrated above. The signal light passing through the objective lens and the aperture of mirror M, is received by the photomultiplier tube which measures its luminance. The observer viewing through the eyepiece can see the signal and its background and make adjustments. Radioactive phosphor of 60 microcuries is used as a reference source for calibrating the instrument. The photometer is also standardized for red light by interposing a red Wratten filter between a standard lamp and a test plate and measuring the luminances of the plate.

Measurements are made in the field of viewing through the instrument from the rear window of a van. Measuring difficulty in aligning the photometer requires that the distance of the instrument from the signal be 37 m so that the sampled area of the signal is 6.35 in. in diameter. The nominal aiming point, A, is located as the midpoint of the road section 150 feet from the signal. The general layout is illustrated below.



A total of 34 signals are measured with 3 readings for each signal indication. No attempts are made to ensure that the sample selected is unbiased due to difficulty with traffic and complexity of some intersections. Other sources of error expected are:

(1) Errors in laying out the positions of the photometer and the nominal aiming line--unlikely to exceed ±1°.

(2) Error in vertical alignment of aiming line--less than 0.5°.

(3) Error due to nonuniform luminance of signal--the sampled area of the signal is supposed to be large enough to introduce only small errors.

(4) Error due to variation in signal area--less than ±0.5%.

(5) Error due to oblique observation--less than ±0.5% for angle of observation up to 20° from the normal.

Measurements are made at night to minimize error due to stray light.

The results show an average of on-axis intensity much lower than the optimum argued by many. About 1/3 of measured in-service signals have intensities so low that they lead to high uncertainty whether they can be seen against a bright sky. They seem to be adequate but afford no margin of safety under adverse conditions.

As more data became available, the author intended to include further analyses in this study, such as:

"(1) Correlations of two or more variables;

(2) An expansion of the different highway system's accidents, injury, and fatality rate trends;

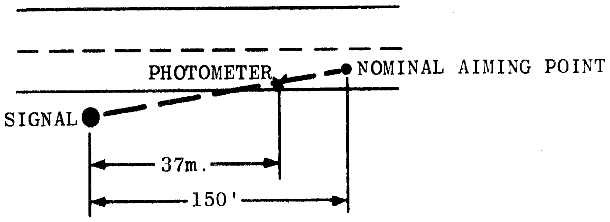
(3) A more comprehensive cost analysis as accident cost study data are refined; and

(4) Development of accident, injury, fatality, and cost equations in conjunction with futher travel forcasts."

Cole, B. L., and B. Brown, "Intensity of In-Service Road Traffic Signal Lights," Australian Road Research; Journal of the Australian Road Research Board, Vol. 2 (6), December 1965.

The color of a traffic signal seen against a bright sky is less easily distinguished than that of one seen against a darker background. The red signal lens, probably more important than the others, has lower transmittance than the green or yellow lenses. There are differing opinions about standards of minimum intensity for the red light against a bright sky. The highest sky luminance cited by Boisson and Pages is 10^4 cd/m^2 near the horizon on a clear, bright day. For a sky luminance of 3000 cd/m^2 an intensity of 200 cd is considered sufficiently striking. The same value is also recommended by the ITE. This represents a minimum intensity for a clean and properly adjusted signal. Consideration must be given to misaimed signals, damage, dirt, tarnishing of the reflector, and aging of the lamp. Because about 2% of all male drivers can be expected to exhibit either protanopia or protanomaly, a higher minimum intensity for the red signal seems desirable. A value of 800 cd is suggested, but this will create glare at night, and thus would necessitate the use of dual-intensity signals.

In spite of all, existing signals are proved to control traffic movements 'reasonably well.' From this it is seen that a knowledge of the photometric characteristics of these signals may provide certain valuable information.



Cole, B. L., A. M. Forbes, H. J. Turner, and J. H. Leevers, "Transport Lighting," Australian Road Research; Journal of the Australian Road Research Board, Vol. 2 (6), December 1965.

This article is a summary of a report on transport lighting at the 23rd ordinary meeting of the Australian National Committee on Illumination.

The activities and aims of those four committes of the International Commission on Illumination concerned with transport lighting are reported. Work on transport lighting in Australia is also summarized in the report. The activities of the four committees are listed below.

1. STREET LIGHTING

The work of this committee includes:

(1) International recommendation on street lighting

(2) Classification of road surfaces from the standpoint of photometric characters

(3) Problems of lighting of motorways

(4) Luminous intensity of street lanterns, its distribution, measurement, and relation to type of surface

(5) Measurement of luminance of street-lighting, the design of luminance meters and their precision

(6) Study of special installations, especially those operating at high or low levels and junctions at split levels

(7) Compilation of references

In addition to the above, the work in Australia also includes:

(1) Study of the influence of wet weather on street lighting techniques

(2) Classification of road surfaces for street-lighting by simplified measurements

(3) Accident statistics

(4) Entries and exits to motorways and transition zones

(5) Headlight performance in built-up areas

(6) Field tests of lantern performance for a representative bituminous concrete pavement.

2. VEHICLE HEADLIGHTS AND SIGNAL LIGHTS

The work includes:

(1) Research of the simple reflex reflector: color, intensity, distribution of reflected light, and maintenance. The height of reflector is recommended not to exceed 90 cm.

(2) Research of turning directional indicators: its distribution of light vertically and horizontally; its intensity, and the adoption of a two-level system for use in daytime and at night; its mounting height; its location with respect to other lights; its flashing characteristics and color; and its standardization for road safety.

3. FUNDAMENTALS OF TRAFFIC SIGNAL LIGHTS

(1) Study of the physiological and psychological observations which influence the recognition of signals, e.g., background luminance and the effect of background screens; and angle of observation and movement of the driver, and the effects of the degree of attention the observer can give the signal.

(2) Preparation of recommended signal threshold values as functions of background luminance and angular subtense of the signal. The published recommendations give threshold values modified by a safety factor. This safety factor is ill-defined, and further research is needed for a more precise specification

(3) Study of the optimum intensity and luminance of a red signal as a function of angular subtense, background luminance, and node of attenuation.

4. AVIATION GROUND LIGHTING

(1) Examination of the results of actual operations using various systems of visual aids;

(2) Formation of a method for calculation of conspicuity and glare of flashing lights, exchange of information, particularly on landing mats;

(3) Methods of indicating angles of aircraft approach;

(4) Development and assessment of the "visual approach slope indicator."

Crosby, J. R., "Accident Experience on the New Jersey Turnpike, Traffic Engineering, Vol. 29, February 1959, p. 18-23, 34.

The accident study on the New Jersey Turnpike covered a six-year period from 1952 to 1957 during which 6,000 motor vehicle accidents occurred.

In the analysis of accident data, all vehicles are broken down into seven classes. The number of accidents, vehicle mileage, and the frequency ratio of each class is tabulated. The frequency ratio is calculated as the ratio of the number of Class I vehicles (passenger cars and single-tire trucks) involved in accidents to the total number of vehicles involved in accidents divided by the ratio of the vehicle-miles traveled by Class I vehicles divided by the total vehicle-miles for all classes. Frequency indices are then calculated by dividing the frequency ratios of each class by the Class I ratio. The frequency index for Class I itself is 1.0. The frequency index for the Truck and Truck Combinations Class is 1.7, which means that 70% more trucks are involved in accidents as compared to passenger cars (Class I), having taken into account differences in mileage. Since truck traffic accounts for only 10% of the vehicles using the Turnpike and 9.8% of the total vehicle-miles, this means that trucks share a proportionately higher number of accidents.

Further study into the statistics of truck accidents reveals some details. The available data indicate that trucks share 25% of all accidents, 25% of injury accidents, and 39% of all fatal accidents. The severity of truck accidents is also greater, averaging 1.4 fatalities per accident as compared to 1.2 for fatal accidents not involving trucks. Rear-end collisions involving trucks account for 65.5% of fatal accidents, 75% of which involve a vehicle striking the rear of a truck, and the other 25% is due to trucks striking other vehicles. 75% of the rear-end fatal accidents occurred during hours of darkness. Some of the fatalities were caused by the shifting of truck load which crushed the occupants of the truck cab. In these instances it was found that the trucks did not comply with the Interstate Commerce Commission requirements for securing the load or for a substantial backboard between the load and the driver compartment. Many severe accidents have resulted from slow-moving vehicles pulling into speeding traffic. Many truck drivers involved in accidents reported they blacked out or fell asleep in night driving under schedule. Tests made on truck tail lights on the Turnpike showed that 88.5% of them were below 3.0 candle-power. A large number of trucks had tail lights of 1/4 candle-power only.

It is concluded that the higher proportion of truck accidents as revealed from the accident statistics may be due to the following causes:

(1) Inadequate rear light equipment, especially the clearance and tail lights

(2) Lack of suitable rearward warning equipment for distinquishing slow-moving vehicles from vehicles which are stopped

(3) Lack of adequate design of rear and front bumpers of trucks

(4) Lack of enforcement and control over truck service schedule at night

(5) Poor operating characteristics of trucks

(6) Improper loading and securing of truck loads

Dent, George H., "Design of Shoulders for Flexible Pavements," HRB Bulletin 151, Symposium on Highway Shoulders, 1957.

The authors of this article emphasized that for heavy-duty highways, shoulders should consist of either hot-mix asphaltic concrete or penetration macadam. The minimum width should be 5 ft. or more with places of refuge for disabled vehicles. For secondary roads, shoulders may be surface treated or consist of mixed-in-place surface courses.

There are several ways to delineate the shoulders from traffic lanes: (1) by a painted stripe, (2) by using different colors, (3) by using a coarser texture or surface for the shoulders, creating a drumming sound when vehicles travel on them, and (4) by using 3/4 aggregate in a thin ridge or line along the pavement edge.

The base course width of roads varies from state to state. Some use full width of section, or from ditch line to ditch line, and some extend 0.25 to 1.5 ft. beyond the traffic lanes. Other design practices also vary from state to state; details are reported in HRB SR 22 and WASHO Road Test, Part 2: Test Data, Analysis, Findings.

Frye, Fred, "The Effect of an Expressway on an Area," Accidents CATS Research News, Vol. 4 (3), December 1, 1961, p. 9-12.

The paper summarizes the results of investigations on the effects of expressways upon accident rates and types.

The Congress Expressway Area under study comprised 12 Chicago Police Districts. The two years chosen for comparison were the year before the first segment of the expressway was opened (1955) and the year after it was open for its full length (1959). Accidents were classified as property damage only, injury, or fatal. The results showed that of the total number of accidents in the city as a whole, the percentage of accidents of all three types in this study area decreased from 1955 to 1959, while the percentage increased in remaining areas of the city. Before the opening of the Northwest Expressway, the accident rate in the Northwest area was increasing at a rate of about 14%, similar to other areas of the city. After the expressway was opened it was reduced to 3.8%.

A comparison of accident experiences for periods of four months before and after the opening of the suburban section of the expressway showed a 6.5% reduction. An attempt was made to relate the types of streets: a preliminary survey revealed 8.4% and 12.1% reductions on local streets and on arterials, respectively.

Frye, Frederick F., "Eisenhower Expressway Study Area," CATS Research News, Vol. 6 (4), October 1964, p. 7-13.

A study was made on the effect of expressway construction on traffic flow and accidents. The data collected were compared with three control areas without expressways. The research arrives at these conclusions:

(1) The shifting and rearrangement of traffic to the expressway reduced accidents by 25% in the study area. Accident rates on arterials also decreased.

(2) A reduction in traffic on the arterial network resulted in a greater percentage reduction in accidents. In the study area traffic increased 21% while the accident rate decreased 8%; but in the three control areas, accidents increased 14% while traffic increased 14%.

(3) Fifty percent of the 21% increase in traffic in the study area was a result of traffic diverted from other areas. Another 20% of the 21% increase was caused by adverse travel necessary to gain access to the expressway. The remainder of the increase was due to natural growth of traffic and was at the same rate ($3\frac{1}{2}$%) as the observed trend from 1953 to 1959.

(4) The hourly distribution of traffic on an average weekday showed a marked difference between 1961 and 1964. The peak hour of traffic remained constant while the off-peak periods accounted for the increase in total traffic. A greater proportion of the total traffic was carried in 1964 than in 1961.

(5) The change in the hourly distribution of traffic in a day between 1961 and 1964 (increase occurring during the non-peak hours of traffic) resulted in an increase in accidents of about 7% even though traffic volumes increased 10%. The number of fatal accidents doubled from 1959 to 1963 in the control areas, but in the study area, it remained almost constant.

Lind, Bruce A., Hyoungkey Hong, "Traffic Accident Study on Milwaukee Expressway," Journal of the Highway Division, Proceedings of the American Society of Civil Engineers, Vol. 91 (1), January 1965, p. 25-48.

The purpose of this study was to establish some relationship between various types of traffic accidents and various geometric elements of the expressway, and to make necessary improvements in the geometric design features.

The study covered a total of 8.5 miles of the Milwaukee County Expressway. Accidents were recorded from January 1962 to October 1963.

The first part of the accident analysis applies Chi Square tests to the accident data to test the independency of variates. These variates were: types of accidents; weather; pavement surface and light conditions; causes of accidents; property damages; and accident distributions.

The most prominent type of accident under all weather conditions was the rear-end collision. Fixed-object collisions occurred more frequently in snowy weather whereas in clear weather sideswipe collisions were more frequent. In rainy weather they were equally frequent. This suggests that slippery pavements resulting from foul weather conditions caused more fixed-object collisions.

The data also showed that many fixed-object accidents occurred during or after snowfall, indicating the desirability of early snow removal or de-icing measures. Rear-end collisions during rain were only half as many as on wet pavements. The same was true for running-off-roadway accidents. These indicate the need for careful driving. The authors conclude that pavement conditions, rather than weather conditions, would be a better measure of accident likelihood.

It is noted that during the night fixed-object and running-off-roadway accidents were observed more frequently. Also, 41% of the total accidents occurred at night for only 29% of freeway usage, indicating that in spite of lower vehicle-mileage, there were more night accidents. The Chi Square test also showed that there were more fixed-object and running-off-roadway accidents than expected.

The chief cause of accidents was "following too closely." Other causes were high speed, improper overtaking, inattentiveness, and slippery pavement.

As a second part of the analysis, locations with high accident frequency were examined and the types of accidents were correlated with various geometric design features.

It was found that on the through lanes of the expressway, bottlenecks resulting from narrowing of right-of-way or from heavy merging or diverging maneuvers in short distances, sideswipes, and rear-end collisions were significantly frequent. There was no clear indication of any correlation between types of accident with either horizontal or vertical alignments. However, it appeared that many sideswipes and rear-end and fixed-object collisions occurred because drivers failed to stay in their own lanes on sharp horizontal reverse curves, and because cars either followed too closely or the lead vehicle reduced speed abruptly on the downgrades of vertical curves.

Accidents in the interchange area were mainly fixed-object, sideswipe, run-off-roadway, and rear-end collision, revealing some problems in the geometric design of interchanges. Three of the interchange ramps had a ratio of radii of 3 so that drivers had to reduce speed by 5 to 10 mph to negotiate the sharper curves of the ramps, contrasting to the 35 mph speed sign which indicated only a uniform speed. Some noses were not clearly visible and contributed to some accidents. Some accidents could be attributed to confusing geometric layouts. For example, where right and left ramps were closely located, and where the junction of through lanes with interchange ramps was located near the downgrade side of a vertical curve, or located on a horizontal curve, accidents frequently occurred.

Twenty-one percent of the total accidents occurred at the entrances to ramps. Width of the ramp noses and lane width were significant factors. Sideswipes and rear-end collisions were often noted where a two-lane ramp narrowed down to a single lane while the sign "Follow Single Line" did not clearly indicate who should yield the right-of-way in merging maneuvers. The ramp length, and horizontal alignment of entrance ramp did not show any direct correlation with accidents.

Direct entry and short taper acceleration lanes were subject to considerable conflicts. The direct entry acceleration lane (less than 200 ft.) was the primary cause of rear-end collisions.

Exit ramps accounted for 16% of the total accidents. Width and length of exit ramps did not indicate any correlation with the type of accidents. Short ramps (less than 400 ft.) caused sluggish traffic operation during the peak period.

There was no indication of any specific correlation between direct-exit deceleration lanes and accident types. The observation agreed with the proposition that direct-exit deceleration lanes were more efficient than parallel deceleration lanes.

No relationship between vertical alignment of exit ramps and accidents could be found. It was noted that 51% of all exit ramp accidents occurred on sharp vertical curves on exit ramps.

Other geometric features of exit ramps or entrance ramps, such as horizontal curve, relative gradient between ramp and freeway, difference between elevation of ramps and freeway, and location of structure relative to the acceleration or deceleration lanes, did not indicate any direct relationship with any particular type of accidents.

The conclusions drawn by the authors are as follows:

Geometric Design

1. Inadequate width of entrance ramp for two-lane operation indicated correlation mainly with sideswipe accidents.

2. For two-lane operation the entrance ramp width at the nose was not sufficient, resulting in sideswipe accidents.

3. The length of entrance ramp did not indicate any correlation with any particular type of accident. However, a relatively long entrance ramp (>700 ft.) was beneficial in reducing sluggish traffic operation on the ramp, where only one column of traffic was allowed to enter the through lanes.

18. Because of sharp compound curves on interchange ramps, drivers had difficulty adjusting speed and direction on the ramps. As a result, striking-fixed-object accidents were the most common.

19. In designing a freeway with respect to adequate visibility, especially at points of ingress and egress, a revised concept of visibility is needed. The conventional concept of visibility on a vertical or horizontal curve is in terms of a distance (sight distance). Although sight distance is an important factor to consider, it is not sufficient enough for providing adequate visibility on urban freeways. In addition to how far a driver can see, what a driver can and cannot see with respect to various geometric features (low or high, small or large) is an important factor to consider in defining and establishing safe visibility.

Traffic Accidents

1. On entrance ramps, the most predominant type of accident was sideswipe; rear-end collisions represented the second highest in frequency.

2. On freeway through lanes and exit ramps, the rear-end collision was the most predominant type of accident; sideswipe was the second highest in frequency.

3. In all weather conditions, the rear-end collision was the most frequent. In snowy weather, more fixed-object accidents than sideswipes were observed, whereas in clear weather the reverse was true.

4. The frequency of fixed-object accidents on snowy or icy pavement was twice as high as that in snowy weather.

5. The frequency of rear-end accidents in rainy weather was only one-half as many as that on wet pavement.

6. The probability of being involved in accidents on snowy or icy pavement during the night was only slightly higher than that for during the day.

7. Driving on wet pavement caused by rain requires caution regardless of day and night. The probability of getting into accidents on wet pavement was nearly the same for day and night.

8. The most predominant cause of accidents was "following too closely." Other causes were, in descending order, excessive speed, improper overtaking, and inattentive driving.

Loutzenheiser, D. W., "Geometric Design of Highway Shoulders," HRB Bulletin 151, Symposium on Highway Shoulders, 1957.

The design of shoulders should be such that efficiency, safety, and mobility are ensured both where traffic is flowing smoothly and when emergencies arise. Several geometric requirements have to be met, i.e. width, continuity, distinctiveness, direction and amount of cross-slope,

4. Direct entry and short taper types of acceleration lane were subject to a considerable internal conflict--mainly rear-end and sideswipe collisions.

5. Long taper and auxiliary types of acceleration lanes were favorable, especially when the entrance volume was relatively high.

6. For two-lane operation on some of the exit ramps, the width of the ramp was not adequate, often resulting in sideswipe accidents.

7. Long exit ramps were favorable for high exit volume, so that traffic would not back into the through lanes to hinder through movement.

8. Direct exit deceleration lanes were the most efficient in operation, because they follow the general travel path of the majority of exiting vehicles.

9. It was observed that the majority of exiting drivers would not use the first one-third of a long deceleration lane, unless the exit traffic was heavy

10. A sharp summit vertical curve on an exit ramp was more responsible for rear-end collisions because of limited visibility than any other geometric feature.

11. It is desirable that the ramp noses of both exit and entrance ramps be well marked at all times.

12. Improvement of weaving section design is desirable. Existing design methods are unrealistic with respect to operating speed.

13. Incorporation of a reverse curve on through lanes requires some caution. The optimum horizontal curvature and curve length must be determined in relation to the vehicle speed.

14. In a directional interchange, where the interchange ramps branch off both left and right at approximately the same location, a considerable amount of confusion and internal conflict of diverging traffic was observed. It is desirable to offset these diverging ramps by some distance.

15. At the merging end of the interchange, it is desirable to offset the left and right acceleration lanes of interchange ramps so as to reduce possible internal conflicts.

16. It is desirable that the junction of through lanes and interchange ramps (in the immediate vicinity of ramp nose) be located so as to be seen by the drivers approaching the interchange. One of the undesirable locations of this ramp nose would be on the down grade side of a summit vertical curve, where the nose cannot be seen until the vehicle is nearly at the summit.

17. When the junction of through lanes and interchange ramps (the immediate vicinity of the ramp nose) is on a horizontal through lane curve greater than 3^o, considerable confusion was observed as a result of optical illusions created by the horizontal curvatures of both through lanes and interchange ramps. It is desirable to provide a tangent section or a flat horizontal curve for through lanes approaching an interchange.

inclusion of curbs and drainage inlets, outer edge rounding, and the slopes beyond. Our basic question concerning the width and continuity of shoulders as set forth by the author is, "Under what traffic volumes and operating conditions are non-continuous and partial width shoulders adequate?" This is a problem of economy. One proposed method is to provide a continuous but partial width (4-7 ft) shoulder where high expenses for shoulders are involved. The other method is to provide a discontinuous shoulder with emergency parking bays at intervals. A third method is to combine the preceding two.

The author asks whether any of these can be satisfactory with regard to traffic operations and safety. If satisfactory, what then will the required dimensions be? He commented on the need for such research studies.

Shoulders on heavily traveled routes have been used as extra lanes during peak hours of flow. Several methods have been used to correct this malpractice, yet more research is needed. For example, the author wonders whether such malpractice is a general problem of any magnitude on expressways or whether it occurs only in special cases. If the problem is general, then what could be done to the shoulders to discourage this malpractice while retaining adequate drainage functions?

The author also raised questions as to the required width of the left shoulder.

Problems also exist as how to provide distinction along continuous stabilized shoulders at the ends of acceleration and deceleration lanes.

"Particularly lacking are studies on vehicle breakdowns, frequency and extent of use of shoulders, the resultant effects on traffic operation and capacity." Statements of research on highway shoulders are contained in HRB SR 12, No. 13.

Moore, R. L., and V. J. Jehu, "Safety Fences," Traffic Engineering and Control, Road Research Laboratory, Vol. 6 (3), July, 1964, p. 180-183.

Safety fences are used to prevent vehicles from crossing the median of a two-way road, to prevent the vehicle from leaving the road at sharp bends and steep slopes, and/or to protect vital roadway structures such as bridge piers. Their functions are to keep vehicles on the road (as on a bridge); to stop vehicles from rebounding onto the road; to stop vehicles from destroying themselves and their occupants; and to serve as anti-dazzle devices. "Basically a vehicle safety fence consists of a continuous horizontal rail which presents a smooth face to an impacting vehicle such that it is redirected, without overturning, to a course nearly parallel to the rail, and with a lateral deceleration that is tolerable to the occupants."

Lateral deceleration is determined by the average distance the center of gravity moves upon impact. The effective height of the guardrail must be equal to the height of the center of gravity if the vehicle is not to overturn. A 2 ft height is sufficient for passenger cars. The greater the lateral acceleration the more likely it is to overturn. The horizontal force that results from the forces perpendicular to the fence, sliding friction, and other effects, must be directed ahead of the center of gravity of the vehicle if the vehicle is to be smoothly redirected. The maximum likely angle of impact is 30° with 20° being the most representative.

A reinforced concrete guardrail with a 12-inch deep convex beam overturned a 3000-lb test car at 46 mph. The rail's lower edge was 10 in. above ground; the vehicle's center of gravity 23 in. above ground.

Corrugated steel rail, bolted directly to posts, pocketed upon impact from a 4000-lb car at 60 mph, 30° impact angle. This pulled the rail below the vehicle's center of gravity, causing it to roll over. This fence is effective up to 35 mph, 20°, with 100 ft minimum length of rail installation.

Double-sided steel guardrails for narrow medians, 30 in. high, blocked out 8 in. from the posts, withstood 60 mph, 32° impact. Blocking out from the posts increases vehicle clearance from posts and maintains rail height when posts are driven back.

Steel cables clamped to I-section posts, 30 in. high, covered with chain link mesh, cushions impact: posts ahead of vehicle are bent back while the mesh crumples ahead to provide resistance and makes a barrier between vehicle and post. It is most effective at large entry angles. At high speeds and angles the vehicle spins out onto the road.

Moore, R. L., "Single-Vehicle Accidents in Relation to Street Furniture," Traffic Engineering and Control, Vol. 4 (7), November 1962, pp. 410-417.

A special study of collisions with lighting columns was done to test three types. A prestressed concrete beam, tubular steel, and thin sheet steel lighting columns, each plated 4 ft into the ground with 75 ft mounting height, were subjected to head-on collisions at approximately 22 mph by 1750-lb pre-war 8-hp cars adapted for the initial towing to the desired speed and for data recording. The degree of deceleration offered by each of the three types of columns decreased from concrete down to thin sheet steel. Car damage was most severe with the concrete, decreasing down to the thin sheet steel type. All three column types required replacement. The resistance of a column to impact is dependent upon type of anchorage to the ground.

The important variables for these tests were the velocity immediately before impact, the deceleration of the cars as a function of time, and the position of the car as shown by high-speed film. Distortion of the structure was determined by before and after the tests.

Petty, D. F., Michael, H. L., "An Analysis of Traffic Accidents on County Roads," Traffic Safety Research Review, Vol. 10 (2), June 1966, p. 44.

The purpose of the investigation was to find the major causes of accidents at low accident frequency locations and to determine possible remedies for the rising number of county road accidents in Indiana. All county road accidents in a sample of 10 counties were analysed from accident reports covering a period of two years (1960-61).

The data showed that 19% of the total accidents were caused by road defects (loose surface materials, gravels, holes, ruts, bumps, defective shoulders, etc.). Skidding was shown to be an important factor in road accidents on both hard and granular surfaces. The frequency of accidents was the same for both types of surfaces. This indicated that if a granular surface were converted into a hard surface, no reduction in accidents could be achieved. The analysis also indicated a greater number of accidents at higher speeds on hard-surface roads than on granular-surface roads. The differences in mean speeds on curved and tangent sections of the hard and granular surfaces respectively were 4.6 and 1.2 mph; thus it was concluded that accidents happened at significantly higher speeds on curves than on tangent hard-surface sections, but not on granular-surface sections. In other words, this indicated that drivers were traveling at speeds too high on curves on hard-surface roads. Accidents due to vision obscurements at intersections totalled only 0.5% out of the total 2,600 accidents. Vehicle obscurements accounted for 9.3%. The safety ratios (accidents divided by intersections) were respectively 0.055 and 0.220 for 3-way (T and Y) intersections and 4-way intersections, indicating that 3-way intersections were about 4 times safer than 4 way intersections. (Even differences in traffic volumes could not account for the total differences in accidents.)

Of all accidents, 85.4% occurred at county road intersections without any traffic control--an evidence that traffic controls were not sufficiently utilized on the county roads. Three types of accidents were more frequent on county roads than other highways and streets. They were the sideswipe collisions, driveway accidents, and running-off roadway accidents on curves. Driver violations were chiefly that of "did not give right-of-way," "followed too closely," "drove to left of center," and "exceeded legal or safe speed." Owing to lower volumes on county roads, the first two types of violations were less frequent than compared with the state as a whole. The third type was considerably higher and the fourth type only slightly higher on county roads than the averages for the whole state. Reported speeds showed that 50% of the accidents occurred below 22.5 mph, 90% at less than 45 mph, 10% above 45 mph, but only 1% above 60 mph. Inclement weather was shown to have no statistically significant effect on accident frequency. Instead, the analysis showed that snow, fog, and sleet conditions resulted in fewer accidents, and rain had no effect. In the analysis the authors used the Percent A/H ratio which they defined as "the percentage of the total number of accidents in a given type of weather (clear and cloudy; rain; snow and sleet; and fog) divided by the percentage of the total number of hours of the type of weather." The ratios for the four types of weather conditions were:

Clear and cloudy	1.12
Rain	1.11
Snow and sleet	0.73
Fog	0.20

Statistical analysis of the data also showed that the type of time (Daylight Saving Time and Central Standard Time) in these counties had a significant effect on the distribution of accidents over the entire day, and also on the distribution during the periods of the day when both light and traffic conditions were most variable. There was no evidence of any change in total accidents associated with the type of time.

The averages of reported pre-accident speeds were 35.3, 27.9, and 23.0 mph for fatal, nonfatal-injury, and property-damage-only accidents respectively. These averages showed statistically significant differences. It was also significant that fewer fatal accidents per 100 accidents occurred on county roads than on state rural roads.

The authors conclude that "The analysis of the data... emphasized the well-known fact that the driver is responsible for a major share of county road accidents." They recommend that "more attention be given in county safety programs to the driver and that he be continually informed and educated concerning accident causing conditions and the personal impact of having an accident."

Shepard, C. H., "Highway Shoulders Construction Practices," HRB Bulletin 151, Symposium on Highway Shoulders, 1957.

The policy for new construction provides for shoulders along the outside pavement edge of all divided pavement highways, and along each edge of all two-lane pavements with more than 200 commercial vehicles per day.

Present design practice requires a 4-ft stabilized aggregate shoulder of 6-in. compacted depth, finished flush, with the pavement surface sloping 3/4 in. per foot.

Some current construction methods are described, with the following points emphasized: Continuous inspection and control are needed to ensure that shoulders are adequately compacted. Side ditches should be properly shaped. Trenches across shoulders should be constructed to drain surface water from trenched pavement and shoulder sections. Consideration should be given to highway traffic on shoulders during construction which interferes with the maintenance of these trenches and creates rutting of shoulders.

Solomon, David, "Traffic Signals and Accidents in Michigan," Public Roads, Vol. 30 (10), pp. 234-237.

Accidents were recorded in Michigan at 89 intersections of all road types from 1946 to 1957. Studies were made one or two years before the installation of traffic signals and for an equivalent period after installation. Traffic signals studied were the stop-and-go and beacons which flashed yellow on the main highways and red on the minor highways. The results were rather surprising, contradicting what might be expected:

(1) Accident Rate. Where stop-and go signs had been installed, the simpler the intersection the greater the increase in accidents. On the other hand, installation of flashing beacon's reduced accidents for all types of intersections.

(2) Injuries. The number of persons injured decreased by 20% with stop-and-go signals and by 50% in the case of flashing beacon installations. The number of fatalities in both cases decreased.

(3) Type of Accident. With stop-and-go signals rear-end, head-on and sideswipe collisions increased 200%, 157%, and 74% respectively. Angle collisions decreased. With beacons, all types of collisions decreased by 25%. The head-on collision is still the predominant type.

(4) Light and Weather Conditions. With the stop-and-go the number of accidents during inclement weather increased by a factor of 4 compared with accidents during all weather conditions, both in daytime and at night. With beacons daytime accidents decreased by 1/5 during both inclement and all weather conditions. Nighttime accidents during inclement weather decreased 1/6, and by 1/3 for all weather conditions.

(5) Traffic Volume. The greatest reduction of accident rates occurred at the higher-volume intersections for stop-and-go signals and at the lower-volume intersections for flashing beacons. Installation of both signals increased the volume by about 11%. Traffic volumes at stop-and-go signals were an average of 2 to 3 times greate than the volume at intersections having flashing beacons.

(6) The Effect of Number of Intersecting Legs. For stop-and-go signals, the greatest increase in accidents is at the 3- or 4-leg undivided intersections. There is no change on the 4-leg divided and on those having 5 or more legs accidents decreased. With beacons the accident reduction is directly proportional to the number of intersecting legs.

Stonex, K. A., "Vehicle Data; Driver Eye Height and Vehicle Performance in Relation to Crest Sight Distance and Length of No-Passing Zones," HRB Bulletin 195, Relation Between Vehicle Characteristics and Highway Design, a symposium, 1958.

The facts presented by the paper can be summarized as follows.

(1) The trend to reduce the overall height of vehicles tends to move the design eye height to a lower level, thus invalidating the present design criteria for vertical curves.

(2) Observations by General Motors in 1936 showed that (a) the seated eye height of males was about 28½-in. rigid seat, and (b) the seat cushion was depressed an average of 2 inches. The same measurements were continued to 1957. The fleet of test cars consisted of passenger cars from each make and model each year, except sports cars and foreign cars.

(3) Several phases of styling changes from 1939 to 1957 were noted to have significantly changed the median value of eye height, the lowest value being 51 inches (1957) and the highest, 57 inches (1936-1939).

(4) Seat cushion depression varied widely. Its median value ranged from 4½ inches (1956) to 4.2 inches (1957).

(5) Predicting future reduction of eye height is difficult because of the dependance on customer acceptance and design skill.

(6) On the average, the driver's eyes are approximately 10 inches below the highest point of the car.

(7) Ten years from 1957, the practical ultimate height of GM cars may be just 53 inches. However, the author believed that the trend of height reduction was nearing an end. He pointed out that the reduction in height was being compensated for by better operating performance, greater rated horsepower, and greater transmission flexiblity.

Taragin, A., "Role of Highway Shoulders in Traffic Operation," HRB Bulletin 151, Symposium on Highway Shoulders, 1957.

This research on the effect of highway shoulders on traffic operations is summarized as follows.

(1) Effect of shoulder width and type on traffic when shoulder is not occupied by parked vehicles or other objects: Extensive studies in 15 states showed that the speed and lateral position of vehicles are not affected by shoulder width if it is clear and at least 6 ft wide. Bituminous-treated shoulders 4 ft in width increase the effective surface width of adjacent two-lane concrete roads less than 20 ft wide by about 2 ft. Shoulders at least 4 feet wide have no substantial effect.

(2) When the shoulder is occupied by parked vehicles or other objects (parapets, bridge piers, abutments, guardrails, utility poles) a clear distance of 4 ft is required from pavement edge to object. The capacity is affected inversely as the width is decreased from 6 feet. Without shoulders, one disabled vehicle can reduce the capacity by more than the capacity of one lane.

(3) There is yet little information on the extent of shoulder use by parked vehicles, Information of this type is needed to determine the frequency of parked vehicles per mile of highway at various traffic volumes and on different types of highways. The author suggests several questions for research:

1. How is the frequency of parking related to trip length and to distance from urban areas?
2. Do turn-outs on a highway decrease the frequency of shoulder use?
3. Do drivers take advantage of wider and better stabilized shoulders to a greater extent than they do narrow and unpaved shoulders?
4. Can the parked vehicle be ultimately related to accident causation?

(4) Intensive studies by several states on the relation between shoulder width and type and motor vehicle accidents produced contradictory results. More information was being sought by several states.

Waldram, J. M., "Visual Problems on Motorways," Illuminating Engineering Society Transactions, London, Vol. 26, 1961, pp. 66-75.

This study attempts to answer two questions:

(1) What does the driver need to see? With what emphasis?

(2) By what means is this information presented to drivers in daylight or artificial light?

For the first part of the study, a vehicle was driven on the roadway as usual and a continuous account of the driver's visual attention and road features is recorded.

In the second part, the driver comments on road features and obstructions, their visibility (whether clearly or badly revealed) and the reasons for it together with comments on the effects of background, color, details, and so forth.

Both series of records are supplemented with films taken from the driver's viewpoint and synchronized with the driver's comments. The driver's operation of controls is also filmed.

Further study is made of the movement of the driver's eyes under various traffic and weather conditions, and during both day and night.

Films are also taken on the M1 motorway by day and at night by the light of headlamps.

The author emphasizes that headlamps alone are inadequate for night driving. The driver cannot see distinctly and easily the road features ahead. Delineators are not efficient enough to guide the driver traveling at high speed. Taillights of preceding vehicles may be confusing and intermingled. The closing rate and speed changes of leading vehicles are difficult to estimate. Glare of headlamps at intersections and on grades presents a problem. Rear mirrors or wing mirrors cannot provide an estimation of the closing rate of overtaking vehicles, and may also reflect headlights, causing blinding and discomfort glare.

The author emphasizes that vehicle headlights are inadequate. Tail lights are insufficiently standardized in relative positions of signal and rear lights. The variability of intensity is too great. Variation of obligatory signals in height causes confusion. Color and flashing characteristics are not standardized. Brake lights provide no distinction between a slowing down and an emergency stop.

The author concludes that fixed lighting is needed as the density and speed of traffic increases, and that fixed lighting is adequate to provide night-driving with as much safety and speed as daytime driving. He suggests points on lighting requirements and vehicle signal improvements.

Westland, J. G., "Some Principles and Problems of Traffic Signal Design," Australian Road Research; Journal of the Australian Road Research Board, Vol. 2 (6), December 1965.

This paper outlines the functions of traffic signal installations and discusses some of the management and technical decisions involved in designing local and system installations. Some points of interest are noted below.

(1) A minimum traffic volume of 600 vph is adopted by the Victorian State Traffic Commission as the warrant for pedestrian and school crossing signals.

(2) The determination of priority of a particular signal installation is based on a "delay index" which measures the costs of delay incurred vs. accidents prevented by installing signals.

(3) The aiming of signals affects the observed intensity of the signals seen by the driver. The aiming rules used in Australia are as follows:

LANTERN LOCATION	AIMING POINT
At left of stop line	At drivers 150 ft from stop line.
At right of stop line	At drivers at a distance which would be traveled in 10 sec at the 85 percentile approach speed.
At a far right location	At drivers stopped at the stop line if no far left signal is provided; if one is provided, aim as at right of stop line location.
Far left	At drivers 150 ft from stop line.
Overhead	As for signals at the right of a stop line.

(4) Measurement of luminous intensity through a clear glass refractor is compared to one of clear plastic material. The former has a specific luminous intensity of 1.70 on the axis while the latter yields a value of 1.35. However, for angles of observation away from the optical axis, they show approximately the same specific luminous intensity.

(5) The intensity of signal is such that it is recognizable at a distance corresponding to a period of 10 seconds at the design approach speed. For example, if the speed is 40 mph, signals should be recognizable at a distance of 600 feet.

(6) With sodium vapor lanterns for lighting, the red and amber signals become difficult to distinguish. Mercury luminaires create similar problems with green signals.

(7) The use of amber and red arrows in place of circular shapes was being investigated by the Australian Road Research Board in 1965.

(8) A problem is present when signals face a low winter sun or when they face a low morning or evening sun during the summer. Long visors and horizontal louvres are generally needed.

(9) The green signal period for pedestrians is calculated according to the Victorian rule, as follows: (a) To calculate the 'walk interval,' take 4/5 the distance between curbs to be crossed in one bound, divide by the walking speed 4 ft/sec, and add 2 sec starting lag. If a refuge is less than 8 feet wide, pedestrians are supposed to cross the road in a single bound. (b) To calculate the 'clearing interval,' take 4/5 of the distance between curbs to the nearest refuge and divide by a walking distance of 5 ft/sec.

(10) In Sydney a master-controlled linear progression system is used to supervise and ensure that adjacent controllers operate with a uniform cycle length, each commencing at a specified offset from the next controller in the system. The split and sequence at each intersection can be varied from the central station as part of the program being timed. In addition to this, television monitors also enable certain selected approaches to be supervised and manual adjustments to be made. A total of 50 programs can be set up and selected at will by time switch. A

program selects one of a continuous series of accurately timed pulses produced at 2-sec intervals. These pulses are then transmitted to a local controller, either terminating the local green or initiates a local function. Cable linking allows the response of each local controller to be indicated at the master station. The ultimate capacity of the system is 200 intersections (100 is presently used).

Webster, F. V., and P. B. Ellson, "Traffic Signals for High-Speed Roads," Road Research Technical Paper #74, Road Research Laboratory, Harmondsworth, England, 1965.

At signalized high-speed road intersections, dangerous situations sometimes occur because vehicles traveling at high speed are unable to stop. A driver seeing a yellow signal and unable to stop may not have time to clear the intersection before the signal changes to permit traffic from the intersecting road to proceed through.

Several proposals have been suggested to eliminate this difficulty:

(1) Reduce the speed of approach vehicles.

(2) Give advance warning time or extra warning of an impending change of right-of-way.

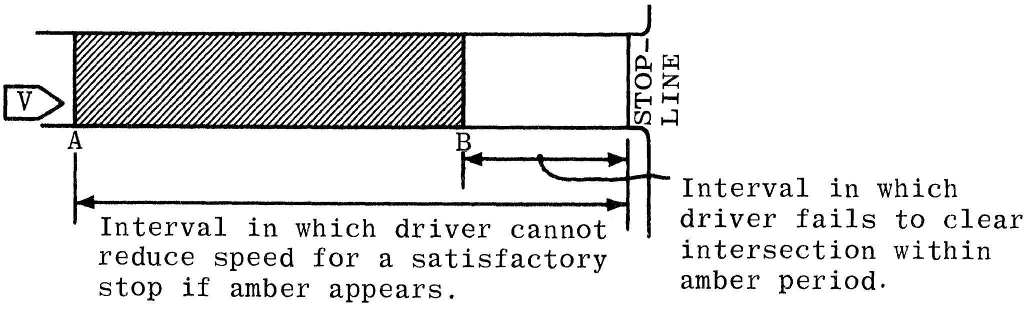
(3) Provide longer green time.

(4) Modify the signal system so that the driver can either stop safely or drive on safely.

Since the last method appeared to be most practical, the Road Research Laboratory carried out the following investigation of this possibility.

First, data were collected to show what percentage of drivers stop at the stop-line when they see the amber signal appear at a certain distance from the intersection. Data were recorded as to the speed of vehicles and their distances from the stop-line when amber appeared, and whether they stopped satisfactorily at the stop-line.

From the data the drivers' behavior could be analyzed. It was possible to locate distances from the stop-line at which a vehicle traveling at a given speed would fail to stop satisfactorily, or to clear the intersection before the red signal appeared. This is illustrated below.



It can be seen that any vehicle traveling at the given speed will fail to stop safely or to clear the intersection if the vehicle is within the shaded portion AB when amber appears.

Therefore, the problem is reduced to one of detecting whether or not a vehicle traveling at its speed is in such a critical position and deciding what to do about it.

Five different makes of cars averaging 3 years old with good hydraulic brakes were used. The test drivers were selected from the age group 23-49 having varying degrees of driving experience.

The instruction given to the driver was:

"Drive at the set speed until either the amber signal appears or you have passed the intersection. When you see the amber signal, act as if you were driving on a public road. Sometimes you will have plenty of time to stop; at other times you will be too near to the junction to stop. 'Emergency' or 'crash' stops are not required."

The speed and amber-distance (the distance of a vehicle from the stop-line when amber signal appears) for each test run was selected at random. Photo-electric vehicle detectors located at the predetermined amber-distances change the traffic signal to amber when vehicles pass them. The number of 'satisfactory' stops were recorded.

The amber-distance corresponding to a probit Y and necessary for a satisfactory stop is expressed by the formula:

$$\log d = 0.017\ V + 0.080\ Y + 1.12 \qquad \text{if } 30 \leq V \leq 70 \text{ mph}$$

where d is the amber distance in feet

V is the speed in mph

Y is probit (a table of conversion from percentage to probit is provided on page 7, Table 3).

From the findings on drivers' performance it was confirmed that critical sections on the roadway did exist within which the vehicle traveling at that particular speed will be in a dilemna when arriving at the intersection. These critical sections were defined as the portions of the roadway between the 90th percentile value of amber-distances for satisfactory stops at the given speeds and the distances from the stop-line such that the vehicle could clear the intersection in the set amber period. These values are listed on page 11, Table 5.

Several methods are examined in part II to alleviate the difficulties of drivers traveling at moderate and high speeds. One of the methods is to locate the detector farther away from the stop-line, but this will cause unnecessary delay due to longer green time. Another method is to provide double-detection but it also increases delay to cross-traffic. Alternatively a number of detectors could be installed at short intervals with each detector so arranged that it responds only to vehicles within certain speed ranges. But this is far from being practical. The Laboratory proposes a method by which only one additional detector is required. In this proposal details of techniques, design, of speed-assessor and timer are illustrated, charts and design graphs are provided, and problems on errors in detection, change of speed vehicles, and tolerance of equipment design are discussed. At the time of publication of the proposal, prototype models were being constructed.

4 CASUALTY RECOVERY

John M. Armstrong

INTRODUCTION

THE EXISTING SYSTEM

At present, initial treatment and transportation for traffic accident victims is usually provided by motorized ambulance service. The history of the ambulance goes back to the Napoleonic Wars when it was given its name, meaning "mobile hospital," by a French army surgeon.

Present-day ambulance service is provided by a variety of sources. Ambulances are operated by both public and private agencies, e.g., police departments, municipal hospitals, health departments, fire departments, funeral homes, private ambulance companies. Although the quality of equipment and service varies significantly among these agencies, basic operating procedures seem to follow a generally consistent pattern regardless of the financial or organizational structure of the service. The ambulance is generally on call 24 hours a day or at least some specified portion of the day. Calls for ambulance service seem to originate from two basic sources: telephone calls from witnesses or bystanders at the accident scene, and radio communication from police units on the scene. The nature of the communication system involved in ambulance services depends upon the type or organization; for example, police ambulances usually operate within the police department's communication network.

Ambulance design varies widely both in the medical equipment carried and in the vehicle itself. Nonuniform laws regulate design capability, training of crews, size of crews, area of coverage, etc., permitting great disparities from one locale to another in the type of emergency treatment and transportation available. Some areas, for example, do not require ambulance crews to have first-aid training and do not require ambulances to carry a minimum complement of medical treatment equipment.

In discussing present casualty recovery systems, it is interesting to consider the analogy between the traffic casualty problem and the military battlefield casualty recovery problem. The rationale for this analogy is quite straightforward: as a motorist moves along a network of roads and highways (the battle zone) at any given time he is faced with the possibility of an accident (enemy threat). If involved in an accident (enemy engagement) the odds of his injury and/or survival (battle outcome) will depend on many variables, such as the type or location of the accident (conflict level). If he is injured, an assistance procedure must be initiated (medical treatment and evacuation), and the probability of his survival will depend not only upon the type of injury but also upon the effectiveness of the treatment and transportation system available to him.

It is interesting to note that in the military area the case fatality ratio[1] has dropped by a factor of 0.5 in each of the last four major conflicts (World War I, World War II, Korea, and Viet Nam). Thus, at least at first glance, it would appear that there may be some merit in studying the military casualty recovery system. However, it should be noted that injury types in the military situation may be significantly different from automobile accident injuries. It would also appear that, at least in Korea and Viet Nam, the rapid transportation provided by the helicopter has played a key role in lowering case fatality ratios.

There is at present no unified "medical corps" for treatment of crash injuries. Just as the military establishment would not consider a battlefield campaign without a medical corps, it seems reasonable that a system analysis of traffic safety must include casualty recovery as a significant factor in system design and operation. As long as traffic accidents result in injuries, recovery and treatment systems will be necessary.

The following sections present (1) a general description of the problems, (2) the relevant variables involved in analysis, (3) a possible model, (4) a literature review, and (5) research recommendations.

---

[1] The case fatality ratio is defined as the number of deaths per 100 injuries sustained in vehicle accidents.

CASUALTY RECOVERY: A RESEARCH PROBLEM

THE INJURY PROCESS

The occurrence of an automobile accident often results in personal injury.[2] The trauma[3] that results from a vehicular accident is primarily a function of an energy exchange between occupant and vehicle or occupant and external environment.

In the time immediately following an accident, the victims may be in any one of many possible states of trauma. Definition of these states is somewhat arbitrary and depends upon the particular setting with which we are concerned. The limiting or bounding states are death (or fatal trauma) and noninjury.

Thus, in the period following the accident epoch (the accident epoch designates the point at which injury may occur) there is a probability of 1.0 that the occupants of the vehicle will be in some as yet undefined state of trauma. It is readily apparent that any given state of trauma is subject to transition with time, and that in the absence of treatment this transition is most likely to be in the direction of increasing trauma level and decreasing probability of survival. Thus it may be stated that the degree of trauma varies with time, and that the transition may be continuous.

PROBLEM STATEMENT: ASSUMPTIONS AND HYPOTHESIS

We are now ready to state more formally one of the major questions to which this study will address itself: Given an individual who has been involved in a vehicle accident, and who is in an initial state of trauma (excluding death) resulting from this event, what are the various relationships of factors that will determine whether he eventually recovers or suffers possible permanent damage?

Our first assumption is that for casualty recovery systems the major measure of effectiveness is the case fatality ratio. If we were dealing with a model of the recovery process, we would speak of

---

[2] In 1964, there were 1,200,000 accidents resulting in 47,000 deaths and 1,700,000 injuries.

[3] Trauma is defined as a state or condition of injury violently produced.

the fatality-injury probability as a measure of effectiveness and define it as the conditional probability of death given the occurrence of an injury.

From the discussion of the injury process in the preceding section, we make the following major hypotheses: two major variables which affect the case fatality ratio are time and medical treatment. As will be shown in the next section, two major factors in time are important in determining the case fatality ratio: the response time of the recovery system and the survival time associated with a given trauma level.

Basically, this type of analysis assumes a time-varying random function H which describes the physical health of a human at a given time. There are then several processes which govern the time-evolution of the H associated with a particular human:

The normal distribution of health
The injury process associated with an accident
The post-injury, pre-treatment process
The treatment process
The final recovery process.

The injury process associated with an accident is described in the accident prediction model (for timing) and the crash dynamics model (for type of injury). The time to treatment is the response time of the recovery system. The treatment process is dependent on the medical and equipment description of the recovery system.

This H-process is a full description, in stochastic terms, of the injury and recovery processes. The case fatality ratio and all other measures of interest may be put in terms of probabilities associated with the process. Specifically, the case fatality ratio (CFR) is:

$$\text{CFR} = \Pr\{\text{H in death state after t, H never in normal health state after t} \mid \text{accident at time t}\}$$

RECOVERY SYSTEM ANALYSIS: A PRELIMINARY MODEL AND RELATIONSHIPS

In this section, we shall express the two major time factors, recovery system response time and injury survival time, in functional form and relate them to the effectiveness measures, the performance measures, and (where possible) the hardware characteristics of the recovery system.

Recovery system response time is defined as the time in which an operational recovery system detects, reaches, and treats or transports to treatment, an individual who is in a trauma state as the result of a traffic accident. At this point, we should note that in examining response time of existing or future recovery systems, two major response strategies must be considered: transportation of the individual to a treatment center, and treatment of the individual on the scene of the accident; as well as numerous combinations of the two strategies (e.g., provide treatment on scene [and/or en route] and/or transport to a higher-level treatment center).

Response Time. The system response time is taken to be a random variable and is expressed as:

$$T_R = T_d + T_r + T_s + T_t + T_a + T_f \qquad (1)$$

where T_d = the time to detect the accident event after it has occurred

T_r = the time to report or communicate the detection

T_s = the time required to search for, locate, and activate a mobile unit for dispatch to the event scene

T_t = the time required for the mobile unit to reach the scene

T_a = the time the unit spends at the scene extricating, treating, and loading victim

T_f = the time required to reach a higher-level treatment center (if necessary)

Note that for systems in which the mobile unit has a high or sufficient level of on-scene treatment capability, loading and transporting the victim would be unnecessary and the response time might be represented as

$$T_R = T_d + T_r + T_s + T_t + T'_a \qquad (2).$$

where T'_a = time required for evaluation and/or treatment.

Injury Survival Time. The injury survival time t_s is defined as the time an accident victim can survive (i.e., remain in any but the fatal state, as discussed in the earlier section on the injury process) without treatment, where treatment is defined as any action by the recovery system to deal with the injury or to transport the victim to a treatment facility.

It seems reasonable to regard injury survival time as a function of both the type and degree of injury, an approach which suggests the construction of a taxonomy of injuries. For the purpose of analysis, it may be helpful to express injury degree in a particular way: we define the injury level, for a given type of injury, as t_s, the survival time defined above. It is now possible to relate the degree of injury to survival time and thus establish one of the major variables required in the design and evaluation of a recovery system.

This survival time t_s must be presumed to be a stochastic function of the actual physical descriptors of the injury, i.e., type and degree. If this is the case, then injury survival time can be expressed in terms of a probability distribution function with parameters (in some form) I and θ (I = type, θ = degree).

In Fig. 4-1, for a fixed $\{I, \theta\}$ the density function of t_s is shown conceptually.

Analyzing Effectiveness. Thus far, we have structured a preliminary model of the accident injury process and the recovery system response. We have assumed that the processes governing the injury event and recovery response are stochastic in nature and therefore that injury survival time and recovery response time are random variables.

Earlier, it was assumed that the measure of effectiveness for recovery systems is the case fatality ratio.[1] We also hypothesized that recovery system response time T_R and injury survival time t_s were two major factors in determining case fatality ratio. We will now introduce one additional variable to the analysis and then develop a relationship for evaluating the effectiveness measure in terms of the three system performance measures.

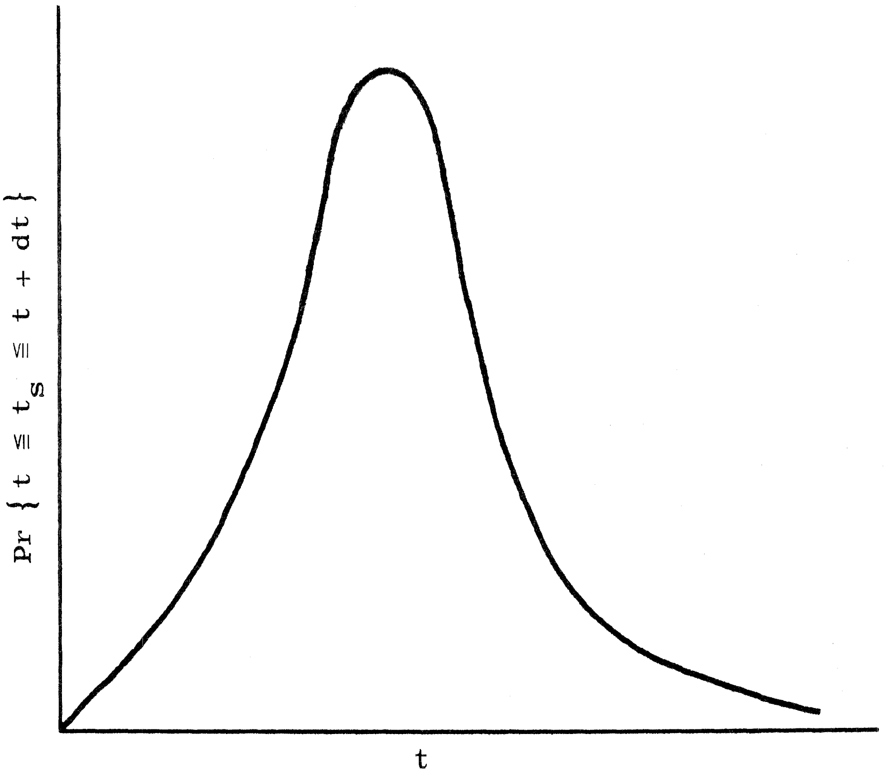


FIGURE 4-1. CONCEPTUAL ILLUSTRATION OF THE SURVIVAL TIME DENSITY FUNCTION

The third performance factor is concerned with the ability of the recovery system to give medical treatment for injuries of a given type and degree. This medical treatment level, M, may be thought of in terms of the equipment and personnel qualifications necessary to treat a given injury level and type. Whether or not the variable M is stochastic is not clear. For evaluating existing systems, one can probably determine explicitly the levels of M that the recovery system is capable of providing. For the following discussion, we will, however, treat M as a random variable.

Given the condition that an accident has occurred, the probability of survival is then the joint probability of two events: the probability of sustaining a specific type and degree of injury <u>and</u> the probability of survival given this injury level and type has occurred. If we sum this joint probability over all possible degrees and types of injury, we will obtain $\Pr\{S|A\}$, the probability of survival given an accident has occurred. In mathematical terms, we have:

$$\Pr\{S|A\} = \sum_{\theta_i}\sum_{I_1}\Pr\{S|I_1\cdot\theta_i\cdot A\}\Pr\{I_1\cdot\theta_i|A\} \qquad (3)$$

The left-hand side of the expression is the desired conditional probability of survival, and the right-hand side is as described above.

In developing an expression for the recovery system effectiveness measure in terms of performance measures, we will consider the conditional probability of survival given an individual has been involved in an accident. Note that we must now deal with probabilities rather than fatality ratios. In addition, the probability of survival is simply 1 - P_d, the probability of death. Thus, the measure of effectiveness in our model becomes a probability of living given an accident has occurred, rather than a ratio of death per 100 injuries. The mathematical formulation is given in Appendix 4-A. The resultant expression is:

$$\Pr\{S|I_1\cdot\theta_i\cdot A\} = \left[\int \Pr\{T_r \leqq t\}\, d\Pr\{t_s > t|I_1\cdot\theta_i\}\right] \Pr\left\{M \geqq M_{(I_1,\theta_i)}\right\} \qquad (4)$$

where $M_{(I_1,\theta_i)}$ is the least M which successfully treats injury level I_1 and degree θ_i. Refinement of the model will allow the medical level M to have probabilistic rather than binary (save-fail) effect.

The right-hand term expresses the performance capability of the recovery system (T_r, M) and the survival time t_s.

SUBSYSTEM PERFORMANCE MEASURES, HARDWARE CHARACTERISTICS, ENVIRONMENTAL PARAMETERS, AND OTHER ANALYSIS VARIABLES

RECOVERY SYSTEM VARIABLES AND PARAMETERS

In the previous section, an expression for subsystem effectiveness was developed. Evaluation of this expression requires inputs that describe the physical and performance characteristics of the recovery system, the environment in which it operates, and its operational strategies. In this section, these elements will be (where possible) identified and discussed in relation to subsystem effectiveness.

Recovery System Effectiveness. Under Injury Survival Time the recovery system effectiveness measure $(e_1) = \Pr\{S| I_1 \cdot \theta_i \cdot A\}$ and its relationship to the overall survival function $\Pr\{S|A\}$ was discussed. It is again noted that $\Pr\{S| I_1 \cdot \theta_i \cdot A\}$ is related to the case fatality probability by the expression $1 - \Pr\{D| I_1 \cdot \theta_i \cdot A\}$ where $\Pr\{D| I_1 \cdot \theta_i \cdot A\}$ is the conditional probability of death given an accident and injury has occurred.

Recovery System Descriptors. The major measures are:

T_R = recovery system response time (y_1)

M = medical capability level (y_2)

I_1 = injury level (in terms of survival time)(y_3)

0_i = injury type (y_4)

There are, of course, many other parameters that relate to the four major types given above. These measures mainly affect T_r:

Mobile unit reliability (y_5)

Mobile unit speed (y_6)

Communication channel performance: e.g., range, reliability, etc. (y_{7_i})

Accident detection and reporting performance: e.g., search-area coverage, scan rate, resolution level, discrimination capability, etc. (y_{8_j})

Mobile unit navigational performance: e.g., position accuracy, beam receiving ability, etc. (y_{9_1})

Mobile unit handling or maneuverability (y_{10})

Payload capability of mobile unit (in terms of number of accident victims) (y_{11})

Accident location coordinates (y_{12})

Mobile unit location coordinates (y_{13})

Operational Strategies. Operational strategies act upon the performance capabilities of the subsystem and thus affect the value of effectiveness for the subsystem. Parameters of operational strategies are:

Accident-detection search rate (S_1)

Mobile unit patrol assignment (S_2)

Receiving hospital assignment policy (S_3)

Mobile unit readiness status (S_4)

Regional coverage policies (involving other recovery systems) (S_6)

Mobile unit velocity mode (S_6)

Replacement policy for mobile unit (S_7)

Location of mobile unit garage (S_8)

Hardware Performance Characteristics. These characteristics are measures of the physical component elements of the subsystem in relation to subsystem performance.

Mobile unit engine horsepower or thrust (X_1)

Mobile unit weight (X_2)

Mobile unit volume (X_3)

Traction factor of mobile unit (X_4)

Moment of inertia (X_5)

Other vehicle hardware characteristics related to performance (X_6)

Environmental Parameters

Visibility (0_1)

Precipitation (0_2)

Terrain characteristics, grade, soil-bearing characteristics, etc. (0_3)

Humidity (rotating or fixed wing units) (0_4)

Temperature (0_5)

SUBSYSTEM RELATIONSHIPS

As described in Chapter 1 various parameters, characteristics, and performance measures must be established in order to determine system effectiveness. This, of course, is true in each of the hierarchial subsystem studies. In the section on Analyzing Effectiveness, a relation for recovery system effectiveness was given in terms of the major performance measures (Eq. 2 and 3). Each of the performance measures can be expressed in terms of the environmental parameter, and hardware and physical characteristics.

Thus, for example, the system response time T_R (Eq. 1) is dependent on the recovery system descriptors y listed above, since each of the various elements of T_R is directly affected by the y variables.

$$T_R = f_1(y_1, y_2, \dots, y_n)$$

Each of the y variables is a function of a subset of hardware characteristics. The control strategies are functions of environmental parameter, performance variables, and hardware characteristics.

INTERSECTION WITH OTHER SUBSYSTEM OUTPUTS

The operational and performance characteristics of the various other subsystems will affect the recovery subsystem effectiveness. The best illustration of this can be seen in Eq. (3), where the term $\Pr\{I_l \cdot \theta_i | A\}$ is the probability of sustaining an injury state (I_l, θ_i) given the occurrence of an accident. Several factors enter into the evaluation of this term. We assume that a taxonomy of injury types and levels has been established. The results of various other subsystem studies, particularly the ones dealing with vehicle design and crash dynamics, will tell us what types of initial collisions produce second-collision forces that result in the given injury types and levels. An accident prediction model will give the distribution in time and the spatial coordinates of the collision types. With this information we can evaluate the probability of sustaining (I_l, θ_i) in any given accident. Thus, vehicle and highway design will have a direct bearing on $\Pr\{I_l \cdot \theta_i | A\}$.

Another interesting kind of subsystem interaction is that in which modification to improve the safe performance of one subsystem element results in a heavier demand upon any of the others. For example, a modification of vehicle design which reduces instantaneous fatalities may produce an increased number of nonfatal injuries requiring medical treatment by the recovery system.

In general, the most significant of the interactions which will concern us are those in which increasing the "safety" of one subsystem affects the performance of the casualty recovery system. For example, in highway design, the introduction of electronically controlled lane occupancy without provision for the rapid passage of emergency vehicles such an ambulance would seriously affect the performance of the casualty recovery system.

SUMMARY

A recovery subsystem effectiveness measure and related performance measures have been discussed. Each class of performance variables relates directly to recovery system design and operation. The performance of the recovery system determines in turn the worth of the subsystem. Note that the effectiveness measure of the recovery system becomes a performance measure of the total highway safety analysis.

The purpose of this chapter is to form a logical framework to determine what a "good" recovery system is, and what variables affect its "goodness." For example, how fast is fast enough for effective recovery; how are other subsystems related to recovery effectiveness; and how might new technology (e.g., new accident-detection techniques) relate to recovery system performance.

Other measures of recovery system effectiveness no doubt exist and could be considered; however, those presented here appear physically logical as well as intuitively reasonable. The concept of reducing the case fatality ratio (or increasing the number of injury survivals) is consistent with the general goal of highway safety; to save lives. Other considerations might include the degree to which physical disabilities and human suffering are prevented or reduced. It should be emphasized that a recovery system is not concerned with the prevention of accidents or the reduction of instantaneous fatalities, except as they relate to demand on the system, and affect the system response. Thus, their design and operation are not positive in the sense of accident prevention as are other elements in the highway safety analysis.

In the following two sections the literature on recovery systems and some recommended areas of research are discussed.

THE LITERATURE

A general search was made in the literature covering ambulance services, case fatality ratio statistics, and other topics related to traffic casualty recovery. This search is not complete, but it is felt that the sample is representative of the general field. Specific references are listed, and a bibliography of additional material is also presented.

THE AMBULANCE: SERVICE AND UTILIZATION

Few studies of ambulance service have been made. Waller (1964; 1965; $1966_{a,b}$) has published a number of constructive reports on the ambulance; the second of these (1965) contains an excellent survey of current ambulance practices, organizational structure, equipment and crew training, financial considerations, speeding ambulance policies, etc. The major problem areas reported in current ambulance service were:

1. Lack of training and low salaries for ambulance personnel; absence of state regulations governing personnel.

2. Equipment deficiencies, in terms of minimum requirements by law (corresponding to our $M_{I_1\theta_i}$); lack of regulation of ambulance design variations.

3. Financial patterns, a major problem contributing to lack of adequate ambulance service. It has been estimated (Waller, 1966_a) that a minimum of 750 to 1000 calls per year are required before an ambulance can fully sustain itself financially without volunteer personnel. This may represent three to five times the annual caseload of the average rural unit. In addition, bills for long ambulance runs, particularly those involving nonresidents, are less often paid. It has been further estimated that it would cost most rural communities about $20,000 to $25,000 annually per ambulance to run their own service, but that they can improve existing private or voluntary services with an annual subsidy of $3,000 to $5,000 per unit.

4. Regulation of ambulance operating procedures is inconsistent and has not been sufficiently studied.

A significant and interesting problem was brought out by Waller (1966_b) in his discussion of the utilization of ambulances. In that study, only 33% of the ambulance trips investigated were related to traffic accidents. Thus, in any analysis aimed at improving recovery systems for traffic accident victims, other patterns of ambulance use must be considered, particularly when operational strategies are involved.

Irma West (1964) presented a study of 5000 ambulance trips which shows the same general trend in ambulance utilization, and also provides some interesting insights into the question of speed, nature of emergency, etc. The information gathered included: time of day, origin, destination, miles traveled, speed limits in effect, maximum speed over these limits, circumstances requiring extra speed, authorization for speed, number of ambulance accidents. Her findings are summarized below.

1. For about 8% of the patients carried by ambulances at speeds above prevailing limits, the attending physician felt the time saved was of benefit. In most of these cases, the patient was seriously poisoned.

2. The average ambulance trip reported was 8 miles and the average time saved by exceeding the speed limits was 2.7 minutes. In about 80% of the trips, the speed limit was exceeded by less than 16 miles per hour.

3. Police directed the ambulance to exceed the speed limit for 36% of the trips, the ambulance crew for 34%, physicians for 17%, and others for 13%.

4. About two-thirds of California's ambulance operators did not report any trips where speed limits were exceeded while patients were transported.

5. Most of the patients carried by ambulances exceeding the speed limit were <u>seriously</u> ill or had been <u>seriously</u> injured. About one-fourth of them were too ill to survive.

6. Almost half of the patients carried by ambulances exceeding the speed limit had been injured, usually in traffic accidents; about 17% were too badly injured to survive. Head injuries were the most frequent kind of serious injury.

7. About 15% of the patients carried by ambulances exceeding the speed limit were heart cases and almost 60% of them were too ill to survive.

8. In most instances, ambulance personnel were accurate in deciding the seriousness of the patient's condition.

9. During the six months of the study 61 traffic accidents occurred to ambulances. Seventy-eight people were injured and one killed. The majority occurred at intersections. However, none of the injuries occurred as a result of ambulances exceeding the speed limits while carrying patients.

In regard to the effect of demand on ambulance service in terms of accident injury distributions, Waller (1963) provides a comparison of urban versus rural environments. In his report, an analysis by the California Department of Public Health of reports filed by the Highway Patrol in 1961, it was stated that traffic accidents injured one and one-half times as many people per 1000 population in rural California counties (under 50,000 people) as in urban counties (over 500,000 people), and that persons injured in rural counties were almost four times as likely to die of their injuries as those injured in urban counties.

> A death certificate study was undertaken of 782 traffic deaths (excluding pedestrians) occurring in rural and urban California counties during 1961. Accidents occurring in rural counties tended to be single vehicle accidents which resulted in less severe injuries, while those in urban counties tended to be two vehicle and multiple vehicle accidents resulting in more serious injuries. The anatomical distribution of injuries was the same for both urban and rural accidents. However, people dying in rural accidents more frequently died at the scene of the accident, died sooner after injury, and died of less serious injuries than did those injured in urban accidents. For injuries where theoretically few lives could be salvaged by prompt emergency care, the time between injury and death was about the same in urban as in rural counties. Where such care should delay or prevent death because the injury was possible or probably salvageable, those injured in rural counties died more quickly [sic].
>
> Thirty-two percent of fatalities in rural counties happened to urban and out-of-state residents, while only 12 percent of fatalities in urban counties were to rural or out-of-state residents, suggesting that traffic accidents to nonresidents may place an excessive load upon medical care resources in rural areas (Waller, 1963).

INJURY INVESTIGATIONS

The measure of effectiveness for recovery systems, as discussed in previous sections, will require quantification of injury level and type. A preliminary search of the literature for a "taxonomy of injuries" showed that little has been published in this area. Braunstein (1957) attempts a semiscientific categorization of traffic injuries but does not indicate any quantitative relation between injury severity level and chance of survival. There are numerous other papers filled with statistics on various classes of injury and their frequencies; none, however, would seem to be of use in our model.

Most of the medical articles dealing with injuries discuss treatment of injuries in terms of techniques of treatment, but again do not indicate a relative urgency of treatment for specific injuries (type and level) except to note that "treatment should be given as soon as possible." The bibliography cites several of these articles.

SUMMARY

It was concluded from the preliminary literature search carried out in this study that there have been relatively few useful studies of casualty recovery as an integral part of traffic safety. By useful, it is meant that no attempt has been made to analyze the recovery problem in the context of a general systems approach to determine the underlying characteristics of the recovery problem, and to establish techniques that will provide answers to such questions as posed previously in this report. It seems, in fact, that there has been surprisingly little activity in determining whether we have adequate ambulance service, or in defining what adequate service is.

The major points covered in the literature are concerned with better crew training, standardized laws for regulating ambulances, and equipment for ambulances. Discussions of new techniques proposed for recovery system operation are offered in Maynard (1965) and Lee (no date available); these techniques include electronic intersection accident detection systems and traffic signal devices for passage of emergency vehicles.

The literature had little to offer in the area of relating injury level to survivability. No attempt seems to have been made by the medical profession to quantify the relationship between trauma, survival, and treatment level.

While excessive speed in current ambulances has been reported as generally unnecessary, this does not imply that time is a secondary factor. In many cases, the fact that the ambulance arrived at the scene and possibly administered some "delaying action" treatment was a time element itself, and is compatible with our model in which T_R and $M_{I_1 \theta_i}$ both contribute to survival.

One of the most significant existing problems pointed out in the literature is the sharp difference in case fatality ratios between urban and rural areas. This would seem to be a current problem where attention might be focused.

In addition to activities reported in the published literature, there are some current projects underway that directly relate to casualty recovery. One of these is a study being conducted by the National Research Council (NRC) of the National Academy of Sciences under Dr. S. Seeley. This effort is a survey and commentary on emergency transportation and medical treatment facilities in a national setting. The NRC will release a "white paper" regarding this subject and will point up areas where improvements are needed.

It was recently noted in Automotive News that the University of North Carolina has received a $400,000 grant to study accident detection and emergency centers. At this writing, no further information was available.

In addition to the listed references, a general bibliography is presented at the end of this chapter. Further data and a more complete discussion of the topics described above may be found in these source citations.

CONCLUSIONS AND POSSIBLE AREAS OF RESEARCH

CONCLUSIONS

The ability to provide medical treatment and transportation for traffic accident victims is a major element in the health resources of a region. In view of this need, it is surprising that so few professionals in the health fields have actively concerned themselves with the problems of casualty recovery, and particularly ambulance service. Medical doctors are seldom carried on ambulance runs; indeed, if other emergency service personnel were adequately trained for this service, the physician's skills would be better utilized in the treatment center setting. However, the fact remains that in general the health professions have not expressed an active concern with developing casualty recovery resources.

It is reasonable to state that there is merit in a systems approach to the casualty recovery problem. At present, the situation is not clear in terms of system performance, for there are no performance criteria by which to judge emergency service. Various writers in the professional and public press have called for improved ambulance service, yet few seem to have evaluated, in any scientifically sound sense, the performance of existing services. The question of what constitutes good emergency service has not been answered. Most of the difficulty in the present system seems to rest in a loose and ill-defined structure of ambulance services operating under a multitude of regulations and situations. Principles of management science and new technology have apparently not been applied to the design and operation of ambulance services, and the entire system operates under limiting financial constraints.

It is concluded that in order to determine the effectiveness of present emergency medical recovery operations, and to evaluate design of future systems, performance criteria must be established.

PROPOSED RESEARCH AREAS

The present investigation has attempted to outline some basic approaches for analyzing the casualty recovery problem. Some assumptions regarding system effectiveness and performance have been made in terms of various recovery system characteristics. In order to test the validity of these assumptions and the major hypotheses that accompanied them, some specific areas of research are suggested. In addition, more general research topics, particularly concerning the operations of present systems, are discussed.

In outlining the research areas, the major objective is twofold: (1) to develop the proper data and relationships for testing the hypotheses, and (2) to use the composite hypothesis (if accepted; if not, hopefully to generate a new one) for evaluating present recovery systems and for designing new ones.

Relationship Between Response Time and Case Fatality Ratio. A hypothesis was made that the case fatality ratio was related to recovery response time. In addition, it was stated that case fatality ratio is also related to accident type and medical treatment level. Testing these hypotheses would be a major element in a casualty recovery research program.

Two separate but related approaches present themselves. The first might be viewed as macroscopic, the second as microscopic.

(1) Sample Measurement of Case Fatality Ratio and Response Time. This study would effectively be research into the M-process in the post-injury, pre-treatment process, conditional on a specific time lapse (T_R). If T_R has strong influence on the value of H at the time (injury time + T_R), T_R has an effect on the CFR, which is determined wholly by H(injury time + T_R) and M, the treatment provided.

For several presently operating ambulance services, the response time T_R distribution and case fatality ratio would be determined either from record data, if available, or from actual on-site measurement, if possible. Different types of ambulance services would be studied (e.g., those operated by private firms, by police, by volunteers, etc.) in an attempt to obtain a wide variation in response time, if one exists.

In order to isolate the effect of accident type distributions, systems with comparable distributions would be studied. The effect of comparing two areas where the number of ambulances are different would be reflected in response-time measurement.

From the literature, it was noted that a significant difference in CFR existed between injuries incurred in an urban environment as compared to a rural setting. It well may be that here the major effect of time is most apparent, e.g., urban recovery systems might all fall within the same response-time interval and CRF level, whereas rural systems may fall within a higher T_R interval and comparably higher CFR level.

The effects of different "on-board" treatment levels will have to be determined. At present, it is believed these effects can be ignored, at least for first-order effects. In addition, the influence of variation in hospital emergency receiving facilities must be accounted for.

In regard to the response-time model, the various time elements in the model (Eq. 1) form the basis for the data-gathering task. This is not to say information will be readily available for all elements; for example, detection time may be a problem, particularly for rural environments. It is thought that trip times to and from the scene and the time spent on scene might be available from ambulance firms and police records. The difficulty in this approach is, of course, that one must be extremely careful in interpreting CFR and T_R data for two different areas due to possible hidden effects in the data such as the injury level and type distribution. A regression analysis of the data would only indicate a trend, and not necessarily a cause and effect, in the relation. This leads us to the second possible area of investigation.

(2) Medical Establishment of Injury-Survivability Times. The actual effect of time upon survivability of an injured person is a medical problem. Aside from superficial cuts and abrasions, most accident injuries require some medical attention to ensure satisfactory recovery.

It is suggested here that a medical evaluation of the effect of time on various injury types be carried out. A taxonomy of accident injury types, relating to body area, will be necessary. Next, an experimental program to determine medical relationships involving injury level and time for given types would have to be conducted. An experimental program in this area may be very difficult.

One possible approach might be to survey qualified physicians to obtain estimates of survival time for a taxonomy of injury types. Such a survey, of course, would be based on the physician's individual experience with auto accident victims and would at best be approximate. However, it could provide correlation to a study of the type described at the outset of this section, where injury type distribution could be obtained along with T_R and CRF.

In Appendix 4-A the relationship between injury level and impact level is developed in a functional form. The required relationship for evaluating the injury level from accident "type" is described by the functional equation.

Design and Operations Oriented Research. Additional areas of research necessary for casualty recovery systems analysis are listed below.

(1) Determine useful techniques for studying the location of ambulance services and emergency clinics: Treating the recovery system as a resource, allocation studies to meet a predicted demand might reveal more efficient patterns of ambulance center location and distribution. Consideration of available hospital emergency treatment facilities would be an important element in determining location of ambulance services.

Closely related to the design of an effective system is the hospital's organization of resources for the arriving emergency patient. Although in most areas the hospital is adequately prepared to meet the patients' needs as soon as he arrives, this is not always the case, and rapid and efficient detection, recovery, and transportation may end at the hospital with a long wait for proper diagnosis and treatment. Thus hospital response must be integrated with the recovery system if the overall goal is to be achieved.

(2) Analyze regional distribution of accident injury demand in terms of the possibility of providing regional ambulance coverage with a single centralized communication center to provide optimal scheduling: Allocation of the ambulance service to other demand sources would undoubtedly be a significant factor, since traffic accidents are only a portion of total demand. Studies of the cost effectiveness of including several communities within a single ambulance coverage area policy would be required. The development of objective functions that relate fixed and operational costs to some desired performance level would be necessary to determine optimal coverage areas and equipment acquisition.

(3) Investigate the feasibility of applying new technologies to the casualty recovery problem. For example:

(a) Feasibility studies of creating mobile field emergency hospitals such as the Army's Mobile Unit for Surgical Treatment (MUST).

(b) Detailed cost effectiveness studies of replacing ambulances with helicopters in certain areas.

(c) Feasibility studies of using electronic or other sensing devices for accident detection and reporting.

The research strategy suggested in this chapter is illustrated by a developmental flow sketch presented in Fig. 4-2.

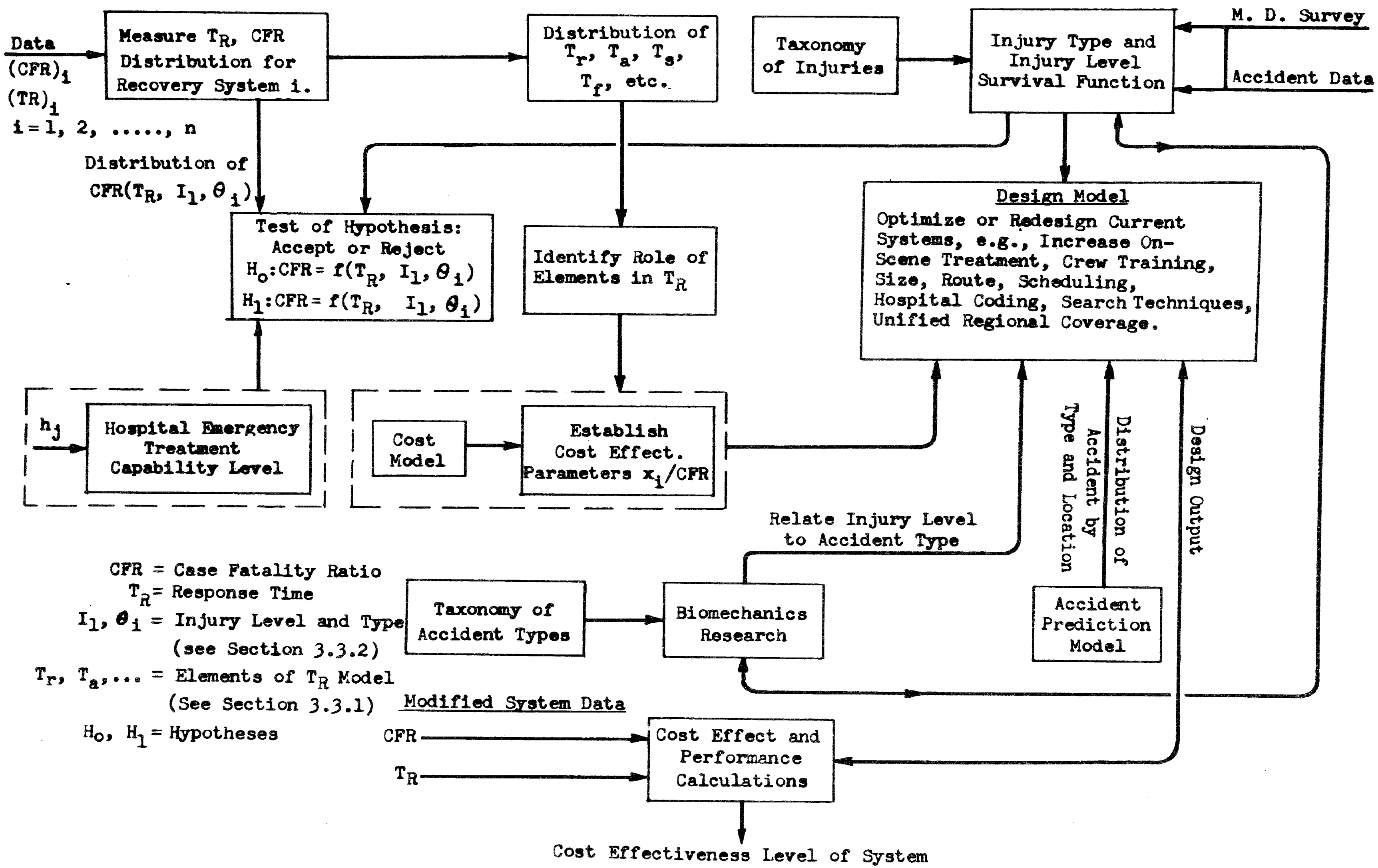


FIGURE 4-2. DEVELOPMENTAL FLOW OF CASUALTY RECOVERY RESEARCH

Appendix 4-A

PROBABILISTIC STRUCTURE FOR RECOVERY SYSTEM MODEL

The relationship between survival and recovery system performance is developed as follows:

$\Pr\{S | I_l \cdot \theta_i \cdot A\}$ = the conditional probability of survivals given an accident has occurred and injury level I_l and injury type θ_i have been sustained.

$\Pr\{I_l \cdot \theta_i | A\}$ = the conditional probability of sustaining injury level I_l and type θ_i.

The marginal probability of survival given an accident is found by summing the joint conditional probability of survival over all possible values of I_l and θ_i.

$$\Pr\{S|A\} = \sum_{I_l}\sum_{\theta_i} \Pr\{S|I_l \cdot \theta_i \cdot A\}\Pr\{I_l \cdot \theta_i | A\}$$

For a given injury type θ_i, it is assumed that survival time (without treatment) is a function of I_l, and thus injury level could be defined by survival time. If we treat injury level as a continuous function, we can write the following, assuming an inverse relation with survival time:

$$\Pr\{t_s \geqq t | \theta_i\} = \Pr\{I_l \geqq y | \theta_i\}$$

where t_s = survival time
t = time
y = a fixed injury level.

Using T_R as the random variable for recovery system response time and M as medical treatment capability, we can write the following:

$$\Pr\{S| I_l \cdot \theta_i \cdot A\} = \Pr\{T_R \leqq t\}\Pr\{t_s > t|\theta_i\}\Pr\left\{M > M_{I_l \theta_i}\right\}$$

equivalently we can write

$$\Pr\{S| I_l \cdot \theta_i \cdot A\} = \Pr\{T_R \leqq t\}\Pr\{I_l \leqq y|\theta_i\}\Pr\left\{M > M_{I_l \theta_i}\right\}$$

Note that we are not considering those injury levels where the probability of survival without treatment is 1.0, or where the probability of death with treatment is 0. These states represent elements in the set of possible states that are limiting in nature and are not treated in the model.

To relate injury level I_l to an accident impact model, the following is presented.

Let:

A = accident event

I_l = lth level of injury

t_k = kth level of force transmitted to passenger

x_i = ith level of total impact force

then:

$$\Pr(I_l|A) = \sum_k \Pr(I_l|A\cdot t_k)\Pr(t_k|A)$$

where,

$$\Pr(t_k|A) = \sum_i \Pr(t_k|A\cdot x_i)\Pr(x_i|A)$$

thus,

$$\Pr(I_l|A) = \sum_k \Pr(I_l|A\cdot t_k)\left\{\sum_i \Pr(t_k|A\cdot x_i)\Pr(x_i|A)\right\}$$

Thus, we can now see the contribution required to evaluate the injury level sustained in an accident A.

$\Pr(I_l\|A\cdot t_k)$	is a biomechanical measure relating physical injury to the forces sustained by a passenger in an accident.
$\Pr(x_i\|A)$	is a mechanical measure relating to the magnitude of forces incurred in an accident.
$\Pr(t_k\|A\cdot x_i)$	is a mechanical measure of forces sustained by a passenger when force x_i is applied to the vehicle.

REFERENCES

Braunstein, Paul W., "Medical Aspects of Automotive Crash Injury Research," Journal of the American Medical Association, V. 163, No. 4, Jan. 26, 1957.

Lee, Richard C., "Safe Movement Plans Are Discussed for Emergency Vehicles," Traffic Quarterly.

Maynard, H. L., Accident Sensing and Surveillance System, Cornell Aeronautical Laboratory, Inc., Buffalo, N. Y., Oct. 1965.

Waller, Julian A., California Progress Report, speech to the California Ambulance Association, July 17, 1963.

Waller, Julian A., Robert Curran, and Frank Noyes, "Traffic Deaths: A Preliminary Study of Urban and Rural Fatalities in California," California Med., V. 101, 1964, p. 272.

Waller, Julian A., "Ambulance Service: Transportation or Medical Care," Public Health Reports, V. 80, No. 10, Oct. 1965, pp. 847-853.

Waller, Julian A., "The Control of Accidents in Rural Areas," Presented at the 19th National Conference on Rural Health, Colorado Springs, Colorado, March 19, 1966_a.

Waller, Julian A., "Utilization of Ambulance Services in a Rural Community," American Journal of Public Health, V. 56, No. 3, Mar. 1966_b, pp. 513-520.

West, Irma, et al., "Study of Emergency Ambulance Operations—A Preliminary Report," International Rescuer, Nov.-Dec. 1964, p. 9.

SELECTED BIBLIOGRAPHY

The Ambulance and Related Topics

Bold, J. D., Emergency Action for Motorists, South African Road Safety Council, 1963, 33 pp.

Brennan, William T., Guideline Inventory, A Survey for the Emergency Aid and Transportation Program, Department of Health and Safety, Indiana University, 1964, 29 pp.

Hampton, Oscar P., Transportation of the Injured, A Report, American College of Surgeons, 1960.

Transportation of the Injured, Carl B. Young, Jr., Ed., Charles C. Thomas, pub., 1958, 238 pp.

A Guide for Operating Ambulance Fleets, National Safety Council, 1965, 32 pp.

Final Report on Ambulance Traffic Accidents, National Safety Council, July-Dec., 1961, 10 pp.

Fundamentals of First Aid for Ambulance Attendants, Committee on Trauma, North Carolina Chapter, American College of Surgeons, 1963, 14 pp.

Health, Medical Care and Transportation of the Injured, U. S. Government Printing Office, 1966, 26 pp.

Immediate Care and Transport of the Injured, George C. Curry, Ed., Charles C. Thomas, pub., 1965, 187 pp.

Regulations and Laws Relating to Ambulance Operation in California, California Department of Motor Vehicles, 1964, 28 pp.

Bothwell, P. W., "Ambulance Design," 7th Stapp Car Crash Conference Proceedings, 1965, pp. 324-344.

Curry, George J., and Lyttle, Sydney N., "The Speeding Ambulance— Does It Save Lives?," *Traffic Safety*, Aug. 1958, p. 7.

Kearney, Paul W., "After the Accident What Do You Do?," *Family Safety*, Summer, 1964, pp. 12-15.

Wade, Preston A., "Care of Injured in Traffic Accidents," National Safety Congress Transactions, V. 28, 1958, pp. 78-81.

Wade, Preston A., "What's Wrong with First Aid?," *Traffic Safety*, Jan. 1959, pp. 8-10.

Waller, Julian A., "What's the Real Trouble with Ambulance Services?," *Traffic Safety*, Oct. 1965, pp. 10-11, 35.

Wolfrum, C. E., "Police Helicopter Operations," *Traffic Digest and Review*, Sept. 1957, pp. 1-3.

Young, Carl B., "Traffic Accidents, First Aid, and the Police Officer," *Police*, Jan.-Feb. 1957, pp. 33-37.

"Balm for the Victims," *Journal of American Insurance*, Aug. 1964, pp. 24-25.

"Care of the Injured," *National Safety News*, Mar. 1959, p. 244.

"Emergency Road Service - What it Means," *American Motorist*, Jan. 1965, p. 12.

"Hot Line to Rescuers," *Journal of American Insurance*, Nov. 1964, pp. 18-20.

"The Handling and Treatment of Road Accident Casualties: A Survey Carried Out in South Africa," *International Road Safety and Traffic Review*, Winter 1964, p. 11.

"Mercy Missions or Mayhem?," *Journal of American Insurance*, March 1964, pp. 28-29.

West, Irma, "A Look at Ambulance Standards and Services in California," *California's Health*, Vol. 21, No. 3, August 1, 1963.

Merrill, Malcolm, Opening Address, California Ambulance Association Convention, Anaheim, California, Apr. 3, 1963.

Accident Injury Studies

DeHaven, H., *Accident Survival* - Airplane and Passenger *Automobile*, Cornell University Medical College, Jan. 1952.

DeHaven, H., *Crash Deceleration, Crash Energy, and Their Relationship to Crash Injury*, Wright-Patterson Air Force Base, Ohio, Dec. 1950.

Kulowski, Jacob, Crash Injuries: The Integrated Medical Aspects of Automobile Injuries and Deaths, Charles C. Thomas, Pub., Springfield, Ill., 1960, 1080 pp.

Mathewson, J. H., The Principle of Accident Effect Reporting, Institute of Management Sciences and The Operations Research Society of America, May 1, 1959, 23 pp.

Roth, H. P., Impact and Dynamic Response of the Body, Protection, Inc., Los Angeles, 1952.

Roth, H. P., Physical Factors Involved in Head-on Collisions of Automobiles, Institute of Transportation and Traffic Engineering, Los Angeles, 1952.

Schwimmer, Seymour and Wolf, Robert, Leading Causes of Injury in Automobile Accidents, Cornell Aeronautical Laboratory, 1962.

Stapp, J. P., Tolerance to Abrupt Deceleration, Butterworth's Science Publications, London - AGARDograph No. 6, 1955.

Starks, H. J. H., Research on Injuries Sustained in Road Accidents, Road Research Technical Paper No. 37, 1956, 14 pp.

Taylor, E. R., Physiologic Effects of Abrupt Deceleration, Aerospace Medical Association, Elmhurst, Ill., Apr. 1962.

Young, Mary F., Medical Approach to Intensive Investigation of Motor Vehicle Accidents, Northwestern University Traffic Institute, 1958, 49 pp.

Medical Aspects of Traffic Accidents, Proceedings of the Montreal Conference, Harold Elliott, Ed., 1955, 505 pp.

On Medical Aspects of Motor Vehicle Accidents, Bulletin of the Buchanan County Medical Society, Oct. 1956, 24 pp.

Aldman, B., "Biodynamic Studies on Impact Protection," Acta Physiologica Scandinavica, V. 56, Supplement 192, 1962, pp. 1-80.

Bierman, H. R., "Protection of the Human Body from Impact Forces of Fatal Magnitude," The Military Surgeon, V. 100, No. 2, Feb. 1947, pp. 125-141.

Curry, George J., "Symposium on the Medical Aspects of Traffic Safety and Related Injuries," American Journal of Surgery, Oct. 1959, pp. 517-583.

Gessane, William, "The Prevention of Injury and Death from Traffic Accidents," Police, Jan.-Feb. 1961, pp. 51-54.

Kulowski, J., "Medical Aspects of Motor Vehicle Accidents," Bulletin of the Buchanan County Medical Society, V. 9, 1956, pp. 5-22.

Moseley, Alfred L., "Variables in Automobile Death Cases," Technical Aspects of Road Safety, No. 19, 1964, pp. 2.1-2.5.

Rutley, K. S. and Cook, J. A., "Crash Injuries," Automobile Engineer, Feb. 1962, pp. 55-57.

Stapp, J. P., "The Work of the Aero-Medical Laboratory, U. S., on Crash Forces and Injuries," *Research into Road Safety*, Organization for European Economic Cooperation, 1960, pp. 19-25.

Stewart, Rober G., Hill, Vern L., and Warmoth, Edward J., "Non-Collision Fatal Accidents: How Many can be Prevented," *Traffic Safety*, Research Review Supplement, Dec. 1959, pp. 26-28.

Detection and Communication

Kuprijanow, A., Pogust, F., and Forster, H., *Means of Locating and Communicating with Disabled Vehicles*, Interim Report, Highway Research Board, NCHRP Report No. 6, 1964, 56 pp.

Hulthen, Lamek, "A Note on Detection Probability and Observance in Road Traffic," *Aspects Techniques de la Securite Routiere*, No. 9, Mar. 1962, pp. 4.1-4.8.

Rostkowski, Edmond, "Hospitals on Wheels," *National Safety News*, June 1959, p. 70.

"The Reporting of Accidents for Recording and Information Purposes," *Ninth Pan American Highway Congress*, Chapter V, Topic 17, Document No. 171.

Hospital Emergency Facilities

A Model of a Hospital Emergency Department, The Committee on Trauma, American College of Surgeons, Cornell Aeronautical Laboratories, 1961.

Gissane, W., "Birmingham Accident Hospital," Proceedings of the Royal Society of Medicine, V. 1, 46, 1953, pp. 449-456.

5 DEGRADATION

JOHN DUNCAN

INTRODUCTION

The general difficulty with a description of performance degradation is that the various capabilities and performance measures of the highway transportation system are complex stochastic functions of time and usage. To determine how degradation affects accidents, then, one must know the extent to which a subsystem has been degraded. The related problem of maintenance and control is then expressed in the following terms: given a particular degradation process, we may rebuild or otherwise alter the physical system at times determined by age, usage, or possibly by some measure of performance degradation (an inspection problem). Research in this area should answer such questions as What would be the outcome of such considerations? and What is the best way to implement them? Various portions of the mathematical model necessary to study these problems are outlined in the following sections. Some discussion of the various components and subsystems of a transportation system and associated degradation processes are presented.

Most degradation studies are actually studies of hardware and performance measures for individual components or component groups. The only general topic needing additional research is the relation of subsystem performance degradation to component degradation. This area has had almost no study for any but the simplest subsystems in which performance is rated either adequate or inadequate and in which the failure of any component means the entire subsystem fails.

STOCHASTIC MODELS FOR DEGRADATION

A very general notation for the temporal behavior of component performance degradation is the real-valued function y(t). Such functions may be thought of as generated by a stochastic process. The process associates with an event (i.e., aging or operation), a function y(t). The value of y(t) for any t is thus randomly generated by the process. In order to describe completely the realization of the event we must specify the entire function y(t); i.e., the ensemble of values generated for the time domain over which the process is defined. Since the function is random the likelihood of a particular value $y(t_i)$ at time t_i may be described by some distribution, e.g.,

$$F(y_i/t_i)$$

The likelihood of a particular value $y(t_i)$ at t_i and a particular value $y(t_j)$ at t_j may be described by a joint distribution:

$$F(y_i, y_j | t_i, t_j)$$

One could continue in this fashion to develop distributions of any order

$$F(y_1, \ldots, y_i, \ldots | t_i, \ldots, t_i, \ldots)$$

Within this general framework there are special processes which have been studied intensively in connection with reliability and other engineering problems.

THE GAUSSIAN PROCESS

The distribution $[y(t_1), \ldots, y(t_i), \ldots, y(t_h)]$ is multivariate normal and is completely characterised by $E\{y(t_i)\}$, where $i = 1, \ldots, n$.

$$\text{Cov}\ \{y(t_j), y(t_k)\} \qquad \begin{array}{l} j = 1, \ldots, n \\ k = 1, \ldots, n \end{array}$$

The wide applicability of this process derives from two inherent properties:

(1) The joint distribution for $y_1, \ldots y_i, \ldots, y_n$ can be constructed knowing only the pairwise distributions for all combination (y_i, y_j).

(2) Processes constructed from linear functions of $y(t_1), \ldots, y(t_n)$ are themselves Gaussian.

The Gaussian process in its full generality is of limited practical importance since one cannot estimate the infinite number of parameters

$$E\{y(t_j)\},\ \text{Cov}\ \{y(t_j), y(t_k)\}$$

Instrumentation and control engineers began work on the wide-sense stationary version of the process in the years shortly before World War II.

$$E\{y(t_j)\} = \mu \text{ for all } t_j$$

$$\text{Cov}\ \{y(t_j),\ y(t_k)\} = \sigma^2(t_k - t_j)$$

The process in this form has limited application to reliability because the performance measure, y, does not degrade with time.

Some work by C. Hamilton (1963) indicates that the nonstationary process

$$E\{y(t_j)\} = \beta e^{-\gamma t} \text{ for } \alpha \geqq 0,\ \beta \geqq 0,\ t > 0$$

$$\text{Cov}\ \{y(t_j),\ y(t_k)\} = \sigma^2(t_k - t_j)$$

is useful for predicting degrading performance measures.

Gaussian Markov Process (Autoregressive Processes)

$$y(t_j) = \lambda y(t_i) + Z(t_j)$$

where $Z(t_j)$ is Gaussian

λ is a regression coefficient for the sequence $t = t_0, t_1, \ldots, t_i, t_j, \ldots$

Properties

(1) $y(t_i)$ is a continuous random variable

(2) $\Pr\{y(t_j) < Y | y(t_i),\ i = 0, 1, \ldots, j - 1\}$
$= \Pr\{y(t_j) < Y | y(t_{j-1})\}$

This process has been assumed for many signal detection and sales forecasting problems. It might be worth its salt for modeling some of the continuous performance measures which are periodically inspected. Hamilton (1963) found this to be the case for electronic components. Possible applications in highway maintenance could be to bearing wear and brake shoe wear.

Semi-Markov Processes

$y(t_i)$: has a finite number of values, each of which is called a state.

$F_{ij}(t)$: is the waiting time distribution for the process to pass from state i to state j.

P_{ij}: is the probability that a system in state i passes to state j.

Heuristically, the process begins in state i and passes to state j in time $t_j - t_i$ with probability $P_{ij} dF_{ij}(t)$, $t = t_j - t_i$. A realization for such a process would look like Diagram A.

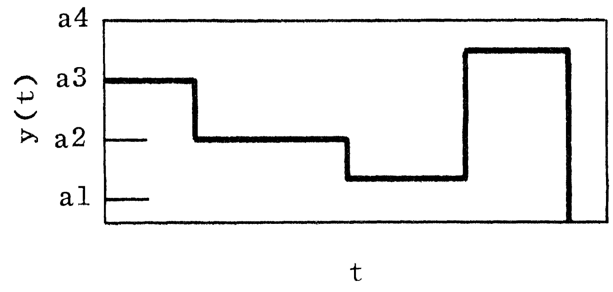


Diagram A.

This process has the important property that

$$\Pr\{a_i \rightarrow a_j \mid t_i, \ldots, t_i, a_1, \ldots, a_i\} = P_{ij}$$

and therefore the probability of a transition from i→j is dependent only on the state i and not upon prior states.

Poisson Process. There are a number of counting processes or point processes which have been used in maintenance models. The processes generate events or epochs at specific times, say $t_1, \ldots, t_n, \ldots$. In the maintenance problem these might be equipment breakdown times. For the Poisson processes, the intervals between these epochs are independent and identically distributed according to

$$F(t) = 1 - e^{-h(t)t}, \qquad h(t) \geqq 0$$

DEGRADATION MEASURES

The measures given in this section are functions of stochastic process realizations or distributions. The processes under consideration depend upon the parameters age or usage. Both the range of these functions (the sample space) and parameter may be either discrete or continuous.

The purpose of these measures is to provide an ordering of stochastic processes which indicates how component service may vary with one of the two parameters. In this way one is able to compare the degradation of different types of components as well as the resistance to degradation provided by different maintenance policies.

It should also be possible to relate these measures to collision or breakdown frequencies. Ultimately it may even be possible to predict collisions or breakdowns using such measures. The choice of these measures for each task will no doubt be based upon trial and error.

A brief description of special stochastic processes that generate the random functions y(t) was given in the preceding section.

MEANS

(1) Expected Pointwise Performance, EPP

$$y(t) = E\{y(t)\} \tag{1}$$

A measure of the average performance of identical components at a given age, t.

(2) Expected Interval Average Performance, EIAP

$$\overline{\langle y(t)\rangle}_{ab} = E\left\{\frac{1}{b-a}\int_a^b y(t)\ dt\right\} \tag{2}$$

For each member of a class of degrading components its time average performance over the interval [a, b] is the random variable,

$$\overline{\langle y(t)\rangle}_{ab} = \frac{1}{b-a}\int_a^b y(t)\ dt \tag{3}$$

which will be called the interval average performance, IAP. The expectation of this random variable is a measure of the average over all members of the class of the IAP.

(3) Interval Average EPP, IAEPP

$$\overline{\langle y(t)\rangle}_{ab} = \frac{1}{b - a}\int_{a}^{b} \overline{y}(t)\, dt \qquad (4)$$

The distinction between IAEPP and EIAP should be made. The former is the interval average of a process measure already defined. Hence for some class of degrading components we have its EPP. The average of EPP over the interval [a, b] for this class is the IAEPP. Only in special cases will IAEPP = EIAP.

VARIANCES

(1) Variance of Pointwise Performance, VPP

$$\overline{\overline{y}}(t) = E\{[y(t) - \overline{y}(t)]^{2}\} \qquad (5)$$

The component performance for a specific time t is the random variable y(t). The variance of this random variable measures the performance spread among members of the class of components.

(2) Variance of IAP, VIAP

$$\overline{\overline{\langle y(t)\rangle}}_{ab} = E\left\{[\langle y(t)\rangle_{ab} - \overline{\langle y(t)\rangle}_{ab}]^{2}\right\} \qquad (6)$$

LOSSES

Up to now the measures have not strictly described the loss in service. It is a simple matter to define other functions of the former degradation measures which do.

(1) EPP Loss Ratio, EPPR

$$D(a, b) = \frac{\overline{y}(a) - \overline{y}(b)}{\overline{y}(a)}, \quad a > b \qquad (7)$$

The numerator is the difference between the EPP's for a class of degrading components at two points in time, a, b. The D(a, b) for a component class has an interpretation quite similar to the failure rate discussed later under Failure Distributions. Consider a class where $y(t) \geqq 0$ for all t. Then D(0, b) is the fractional loss in the expected pointwise performance function of time b.

(2) EIAP Loss Ratio, EIAPR'

$$D_{12} = \frac{\overline{<y(t)>}_{ab} - \overline{<y(t)>}_{cd}}{<y(t)>_{ab}} \tag{8}$$

where $a < b < c < d$.

RELIABILITY FUNCTIONS

With the exception of the loss ratios defined above, the measures given so far do not necessarily measure loss but rather certain average properties of degrading performance. A common method of measuring loss requires a partition of the PM sample space into two regions, satisfactory performance, S, and unsatisfactory performance, S'. This partition is defined mathematically by what is termed a reliability function, g. The function has the property that

$$\begin{aligned} g(y) &= 1 \qquad PM \in S \\ &= 0 \qquad PM \in S' \end{aligned} \tag{9}$$

The reliability function transforms a continuous or discrete PM process into a two-state process. Of course, there are many natural two-state processes in a highway system; for example, all light bulbs which are considered functional if they shine at all. For these processes there is no need to define a reliability function.

POINTWISE RELIABILITY, PR

$$r(t) = E\{g(y(t))\}$$

This measure is sometimes called pointwise availability since it is the probability at time, t, that a component is in S.

RELIABILITY

$$r(t_1, t_2) = Pr\{g(y(t)) = 1 | t_1 \leqq t < t_2\}$$

In other words, the probability that y(t) is in S for the interval (t_1, t_2); usually, $T_1 = 0$.

COMPONENT LIFE

When one is interested in component life, the above measures do not suffice. Suppose initially the component is operating satisfactorily, i.e., $y(0) \in S$. If we watch the component until it fails,

$y(t) \in S'$, we may construe $\tilde{t}$ as the random component life and $g(y)$ as the life function. Note that

$$\begin{aligned} &\Pr\{g(y(x) = 1) \mid x \in (0, t)\} \\ &= \Pr\{\tilde{t} > t\} = 1 - L(t) \end{aligned}$$

and expected life may be defined as

$$\bar{t} = \int_0^\infty t dL(t)$$

There are certainly other degradation measures that can be constructed as the need arises. Ultimately, however, the importance of each of those mentioned will depend upon the role of the component in the operations model, upon the actual character of the underlying processes, and upon the ease with which they can be estimated and manipulated.

To illustrate the use of these degradation measures, consider as the PM the coefficient of friction, μ, of a pneumatic tire with dry concrete. Assume μ varies with tire mileage t, a usage parameter, and that tires are considered dangerous or unsatisfactory if $\mu < 0.1$. When a batch of tires has just rolled off the assembly line, one might be interested in describing the initial performance characteristics of the batch. To accomplish this, one could use[1] $\overline{\mu(0)}$ or the average μ over the new batch. One may also describe the quality control of the manufacturing process by the PR of new tires, $r(0)$. This measures the percentage of new tires which are satisfactory.

After the tires have been put to use, degradation could be measured in many ways. For example, one might measure tires for a period of 20,000 miles ($b = 20{,}000$ miles) and describe degradation by estimating

$$\text{EPP}, \ \overline{\mu(20{,}000)}$$

$$\text{EIAP}, \ E\left\{\frac{1}{20{,}000}\int_0^{20{,}000} \mu(t)\,dt\right\}$$

$$\text{EPPR}, \ \frac{\overline{\mu(0)} - \overline{\mu(20{,}000)}}{\mu(0)}$$

---

[1] Actually, the estimate of $\overline{\mu(0)}$.

and reliability, $r(\partial, 20{,}000)$. Notice that an estimate for $r(\partial, 20{,}000)$ is the percentage of tires which performed satisfactorily for 20,000 miles. Finally, if one observed all tires until $\mu < 0.1$, one could estimate the expected life, $\bar{t}$.

PREDICTION AND STATISTICAL MODELING

The idea of model building is to abstract a real physical entity or phenomena so that it might be studied more conveniently. The modeling procedure is more of an art than a science since it strives for simplicity without sacrificing the essential character of the real entity. By modeling a real thing one hopes to expose the important characteristics while suppressing the unimportant details.

The statistical modeling of component degradation has two major uses:

1. To provide input data for the collision model and traffic flow model via simulation or to estimate the measures of component degradation for them. It would of course be possible to treat these models as part of the larger models.

2. To become part of the models for maintenance optimization.

For any subsystem which contains more than one degrading hardware component the modeling task can be approached in three basic ways. In each of these the first step is to define the PM.

Procedure 1

A. Select a DM for the entire subsystem which describes what is thought to be an important characteristic of its degradation.

B. Estimate this DM by building the several subsystems and observing their temporal degradation.

C. Most of the DM's given in the section on Degradation are statistics of the realizations (i.e., the observed PM's vs. time).

Provided the number of subsystems sampled is sufficiently large, good estimates of the DM's will be available by averaging over the realizations. When the samples are small it may be necessary to hypothesize the stochastic process to get good estimates. This procedure provides little insight into the operation of individual components of the subsystem and has little predictive value.

Procedure 2

This procedure is essentially the same as the first except that the entire stochastic PM process is sought (or perhaps the distribution of some random variable such as subsystem life).

A. Hypothesize a stochastic process for the PM

B. Design an experiment to estimate the unknown parameters of the process

C. Sample the temporal degradation for several fabricated systems

D. Estimate the unknown parameters and perform tests to determine the validity of the original hypothesis

The procedure outlined is quite heuristic since it begins by hypothesis and ends with a test. There is no guarantee that this procedure will ever converge to a suitable model, but it is the standard approach. In general this procedure will require more data than the first; however, it also provides more information. In fact, any of the DM's can be predicted (estimated) from the model process. However, since we have only modeled the PM process for the subsystem there is no way to relate this model to individual components.

Procedure 3

In some instances the degradation of an ensemble of components is desired when only the degradation of individual components is known. This would be the case before the design or change in a complex system. The steps in this procedure are:

A. Specify the components of a system, select their PM's, develop functional, probabilistic, or logical qualitative relations among them (i.e., find the paper subsystems model).

B. Carry out procedure 1 or 2 for each component.

C. Simulate, using A and B, data describing the subsystem PM degradation.

D. Carry out procedures 1 and 2 for the simulated data. There are alternatives to steps C and D which have been employed for simple system structures. The first is to find analytically the relation between subsystem DM and component degradation. This has been done, as will be discussed, for series and parallel "block box" models using the reliability DM. The second is to find analytically the subsystem PM process in terms of the component processes. This is possible for some of the so-called coherent models of Birnbaum, et al. (1961). The predictive value of procedure 3 is quite large indeed, since we measure the degradation of components and deduce the degradation for the subsystem.

A typical box model is discussed by Rosenblatt (1957). Her model assumes independent failure in models of a series of components such that the failure of any component causes failure of the entire system. In addition, she allows environmentally induced failures of the entire system. If all components have an individual reliability, r, and induced failure probability, t, the reliability for a series chain of n components is

$$R = \frac{r^n}{(1 - t)^{n-1}}$$

where $0 < r + t < 1$

The papers of Von Neumann (1956), Moore and Shannon (1956), Birnbaum et al. (1961) develop a theory for what has been called coherent systems. The theory results in reliability bounds on the coherent structure. These systems have components which exist as usual, in only two states, $X_i = 0, 1$. The system itself also exists in only two states, 0, and 1. The relation between the components and the system is given by a structure function $\varphi(X_1, \ . \ . \ . \ , X_n)$ with the properties

$$\begin{aligned} \varphi(\underline{1}) &= 1 \\ \varphi(\underline{0}) &= 0 \\ \varphi(\underline{X}) &\geqq (\underline{Y}) \text{ whenever } x_i \geqq y_i \\ & \qquad \text{for all } i \end{aligned}$$

The last property indicates that operating components do not hinder the operation of the system. One of the bounds developed for this structure is

$$g[1 - F_1(t), \ldots, 1 - F_i(t), \ldots, 1 - F_n(t)]$$

$$\geqq g\left(e^{-t/\mu_i}, \ldots, e^{-t/\mu_i}, \ldots, e^{-t/\mu_n}\right)$$

where $g(P_1, \ldots, P_i, \ldots, P_n) = F\{\varphi(\underline{X})\}$,

$t < \min(\mu_1, \ldots, \mu_i, \ldots, \mu_n)$,

F_i is a IFR failure distribution for component i.

There exist several other bounds and theorems on coherent structures (Barlow and Proschan), not the least of which is a theorem on the S-shaped nature of system reliability for identical components (see Diagram B)

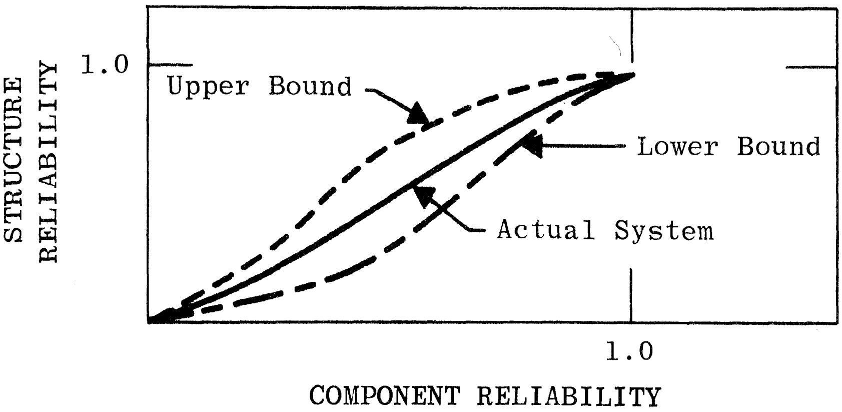


Diagram B

A special case of the coherent structure has received some attention because in the case of independent and identical components the system reliability can be found analytically. This is called the k out of n structure (a binomial structure) where the system fails if k components fail. The system reliability is given by the binomial distribution.

An important facet of prediction, completely overlooked in most papers on reliability, is the role of predictors other than historical usage, age, and system status. A predictor is a measurable factor or variable which is somehow related to degradation. More specifically, any parameter of the distributions associated with a degradation process that can be expressed as a function of other variables can be estimated or predicted in terms of the latter variables (the predictors).

Environmental variables such as temperature, atmospheric water vapor content, audio noise level, irradiation levels etc., may influence degradation. Knowing the values of such variables may allow more accurate prediction of distribution parameters.

The heuristic search for these parameters is well described in most texts on statistical linear models or experimental design.

FAILURE DISTRIBUTIONS

This section deals with two-state processes only. However, using the notion of a reliability function, this restriction does not severely limit the application of the ideas. If a component is found initially to be in the satisfactory state,

$$y(0) = 1$$

a natural question to ask is "how long will the component remain in that state?" This time is said to be the component life or time to failure, and is a random variable denoted by $\tilde{t}$. Commonly hypothesized distributions for failure times are

Exponential:

$$f(t) = \lambda e^{-\lambda t}, \qquad \lambda > 0,\ t \geqq 0 \tag{10}$$

Weibull:

$$f(t) = \lambda \alpha t^{\alpha - 1} e^{-\lambda t^{\alpha}}, \qquad \lambda > 0,\ \alpha > 0,\ t \geqq 0 \tag{11}$$

Truncated Normal:

$$f(t) = \frac{1}{\sqrt{2\pi\sigma a}}\, e^{-\frac{(t-\mu)^2}{2\sigma^2}} \qquad \begin{matrix} \mu > 0,\ \sigma > 0 \\ a > 0,\ t \geqq 0 \end{matrix} \tag{12}$$

Geometric:

$$P_t = P(1 - P)^t \qquad 0 < P < 1,\ t = 0, 1, \ldots \tag{13}$$

Poisson:

$$P_t = \frac{\lambda^t e^{-\lambda}}{t!} \qquad \lambda > 0,\ t = 0, 1, \ldots \tag{14}$$

Other distributions which have been used are:

Gamma	Pascal
Extreme Value	Binomial
Log Normal	

An important notion which can be expressed in terms of these distributions is that of <u>failure rate</u> or <u>hazard rate</u>

$$h(t) = f(t)/1 - F(t) \tag{15}$$

$h(t)dt$ is interpreted as the failure probability in the interval $[t, t + dt]$. The failure rate function places some distributions into two important groups: monotone increasing failure rate, IFR, and monotone decreasing, DFR.

The failure rate of the exponential is constant; it would be used for components whose failure is in no way related to past service or age. The gamma is used when failure is the result of a specified number of Poisson epochs. For example, a vehicle supplies no light when four filaments have failed.

The Weibull family (IFR for $\alpha > 1$) has been used to describe fatigue failures (Weibull, 1939), vacuum tube failures (Kao, 1958) and ball-bearing failures (Lieblein and Zelen, 1956).

The normal (IFR) has been used (Davis, 1952) to describe electronic component failures.

The importance of these distributions is enormous for if one of them can be fit to component failure data, then degradation measures such as expected life can be estimated. Also, these laws are used for systems simulation to predict accidents and traffic patterns and are used for economic maintenance design.

The statistical procedures for fitting these distributions to failure data will not be discussed. Be assured, however, that for the simple failure distributions mentioned, good parameter estimators are available, as are statistical methods for testing the goodness of a fit.

A SYSTEM FAILURE DISTRIBUTION

For a system containing degrading components where the failure of any one causes failure in the system (not strictly a valid assumption for highway systems) and where components are immediately replaced, the steady-state time between failures for this system is distributed exponentially under the following conditions:

(1) The point process for each component, i, $N_i(t)$ is a stationary renewal process.

(2) The process obeys superposition, i.e.,

$$N(t) = \sum_{i=1}^{N} N_i(t) \tag{16}$$

(3)
$$\Lambda \cdot \sum_{i=1}^{n} \frac{1 - F_i(t/\Lambda)}{m_i} \rightarrow 1, \quad \begin{matrix} n \rightarrow \infty \\ t \rightarrow 0 \end{matrix} \tag{17}$$

where

$$\Lambda = \sum_{i=1}^{n} \frac{1}{m_i}; \ \lim_{t\to 0} \frac{P[N_i(t) > 0]}{t} = \frac{1}{m_i} \tag{18}$$

and $F_i(t)$ is the failure distribution for component i. The resulting time to failure is

$$f(t) = \frac{1}{\Lambda}e^{-t/\Lambda} \tag{19}$$

This theorem, due to Ososkov (1956), might find substantial application in both traffic flow and collision prediction. For example, suppose the number of collisions over a stretch of road for some time interval is due to the number of blowouts and headlight failures. If the number of vehicles in the system at any time is constant and if tires and headlights are replaced within short periods after failure, the system may qualify. Since there are many vehicles each of which could independently cause an accident or traffic tie-up, the system will qualify if the failure distributions for each tire and light have the appropriate properties.

A similar theorem has been developed by Grumel (1958) for survival probabilities. Under certain conditions if components fail as

$$\lim_{t\to 0} F(t) \rightarrow \beta t^2, \qquad \beta > 0 \tag{20}$$

then

$$G(t) \rightarrow 1 - e^{-n\beta t^{\alpha}} \qquad (21)$$

where n is the number of components and G is the failure distribution for system. Using this theorem one might describe the survival of a single vehicle having many components.

The above two theorems intuitively supply an approach to highway system degradation based on superposition. If the two outputs of the highway system, traffic flow and collisions, can be described as point processes and if the epochs of each process are functionally or probabilistically related to the degrading point processes, one should investigate the properties of a system having an infinity of components. Such an approach may provide a failure law for the entire system which is related to the components.

MAINTENANCE POLICIES

The special problem of replacement in the overall maintenance activity (design, inspection, repair-replacement) has been recently studied by Flehinger (1962a), Weiss (1956), Herd (1955). There are three policies which have been thoroughly treated:

1. Age replacement
2. Random age replacement
3. Block replacement

The age replacement policy calls for component replacement immediately upon failure or when a component reaches a certain age, T. If we allow T to be a random variable, then the policy is called random age replacement. Finally, if we have a system of components, we can replace them on an individual basis or in blocks. If replacement involves more than one component, the policy is called a block replacement policy.

An important fact about such policies is that their use is profitable in terms of improving reliability only when the failure process is IFR. (That is, failure is more likely as the component ages; if the component failure process is DFR, then older components are more reliable than new ones.)

Another characteristic about each policy is that the point process of replacement times is a renewal process. Roughly, this means that just after any replacement, the time to the next replacement is distributed identical to all replacement epochs.

A general statement can be made concerning these replacement policies (Barlow and Proschan, 1965). For any system of components having an IFR, block replacement requires a larger number of replacements but provides fewer in-service failures than does individual replacement. Also, age replacement requires a larger number of replacements and provides fewer in-service failures than random replacement.

AGE REPLACEMENT

Some of the important characteristics of an age replacement policy at T are determined from the renewal process that it defines. Consider a component with failure distribution F(t). The average time to replacement is

$$\mu_T = T \int_T^{\infty} dF(t) + \int_0^T t \, dF(t) \tag{22}$$

where $P_T = \int_0^T dF(t)$ is the probability that the next replacement is an in-service replacement.

The steady-state expected time to an in-service failure is

$$\ell_T = \frac{\mu_T}{\int_0^T dF(t)} \tag{23}$$

As $T \to 0$, $\mu_T \to 0$, $\ell_T \to \infty$, which is to be expected since components are being replaced extremely often.

RANDOM REPLACEMENT

One form of random replacement is to replace only when an item has failed. This is described by the above equations when $T \to \infty$. Comparing random replacement to periodic replacement

$$\mu_{\infty} - \mu_T = \int_T^{\infty} (t - T) \, dF(t) \qquad t \geqq T \tag{24}$$

$$\ell_T - \ell_\infty = \left(\frac{1}{P_T} - 1\right)\int_0^T t\ dF(t) - (\mu_\infty - \mu_T) \quad (25)$$

These equations show that random replacements require fewer replacements but the time to in-service failure is greater.

BLOCK REPLACEMENT

In complex systems there are many types of block replacement. For this report we shall only consider the series system where any component failure causes system failure, all components have identical failure distributions, and the entire system is replaced when a component fails. For such an n-component system, the failure distribution is:

$$F_n(t) = [1 - (1 - F(t))^n] \quad (26)$$

with mean

$$\mu_s = T(1 - F(t))^n + \int_0^T nt[1 - F(t)]^{u-1}\ dF(t) \quad (27)$$

If components for such a system are replaced on an individual basis

$$\mu_I = \frac{\mu_T}{n} \quad (28)$$

It can be shown that for any T

$$\mu_s > \frac{\mu_T}{n}$$

which shows that the system fails less often if all components are replaced. To find the exact gain, Eq. 27 must be used.

OPTIMUM MAINTENANCE

In this section we will seek information on the tradeoffs in maintenance which minimize cost. Beginning with the in-service failure costs C_f and a planned replacement cost, C_r, the total cost for $[0, t]$ is

$$C(t) = C_f E\{N_f(t)\} + C_r E\{N_r(t)\} \quad (29)$$

Barlow and Proschan (1965) show that planned age replacement is always superior to random replacement for IFR distributions. The necessary condition that must be satisfied for the optimal interval is

$$h(T)\int_0^T [1 - F(x)]\, dx - F(t) = \frac{C_r}{C_f - C_r} \tag{30}$$

Weiss established an important result concerning planned replacement when F is not known. Suppose we want to minimize the maximum expected cost of maintenance. If we know only the mean failure time, we should replace only when failure occurs. Hence, for the highway problem, planned replacement should be considered only when the failure distribution has been established as IFR. Another theorem states that the optimum interval is finite if and only if for the F distribution

$$\frac{\sigma}{\mu} < 1 - \frac{C_r}{C_f} \tag{31}$$

FINITE TIME SPAN

Up to now we have assumed $t \to \infty$, but there may be instances when a highway subsystem such as a road will be retired permanently. In this case, the above remarks do not apply. Barlow and Proschan (1962) have applied dynamic programming to achieve optimal maintenance policies for finite planning horizons.

INSPECTION

In most realistic system models detection of a failure does not occur instantaneously, nor is it free (Brender, 1963). Consider the problem when there is a cost of inspection, C_1, a cost per unit time for an undetected failure, C_2, and an average repair-replacement cost, C_3.

After replacement (assumed to be a renewal point) his policy consists of a set of checking times, $\underline{T} = (t_1, \ldots, t_i, t_j)$, $t_j > t_i$. The expected cost per cycle is

$$C(\underline{T}) = \sum_{i=0}^{\infty} \int_{t_i}^{t_{i+1}} C_1(i + 1) + C_2(t_{i+1} - t)\, dF(t) + C_3 \tag{32}$$

The expected cycle length is

$$T(\underline{T}) = \mu + \sum_{i=0}^{\infty} \int_{t_i}^{t_{i+1}} (t_{i+1} - t)\, dF(t) + T_r; \tag{33}$$

where T_r = repair time

μ = mean component life

Brender provides a numerical algorithm for this problem.

MAINTENANCE POLICIES FOR MARKOV DEGRADATION

In the last section the PM was allowed to take on only two values. Here we will attempt greater accuracy by allowing the PM any one of a finite number of values or states, $Z = 1, \ldots m$. When $Z = 1$, the system is new and when $Z = m$, the system has failed and must be replaced. The parameter of the Markov process, t_n, will be discrete, $t_n = T, \ldots nT, \ldots$.

It will be assumed that for any t_n one inspects the system and can either replace the component or do nothing. The policy is denoted by X(Z), where X(Z) = 1 if replacement and X(Z) = 0 otherwise.

Using the notation from Prediction and Statistical Modeling

$$P_{ij} = \Pr\{Z_{n+1} = j \,|\, Z_n = i\} \tag{34}$$

The notion of a conditional period reliability can be defined as

$$r_n(i) = 1 - P_{im} \tag{35}$$

which is the probability that the system will fail during the interval $[t_n, t_{n+1})$ when it is found in state i at t_n. If the state probabilities, $\Pi_i^{(n)}$, for t_n are known, the unconditional reliability is defined by

$$r_n = 1 - \sum_{i=1}^{m} \Pi_i^{(n)} P_{im} \tag{36}$$

An example of such a system could be the brakes, which contain many components, e.g., $Z_1, \ldots, Z_q$. By dividing the states of each component into three states (essentially new, marginal, and failed), there would be a total of 3^q states. At any step in our chain let the state of the system be described by g tuple $Z = (Z_1, \ldots, Z_i, \ldots, Z_q)$ where $Z_i = 1, 2, 3$.

In terms of the total system, one could define classes of states similar to the three mentioned by using the following rules. If all components are essentially new, then the system is new; if one component has failed, then the system has failed. All other states are marginal states. Note that of the 3^q states, one would be new, $2^2 - 1$ would be marginal, and $3^q - 2^q$ would be failing. By renumbering these states we defined for any n ($Z_n = 1, . . . , 3^q$) such that $Z_n = 1$ implies a new system

$Z_n = i$, $i = 2, . . . , 2^q$ implies a marginal system

$Z_n = i$, $i > 2^q$ implies failure.

With the aid of a structure function, the rules relating the system's state to the states of the q components, the Markov process for the system can be defined in terms of the Markov component processes. Such models have been used to determine the policy $X_n(Z)$ which minimizes the expected cost for both steady-state and finite planning horizons.

The Markov maintenance models uncovered in the literature seem somewhat simplified for the highway systems problem. Flehinger (1962b) treats a system which is in one of $m + 1$ states $(0, 1, . . . , m)$. When the system enters state m, it is immediately replaced with a new one in state 0. He finds a set of optimal states (K, $K = 1, . . . , m - 1$) called marginal states where the component is also replaced. This set minimizes the expected cost per period when the cost of in-service failure exceeds the planned replacement. Along the same lines Derman (1962) has proven that for IFR transition matrices, an optimal control rule indeed exists.

Drenick (1960) treated the problem of finding the optimal checking interval for a marginal test. Using dynamic programming he was able to find an algorithm which selects both the optimal marginal states and optimal checking interval for minimizing the expected cost. The costs considered are such that inspection cost is less than planned replacement cost which is less than in-service failure cost.

A cost-oriented maintenance problem has been formulated by Derman (1962). His model again assumes system degradation to be a finite discrete parameter Markov process with inspection at every step in the chain and considers a randomized repair-replacement rule dependent upon the current system state. Let

$$X_{ik} = \Pr\{Z = k | i \text{ and immediate repair}\} \tag{37}$$

with cost C_{ik}. Since there are m + 1 states before repair and m + 1 after repair, there are a total of $(m + 1)^{m+1}$ policies even if X_{ik} = 0, 1. For each policy matrix X there is a new transition matrix

$$P^* = PX \tag{38}$$

For any policy the steady-state probabilities, π_i, are defined by

$$\pi_j = \sum_{i=0}^{m} \pi_i P^*_{ij} \tag{39}$$

The steady-state probabilities are thus a function of the policy X. The expected cost per interval of a policy is

$$\sum_i \sum_k \pi_i^{(n)} C_{ik} X_{ik}^{(n)} \tag{40}$$

The task of finding the optimal policy may be approached by Howard's policy space method or by the linear programming method of Derman (1960).

AN EXTENDED MARKOV MAINTENANCE MODEL

An extension of Derman's model provides a mathematical formulation which is more appropriate for the highway system. This extended model yields a mathematical formulation for the optimal maintenance policy that minimizes the expected combined costs of design, inspection, repair and replacement for either finite or infinite planning horizons, with reliability constraints. Hence, for a fixed level of reliability we seek the control policy which minimizes total (finite horizon) or period average costs (infinite horizon). The model is based on these assumptions:

1. A planning horizon of N periods for a new subsystem
2. A discrete parameter space

   $t_n = n\tau$, $n = 0, 1, . . ., N$

3. A discrete state space $Z = 1, \ldots, a, \ldots, b, \ldots,$ m such that the component fails in any state $b, \ldots, m$ and is new in any state $0, \ldots, a$

4. Markov degradation with transition matrix (P_{ij})

The cost of repair and replacement in this model will be denoted by C_{ik}. The cost is incurred when the system in inspected, found to be in state i, and is repaired or replaced to state k. The policy, X_{ik} which is allowed to be random, is the probability that the system is repaired or replaced to state k when it is found to be in state i. For any period, whether or not the component is inspected, there is a probability vector $\pi^{(n)}$ for the state probabilities at the beginning of any period. Applying the policy $X^{(n)}$, the expected cost of replacement and repair for the period is (Eq. 40):

$$\sum_i \sum_k \pi_i^{(n)} C_{ik} X_{ik}^{(n)}$$

The choice of the initial design is actually no different than the choices of replacement after the system has reached a failure state. If there is a choice among $(1, \ldots, a)$ new designs, then assume $\pi^o = (1, 0, \ldots, 0)$ and restrict $\sum_{k=o}^{a} X_{ik}^{o} = 1$ for all i. In this way, at the beginning of operation only replacement (initial design) is possible although specific design can be selected with probability less than 1.

The inspection cost will be assumed the same for each period (except perhaps for the first period) and will be denoted by φ. Let the inspection policy for period n be random and denote its probability by Y_n. The expected cost for period n would be φY_n. There might be an advantage in allowing this probability to depend on the status of the system after the last inspection. For the time being, this will be ignored.

The total cost of inspection, replacement, repair, and design can now be written for n periods. To compute the state probabilities for the beginning of any period, a new transition matrix will be defined.

$$P_n^* = Y_n X^{(n)} P + (1 - Y_n) P \tag{41}$$

The elements of this matrix are

$$P_{ij}^{*(n)} = Y_n \sum_{k=1}^{m} X_{ik}^{(n)} P_{kj} + (1 - Y_n) P_{ij} \tag{42}$$

Literally, the probability in period n that the system will pass from state i to state j is the sum of two probabilities. The first is the joint probability that the system will be inspected and that the system will pass to j using policy (X_{ik}) while the second is the probability that the system will not be inspected and that it will pass to j naturally. The state probability vector for any n is

$$\Pi^{(n)} = \Pi^{(o)} \prod_{i=1}^{n} P_i^* \tag{43}$$

This transition matrix defines a modified Markov Process which depends on the policy choice.

The expected cost for period n is

$$\varphi Y_n + \sum_{jk} \left[\Pi^{(o)} \prod_{i=1}^{n} P_i^* \right]_{j \text{ component}} C_{jk} X_{jk}^{(n)} \tag{44}$$

and the sum for N periods is

$$\varphi Y_n + \sum_{n=1}^{N} \sum_{jk} \left[\Pi^{(o)} \prod_{i=1}^{n} P_i^* \right]_{j \text{ component}} C_{jk} X_{jk}^{(n)} \tag{45}$$

The reliability constraint is based upon a special form of reliability function. Since states b + 1, . . . , m are failing states, the failure event is the occurrence of any state in this class. Hence a pointwise reliability constraint can be defined by

$$1 - \sum_{j=b+1}^{m} [\Pi^{(n+1)}]_{j \text{ component}} > r_n \tag{46}$$

The simplest constraint is

$$r_n = R \text{ for all } n \tag{47}$$

The optimization problem for finite horizons is specified by minimizing Eq. 45 by choosing Y_n, $X_{ij}^{(n)}$ for all n subject to the constraint specified in Eq. 47. A good approach to this problem is that of numerical solution by dynamic programming.

The optimization problem for an infinite planning horizon could be stated as

$$\min_{Y,X_{ik}} \quad \varphi Y + \sum_{ik} \Pi_i C_{ik} X_{ik} \tag{48}$$

subject to

$$1 - \sum_{i=b+1}^{m} \Pi_i \geqq R \tag{49}$$

$$\Pi_j - \sum_{i=1}^{m} \Pi_i P^*_{ij} = 0 \tag{50}$$

$$\sum_{j=1}^{m} \Pi_j = 1, \qquad \Pi_j \geqq 0 \tag{51}$$

$$\sum_{j} X_{ij} = 1, \qquad X_{ij} \geqq 0 \tag{52}$$

$$0 \leqq Y \leqq 1 \tag{53}$$

Howard's policy space method should provide a good approach to the numerical solution of this problem.

There are other formulations of related problems worthy of note. One of these was mentioned in passing, namely a randomized inspection policy dependent upon the state at the last inspection. One might argue intuitively that this should be profitable since when the system is found to be in good shape the time to the next inspection should be greater than if the system was found to be in poor shape (assuming that in both cases no action was taken).

There are other reliability constraints which for some problems would be more realistic. Suppose one is interested in constraining the reliability for each of several time intervals instead of one interval. This would constrain the sum of the first-passage-to-failure probabilities rather than constraining the state probabilities themselves.

Another formulation worthy of note is that of minimizing maintenance cost when either the expected component life or first-passage-to-failure probability is constrained.

The policies discussed so far are all randomized policies which admit as special cases deterministic inspection and repair policies. An interesting problem is to determine under what conditions $X_{ij}^{(n)} = 0, 1$.

The Markov process emphasized so far assumes a discrete parameter. It may be necessary to introduce the notion of semi-Markov processes when such an assumption is not warranted.

Appendix 5-A
DEGRADING HIGHWAY COMPONENTS

A transit system which covers even moderate geographical regions is a most complex system owing to the large number of vehicles, the dissimilarities among vehicles and road designs, and the driving habits and objectives of operators. An essential objective in abstracting this system is to maintain the interactions among vehicles (and pedestrians), roads, and drivers.

One way to accomplish this is to view the trajectory of a vehicle as the output of a feedback control drive consisting of:

1. Vehicle
2. Driver
3. Road surface
4. Information sources and channels providing the driver with navigational information

Interaction among vehicle trajectories is governed by:

5. Highway layout
6. Traffic control equipment
7. Traffic laws

A collision is a system output and could be defined as an intersection of two or more trajectories, or the intersection of one or more trajectories with a fixed object. Traffic flow intensity would be the number of trajectories passing over a road cross section per unit time. Transit time would be the time difference between two points on the trajectories.

It is possible to infer a great deal about such system outputs by studying the trajectories of vehicles separately. With this in mind, one might abstract the operation of vehicle as follows: the (human) vehicle controller takes in navigational information from various sources, processes it, and feeds the output into the vehicle's system via the vehicle actuators. As the vehicle responds, new information is presented to the controller. If a collision occurs, its severity is determined by the final trajectory conditions, the relative strengths of the bodies involved, and the operation of safety equipment. Although this is a greatly simplified view (the stages in this process are neither lineal nor discrete), it suggests the complex interactions that exist among groups of components and the difficulty of finding a single cause for an accident. Since many of the vehicle components change with age, it stands to reason that vehicle trajectories also change. The components listed below are grouped according to their natural connections in this feedback loop. Where possible, degradation and failure (limiting degradation) are qualitatively described. (Note that this list is not a complete list of degrading components.)

SELECTED DEGRADING SUBSYSTEMS

	Type of Degradation
Brakes:	
Mechanical linkage	Shear Jamming Loosening by wear
Hydraulic linkage	Pressure losses from wear, contamination, leakage
Shoe and drum	Losses in coefficient of friction from wear, surface contamination
Tires	Losses in frictional force due to blowouts, wear Uneven braking from nonuniform tire pressure

Steering:	
Mechanical linkage	Shear Jamming Loosening by wear Stiffening from lack of lubrication, loss of hydraulic pressure
Tires	Losses in frictional force from blowouts, wear Tendencies for instability from improper tire pressure
Suspension:	Excessive weight transfer from broken or fatigued springs or overloading Sustained angular and vertical oscillation from shock absorber wear
Power Train:	Losses in acceleration from wear, lack of engine maintenance, shear, tire degradation
Road Surface:	Losses in coefficient of friction from wear and surface films Roughness from erosion and cracking
Illumination:	Losses in seeing from burned out street lights, vehicle head lamps, dirty or broken windshields, failing wipers
Information and Control Devices:	
Stop lights Vehicle brake lights Vehicle turn signal lights RR crossing lights Merge lights, etc.	Broken or dirty lenses Filament degradation
Reflecting signs Road edge reflectors Painted center lines Vehicle mirrors	Reflection losses from surface deterioration or dirt

Vehicle gauges	
Speedometer	
Light indicators	
Oil and water pressure	
Generator light	
Horn	Losses in audio level
Safety Equipment:	
Seat belts	Losses in tensile strength
Emergency brakes	
Guardrails and posts	Losses in shear strength
Vehicle cab	Losses in designed crash characteristics
Steering wheel	
Windshield	

Appendix 5-B
VEHICLE AND ROAD INSPECTION

In connection with the maintenance problem about 20 states require periodic vehicle inspections covering the following components:

- Steering
- Wheel alignment
- Emergency and service brakes
- Tires
- Head lamps
- Brake, directional signal, tail lights, reflectors
- Seat belts
- Glass, mirrors
- Windshield wipers, washers
- Exhaust system
- Horn

The cost of these inspection procedures has generally been less than $3.00 per vehicle. At this rate it is likely that brakes, steering, and suspension systems receive a somewhat superficial treatment. Nevertheless, there is little doubt that at their current cost these inspection programs are effective. However, to say that they are really a significant deterrent to collisions is another matter.

The MSU (1965) study on the subject provides excellent cost information but does not spell out inspection or testing procedures for individual components. These procedures must be standardized if uniform inspections are to be achieved.

For greatest effectiveness, timing of inspections should be neither periodic nor random. If owners all comply with the law, only planned periodic inspections would be necessary. These planned inspections should depend upon whether or not failure rates are IFR. If the failure rate increases with age, then inspections should be made more frequently for older vehicles. If a vehicle is tested and found to comply only marginally, or appears to have had extensive use, it should be re-inspected sooner.

Random but very superficial inspections will probably be required to guarantee that stickers are not being falsely obtained, or that equipment such as tires and brakes are not being temporarily repaired or replaced in order to pass scheduled inspection.

No information was found on road inspection methods.

REFERENCES

Barlow and Proschan, 1965.

Birnbaum, Esary, and Saunders, "Multicomponent Systems and Structures and Their Reliability," Technometrics, Volume 3, No. 1, 1961, pp. 55-71.

Brender, D., "A Surveillance Model for Recurrent Events," IBM-Watson Research Report, 1963.

Davis, D., "An Analysis of Some Failure Data," J. Amer. Statist. Assoc., Volume 47, No. 258, 1952.

Derman, C., "On Sequential Decisions and Markov Chains," Management Science, Volume 9, No. 1, 1962, pp. 16-24.

Drenich, R., "Mathematical Aspects of the Reliability Problem," J. Soc. Indust. Appl. Math., Volume 8, No. 1, 1960.

Flehinger, B., "A General Model for the Reliability Analysis of Systems Under Various Preventive Maintenance Policies," Ann. Math. Statist., Volume 33, 1962(a).

Flehinger, B., "A Markovian Model for the Analysis of the Effects of Marginal Testing on System Reliability," Ann. Math. Statist., Volume 33, 1962(b).

Grumel, E., Statistics of Extremes, Columbia University Press, New York, 1958.

Hamilton, C., 1963.

Herd, G., "Failure Rates," ARINC Monograph 2, Aero Radiation Inc., Washington, 1955.

Kao, J., "Computer Methods for Estimating Weibull Parameters in Reliability Studies," IRE Transactions on Reliability and Quality Control, 1958.

Lieblein, J., and Zelen, M., "Statistical Investigation of the Fatigue Life of Deep-Groove Ball Bearings," J. Res. Nat. Bur. Std., Volume 57, 1956.

Michigan State University, Study Report and Plan for Periodic Vehicle Inspection in Michigan, Report of the Highway Traffic Safety Center, Michigan State University, Lansing, 1965.

Moore and Shannon, 1956.

Ososkov, G., "A Limit Theorem for Flows of Similar Events," Theory of Probability and Its Application, Volume 1, 1956.

Rosenblatt, 1957.

Von Neumann, J., Probabilistic Logics, Automata Studies, C. E. Shannon and J. McCarthy, ed., Princeton University Press, 1956.

Weibull, W., "A Statistical Theory of the Strength of Materials," J. Appl. Mech., Volume 18, 1939.

Weiss, G., "On the Theory of Replacement of Machinery with Random Failure Time," NRL Quart., Volume 3, 1956.

6 THE VEHICLE CONTROLLER

Daryl Bem

It is the purpose of this chapter to discuss the ways in which the human controller of the vehicle and his behavior are conceptualized within the total system. Current studies from the literature are discussed in order to suggest the kinds of future research that seem most likely to exploit the advantages of the systems analysis, and a number of functional relations among variables that have been explored experimentally are listed in Appendix classified into the categories provided by the present analysis. The chapter, however, is not intended to serve as an exhaustive review of the literature, nor is the designation of a piece of research as not maximally useful within the systems framework intended to be a criticism of the work by any other criterion. A critical review of the relevant literature will be found in the A. D. Little report (1966), especially in the section entitled "Driving as a Skill."

It will be recalled from Chapter 1 that the total system is characterized by the following five equations:

$$e_\ell = F_\ell(Y, S, t) \qquad [1]$$

$$s_k = \Psi_k(Y, x_h, \theta_d) \qquad [2]$$

$$y_j = f_j(x, \theta) \qquad [3]$$

$$x_i = s_i(x) \qquad [4]$$

$$x_i = h_i(Z, \theta_s) \qquad [5]$$

Equations [1] characterize the operations model and assess the worth of a proposed system in terms of its service, safety, and cost as a function of the subsystem performance capabilities Y, the control strategies S, and time t. Equations [2] characterize the functional relations into which the control strategies themselves enter; these will be taken up first in the discussion.

CONTROL STRATEGIES (s_k)

A control strategy refers to any behavior on the part of the vehicle controller that ultimately affects the motion or dynamic variables of the vehicle subsystem. The set of control strategies thus includes not only the controller's "conscious" decisions (such as the decision to pass another vehicle), but other behaviors as well (such as his maintenance of a particular lateral distance between his vehicle and the center line, the frequency with which he looks into the rear view mirror, the number of steering wheel reversals per minute, and so forth). In short, the term "control strategy" should not be taken to imply the existence of a deliberate chain of reasoning which necessarily precedes the behavior so designated. Any index of the controller's behavior is potentially a component of the control strategy vector S.

Control strategies may be usefully divided into a number of major categories. The outline below represents one such division and enumerates some of the subordinate strategies within each category. It will be noted that each strategy could be further subdivided; and, in principle, each control strategy could be ultimately expressed in very molecular terms. For example, the behavior of "passing" could be reduced to the detailed motions of the vehicle at each point in time, i.e., to the controller's exact manipulations of the accelerator, brake, and steering wheel, or even to his muscular responses. As in science generally, however, the level of analysis selected will probably be only as molecular as is necessary to yield the orderly functional relations among the variables which permit the desired degree of prediction and control. As shall be discussed later, it is desirable to conceptualize many of the system variables at a level of analysis which is independent of idiosyncratic features of specific situations with little generality. In addition to the actual behaviors which alter the dynamic variables of the vehicle subsystem, each strategy is assumed to entail such auxiliary behaviors as checking the rear-view mirror, communicating intent to others, and so forth, all of which can be also investigated as control strategies in their own right.

I. Entering a roadway in the presence and in the absence of other traffic
    A. Entering from a merging lane
    B. Entering from a right angle intersection
    C. Entering from a parking space

II. Leaving a roadway
    A. Turning in front of oncoming traffic
    B. Turning in the absence of oncoming traffic
    C. Leaving the roadway via an exit ramp

III. Continuous driving in the absence of roadway discontinuities (curves, grades, control devices, speed zones, obstacles in or near roadway, scenery, etc.)
    A. Speed maintenance. Mean and variance in the presence and in the absence of other cars.
    B. Following behavior (headway maintenance, overtaking, etc.)
    C. Lane preferences and changes
        1. Passing behavior
        2. Exit preparation
    D. Responses to oncoming vehicles

IV. Control strategies in the presence of roadway discontinuities

V. Miscellaneous driving behaviors (sight-seeing, looking for a parking place, etc.)

As indicated by Eq. [2], control strategies are considered to be functions of the capabilities and current states of the controller (X_h), the environment θ, and the vehicle subsystem's capabilities (Y). (It will be recognized that the outline above is itself a categorization of the different operating environments within which specific control strategies may be realized.) Current states of the controller include both characteristics of a particular driver (e.g., risk-taking proclivities) and shorter-range states (e.g., fatigue due to prolonged driving). The environment includes both the natural and man-made environmental setting (e.g., weather, roadway illumination), and the momentary dynamic environment (e.g., velocity of an approaching vehicle). Finally, many of the controller's behaviors, such as the decision to pass another vehicle under certain circumstances, depend upon the subsystem's capabilities Y (e.g., acceleration capability of the vehicle). It will be noted that this includes the controller's own capabilities X_h which contribute to overall subsystem capability via the functional relations of Eq. [3]. Appendix 6-A lists some of the functional relations within these categories that have been investigated.

Although control strategies may be regarded as direct functions of many environmental variables θ as discussed in the preceding paragraph, it is often more useful to consider certain of these variables as the controller perceives them. Stated in a different way, the controller may be viewed as an information processor who performs a transformation on the environmental variables in a systematic way which can be independently investigated. For example, drivers' judgments of the minimum time required to pass another vehicle safely have been plotted as a function of the actual minimum time (Jones & Heimstra, 1964); judgments of the point at which the vehicle would meet an oncoming vehicle have been related to the distance between the vehicles (Smith, 1963); and, the direction and rate of change in the gap between a lead vehicle and a following vehicle have been estimated by subjects and related to the distance and speed of the lead car (Olson, Wachsler, & Bauer, 1961). In each case, it may be argued that a driver's control strategies would be a simpler function of his judgments of the dynamic environmental variables than the objectively measured variables. Experiments which investigate the functions relating the judgments and the objectively measured variables are known as psychophysical studies, and have enjoyed a long and honored history in experimental psychology.

Just as it is often useful to consider the functional relations between control strategies and the controller's judgmental transformation of environmental variables, so, too, a similar logic obtains for the relations between control strategies and the subsystem capabilities Y. That is, a controller's decision to pass another car when there is an oncoming vehicle is a function of his judgment of the vehicle's acceleration capability, and this judgment could well be in error or subject to extraneous factors. Indeed, alcohol could cause the subsystem's capabilities and the controller's judgment of those capabilities to be inversely related by simultaneously degrading the controller's performance and increasing his confidence. In sum, the functional relations between control strategies and the sets of variables designated as θ and Y in Eq. [2] may appear in most meaningful form if the investigator deals with the controller's judgmental transformations of these variables rather than with the objectively measured variables themselves. Such a conceptualization of Eq. [2] suggests that control strategies could be manipulated by providing the controller with different kinds of sensory feedback regarding these variables, a suggestion that will be discussed in the recommendations section of this chapter.

Research that actually obtains quantitative functional relations among the variables discussed is necessarily built on the knowledge that a particular variable is, in fact, relevant to the control strategy. Within the present state of knowledge, however, much of the research is itself directed toward the actual discovery of the variables controlling the strategies. For example, photographs of expressway traffic have been employed to relate acceleration to aspects of the immediate environment (Perchonok & Seguin, 1964; Perchonok, 1964_a, 1964_b; Hurst, 1964). Similarly, Michael (1965) discusses the sensory factors responsible for the characteristic behavior of a driver in veering away from objects adjacent to the path of travel. Preferences in headway maintenance (following distance) have been investigated for their relations to the driver's sex, age, education, driving experience, vision, brake response time, and attitudes (Wright & Sleight, 1962). Some of these investigations, like those of Perchonok cited above, have been guided by explicit theoretical models. This is not a prominent feature of most of the work in the highway safety area, however, and this point will be discussed in more detail later.

SUBSYSTEM CAPABILITIES (Y)

The human controller is a part of the "mobility" subsystem discussed in Chapter 1. As indicated by Eq. [3], subsystem capabilities are considered to be functions of the capabilities X of the subsystem hardware and the environment θ in which the subsystem is operating. Human performance capabilities are thus assumed to be part of the "hardware" capabilities of the "mobility" subsystem. As noted in Chapter 1, the subsystem capabilities themselves are expressed at a conceptual level which is itself hardware independent; that is, capabilities such as speed, stopping distance, etc. are measurable characteristics of any mobility subsystem irrespective of the design features by which it is realized. It is this emphasis on the general nature of its variables that enables a system's analysis to suggest and evaluate proposed changes in the system that are more than trivial variations of existing hardware.

In current mobility subsystems, human performance capabilities play a widely varying role in overall subsystem capability. For example, performance capabilities of the human controller make only a trivial contribution to the speed capability of automobiles, and research on this particular y_j involves very little human engineering (cf. bicycle design). In contrast, the subsystem's ability to maintain a constant headway (following distance) is, in present systems, almost entirely dependent upon the human controller's visual skills; research on this capability of the subsystem is practically indistinguishable from research on human performance in general (e.g., HRB, 1962; Gantzer & Rockwell, 1966; Fenton, 1965). Finally, other subsystem capabilities such as stopping distance currently depend heavily on both vehicle and human capabilities; it is here that analytic and research techniques of man-machine systems become relevant (Rashevsky, 1965; Bergman, 1965; Inst. Mech. Eng., 1957; Segel & Bundorf, 1965).

It is quite probable that major improvements in the capabilities of the vehicle - controller subsystem will be accomplished by altering major design schemes (e.g., by assigning a task now given to the human controller to machine components). Certainly such courses of action are being considered seriously. Any adequate systems structure must then allow for this type of change and be able to predict the effects.

HUMAN PERFORMANCE CAPABILITIES (X_h)

The human controller may be regarded as an information processor: he takes in information through the senses, processes it, and puts out a motor response which either changes the dynamic variables of the mobility subsystem (e.g., he steps on the brake) or enables him to take in new information (e.g., he looks into the rear-view mirror). The performance capabilities X are not indices of what the human controller does do in a particular set of circumstances—those are the control strategies—but what he is capable of doing.

The general system equation $x_i = g_i(X)$ has been termed the interaction model (Chapter 1) and indicates that changes in one of the capabilities may affect other capabilities of the hardware. This same general equation obtains in the special case of human performance capabilities. For example, if training were to improve an individual's ability to discriminate movements of small amplitude, his ability to produce such movements would also be expected to improve because he could monitor them and provide corrective actions. As discussed in Chapter 1, this same general equation also embraces the fact that many indices of performance capability are combinations of other capabilities and can be specified only when the separate components are known. Thus many performance indices of a complex task, like the time required to complete it, cannot be specified until the specific sequence of responses to be employed is specified.

Like control strategies, human performance capabilities can be categorized into a number of general categories. Unlike control strategies, however, more is known about human performance capabilities in terms of the variables to which they are related, and research in the area of highway safety can profit from the general functional relations discovered in the field of engineering psychology (see Fitts, 1963). The following outline provides a general categorization of the human performance capabilities most relevant to vehicle control seen as components of an information-processing system. (In this outline, in accord with recent theoretical work in sensory psychology, the concept of threshold has been replaced with the more precise concept of sensitivity.)

I. Sensory capabilities and limitations
    A. Sensitivity to visual input
        1. Object detection and recognition sensitivities
        2. Visual acuity for object separation
        3. Sensitivity to cues for depth perception
        4. Sensitivity to different wavelengths of light
        5. Sensitivity to visual distance and its time derivatives
            a. Distance perception
            b. Perception of relative linear velocity
            c. Perception of angular velocity
            d. Perception of accelerations
            e. Perception of radii of curvature and other gradients
    B. Sensitivity to input in other modalities
        1. Auditory sensitivities
        2. Kinesthetic and proprioceptive sensitivities
        3. Vestibular sensitivities (middle ear sensitivity to motion, balance, etc.)
        4. Time judgment capabilities
    C. Special case of above: Ability to discriminate the force, amplitude, speed, and timing of one's own movements (often kinesthetically controlled but not necessarily).

II. Motor capabilities and limitations
    A. Limitations on forces and torques that can be produced
    B. Limitations on abilities to produce movements of specified amplitudes, speeds, and at specified moments in time

C. Refractory period limitations: inability to produce repeated motions rapidly in succession

D. Limitations on ability to automate complex sequences of movements into integrated chains which can be executed without "attention"

E. Limitations on ability to execute more than a few chains of motor responses simultaneously.

III. Information processing capabilities and limitations

A. Processing time (also called transmission delay; combines with the time to execute the motor response to yield reaction time)

B. Limitations on short-term memory capacity

C. Limitations on the amount of information that can be processed per unit time

D. Limitations on ability to process more than one input simultaneously either from a single modality or from more than one modality

E. Limitations on ability to retrieve information from long-term memory

IV. Tolerances for various stresses: ability to function at all

A. Tolerances for extremes of acceleration and other dynamic variables

B. Tolerances for extremes of temperature and other environmental variables

C. Tolerances for sleep deprivation, drugs, and physiological stresses

D. Tolerances for restrictions of sensory input (including "highway hypnosis," etc.)

E. Tolerances for psychologically traumatic events (e.g., accident scene on the highway, personal emotional events, etc.)

F. Tolerances for prolonged periods of task performance.

It will be recalled that the subsystem performance capabilities Y were conceptualized at a very general level so that they would be independent conceptually of particular design features by which they were realized. It will now be noted from the outline above that the human performance characteristics X_h have also been conceptualized in ways that could be realized by vehicle controllers of many types. That is, we have used the general task requirements of vehicle control to define the capabilities whose limitations must be considered rather than the specific physical characteristics of the human controller or the present automobile. Except for tolerances for psychologically traumatic events, every item in the outline would still be descriptive of the controlling component of the mobility subsystem even if the controller were non-human. This is one of the advantages of the recent conceptualization of the human as an information-processing system;

it is an analytic framework ideally suited for systems analysis. It is not surprising that engineering psychologists were among the first to adopt the framework (see Pew, 1965).

All of the general capabilities listed in the outline are functionally dependent on specific variables of design hardware of course, and Eq. [5][$x_j = h_i(Z, H, \theta)$] takes the final step toward specificity and relates each capability index to specific hardware Z (e.g., distance from accelerator to brake), specific indices of human states (e.g., sex, age, etc.), and particular environmental conditions (illumination of the highway). Not unexpectedly, the kinds of functional relationships described by Eq. [5] are those which have been investigated most often. Appendix 6-A lists a number of those studies which deal most specifically with vehicle control.

RECOMMENDATIONS

RESEARCH ON THE CONTROL OF PERFORMANCE CAPABILITIES AND CONTROL STRATEGIES

The human controller has traditionally been given the control tasks of the vehicle without much regard for his suitability for such tasks and with even less regard for compatibility between the way in which the task is to be performed and his capabilities. The field of engineering psychology has begun to treat this neglect (Fitts, 1963), and it is now common to question whether or not a particular task should even be given to the human controller. It is suggested that more of this kind of task analysis is needed in the field of highway safety, particularly within the mobility subsystem. If a particular task can only be performed by the human controller, then there should be more systematic analysis of the effect on performance of the kinds of information and modes of data presentation. The problem of headway maintenance provides an illustrative example. In order to maintain headway at a constant value, the human controller must presently rely on his visual ability to detect relative linear velocity and estimate small variations in distance. These are tasks that he can perform with only limited success, and research has explored the possibility of using visual- and time-judgment aids for improving headway maintenance (HRB, 1962; Gantzer & Rockwell, 1966). Along similar lines, Fenton (1965) has designed a "tactile" controller which gives the human controller the required information through kinesthetic rather than visual feedback. Performance has become virtually perfect with this device under certain conditions, and this kind of research is illustrative of the kind of work that seems most promising in providing useful knowledge about possible changes in behavior. Note that a subsystem capability Y (headway maintenance) was vastly improved by utilizing a different human performance capability K (kinesthetic rather than visual sensitivity) rather than simply attempting to perform minor improvements on the capability itself through training. The field of engineering psychology now has a number of very general principles concerning human task performance which can and should be more fully related to vehicle subsystem design.

Just as subsystem capabilities can be improved by altering the nature of the feedback to the controller, it is also possible to manipulate the controller's strategies (S) in the same way. It will be recalled that control strategies were said to be a function of the controller's perception of the environment and subsystem capabilities. This implies that control strategies could be affected by presenting the controller with different sets of data about the current states of the environment, his vehicle, and himself. There is already some evidence that driving performance is affected by permitting the driver to monitor indices of his steering wheel manipulations.

It is even possible that data could be presented through sensory channels not presently manipulated for this purpose. The vestibular system of the middle ear is a principal source of any driver's present information on vibration, and possibly on horizontal velocity. A mechanical system for stimulating this system directly is within the range of present technology. It should be noticed that in this case one might well feed essentially "false" data to the controller. Research might be done on this and similar possibilities.

Thus the recommendation is that all studies should aim toward a level of conceptualization beyond raw empiricism if the advantages of systems analysis are to be realized. Simply categorizing certain driving strategies aggressive or regressive (Mierke, 1955) is a first step toward generality. A more sophisticated analytical approach has been taken by Senders, et al. (1966). Employing an information-theoretic framework, these investigators present experimental methods for analyzing disparate aspects of the driving environment into the single conceptual variable of attentional demand on the controller, a variable which might eventually even be quantified into bits per second of attentional demand. With such an approach, proposed design changes in the total system could be rationally proposed, tested or simulated, and evaluated on a systematic basis. A proposal to change lighting fixtures at an interchange would be conceptually characterized, for example, as an increase in visual feedback to the controller at a high-information-density highway discontinuity; and, the proposal could be evaluated analytically and compared by some meaningful criterion with other proposals. None of this can be done, however, without explicit attempts to develop and test particular theoretical models.

THE NEED FOR THEORETICAL MODELS

Implicit in the entire discussion of this chapter has been the recommendation that the variables of any research investigation be conceptualized in a way that does not limit the findings to specific hardware components. The bulk of research to date either explores some isolated performance without relating it to subsystem performance capability, or examines independent variables defined only in pure hardware terms (e.g., reflectorized versus illuminated traffic control devices as they affect the number of steering wheel reversals per unit time). Many of such studies, of course, were designed to solve particular practical problems, but they are of little utility beyond the specific situations they explored. It is a major purpose of the systems analysis itself to provide a very general framework for conceptualizing variables at a higher level of generality, but a great deal of analytical work is required to add substance to such a framework. The systems analysis itself does not dictate the particular theoretical models to be employed in analyzing the subsystems, and it only requires that the input and output variables of each subsystem be commensurate with those from other subsystems. Even in the absence of explicit theoretical models, all studies can aim toward a conceptualization of their independent and dependent variables that reach beyond the confines of the hardware employed.

Appendix 6-A

A SELECTED LIST OF FUNCTIONAL RELATIONS FROM THE LITERATURE

I. Control strategies (S) as a function of:

   A. Current states of the controller (X_h)

      1. Preferred headway as a function of controller's age, sex, education, driving experience, attitudes (Wright & Sleight, 1962)
      2. Manipulations of accelerator, steering wheel, and brake pedal as a function of hours of driving and past record of driving (Snider & Rockwell, 1963; Greenshields, undated; Greenshields & Platt, 1964; Platt & Fedderson, 1964)
      3. Speed variance and number of steering wheel and accelerations reversals per unit time as a function of hours of driving (McFarland & Moseley, 1954; Safford & Rockwell, 1966; Platt, 1964)
      4. Simulator driving performance as a function of blood sugar level and sleep deprivation (Halbert & Wojcik, 1965; Halbert et al., 1963)
      5. Simulator performance as a function of rest breaks and prolonged periods of driving (HRB, 1957, 1959)

   B. Environmental variables (θ)

      1. Acceleration as a function of the immediate environment (using photographs of expressway traffic and an analytic approach) (Perchonok & Seguin, 1964; Perchonok, $1964_{a,b}$; Hurst, 1964)
      2. Passing behavior as a function of the interval between the vehicle and an oncoming vehicle (Crawford, 1963)
      3. Accelerator, brake, and clutch manipulations and steering wheel movements as a function of music and speech radio programs (Brown, 1965)
      4. Speed as a function of illumination (Wright, 1963)
      5. Lateral position and speed of vehicle as a function of edge marking on the roadway (HRB, 1960)
      6. Speed as a function of speed control (RRL, 1963)
      7. Behavior of veering from obstacles beside roadway as a function of angular velocity (Michael, 1965)

   C. Subsystem capabilities (Y)

      1. Preferred headway as a function of vision and brake response time (Wright & Sleight, 1962)
      2. Preferred headway as a function of speed (HRB, 1962)
      3. Speed as a function of visual capability (Senders et al., 1966)

II. Human performance capabilities (X_h) as a function of:

   A. Physical hardware characteristics (X)

      1. Headway maintenance as a function of various aids (HRB, 1962; Gantzer & Rockwell, 1966; Fenton, 1965)
      2. Time required to recognize a message on a road sign as a function of the amount of extraneous information (Hoyos, 1965)

   3. Reaction time as a function of brake light placement on lead car (Crosley & Allen, undated)
   4. Reaction time as a function of direction indicator placement relative to stop light (Grime, 1954)
   5. Driving performance as a function of stability states in a variable stability vehicle (Segel & Bundorf, 1965)

B. Current states of the controller (H)

   1. Multilimb coordination, spatial orientation, proprioception, response orientation, and reaction time as a function of hours of driving (Herbert, 1963; Herbert & Jaynes, 1964)
   2. "Vigilance" (detection of on-second signals) as a function of hours of driving (Dobbins, Tiedemann & Skordahl, 1963)
   3. Efficiency of judgment and work performance as a function of Dexedrine (Leake, 1957)
   4. Visual sensitivity as a function of age (Baldwin, 1963)
   5. Vision, tracking errors, reaction time and vigilance errors in a simulator as a function of smoking, smokers vs. nonsmokers and deprivation of smoking in smokers (Johansson & Jansson, 1964; Heimstra, Bancroft & Dekock, 1966)
   6. Perception as a function of fear (Hauss, 1962)
   7. Driving ability as a function of practice time as a function of age (Hakkinen, 1958)
   8. [There is a very large literature on alcohol and its effects. See the A. D. Little report (1966) for a review.]

C. Environmental parameters (θ)

   1. Perceptual skills in driving as a function of darkness (Michael, 1965)
   2. [The enormous amount of research on accidents as a function of environmental conditions is relevant here to the extent that accidents imply impaired performance capabilities. Some of these fall into category IIA above. For a review, see the A. D. Little report (1966).]

REFERENCES

Baldwin, David M., "Accident Records and Research," *Traffic Safety Res. Rev.*, Volume 7, No. 3, Sept. 1963, pp. 6-8.

Bergman, Walter, *The Basic Nature of Vehicle Understeer-Oversteer*, Society of Automotive Engineers, Inc., January 1965.

Brown, I. D., "Effect of a Car Radio on Driving in Traffic," *Ergonomics*, Volume 8, No. 4, October 1965, pp. 475-479.

Crawford, A., "The Overtaking Driver," *Ergonomics*, Volume 6, No. 2, 1963, pp. 153-170.

Crosley, John, and Allen, Merrill J., *Automobile Brake Light Effectiveness: An Evaluation of High Placement and Accelerator Switching*, Indiana University, Bloomington (no date).

Dobbins, D. A., Tiedemann, J. G., and Skordahl, D. M., "Vigilance Under Highway Driving Conditions," Perceptual and Motor Skills, Volume 16, 1963, p. 38.

Fenton, R. E., Driver-Automobile Interfaces, Engineering Experiment Station, The Ohio State University, Communications and Control Laboratory of the Department of Electrical Engineering, Sponsored by the Ohio Department of Highways and the U. S. Bureau of Public Roads, Report No. 202A-7, August 1965.

Fitts, P. M., "Functions of Man in Complex Systems," Aerospace Eng., Volume 21, 1962, pp. 34-49.

Fitts, P. M., "Engineering Psychology," in S. Koch (Ed.) Psychology: A Study of a Science, Volume 5, New York, McGraw-Hill, 1963.

Gantzer, Donald, and Rockwell, T. H., The Effects of Discrete Headway and Relative Velocity Information on Car Following, 45th Annual Meeting of the Highway Research Board, January 1966.

Greenshields, Bruce, Changes in Driver Performance with Time in Driving, The University of Michigan, Ann Arbor (no date given).

Greenshields, Bruce D., and Platt, Fletcher N., "Objective Measurements of Driver Behavior" (joint paper including "The Objective Evaluation of Traffic Stream Flow," and "Objective Measurements of Individual Driver Behavior"), Society of Automotive Engineers, Inc., No. 809A, January 1964.

Grime, G., "Vehicle Characteristics and Road Accidents," Engineering, Sept. 10, 1954.

Hakkinen, Sauli, Traffic Accidents and Driver Characteristics— A Statistical and Psychological Study, Finland's Institute of Technology, Scientific Researches (13), 1958.

Hauss, von K., "Zur Psychologie Optischer Wahrnehmungsstorungen" (The Psychology of Disturbances in Optical Perception), Zentralblatt fur Verkehrs Medizin, Verkehrspsychologie, Luft-und Raufahrt Medizin, Volume 8, No. 3, 1962, pp. 156-163.

Heimstra, N. Brancroft, N., and DeKock, A., Effects of Smoking on Sustained Performance in a Simulated Driving Device, Conf. on the Effects of Nicotine and Smoking on the Central Nervous System, New York Academy of Sciences, 1966.

Herbert, M. J., Analysis of a Complex Skill: Vehicle Driving, ASAMRL Report No. 581, 1963.

Herbert, Marvin J., and Jaynes, William E., "Performance Decrement in Vehicle Driving," J. Eng. Psychol., Volume 3, No. 16, 1964, pp. 1-8.

Highway Research Board (HRB), Driver Characteristics, Highway Research Board Bulletin 152, 1957.

Highway Research Board (HRB), *Characteristics of Vehicle Operators*, Highway Research Board Bulletin 212, 1959.

Highway Research Board (HRB), *Pavement Edge Markings, Shoulders and Medians*, Highway Research Board Bulletin 266, 1960.

Highway Research Board (HRB), *Driver Characteristics*, Highway Research Board Bulletin 330, 1962.

Hoyos, Carl Graf., *Psychologie Des Strabenverkehrs* (Psychology of Road Traffic), Verlag Hans Huber Bern und Stuttgart, 1965.

Hulbert, S., Beers, J., Herzog, J., and Blyden, S., *Blood Sugar Level and Fatigue Effects on a Simulated Driving Task*, Department of Engineering, University of California, Report No. 63-53, October 1963.

Hulbert, S., and Wojcik, C., "Research Activities at the UCLA Driving Simulation Laboratory," *Highway Research News*, No. 17, February 1965, pp. 111-114.

Hurst, Paul M., *Errors in Driver Risk-Taking*, Division of Highway Studies, Institute for Research, State College, Pa., Report No. 2, June 1964.

Institution of Mechanical Engineers (I.M.E.), *Research in Automobile Stability and Control and in Tyre Performance*, Papers presented at the meeting of the I.M.E., Automobile Division, London, Nov. 13, 1956.

Johansson, Gunnar, and Jansson, Gunnar, *Smoking and Night Driving*, Department of Psychology, University of Uppsala, Sweden, November 1964.

Jones, Howard V., and Heimstra, Norman W., "Ability of Drivers to Make Critical Passing Judgments," *J. Eng. Psychol.*, Volume 3, October 1964, pp. 118-122.

Leake, Chauncey D., "The Amphetamines and the Sleepy Driver," *The Ohio State Medical Journal*, Volume 53, Feb. 1957, pp. 176-178.

McFarland, Ross A., and Moseley, Alfred L., *Human Factors in Highway Transport Safety*, Harvard School of Public Health, Boston, 1954.

Michaels, Richard M., "Traffic Safety: A Challenge to Systems Engineering," Address to Washington Society of Engineers, Washington, D. C., March 3, 1965.

Mierke, Karl, *Die Ueberförderung von Letztgrenzen der seelisch-geistigen Leistungs und Belastungs-fähigkeit (II) (Over-Taxation of the Psychic Performance Limits)*, Praxis der Kinderpsychologie und Kinderpsychiatrie, Jahrgang 4, Heft 1, 1955, pp. 15-24.

Olson, Paul L., Wachsler, Robert A., and Bauer, Herbert J., "Driver Judgments of Relative Car Velocities," *J. Appl. Psychol.*, Volume 45, No. 3, 1961, pp. 161-164.

Perchonok, Kenneth, *The Measurement of Driver Errors*, Division of Highway Studies, Institute for Research, State College, Pa., Report No. 1, April 1964(a).

Perchonok, Kenneth, *The Determination of Environmental Factors Controlling Driver Behavior, Division of Highway Studies, Institute for Research, State* College, Pa., Report No. 3, July 1964(b).

Perchonok, Kenneth, and Seguin, Edmond L., *Vehicle Following Behavior: A Field Study, Division of Highway Studies, Institute for Research, State* College, Pa., Report No. 5, Sept. 1964.

Pew, R. W., "Human Information Processing Concepts for System Engineers," in R. E. Machol (Ed.), *System Engineering Handbook*, New York, McGraw-Hill, 1965.

Platt, Fletcher N., "A New Method of Evaluating the Effects of Fatigue on Driver Performance," *Human Factors*, Volume 6, No. 4, Aug. 1964, pp. 351-358.

Platt, F. N., and Feddersen, G. J., "Driver Performance: The Search for Objective Measurements," Paper presented at the Symposium and Workshop on the Quantification of Human Performance—Human Factors Society, August 17-19, 1964.

Rashevsky, N., "Man-Machine Interaction in Automobile Driving," *Traffic Safety Res. Rev.*, December 1965, pp. 101-107.

Road Research Laboratory (RRL), *Research on Road Safety*, Her Majesty's Stationery Office*, London, 1963.*

Safford, Robert R., and Rockwell, Thomas H., "Performance Decrement in Twenty-Four Hour Driving," 45th Annual Meeting of the Highway Research Board, January 1966.

Segel, Leonard, and Bundorf, R. Thomas, *The Variable-Stability Automobile* (including "Concept and *Design" [Segel] and "The Use* of a Variable-Stability Vehicle in Handling Research" [Bundorf]), Society of Automotive Engineers, Inc., November 1965.

Senders, J. W., Kristofferson, A. B., Levison, W., Dietrich, C. W., and Ward, J. L., *An Investigation of Automobile Driver Information Processing, Report No. 1335, Final Report on Contract No. CPR-II-0958*, Submitted to U. S. Dept. of Commerce, The Bureau of Public Roads, April 1966.

Smith, R. Dean, "An Exploratory Study of Predictive Judgments in a Traffic Situation," *Scand. J. Psychol.*, Volume 4, 1963, pp. 65-76.

Snider, John N., "The Development of an Instrumentation System to Measure True Driving Performance," *Traffic Safety Res. Rev.*, Vol. 7, No. 3, Sept. 1963, pp. 23-27.

Wright, John Martin, *An Evaluation of Intersection Illumination, Purdue University, July 1963.*

Wright, Stuart, and Sleight, Robert B., "Personal Characteristics of Automobile Drivers Related to Following Distance," Presented at the 33rd Annual Meeting of the Eastern Psychological Association, Atlantic City, New Jersey, April 1962.

7 COST PREDICTION

KERRY KILPATRICK

INTRODUCTION

The preceding chapters described in detail the general systems framework proposed for the study of the highway transportation complex. The functional and theoretical aspects of the major subsystems have been structured, and plausible interrelationships between these subsystems have been indicated. However, in addition to providing the analyst or decision-maker with various performance measures as output, the highway transportation system model must produce an accurate estimate of the cost of either making changes in the present system or introducing a completely new capability into the system. (For verbal economy, both will be referred to as system modifications.) It is clear that to permit realistic comparison among alternative system modifications a cost prediction model must be able to estimate accurately the cost of modifications in all the various subsystems. It is the purpose of this chapter, therefore, to describe what is meant by the costs of system modifications and to outline in detail the techniques which may be used to calculate these costs.

Any actual modification task may be viewed as a project which has some point of conception, a duration, and a terminus (perhaps far in the future). During the course of the project, plans are made, engineering and development work done, physical hardware procured, and the system finally operated and maintained. Thus, as a point of departure, we may broadly define the cost of a system modification project as those expenditures required for the development, procurement, and operation and maintenance of a given system modification, the introduction of or the alteration of a well defined set of system component parts performing certain well defined tasks over some relevant time span.

COST TO WHOM?

Having thus defined in a general way which costs the cost prediction model should be designed to estimate, we must attempt to identify the cost recipient. We can define two main methods of assigning the resultant costs of the various subsystems. The first is to return all costs to a common recipient, such as "the general public." For instance, the cost of structural improvements in auto bodies, while initially incurred by the manufacturers, would be borne by the public through increased automobile costs. Similarly, highway construction costs would be borne by the public through increased taxes, etc. To return all costs to a common base requires that costs (or incremental costs) be carried through a series of assumed relationships which will indicate how they will be redirected from the initial recipient to the common recipient.

The advantage of the common base approach is that the costs of investment alternatives can be easily compared. For example, assuming their necessary functional relationships to reducing traffic deaths were known, one could simply compare the costs of increasing the interior padding in cars with that of removing all hazards (e.g., trees) within 100 feet of the road. Two obvious disadvantages of the common base approach are that (1) a common base (such as "the general public") cannot be precisely defined, and (2) the validity of the relationships through which costs are returned to the common base would at best be doubtful.

The second general approach to assigning costs is to develop certain well defined categories and to assign all costs incurred to one or more categories. For example, the cost of installing seat belts in all vehicles would be categorically assigned to the vehicle owner. The cost of removing roadside hazards would be assigned proportionately to the municipal, county, state, and federal highway departments. Advantages of this approach are that the assignment of costs is relatively straightforward and the categories easily recognized. On the other hand, it is obviously difficult to compare costs of alternative investment decisions when the costs are incurred by different and frequently overlapping sectors of the economy; that is, when the constituent elements of the various recipient categories cannot be clearly distinguished from each other.

Unfortunately the cost of a project might not be the same or even consist of the same elements when viewed with respect to different sectors of the economy. Consider, for example, the cost of automobile insurance. When the cost of insurance is thought of as being incurred only by vehicle owners, then it is clearly and simply defined. However, when viewed in terms of the national economy, the situation is more complicated. Since the cash flow in from the policy holders are offset by the flow out to claimants, one might conclude that the net cost to the nation is zero. But jobs are created, insurance offices built, and investments made by insurance companies which clearly act as stimulants to the economy. Thus one might argue that while insurance may be a direct cost to policy holders, the existence of insurance companies stimulates the national economy and therefore provides indirect benefits to the general public. Attempts have been made to trace such indirect costs through the national income accounts by means of simulation models (Jacobson & McGovern), but the complexities that arise when one attempts to consider transportation system costs in the broad economic sense are far removed from our immediate purpose of cost prediction. In this discussion, the cost recipients are assumed to be only those individuals or agencies incurring direct, differential costs (i.e., costs which would not be incurred if the system modification in question were not made) or direct reductions in revenue. No attempt will be made here to develop relationships by which these costs may be returned to a common base; they will be assigned to categories which closely resemble those currently in use.

UNCERTAINTY

Cost prediction is inherently uncertain. This uncertainty can appear as requirements uncertainty or cost-estimating uncertainty (Fisher, 1962a). Requirements uncertainty refers to possible variations in cost elements resulting from unexpected changes in the physical characteristics of the system being evaluated. Cost-estimating uncertainty refers to statistical variability in the mathematical cost-estimating relationships (CER) used to predict costs of a project.

The cost prediction model proposed in this chapter has been structured so that all costs incurred throughout the life of a project may be estimated after the basic system specifications are formulated but before any research and development has begun. Because of this, our cost prediction model will be particularly sensitive to requirements uncertainty. Unfortunately, most of the literature to date suggests that inaccuracies in cost estimates resulting from requirements uncertainty are far greater than those resulting from cost-estimating uncertainty. Fisher (1962a) states that cost-estimating uncertainty may cause average variations in cost estimates of 20 to 30 percent. However, the total observed variations of cost estimates from actual costs may average 200 percent and higher. Thus, we can conclude that the variation resulting from requirements uncertainty may be an order of magnitude greater than that due to cost-estimating uncertainty.

One attempt to compensate for these variations is the "debiasing" formula developed by Summers (1962). He endeavors to adjust cost estimates (for average systems) to those actually incurred by obtaining a "debiased" cost estimate in terms of the initial cost estimate, the timing of the estimate within the development program, the degree of technological advance represented by the project, the length of the development program, and a measure of the skill of the cost estimators. Further study must be done to see whether a similar means of recognizing cost-estimating errors may be applicable to the cost prediction model proposed here. For the present, however, it is sufficient to keep in mind that inaccuracies do exist in cost estimates and that systematic means for dealing with these inaccuracies have been explored elsewhere.

STAGES OF PROJECT LIFE

With these preliminaries behind us we must now observe that to attempt to estimate the total cost of a project in the single step implied by the preceding discussion would be completely intractable. We need instead to divide the life of the project into stages, and to develop a means of estimating the costs incurred by the activities associated with each project stage. The project stages used here are those usually associated with commercial or military projects, namely, Research, Development, Test, and Evaluation (RDT&E), Procurement, and Operating and Maintenance (O&M).

The costs associated with these three stages are

(1) RDT&E: Outlays for basic research and exploratory development, for model building, test, and evaluation and for detailed engineering of the project.

(2) Procurement: One-time outlays required to alter the present system or to introduce a new capability into the system.

(3) O&M: The recurring differential costs needed to operate and maintain the total system subsequent to the system modification resulting from this project.

EXCLUSIVENESS AND INDEPENDENCE OF PROJECT STAGES

Figure 7-1 shows the general relationships between these project stages (RDT&E, Procurement, and O&M). Clearly, they will overlap to some extent, since it is possible to begin one stage before the preceding stage is completed; it is possible, for example, to begin operating the system before the total quantity scheduled for Procurement (e.g., miles of interstate highway) has been produced. This temporal overlap causes no real problems when assigning costs to the proper project stage. What is troublesome is that some activities are not naturally mutually exclusive with respect to project stages. Such a gray area occurs, for example, in the case of engineering changes initiated by Manufacturing to correct defects that should have been caught by Engineering. Are the costs of re-engineering the job, changing the drawings, retooling, and scrap to be assigned to the RDT&E stage or to the Procurement stage? Questions such as this one must be resolved through accounting definitions so that mutually exclusive cost assignments may be made without difficulty. Once these accounting definitions have been formulated, all costs incurred can then be viewed as mutually exclusive and collectively exhaustive.

The problem of independence, unfortunately, is not so easily resolved. Monies spent in earlier stages of development may have a definite and predictable (at least in direction) effect on costs in later stages. For example, increasing the amount of proto-type testing may reduce procurement costs and will probably reduce maintenance costs. Similarly, increasing procurement costs by tightening quality control standards may reduce wear rates and, thus, operating costs. The development of the exact functional relationships between the levels of activities in one stage and the costs in other stages is an unexplored topic worthy of considerable research.

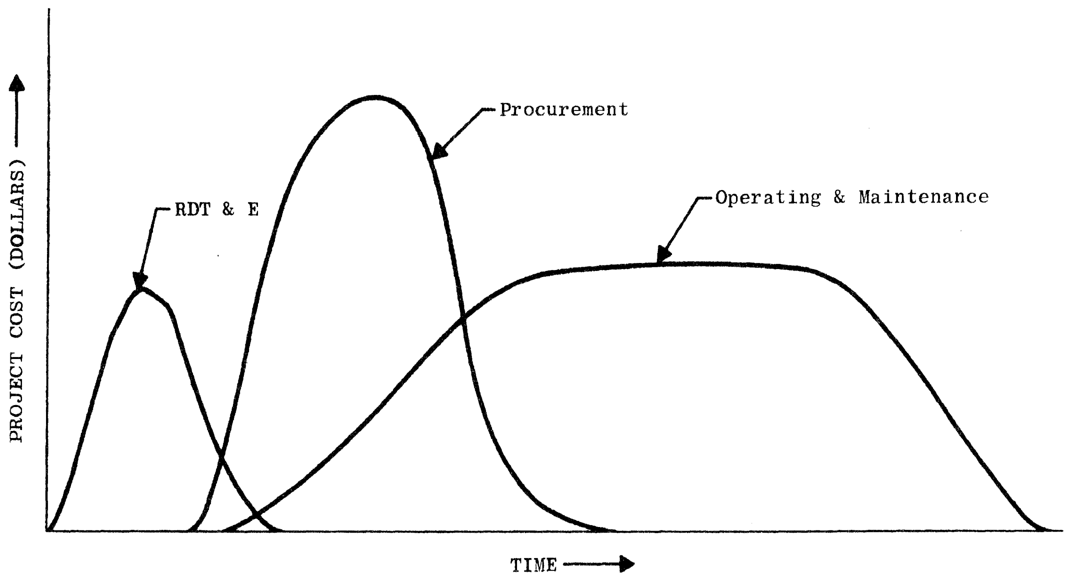


FIGURE 7-1. STAGES OF PROJECT LIFE

Although it is clearly an oversimplification, the detailed discussion of the cost prediction model which follows in the next three sections will assume that the costs incurred in each stage of project life are independent. In the final section of this chapter (Further Research) the problem of project stage interdependence will be discussed further.

THE STRUCTURE OF THE COST PREDICTION MODEL AT EACH PROJECT STAGE

Some general comments on the cost prediction problem for each stage of project may help to chart the course for the detailed development in the sections to follow. The preceding chapters have indicated that the performance of the overall highway complex may be simulated only after the relevant variables have been described and their interactions identified. These specifications freeze the overall design of the highway transportation system so that a specific design may be studied. When the specifications are changed, the system can again be simulated. Thus the resultant performances of different systems may be compared.

The proposed cost prediction model may be structured in an analogous manner. Given a 'frozen' configuration of a system or of some change in the system, the cost prediction model would estimate the costs associated with the three stages of project life. Thus, in some cases, the cost prediction model will use the same inputs as the models of the various subsystems discussed above. In addition, however, the cost prediction model may also base its predictions on the outputs of the subsystem models (y_j).

The precise structure of the cost prediction model will differ somewhat for each stage of project life, because the required activities in each stage differ. During the RDT&E stage certain activities must be performed to bring the idea from conception to the point where production may begin. Hence, the cost of each of these <u>activities</u> must be predicted. As we shall see below, descriptors of the system (such as the relation of proposed system performance (y_j) to the current state of the art) may be important in predicting these RDT&E costs.

The costs of Procurement will be predicted on the basis of the specified hardware for the system. The proposed cost prediction model would estimate the cost of each component part of the system by means of cost-estimating relationships based on performance characteristics of the component parts. The total procurement cost could then be estimated by relating total expected procurement quantity and production rates to unit production costs.

The task of predicting costs for a project is in part a <u>forecasting</u> problem of the same type encountered for RDT&E and Procurement and in part an <u>accounting</u> problem. Forecasts must still be made of personnel requirements, training costs, facility costs, etc. in a manner similar to that employed for the Procurement stage. On the other hand, since such things as wear and failure rates will be produced as output from the system degradation model and the fuel usages, etc., will be output of the operations model, the <u>prediction</u> task has already been completed. To obtain the relevant O&M costs the various usage rates need only be multiplied by the appropriate cost factors.

The differences among the approaches proposed for predicting the costs associated with the three stages of project life will become more clear as the model structure is explained in detail below. The remainder of this chapter is divided in four major sections:

Predicting RDT&E Costs
Predicting Procurement Costs
Predicting Operating and Maintenance Costs
Future Research

The first three of these sections include a brief literature survey and the development of the cost prediction model for the indicated stage of project life. The last section reviews some of the areas indicated in preceding sections where further research may be profitably pursued and presents some recommendations.

PREDICTING RDT&E COSTS

As described previously, one objective of the cost prediction model is to provide a means of estimating the cost of the RDT&E stage of a system modification project. The RDT&E stage can be defined as "a finite sequence of purposeful, temporally ordered activities, operating on a homogeneous set of problem elements, to meet a specified set of objectives representing an increment of technological advance" (Norden, 1963). The associated costs of this project stage are those outlays for the cost elements of labor, material, and overhead, which are necessary to reach the objectives.

Various estimates (Hertz, 1950) place the relative proportion of expenditures for RDT&E cost elements at approximately 75% for labor, 10% for material, and 15% for overhead. Obviously, the prediction of project labor costs is of primary importance. This is especially true if, in a given industry or firm, the remaining cost elements have been observed to be a relatively stable proportion of labor costs from project to project and, thus, can be estimated as a percentage of labor costs.

GENERAL METHODOLOGY

The basic difference between the cost prediction task for RDT&E and that for Procurement is that the latter requires outlays for a well-defined, discrete set of objects or services, whereas the outlays for RDT&E buy solutions to a finite (but unknown) number of problems. The visible outputs of the RDT&E stage are reports, blueprints, prototypes, etc., which in themselves have little intrinsic worth and cannot be costed on a per unit basis. What can be costed are the individual activities necessary to proceed from the conception of the RDT&E stage to its completion. Each activity, such as planning, design, test, etc., has associated with it a starting point and a duration, and over its duration some distribution of effort measured, for example, in man-hours. Thus if a means could be developed to identify all project activities and assign an accurate effort distribution to each activity, the RDT&E cost prediction problem could be solved.

LITERATURE OF RDT&E COST PREDICTION

A review of the relevant literature reveals two principal approaches to the prediction of costs in the RDT&E stage: network techniques and life cycle methods. The network techniques, such as PERT/Cost, require that all activities and events of a given project be described in atomistic detail. The life cycle models, on the other hand, deal with gross statistical aggregates and are based upon the repetitive patterns of effort distributions for major activities as observed over a representative sample of projects.

The literature dealing with network costing techniques is extensive. Consequently no detailed review is attempted here. However, see Norden (1963), Hertz (1950), DOD (1962), Bigelow (1962), and Poletti (1964) for bibliographies and recent articles of particular interest.

The central point to observe with regard to network costing techniques is that (although some theoretical criticisms may be raised) they are unquestionably more accurate cost predictors than life cycle models *if* all the events and activities for the entire project can be determined *before* the project is started and in the *minute detail* required. It is precisely the *inability to identify* the required activities and events, however, that makes the RDT&E cost prediction task difficult in the first place. To assume that the precise nature of the activities and events leading from conception to production is known in advance is to assume that the required solutions to the problems to be solved during the RDT&E stage are also known, an obvious contradiction.

Since our cost prediction model must be able to predict RDT&E costs on the basis of the initial system performance specifications only, it appears that the network techniques will be of little value as cost prediction models. Consequently, as crude as it may be, the life cycle model will be adopted for the proposed RDT&E cost prediction model described below.

The majority of the accessible published work on life cycle models has been written by P. V. Norden (1958, 1960a, 1960b, 1962) and his associates at IBM. Similar studies have been reported by B. V. Dean (1959) at Case and S. Kaplan (1959) at RCA. All these studies have found that the effort (e.g., in man-hours) expended on RDT&E projects tends to be distributed over the duration of the project in regular, repeatable patterns. Also, a set of major activities common to all projects has been observed. The life cycle models attempt to predict future RDT&E effort requirements by extrapolating the effort distributions over a given set of activities to a similar set of activities for future projects.

Since Norden's work is the most extensive and has been heavily borrowed from in our proposed model below, some additional comments on the basic structure of his model will be valuable. Norden is concerned solely with predicting man-hours of RDT&E effort. No direct attempt has been made to predict expenditures for facilities, prototypes, experimental tooling, etc. Through a careful analysis of a project typical of the company's class of work a reference project has been developed. The manhour requirements for activities on future projects are then predicted on the basis of effort expended on corresponding activities in the reference project. In particular, a factor, Δ_a, is calculated which relates the complexity of the project being predicted to that of the reference project. The project complexity is defined as a function of the magnitude and difficulty of the task at hand. Magnitude is measured by a count of the smallest identifiable unit elements comprising the primary task, while difficulty is represented as an index based on previous comparable work. Norden bases his difficulty index on a ratio of some system performance measures to some component descriptors. An alternative method of defining difficulty is developed below.

Once each Δ_a has been determined, the effort distributions are fitted to the activities using the same forms as distributions in the reference project or comparable precious projects. The overlap between activities is then assessed and the total distribution of effort and the duration of the project computed. The exact manner in which these steps are to be carried out is illustrated further below with reference to published work where appropriate.

DEVELOPMENT OF THE RDT&E COST PREDICTION MODEL

To facilitate the construction of the life cycle model, we must first partition the total project into tasks and task components which can be costed individually and then summed to give the cost of the entire project. Figure 7-2 (adapted from Norden, 1958) shows schematically how this breakdown can be structured. If the project, P, for example, required the development of a new passenger vehicle, the tasks, T_i, might be such things as engine development, transmission development, suspension development, etc. The components, C_j, for the transmission development task might be torque convertor development, gear set development, case design, etc. Finally, the subcomponents, S_k, for the torque convertor component might be convertor turbine, stator, convertor clutch, etc. The detail which is possible or desirable naturally depends upon the magnitude of the project, P, and the ability to describe early in the project the relevant components and subcomponents. Norden suggests three guidelines for choosing the smallest subdivisions of a project:

1. They should be large enough to justify the assignment of at least one man for a significant period of time.
2. They should be clearly definable and, if possible, self contained.
3. They should be manageable by small numbers of men or teams.

Having defined the tasks, components, and subcomponents, the next step is to define the activities that must be performed on each T_i, C_j, and S_k to bring the project from conception to completion. A list of these appears in Table 7-I. We can now visualize a matrix with activities labelling the rows and subcomponents the columns. The sum of the effort distributions of all elements in this activity-subcomponent matrix thus gives the total effort for the RDT&E project.

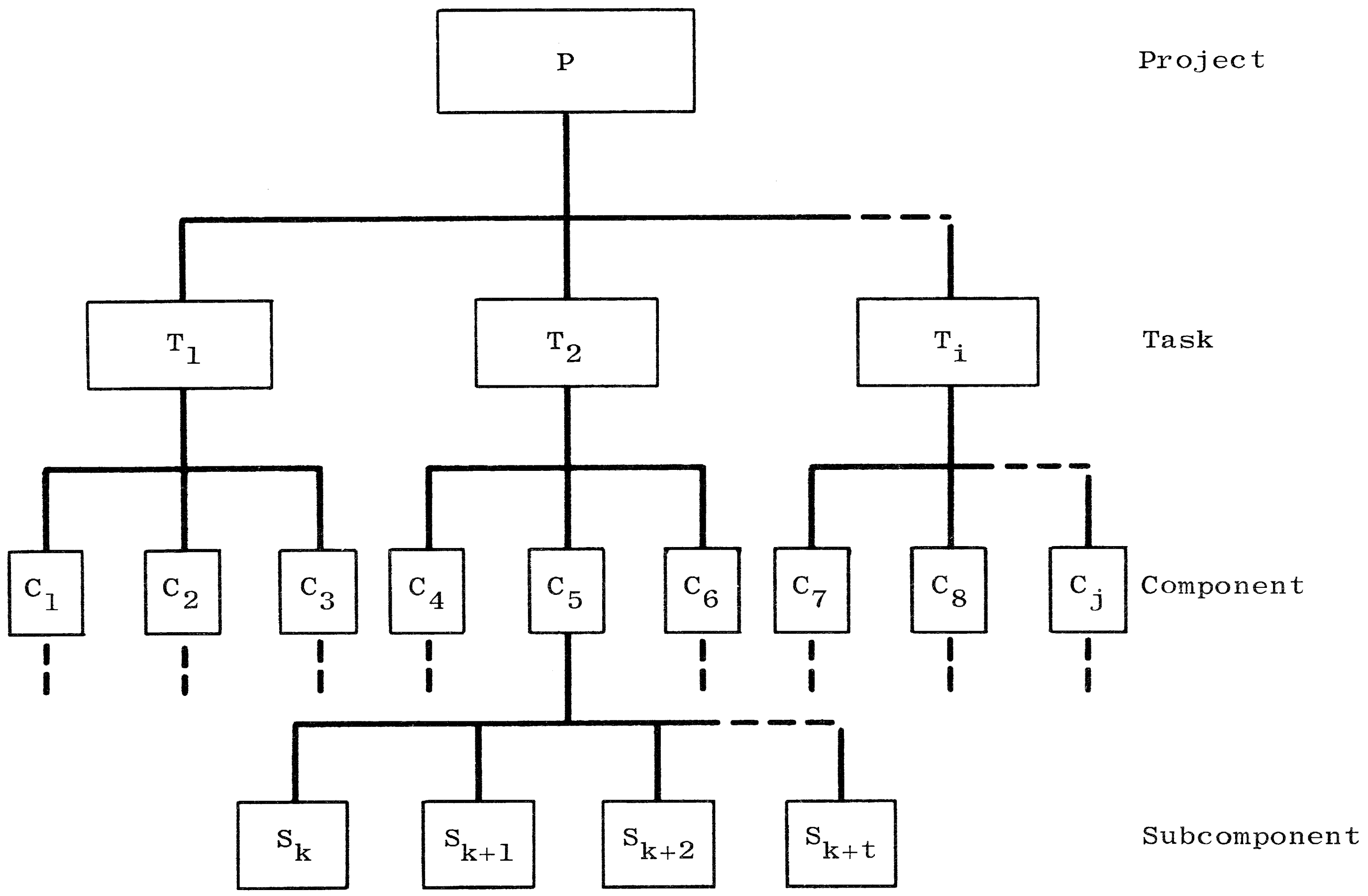


FIGURE 7-2. PROJECT BREAKDOWN INTO TASKS, COMPONENTS, AND SUBCOMPONENTS

TABLE 7-I. ACTIVITIES IN THE RDT&E STAGE

1. Initiation of Project

   Initial formulation of overall design and performance objectives. Processing project requests and obtaining approval and/or developmental funds.

2. Preliminary Analysis

   Problem identification and formulation. Literature search, external contacts. Basic analysis. Formulation of hypotheses and/or possible solutions. Evaluation of possible alternatives—"Which solutions look most promising."

3. Experimental Planning

   Experimental designs. Identify and purchase experimental equipmental equipment and materials. Computation and simulation of overall structure. Preliminary sketches of experimental structures.

4. Materials Research

   Identify possible construction materials. Chemical, metallurgical, physical tests of materials. Material development. Material specifications.

5. Subcomponent Fabrication

   Construct working models of subcomponents. Preliminary machinability and fabrication research.

6. Mock-Up

   Build breadboard models. Construct or purchase test equipment. Feasibility experiments. Analyze results. Evaluate alternative constructions and hypotheses against performance specifications.

7. Design Specifications

   Based on tests, write out detailed material, design, tolerance, and quality control specifications to be followed in subsequent design.

8. Engineering Design

   Engineering design of complete system. Weight calculations and control. Stress calculation. Investigation of alternative production methods.

9. Prototype Design

   Engineering of prototype model.

10. Prototype Layout

    Layout for prototype.

11. Prototype Detailing

    Detail drawings needed for prototype construction.

TABLE 7-1. ACTIVITIES IN THE RDT&E STAGE (Continued)

12. Prototype Construction

    Fabrication of prototype. Test of experimental construction techniques. Design changes to facilitate prototype construction.

13. Prototype Test and Evaluation

    Prototype test program. Environmental tests. Prototype and component evaluation. Design change recommendations.

14. Final Layout

    Preparation of layouts for production model.

15. Detail Drafting and Specifications

    Preparation of detail drawings, specifications, and assembly instructions for production model.

16. Pilot Line Models

    Production of engineering models on pilot lines.

17. Final Performance and Durability Test

    Testing of Pilot Line models against performance and durability requirements.

18. Engineering Releases to Production

    Preparation of production releases for purchasing and manufacturing.

The next step in the life cycle cost estimating process is to determine the starting point, duration, and associated effort distribution for each activity. Norden finds that a Pearl-Reed logistic curve of the form

$$F(t) = K\{1 + \exp(a + bt + ct^2 + dt^3)\}^{-1}$$

provides a good fit to the total cumulative effort distributions for a wide range of projects (Fig. 7-3). Naturally, there are advantages and disadvantages to any equation one employs for curve fitting, and there is no purpose in debating these here. For illustrative purposes, Norden's third-degree logistic curves are adequate.

The central issue of the cost prediction problem is the estimation of total expended effort, K, in the third degree logistic equation; the duration, t_n; and the "shape" parameters a, b, c, and d for a given activity. The shape of the effort distribution is no doubt dependent on a multitude of factors depending on the activity involved.

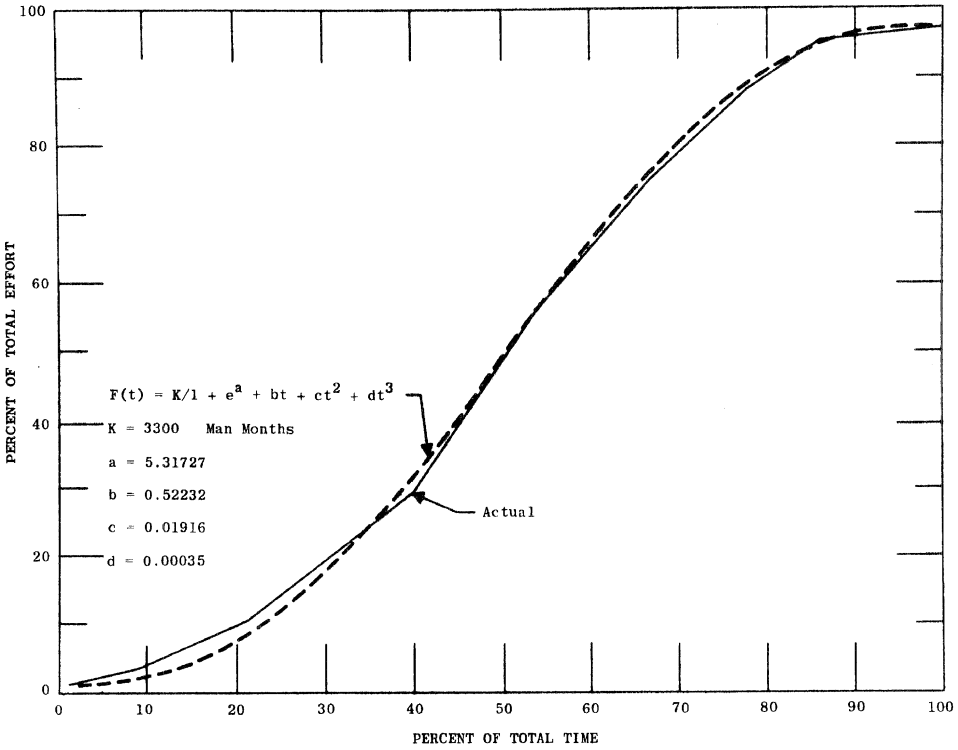


FIGURE 7-3. CUMULATIVE EFFORT DISTRIBUTION FOR THE DEVELOPMENT OF A LARGE COMPUTER SYSTEM (After Norden)

It seems reasonable, then, to assume that the shape of the effort distributions for activities in past projects would provide good predictors of the shape of effort distributions for similar activities in future projects. The total cumulative effort, K, for some activity is related to the duration of the activity, and the rate of application of effort, f(t) (for some arbitrary effort distribution) in the following simple manner:

$$K = \int_{t_o}^{t_n} f(t)\, dt$$

where $f(t) = \frac{dF(t)}{dt}$

Clearly, there are many possible functions, f(t), which when integrated (or, more realistically, summed) over the duration of the activity, $[t_o, t_n]$, will yield the same cumulative effort, K. Hence, the shape parameters of the effort distribution may be relatively unimportant to a decision maker interested only in total cumulative effort required for some activity. On the other hand, if budget allocations (e.g., by fiscal quarter) are required, a more precise determination of the shape parameters is required since the shape of the effort distribution will determine when funds and manpower will be needed throughout the duration of the activity.

To move on with the development of the model, let us assume that the shape of the effort distribution has been determined either by reference to past projects or through some other considerations. This leaves the determination of t_n and K. It is clear that as a project becomes more complex one expects its duration, t_n, and the total cumulative effort to increase in some manner. Also, for a given project, if the pace of development is accelerated (perhaps to meet a deadline) one expects the costs to increase (because of the need for parallel programs, inefficient intra-organizational communication, 'crash' construction of facilities, overtime, etc.). Figure 7-4 attempts to depict these relationships. Each curve is for a constant complexity level (iso-complexity curves). The index C = 1 refers to the complexity determined for the reference project. The other curves are for projects twice as complex, half as complex, etc. as indicated by the indices, C.

We have referred above to the method proposed by Norden to determine these complexity indices, C, i.e., as a function of the magnitude and difficulty of the activity. The notion of magnitude as the number of countable components described by the variables X_i of chapter 1 fits nicely into the general system schema developed thus far for the highway transportation system. An alternative means of measuring difficulty may be proposed, however. Suppose some component has a descriptor, X_i. For example, the component may be an automobile engine and the descriptor the brake horsepower per pound. Engines have undergone a fairly steady development in the last 70 years or so and the rate of change of brake horsepower per pound over time can no doubt be quite accurately computed. We may indicate the present rate of change by $\partial X_i/\partial t_o$. Suppose now the RDT&E project calls for some change in the component descriptor from the existing state of the art, call this change ΔX_i. The predicted duration of the activity may be represented as the difference between t_o and t_n or Δt. We may then construct the ratio $\Delta X_i/\Delta t/\partial X_i/\partial t$. This ratio

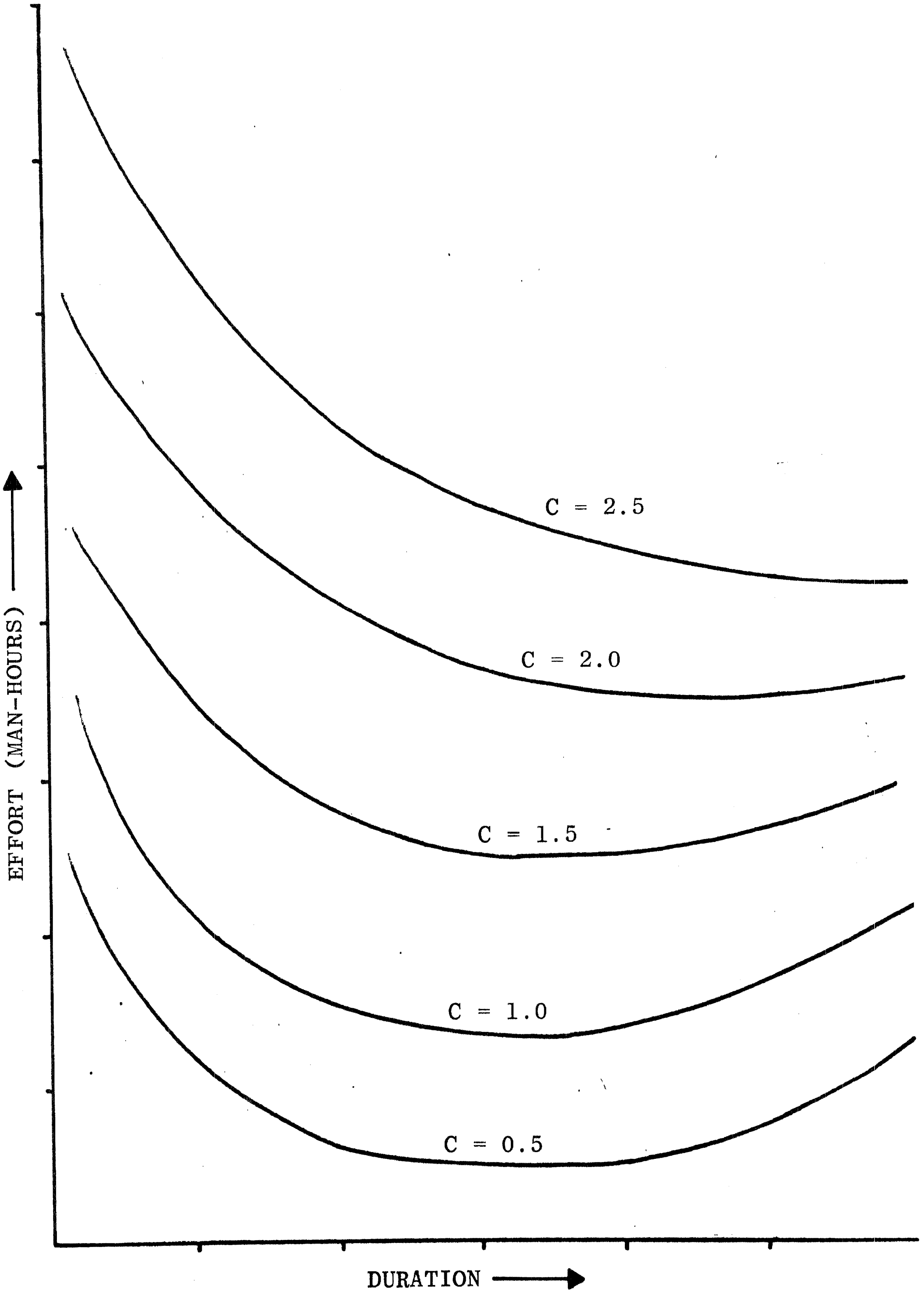


FIGURE 7-4. EFFORT-DURATION FUNCTIONS AT CONSTANT COMPLEXITY

then gives the level of difficulty of the project with respect to the X_i in question (assuming all ratios negative or zero are considered zero).

The project complexity may then be computed as a function of the magnitude, M, the levels of difficulty, d_{ij}, and a measure of the organization's experience in dealing with problems of this type, O. Thus the proposed complexity may be viewed as: $C = f(M, d_{11}, \ldots, d_{ij}, \theta)$. If the X_i's are suitably chosen it may be possible to develop the values, C, independent of a given firm or industry. Thus, the iso-complexity curves may be able to be derived empirically from the experience of a broad cross section of research and development activities.

Theoretically, once the iso-complexity curves are developed it is a simple matter to enter a chart like Fig. 7-4 with a given scheduled duration, t_n, and a complexity, C, and determine the required cumulative total manhours for development of an activity. Conversely, if, for reasons of budget constraints for instance, the available K is known t_n may be determined in an analogous manner.

The total manhours required for the total project is the sum of the manhours required for each individual activity. The total time required to complete the project is, however, not a direct sum of the durations of the individual activities since there may be some overlap between each activity and the one subsequent to it. The amount of this overlap must therefore be predicted as well. The amount of overlap between activity a and a + 1 can easily be described as a percent of total effort K. Thus activity a + 1 begins after X_a% of K has elapsed. The prediction of X_a for each activity would appear to be dependent on the complexity of a and a + 1 and on some measure of the organization's experience in projects of a similar type, thus:

$$X_a = f(C_a, C_{a+1}, \theta)$$

This sounds reasonable since in highly complex activities the results of activity a may well determine the direction of a + 1. On the other hand, in routine activities, or where the organization has more experience, a greater degree of dovetailing may be possible.

Using the preceding techniques we can structure the model to predict RDT&E manpower costs which as noted above have historically amounted to about 75% of the total expenditures for RDT&E projects. To calculate the remaining 25% of the costs there are two choices. Either one can predict the material and overhead cost elements as percent of the predicted direct labor, or material and facilities may be predicted independently of direct labor and overhead. In the latter case one may structure the cost prediction model for material and facilities as a procurement cost prediction. The general development of this structure is considered in the following section.

CONCLUSION

A tacit assumption of the preceding discussion was that the greatest cost prediction accuracy could be obtained from the life cycle technique when the smallest possible breakdown of a project into subcomponents and activities was made. This is theoretically true. However, such a detailed breakdown would also require correspondingly detailed knowledge of the effort-duration functions and the probable shape of the cumulative effort distributions for each activity and subcomponent. The accumulation of such detailed information is practically impossible except in a situation where an organization routinely produces many small RDT&E projects of a similar nature. Hence, what must be done to make the life cycle technique usable is to compact the activity-subcomponent matrix both vertically and horizontally until a workable set of components and activities remain. Norden (1958) has found that grouping all activities into three broad phases, creative (establishing feasibility), design (establishing practicability), and experimental hardware (physical embodiment), yields practical results. He also cuts off the vertical breakdown at the component level.

There can be no general guidelines set down on the amount of vertical and horizontal compaction desirable for a particular RDT&E project for a particular organization. The fineness of component-activity breakdown depends entirely on the overall size of the project and upon the amount of relevant data the organization can bring to bear to develop the effort-duration functions and the cumulative effort distribution.

PREDICTING PROCUREMENT COSTS

In this section a general model will be developed for the prediction of the procurement costs of a system modification project i.e., the costs of introducing into the overall system some new subsystem or some change in the physical arrangement of an existing subsystem. The relevant costs of the procurement of the system modification will again be allocated among certain cost elements such as direct labor, materials, overhead, tooling and equipment, etc.

In the early stages of a system modification project (see Fig. 7-2) the ultimate performance characteristics of the system after modification are specified. Because the complete hardware specifications will not be known with certainty early in the project, the cost estimator is tempted to predict procurement costs on the basis of these performance characteristics alone. Unfortunately this procedure soon encounters severe practical difficulties. For example, while two lengths of road may possess identical performance characteristics with respect to ability to handle traffic, passing sight distance, etc.; their respective costs will vary widely depending on the type of subgrade structure, surface materials, etc. As a further example, consider predicting vehicle procurement costs on the basis of some performance characteristic such as top speed. It is evident that two vehicles can possess the same top speed characteristic but be radically different in design and therefore cost.

The task facing the cost estimator is therefore to associate the required performance characteristics with the physical characteristics of the cost components of the subsystem. In the example above, this would mean that while the performance characteristic of top speed was a poor cost predictor, the gross vehicle weight and brake horsepower of the engine (which, along with a myriad of other factors, determine the top speed) may be better cost predictors. In Appendix 7-A a preliminary attempt is made to identify the major cost components and their respective cost component descriptors for the vehicle (Exhibit I), highway (Exhibit II), and recovery (Exhibit III) subsystems. One example of a cost component for the vehicle subsystem is the engine mentioned above, and one of its cost component descriptors is brake horsepower.

A consequence of the requirement that procurement costs be predicted on the basis of physical cost components is the possibility of errors being made in early estimates because of the imprecision with which the cost components and their descriptors were initially identified. Although the possibility of such errors must be recognized, we feel that the accuracy of the estimate will still be greater than if it were based on performance characteristics alone.

VARIABLE COSTS

All costs associated with a system modification project must be considered as variable costs in the sense that if the project did not exist, then the costs it generated would not be incurred either. A useful distinction can be made, however, between project variable costs and unit variable costs. For example, the differential costs resulting from producing the next unit in an ongoing production process is a unit variable cost, i.e., a marginal cost. The initial investments in plant, equipment, and tooling to produce some scheduled quantity are project variable costs. In either case, the costs incurred will be allocated among the cost element categories of direct labor, materials, overhead, etc.

For cost estimating purposes the distinction between unit variable costs and project variable costs is of fundamental importance since they are influenced by different factors. We have hypothesized that the unit cost of an item depends on its components and their characteristics. The cost also must depend in some way on production characteristics such as the number of units previously produced, the rate of production, and the total scheduled output for some relevant time period. That unit costs depend on the amount of prior production is well documented (Levenson and Barro, 1966) in the aircraft industry where the use of the "learning curve" is standard procedure for cost estimators. While less well documented, it is reasonable that the rate of output and total scheduled output should also affect manufacturing methods and thus unit costs.

On the other hand, the cost components and cost component descriptors of a subsystem, even if known in some detail, do not seem to be very valid predictors of project variable costs. While project variable costs are surely influenced in general by what is being produced, they are even more directly influenced by how something is produced. That is to say, an estimate of project variable costs is very sensitive to the production characteristics which will, in large measure, determine the size of the plant, the capacity of the machinery, etc. Furthermore, it would be unrealistic to expect to have atomistic process engineering data available early in the project. Thus, the initial procurement cost estimate for facilities tooling, etc. must be based on grosser measures such as the production characteristics of quantities, rates, duration, and the product characteristics of weight, height, material, etc.

THE LITERATURE OF PROCUREMENT COST PREDICTION

The majority of procurement cost predictions have been done on an ad hoc basis, and there are consequently very few general formulations of the procurement cost problem. However, various individuals at the RAND Corporation have made some progress in structuring the Air Force weapons system procurement cost prediction problem in a fairly general fashion. Most of this work was originally done for the Air Force's "Project RAND" but is now available in unclassified form through published RAND memorandums. The Levenson and Barro 1966 paper referred to above is one example of the work being done at RAND. More comprehensive formulations are given by G. H. Fisher (1962), David Novick (1961), and Petruschell and Chester (1963). The problems of data collection, cost categories, and the advisability of using a cost model are discussed in these articles. Since much of the RAND work refers specifically to Air Force costing procedures and aerospace systems, a thorough review is not necessary here; however, specific reference to RAND articles will be made where appropriate in the development of our procurement cost prediction model below.

Another body of literature dealing with system procurement problems is represented by the ex post facto studies of Fisher (1962) referred to above and Peck and Scherer (1962). These studies are attempts to determine what went wrong in cost estimates for weapons systems and why. The problem of requirements uncertainty has been discussed above. In addition, Peck and Scherer note that cost estimates based on contractor supplied cost information may be biased on the low side since the contractor was bidding low intentionally to gain a contract. Whether this intentional biasing has any direct bearing upon the highway transportation system costing problem will have to be investigated with further research. The problem of collecting good cost data upon which to base the cost prediction model is, however, of fundamental importance, and any future development of the cost prediction model proposed here must include a thorough evaluation of data sources.

Another relevant work and the direct progenitor of the model developed below is the procurement cost model developed by Bonder (1963) for the main battle tank. Because the model proposed in the following section is in essence identical to Bonder's, a detailed explanation is unnecessary. The major assumptions and proposed methodology carry over directly to the highway transportation system procurement cost model.

TOTAL PROCUREMENT COSTS

The preceding comments were purposely left general under the assumption that the suggested approach to predicing procurement costs would apply equally well to any subsystem, i.e., highway, vehicle, recovery, etc. However, an accurate cost prediction model must obviously take into account all the characteristics of the subsystem whose costs it is to predict. While any system is composed of components in some sense, not all systems require that these components be assembled to form the total system. In the case of the vehicle the assembly process and its associated costs are clearly defined. However, it is not clear that the construction of a highway wherein the subgrade structure, the surface, the shoulders, etc., are combined to form the highway constitutes an assembly process. Further, it is doubtful that the integration of the various components of the recovery system, i.e., the recovery vehicles, surveillance system, communication system, etc., into an effective system is analogous to an assembly process. Nevertheless, it will prove useful to develop a procurement cost prediction model which takes assembly costs, testing costs, repair costs, etc., into account. In the cases where such detail does not apply, the general model is still applicable but the nonrelevant costs are zero.

THE GENERAL PROCUREMENT COST PREDICTION MODEL

The model developed below is based on the following assumptions:

1. In the procurement phase of any system modification project, funds are expended to acquire some entity of measurable value. The total procurement cost is the sum of the unit costs (allowing the possibility that there may be only one unit) and the project costs. A unit cost may be defined variously for different subsystems, e.g., the cost per vehicle in a truck fleet, the cost per mile of highway, etc.

2. The new subsystem or change to an existing subsystem represented by the modification project is procured as an entity which consists of certain well defined components.

3. Where these components must be physically assembled into a workable system, assembly costs are incurred in addition to the cost of the individual components.

4. The production cost elements (direct labor, direct materials, overhead, production engineering, tooling, and general and administrative) can be so defined for each component as to be independent, mutually exclusive, and collectively exclusive.

5. The necessary data from past projects of an analogous nature are available by cost element so that cost estimating relationships for current projects can be derived.

6. Where necessary, costs incurred in different time periods can be summed or compared using present value techniques.

7. A prerequisite for the preceding assumption is that an appropriate discount factor can be computed from interest rates, cost of capital, or other suitable means.

If we adopt the point of view of a purchaser of some hypothetical system, the procurement cost incurred can be represented as:

$$C_{V,R} = V \cdot U_{V,R} + F \qquad (1)$$

where $C_{V,R}$ = total cost of V complete units produced with rate index R

V = number of complete units

$U_{V,R}$ = unit cost at constant V and R

R = rate index which describes the production rate with respect to its mean and variability

F = facilities (land, buildings, etc.) differential with this project

Now

$$U_{V,R} = \sum_{m=1}^{M} c_j + A_{V,R,M} + P + S \qquad (2)$$

where c_j = unit cost of jth component

M = number of components in system

$A_{V,R,M}$ = the unit assembly cost of M components at constant V and R, to include costs of inspection, test, repair, and scrap

P = unit profit received by producer

S = unit shipping and other transfer charges

As noted above, the cost of each component can be allocated among several elements denoted by e_{jm}. We will assume that some unit elemental costs vary in some manner with changes in V and R and that some do not. For example, the "learning phenomenon" may clearly affect unit direct labor costs, i.e., unit labor costs may decrease as volume increases. On the other hand, unit material costs may (assuming no quantity discounts) remain constant over a wide range of V and R. In the case of expendable tooling, the cost per unit may be constant for all V but may vary significantly as R is changed.

Thus, for an arbitrary number, D, of cost elements (labor, material, O.H., etc.) we may form:

$$c_j = \sum_{m=1}^{A} e_{jm} + \sum_{m=A+1}^{B} e_{jm} + \sum_{m=B+1}^{C} e_{jm} + \sum_{m=C+1}^{D} e_{jm} \qquad (3)$$

where the terms e_{jm} for m = 1, . . . , A are invariant under changes in either V or R, i.e., $\partial e_{jm}/\partial V_j = 0$, $\partial e_{jm}/\partial R_j = 0$

the terms e_{jm} for m = A + 1, . . . , B may vary with changes in V but are invariant under changes in R, i.e., $\partial e_{jm}/\partial V_j \neq 0$, $\partial e_{jm}/\partial R_j = 0$

the terms e_{jm} for terms $m = B + 1, \ldots, C$, do not vary with V but may vary with R, i.e., $\partial e_{jm}/\partial V_j = 0$, $\partial e_{jm}/\partial R_j \neq 0$

the terms e_{jm} for $m = C + 1, \ldots, D$ may vary for changes in both V and R, i.e., $\partial e_{jm}/\partial V_j \neq 0$, $\partial e_{jm}/\partial R_j \neq 0$

Since the costs associated with the remaining terms in Eq. 2, i.e., $A_{V,R,M}$, $P_{V,R}$, and $S_{V,R}$, vary so widely from system to system and industry to industry, we will not attempt to deal with them in any detail here. Profit, $P_{V,R}$, may well be calculated as fixed percentage of $\sum_{j=1}^{M} C_j$ in which case it clearly presents no estimating difficulty. Shipping charges, $S_{V,R}$, may be fixed per unit (or zero in the case of a highway). The determination of the assembly costs $A_{V,R,M}$ is a more difficult task and we shall make a few remarks on this problem a bit later.

Returning to the elemental costs, e_{jm}, we note, on the basis of our introductory remarks above, that these costs should be functionally dependent on the performance characteristics of the jth component (i.e., the jth component's cost component descriptors). If we denote the cost component descriptors as x_{jk}, we may now state several hypotheses concerning the nature of these dependencies.

Hypothesis I (e_{jm} for $m = 1, \ldots, A$)

Since we have assumed e_{jm} is not dependent on either V or R but is dependent upon the cost component descriptors, x_{jk}, we may write:

$$e_{jm} = f_{jm}(x_{j1}, \ldots, x_{jk}, \ldots, x_{jm}) \tag{4}$$

An example of such a cost element might be a vendor-purchased part for the vehicle subsystem (nuts, bolts, etc.), the cost of which is constant for any volume and production rate. The function $f_{jm}(\cdot)$ has been referred to previously and is called a cost estimating relationship. A cost estimating relationship then is simply a mathematical statement of how cost is affected by other variables. The exact functional form, $f_{jm}(\cdot)$, is usually empirically derived through regression analysis. Appendix 7-B is devoted to an example of a cost estimating relationship to illustrate its derivation and application.

Hypothesis II (e_{jm} for $m = A + 1, \ldots, B$)

Since we assume e_{jm} may vary with production volume V but not with rate R we may write:

$$e_{jm} = f_{jm}(x_{j1}, \ldots, x_{jk}, \ldots, x_{jn}, V_j) \qquad (5)$$

We might imagine as an example a supplier who can furnish a particular item (e.g., from his inventory) at any rate the system producer may require at no variation in cost. But he may be willing to reduce his price, e_{jm}, if the system producer can guarantee him a large volume, V. Again the form of $f_{jm}(\cdot)$ must be empirically derived. Bonder (1963) suggests that $f_{jm}(\cdot)$ should consist of a constant term plus a term that decays exponentially in V.

Hypothesis III (e_{jm} for $m = B + 1, \ldots, C$)

We assume that the elemental cost does not vary with volume but does vary with production rate, thus:

$$e_{jm} = f_{jm}(x_{j1}, \ldots, x_{jm}, \ldots, x_{jn}, R_j) \qquad (6)$$

By way of example, consider a fixed production facility making widgets at d units per day on straight time. If the production rate R is suddenly increased to fulfill a given volume V (e.g., by a decrease in a delivery deadline) overtime shifts will be required thus increasing the component labor cost element, e_{jm}.

Hypothesis IV (e_{jk} for $k = C + 1, \ldots, D$)

In this case the cost element cost varies with changes in both production volume and rate, thus:

$$e_{jm} = f_{jm}(x_{j1}, \ldots, x_{jm}, \ldots, x_{jn}, V_j, R_j) \qquad (7)$$

This, of course, is the most common case. Consider, for example, the labor costs in an automobile body. If the total anticipated volume V is small, say, for prototype build, then the unit labor costs will be high because mostly hand tooling, general purpose presses and fixtures, etc. are used in the construction. If the production rate R is increased at this low volume, costs will further increase because of overtime, etc. As V is increased so that higher volume production techniques become feasible the unit labor costs decrease for two reasons. First, the increased mechanization requires a smaller per unit expenditure of labor. Secondly, the "learning effect" occurs and worker movements become efficient thereby further reducing the necessary per unit labor content. In the majority of elements one would expect the elemental costs, e_{jm}, to be sensitive to changes in both R and V over some range of values for R and V.

A further development of the preceding ideas in the form of several propositions concerning the relationship of total procurement cost to production volumes, rates, and the timing and duration of the production schedule has been proposed by Alchian (1959). Since these propositions, though relevant, are a bit of a digression they are included as Appendix 7-C.

The testing of the four hypotheses just stated (along with the four additional propositions listed in Appendix 7-C) with respect to the highway transportation system model, of course, constitutes a major research task. Hence, they are introduced here to provide a structure which may be useful for future research. Although this structure appears valid in the light of current economic thought, the true validity of the hypothesized relationships can only be confirmed after the required cost data from actual systems have been collected and the empirical relationships derived.

ASSEMBLY COST

The assembly cost, $A_{V,R,M}$, will now be examined in more detail. We have included as 'assembly' costs not only those costs associated with physically joining the M components into a system but also the inspection, test, repair, and scrap costs associated with the assembly process as well. We may represent the assembly cost as

$$A_{V,R,M} = J_{V,R,M} + I_{V,R,M} + T_{V,R,M} + R_{V,R,M} + S_{V,R,M} \tag{8}$$

where $A_{V,R,M}$ = total 'assembly' cost per unit

$J_{V,R,M}$ = unit cost of joining M components into an assembly

$I_{V,R,M}$ = inspection cost per assembly

$T_{V,R,M}$ = testing cost per assembly

$R_{V,R,M}$ = repair cost per assembly (also called rework cost)

$S_{V,R,M}$ = scrap cost per assembly

and all factors are for a given V,R,M.

The unit cost of joining M components into an assembly $J_{V,R,M}$ may be assumed to be related to the number M and also to some of the component descriptors x_{jk} such as the size, weight, etc., of the component. Thus we assume

$$J_{V,R,M} = f(M, x_n, \ldots, x_{jk}, \ldots, x_{Mn}) \tag{9}$$

Further, it seems reasonable that the cost $J_{V,R,M}$ should increase at an increasing rate as the number of components M increases $\left(\text{i.e., } \partial J_{V,R,M}/\partial M > 0 \text{ and } \partial^2 J_{V,R,M}/\partial M^2 > 0\right)$. It is not possible to make a general comment about the nature of the appropriate x_{jk}'s since these must be discovered through experience with the actual system. However, we feel intuitively that the physical descriptors of the component such as height, length, weight, etc. should be important because there should be some optimum size and weight of a component in a given assembly process above or below which the assembly costs increase at an increasing rate, i.e., the function $J_{V,R,M} = f(\cdot)$ is convex in some x_{jk}'s.

Unit inspection costs, $I_{V,R,M}$, can also be assumed to be a function of M and some x_{jk}'s. In this case, in addition to component weight, size, etc. cost will also be a function of such factors as assembly tolerances.

The nature of the inspection process, e.g., destructive or nondestructive, will also affect costs. Further the probabilities of passing or failing inspection determine not only inspection costs but repair and scrap costs as well. In addition, the probabilities of the final assembly passing the final tests must be considered in assessing the testing cost. Because the development of these relationships becomes involved, the interested reader is referred to Appendix 7-D where they are developed in some detail.

We have assumed that the total unit assembly costs, $A_{V,R,M}$, as represented by Eq. 8, are sensitive to changes in the contemplated required production, V, and the rate index, R. It should be possible, at least theoretically, to predict from Eq. 8 the changes in assembly costs for any given changes in V and R. Furthermore, since the terms on the right of Eq. 8 depend in some way on the characteristics of the individual components, it should also be possible, once the required relationships are derived, to predict how assembly costs will change as the characteristics of the component parts of the system are changed.

Obviously, a great deal of data collection and analysis is required before equations such as Eq. 8 can be derived with any reasonable accuracy. The purpose of the preceding discussion was to indicate, in a general way, the form such equations might take and how they could be used. In Appendix 7-B we show how a cost estimating relationship of the form $C = f(x_{jk})$ is derived, and what some of the uses and limitations of such a relationship are.

PREDICTING OPERATING AND MAINTENANCE COSTS

The task of predicting operating and maintenance costs is simply described: given some well defined configuration of the total highway transportation system, we must discover a means of predicting the costs of operating and maintaining that system over some relevant time period. There are two major aspects to the O&M cost prediction problem. The first is to identify the usage rates and costs associated with the current operation and maintenance of the system. Such things as gasoline, oil, and tire consumption for the vehicle subsystem; patching, bridge painting, and sign repair for the highway subsystem; highway patrol and legal costs for the enforcement subsystem; should be included in this category. The second major aspect is the prediction of the cost of construction and/or procurement of O&M facilities such as service stations and garages, inspection stations, highway department maintenance garages and equipment, and also the initial training costs required by the introduction of a new system.

The general procedure we are proposing is to recognize that the O&M cost of the total highway transportation system is the sum of the O&M costs of the component subsystems. Further, the O&M costs of a subsystem are assumed to be the sum of the costs of its O&M cost constituents. The costs incurred by each constituent can then be allocated among the usual cost elements of labor, material, etc. For example, in the vehicle subsystem a typical cost constituent would be lubrication costs, which could further be accounted in terms of material, labor, and overhead costs.

LITERATURE ON OPERATING AND MAINTENANCE COSTS

In the RDT&E and Procurement sections we noted a scarcity of material relevant to the cost prediction problem. The same is generally true for the prediction of O&M costs. However, there is a great body of literature devoted to actual operating costs of some subsystems of the total highway transportation system. These reports are valuable in helping to identify the relevant cost constituents and to provide preliminary functional relationships necessary for the cost prediction model. The literature deals mainly with three distinct areas: vehicle operating costs, accident costs, and highway maintenance costs. Since the amount of literature is so extensive, it is not feasible to give a detailed analysis here. Instead a general overview is given and some detailed comments made on current articles of particular interest.

The two principal public organizations doing research on vehicle operating costs are the Highway Research Board (HRB) and the American Association of State Highway Officials (AASHO). Although numerous other public and private organizations have conducted such tests from time to time, the literature is dominated by HRB and AASHO. One current study of the HRB (Chaffey, 1965) is typical of the work being done and, in addition, includes an extensive bibliography on vehicle operating costs. The objective of this study is to obtain accurate information on vehicle running costs through road tests. The cost information is being collected to facilitate cost analysis of proposed highway projects. The study relates fuel consumption, tire wear, etc. to factors such as vehicle speed, highway geometrics, and environmental conditions. The instrumentation and methodology of the HRB studies are among the most sophisticated being used in the United States today.

A study that uses as input data similar to those being collected by HRB is being done by Lang and Robbins (1962) at MIT. The Lang and Robbins approach corresponds closely to the systems modelling approach being recommended in this chapter. Using a computer simulation, they have developed a model to predict vehicle operating costs as a function of the highway alignment and operating speeds. Given basic vehicle engineering data such as height, frontal area, tire size, torque curve, air and rolling resistance, brake specific fuel consumption as a function of piston speed, etc., the program computes the required functions relating fuel consumption to operating speeds and horsepower requirements. The vehicle simulation program then "runs" a specified vehicle on a given roadway at a given distribution of speeds and records the resulting fuel consumption. Although the present program does not provide output data on tire wear, oil consumption, maintenance costs, etc., the basic simulation approach could be adapted to do this.

The literature on accident costs is equally as extensive as that for vehicle operating costs. However, as noted in Chapter 11, no comprehensive scheme exists for predicting highway accidents. Similarly, no comprehensive scheme exists for calculating the cost of accidents. The difficulty stems from two sources. First, there is no universal agreement upon what costs to include in the "cost of an accident." Second, the accumulated cost data are incomplete since a large proportion of property damage accidents are never reported. Most reports on the cost of accidents solve the problem of which costs to include by including only so-called "direct" costs, i.e., the differential monetary value of damages and losses to persons and property incurred as a direct result of the accident. Such direct costs include: monetary value of damage to property, hospitalization, lost work time, damage awards, and attorney's and other legal fees. The direct cost approach has been taken in recent comprehensive studies in Massachusetts (Twombly, 1960), Illinois (Billingsley and Jorgenson, 1963) and California (Hosking, 1957).

A more comprehensive viewpoint is adopted in a study done by the Stanford Research Institute (SRI). Essentially the SRI study (1964) is a cost effectiveness study for an expanded national safety program to reduce the costs of traffic, home, and public accidents. In part, however, the study seeks to develop a methodology for assessing the economic costs of traffic accidents, and this is our principal concern. The general approach taken is to relate the various accident cost constituents to the diverse segments of the economy. In addition to the direct costs to those involved in the accidents, an attempt is made to develop an analytical framework to determine the general economic consequences of accidents as represented by foregone consumption and sales losses in various industries, by demographic changes, by increased costs to government and business and other adverse effects on the general economy. To estimate the economic losses (other than immediate direct costs) arising from accidental deaths, a calculation of the total cumulative family income loss resulting from the death of a present or potential head of a household is made. These losses are further assumed to be passed on to business in terms of lost sales and to government in terms of lost taxes. Although the SRI study presents these estimates of the costs associated with deaths from accidents and the number of accidental deaths, it consistently avoids dividing cost by fatalities to obtain an estimate of the average cost to various segments of society of the loss of one human life. Because in effect this procedure places a material value upon a human life, it is frequently rejected without investigating its usefulness for cost-effectiveness analysis.

Another approach to the problem of assessing the human life has been taken by two French highway engineers, Jacques Thédié and Claude Abraham (1961). The central question they pose is: How much should a community agree to spend to save one human life? Their approach is to total the losses, both direct and affective, incurred by the community as the result of one death. The cost categories considered are material damages and medical expenses; administration expenses, i.e., insurance, legal, and enforcement expenses; and production losses. The production losses are calculated by means of complex formulae which take into account the production lost net of probable consumption of services in terms of national income accounts and of free services (e.g., housewives' services). The net result of the calculation is to assign (at 1960 price level) a break-even value of 125,000 NF ($25,320) to the loss of a human life, i.e., the community should be prepared to spend 125,000 NF on a project which will prevent the loss of one human life due to a road accident (in France). Further discussion of the cost of accidents and the desirability of calculating the cost of a human life is given below.

The problem of accounting for costs arising from unreported accidents is handled by the State of Illinois (1962) by using a sampling procedure. Two sources were used: previously filed accident reports and interview forms mailed to a random sample of vehicle owners selected from license plate numbers. From a comparison of the accident reports on file and the returns of the interview forms an estimate of the costs involved in previously unreported accidents was obtained.

Literature dealing with highway maintenance costs for the most part deals with specific costs in specific localities. The bulk of the literature therefore is of little help in developing a general highway maintenance cost prediction model except to indicate which cost constituents to consider. One notable exception to the general trend is a recent study reported by Lawrence Mann, Jr. (1965) at Louisiana State University. The approach presented there meshes nicely with the general approach proposed here. Mann develops a regression model relating predicted expenditures on various cost constituents to such physical characteristics of the roadway as condition of subsoil, surface condition, surface width, width of right-of-way, and traffic volume. Since his procedure is so closely aligned with the one proposed here, further detailed discussion is presented below.

Virtually nothing appears in the literature discussing the O&M costs associated with the legal and recovery subsystems. Another neglected area is the prediction of facilities, manpower, and training costs. Since our aim is to include these costs in our model, a few brief comments will be included below on these topics in the discussion of the O&M cost prediction model.

THE O&M COST PREDICTION MODEL: GENERAL COMMENTS

The ability to predict O&M costs, as we are proposing, presupposes the existence of a model of the total highway transportation system (cf. Chapter 1). Since the inputs to this proposed model have been described in detail previously, it is not necessary to reiterate them here. It should be apparent, however, that the output O&M costs will be sensitive not only to the operating characteristics of the individual subsystems (vehicle fuel consumption rates, highway wear rates, etc.) but also to the interaction between subsystems (traffic congestion, dynamic bridge loading, etc.). For this reason, one cannot speak accurately of the O&M costs of any subsystem independent of all the other subsystems. On the other hand, to attempt to include the description of these interactions along with the discussion of the proposed O&M cost prediction model would hopelessly obscure the basic structure of the model. Hence, for clarity, the discussions of the O&M cost models are kept separate here.

The proposed format for the O&M cost prediction model differs from that for RDT&E and Procurement in that the prediction of current operating and maintenance expenditures is more nearly an accounting problem than a forecasting problem. Since the proposed highway transportation model will yield as output the complete operating profile of the vehicle (miles traveled, engine RPM and power consumption vs. time, wheel slip, fuel consumed, etc.) it will only be necessary to multiply the usage (e.g., fuel consumption in gallons) by the appropriate cost factor (e.g., $0.359 per gallon) to obtain the operating cost. Similarly, the system degradation model (cf. Chapter 5) will predict tire wear and failure; engine, transmission, and driveline wear and failure; etc. The cost model will then account for the cost of such wear and failures by applying the appropriate cost factors. The cost prediction problem with respect to facilities, training and other "one time" costs is however again a forecasting problem in the same sense as that for RDT&E Procurement expenditures. Unfortunately, there is no available methodology for predicting facilities and training costs. Most military predictions (Summers, 1962) are done on a purely historic basis. This indicates that a substantial amount of research must be done before the prediction of facilities and training costs can be placed on as sound a basis as current operating and maintenance costs.

To facilitate the general understanding of O&M cost prediction model, the structure of the model for several major subsystems will be explained individually in the sections that follow. We will be considering vehicle, highway, and recovery system O&M costs in turn. Finally the cost of accidents, since it cuts across several subsystems, will be dealt with as a separate section.

O&M COSTS FOR THE VEHICLE SUBSYSTEM

As we remarked above, the determination of vehicle operating costs is straightforward once the required outputs of total highway transportation system model and the system degradation model are available to the cost model. From a practical standpoint, however, the construction of the cost model should recognize that highway traffic is composed of a widely diverse population of vehicles. Since each type of vehicle has different fuel, oil, tread rubber, replacement parts, etc., requirements it is difficult to envisage an accurate cost model which could attempt to determine the O&M costs of some "typical" vehicle. Conversely, since there are so many different kinds of vehicles, to attempt to keep track of the appropriate cost factors for each individual type would be a tremendous task. What is needed is a classification scheme which is detailed enough to recognize the heterogeneity of the vehicle population but not so detailed that its use would be computationally infeasible. One suggested classification scheme appears in Table 7-II. The further refinement of the classification categories is one of the research tasks that lies ahead.

Assuming a workable classification scheme has been developed, we now proceed to identify the relevant O&M costs to be predicted for the vehicle submodel. Since everyone is familiar with the kinds of costs incurred through vehicle operation the list in Table 7-III will suffice to identify them. To calculate costs associated with depreciation, insurance, garaging and parking fees, statistical averages for each vehicle class may be used. Since tax rates vary according to location, the particular rates chosen will depend on the purpose of the cost prediction being made.

There is no consensus on the appropriateness of including personal time as a cost for non-commercial vehicles. It can be argued (87th Cong., House Doc. 54, 1961, pp. 209-12) that an individual attaches an implicit value on driving time by using a faster toll road as opposed to alternative free routes. However, it is not clear that such driving costs are real when viewed apart from the toll vs. freeway comparison. Nor is it clear that such imputed costs are relevant when viewed in terms of costs to the community. For these reasons, we suggest that driver time costs be considered relevant only for commercial vehicles.

As stated above, no general procedure has yet been developed for predicting operating and maintenance facility costs (see Table 7-III, indirect costs). For the time being we propose that future facility costs be predicted on the basis of historic facility requirements and cost adjusted for future price levels. The development of a general facility cost prediction model is then identified as a proper area for future research.

TABLE 7-II. VEHICLE CLASSIFICATION SCHEME[a]

- Passenger cars:
  - Light
  - Medium
  - Heavy
- Buses:
  - Transit:
    - <16,000 pounds
    - 16,000-19,999 pounds
    - 20,000-25,999 pounds
    - 26,000 pounds and over
  - Intercity:
    - <16,000 pounds
    - 16,000-19,999 pounds
    - 20,000-25,999 pounds
    - 26,000-31,999 pounds
    - 32,000 pounds and over
  - School and miscellaneous:
    - <8,000 pounds
    - 8,000-11,999 pounds
    - 12,000-15,999 pounds
    - 16,000-19,999 pounds
    - 20,000 pounds and over
- Trucks and combinations:
  - Single-unit:
    - 2-axle, 4-tire:
      - <4,000 pounds
      - 4,000-7,999 pounds
      - 8,000-11,999 pounds
      - 12,000 pounds and over
    - 2-axle, 6-tire:
      - <8,000 pounds
      - 8,000-11,999 pounds
      - 12,000-15,999 pounds
      - 16,000-19,999 pounds
      - 20,000-25,999 pounds
      - 26,000 pounds and over
    - 3-axle
      - <20,000 pounds
      - 20,000-25,999 pounds
      - 26,000-31,999 pounds
      - 32,000-39,999 pounds
      - 40,000-49,999 pounds

- Trucks and combinations: (Cont.)
  - Combinations:
    - 2-51:
      - <26,000 pounds
      - 26,000-31,999 pounds
      - 32,000-39,999 pounds
      - 40,000-49,999 pounds
      - 50,000 pounds and over
    - 2-52:
      - <50,000 pounds
      - 50,000-59,999 pounds
      - 60,000 pounds and over
    - 3-52:
      - <60,000 pounds
      - 60,000 pounds and over
    - 2-1:
      - <26,000 pounds
      - 26,000 pounds and over
    - 2-2, 3-1:
      - <26,000 pounds
      - 26,000-49,999 pounds
      - 50,000 pounds and over
    - 2-3, 3-2:
      - <60,000 pounds
      - 60,000 pounds and over
    - 3-3:
      - 60,000 pounds and over
    - 2-51-2:
      - 60,000 pounds and over

---

[a] Based on Table I-7, *Final Report of the Highway Cost Allocation Study*, 87th Congress, 1st Session, House Document No. 54, U. S. Government Printing Office, Washington, 1961.

TABLE 7-III. VEHICLE OPERATING AND MAINTENANCE COST CONSTITUENTS

DIRECT COSTS

Excluding Taxes:

Depreciation
Repairs and maintenance
Replacement of tires and tubes
Parts and accessories
Motor fuel
Engine oil
Transmission oil
Lubricants
Insurance
Garaging, parking
Driver time

Taxes and Fees:

Motor fuel taxes
Registration and misc. fees
Motor carrier taxes (commercial)
Titling
Bridge, tunnel, ferry and road tolls
Special city and county taxes (commercial)
Property taxes
Engine, transmission, and lubricating oil taxes
Social security taxes (commercial)

INDIRECT COSTS

Service stations
Maintenance garages
Parts suppliers (incl. parts inventories)
Terminal facilities (commercial)
Administrative and overhead (commercial)
Inspection stations
Weigh stations (commercial)
Towing service
Salvage service

O&M COSTS FOR THE HIGHWAY SUBSYSTEM

The general scheme proposed for the prediction of vehicle O&M costs is also valid for highway O&M costs. We assume, therefore, the prior development of a highway system degradation model which will predict the incidence and severity of highway maintenance tasks. For reasons similar to those suggested for the vehicle subsystem, a classification by highway types is a potential necessity. One such classification is shown in Table 7-IV. Also, a listing of highway O&M cost constituents is given in Table 7-V. The procurement and maintenance of road equipments as well as equipment maintenance facilities and expenses are considered as properly included in highway O&M costs. Similarly, administrative and engineering expenditures are considered O&M costs on a pro rata basis excluding the shares properly charged to RDT&E and Procurement.

Some essential differences between the costing approach being proposed here and those taken in other recent studies (Mann, 1965, and Baker, 1958) should now be noted. Man has developed a regression

TABLE 7-IV. HIGHWAY CLASSIFICATION SCHEME

Classification by Use[a]

- Freeway
- Expressway
- Major arterial, divided
- Major arterial, undivided
- Collector
- Local residential
- Local industrial
- Local business
- Alleys
- Drives and aprons

Classification by Construction Type[b]

- Primitive road
- Unimproved road
- Graded and drained earth road
- Soil surfaced road
- Gravel or stone road
- Bituminous surface treated road
- Mixed bituminous road
- Bituminous penetration road
- Bituminous concrete, sheet asphalt, or rock asphalt
- Portland cement concrete road
- Brick road
- Block road
- Combination type road
- Bridges

[a] Adapted from National Committee on Urban Transportation, *Determining Street Use* (Procedure Manual 1A; Chicago: Public Administration Service, 1958), pp. 1-2.

[b] U. S. Bureau of Public Roads.

TABLE 7-V. HIGHWAY OPERATING AND MAINTENANCE COST CONSTITUENTS

Highway Maintenance

- Road Surface
- Shoulder
- Curb and Gutter
- Sidewalks
- Drainage Ditches and Culverts
- Fences
- Landscaping
- Traffic Control Devices
- Street Sanitation
- Snow and Ice Control
- Street Lights
- Storm Sewers and Manholes
- Restore Cuts and Subgrade Structure
- Structures

Maintenance Department

- Administration
- Engineering
- Inspection
- Land and Buildings
- Equipment

Operations Department

- Administration
- Engineering
- Sign Shop
- Weighmaster
- Toll Collection
- Policing
- Revenue Collection (titles, licenses, etc.)

Financial Expenses

- Debt Retirement
- Interest on Debts

formula relating highway maintenance cost per mile to several measurable variables which described certain physical characteristics of the roadway. The roadway-traffic interaction is accounted for by Mann in terms of a measure of traffic volume only. The Baker study (1958) concentrated principally on determining the relationship between highway maintenance cost and the size and weight of vehicles comprising the average daily traffic. In contrast to these studies, the approach proposed in this chapter is to base highway maintenance and operating cost predictions on the output of the highway degradation model. The highway degradation model will, in turn, predict the amount of type of maintenance required by considering the known physical characteristics of the roadway and the associated vehicles and their interactions with the road surface.

Prediction of the costs of maintenance facilities and equipment, maintenance administration and overhead, operations facilities, and operations administration and overhead is a problem separate from the estimation of direct maintenance costs discussed above. The prediction of the cost of physical facilities is clearly a procurement cost problem and could presumably be structured in the manner suggested in the procurement section of this chapter. It is not certain at this time what system variables would make good predictors of facilities cost so this is another area for further research. Similarly, further study is needed to develop cost estimating relationships in a wide variety of other operating areas such as the administration of highway departments at the various governmental levels, the enforcement and legal system, driver education, etc.

THE COST OF ACCIDENTS

Of the costs associated with the operation of the total highway transportation system, most difficult to conceptualize is the cost of accidents. In particular, one must be specific in defining the cost recipient. In the case of vehicle, O&M costs could be considered as absorbed by the individual, business, or government owner of the vehicle. Highway O&M costs are borne initially by the various highway departments and agencies and are then redistributed to individuals and businesses through taxes. The recipient is not so clear in the case of costs associated with accidents since accident costs are felt both by private and public sectors of the economy. Some direct costs (e.g., property damage, medical expenses, loss of wages, etc.) are easily accountable in terms of the individual. Other costs, such as unreported damage to public owned property, may be incurred by government at some level. Finally, economic losses due to inefficiencies in production caused by temporary or permanent losses of skilled employees, loss of purchasing power, relief or other compensation, etc., may be an economic loss for the community or nation. A comprehensive cost model should properly allocate the losses due to any accident among the various public and private economic sectors. The SRI study (1964) referred to earlier is one attempt to develop such an allocation scheme.

The present analysis of highway systems does not attempt to solve the problems of associating costs with accidents. Rather, it assumes that measures of accidents and damage will be efficiency measures for the system, not subject to dollar-costing. Only the physical property damage will be costed. A review, however, of the general approaches to cost evaluation of accidents is undertaken below.

The state of Illinois study (1962) illustrates the approach of accounting for accidents in terms of the costs to an individual. Table 7-VI indicates some of the relevant cost constituents for the individual. It should be noted that a large portion of automobile insurance costs could properly be construed as an accident cost rather than an operating cost (Table 7-III). However, the Illinois study adopts the view that only differential accident costs should be considered. This approach also fits best with the cost model proposed here. By recognizing only differential accident costs the costs of accidents can be computed directly from the output of the accident prediction model (Chapter 11). The accident prediction model would supply predictions of the incidence and severity of accidents and the cost model need only extend these quantities by the appropriate cost factors to obtain the accident cost estimate. It is not clear at this time how the resultant costs to the total economy should be modeled or even what costs should be included in this larger category, but a comprehensive cost model should treat these fully.

TABLE 7-VI. ACCIDENT COST CONSTITUENTS FOR THE INDIVIDUAL

Property Damage

- Vehicle
- Contents and/or cargo
- Objects struck by vehicle
- Miscellaneous damage, loss of use, etc.

Medical Expenses

- Ambulance charges
- Hospitalization costs
- Doctor, dentist, nursing, etc. fees
- Medication, drugs, etc.
- Miscellaneous medical expenses

Legal Expense

- Attorney's fees
- Court costs
- Damage awards

Value of Time Lost From Work

- Direct losses for wage earner
- Additional expenses for housekeeping, etc.

(Less) Insurance payments and other compensation for any of the above.

The most difficult cost to assess is the economic loss incurred by the community due to traffic fatalities. While the thought of placing an explicit economic value on human life is somewhat repugnant, society does this implicitly constantly. Safety factors are built into bridges which increase their cost many times. However, trees and drainage ditches line the roadways and continue to take their toll in lives while their presence is justified on the basis of 'economics.' Highway planners are constantly faced with decisions requiring that costs be balanced with safety. The costs are usually well known, but how does one compare an expenditure on safety which may save ten lives against a different expenditure which may prevent one hundred disabling injuries? Are monies spent for stronger guard rails on bridges more effective than those spent for firmer shoulders? To aid the decision makers faced with these kinds of problems, the cost prediction model should be so structured that it could provide estimates of how much a community should be prepared to spend to save a human life, how much to prevent a disabling injury, etc. The precise manner in which this is to be done is an area for a great deal of further study. Both the SRI (1964) and Thédié and Abraham (1961) reports give some partial answers, but neither is directly adaptable to the costs prediction scheme proposed in this chapter.

O&M COST OF THE RECOVERY SYSTEM

The cost of the operation and maintenance of the recovery system might well have been considered as an accident cost. Certainly ambulance costs and various first aid and medical costs have already been considered above. Defining what the O&M costs of the Recovery System are is further complicated by the fact that (at least at present) the recovery system would exist in much the same form (though perhaps at a smaller scale) whether there were highway accidents or not. Ambulances would still be needed, as would hospitals, doctors, highway patrolmen, etc. In short, only the differential cost resulting from accidents of operating and maintaining these facilities and personnel should be considered as the cost of operating the recovery system.

The precise costs to be associated with the recovery system of course depend upon its makeup which is as yet only tentatively defined (Chapter 4). Because of this, few definitive comments can be made with respect to the cost prediction model for the recovery system. One would expect, however, that the recovery system cost model (when considered separately from the accident cost model) would consist of fixed and variable components. The fixed components being emergency facilities, recovery vehicles, etc. The variable components being fuel consumption of recovery vehicles, surgical supplies, etc. The incidence, severity, and location of crash victim with respect to emergency facilities would again have to be predicted by the accident prediction model and the necessary cost calculations carried through as before.

The costs associated with the recovery system as a totality should be viewed as a cost to the community as a whole. In the first place, it is doubtful that fees charged accident victims fully cover the differential costs of the recovery system resulting from accidents. Secondly, because accidents do occur, facilities and personnel are diverted from other medical needs of the community thus resulting in a decrease in the quality of non-accident patient care which is an indirect cost of the accident borne by the community.

Whether the costs associated with the operation and maintenance of the recovery system are viewed as a separate entity or as a part of accident costs is relatively immaterial. The important points to note are that only differential costs of the recovery system are relevant and that clear distinctions must be developed between costs incurred by the accident victim and costs, resulting from the operation of the recovery system, which are incurred by the community in general.

OTHER SUBSYSTEMS

A complete accounting of the operating and maintenance costs associated with the total highway system should include numerous other subsystems only alluded to here. The enforcement and legal subsystem and the insurance structure are among the most important which have not received attention. A recommendation that these systems be explored in depth is included along with recommendations to fully develop the cost prediction models for other areas under Further Research, which follows.

FURTHER RESEARCH

In the preceding sections, preliminary models for predicting the RDT&E, procurement, and O&M costs associated with any system modification project have been proposed. The development of these models was based in part on the meager published work on cost prediction. Principally, however, the development relied on plausibility arguments and a certain amount of intuition. In order to verify that the proposed model does in fact provide a useful model of the "real world" the major thrust of further efforts should be to test the model's assumptions, hypotheses, and structure against actual data.

Confirming that these necessary data are available, of course, constitutes a major research problem in itself. Since there is little immediate benefit to be derived from developing elaborate hypotheses whose testing depends on data not available, a reasonable 'next step' in the development of the cost prediction model would be a compilation of data requirements and an investigation of data sources.

Once having determined the availability of data, a workable cost allocation scheme should be devised to provide an unambiguous answer to the question "cost to whom?". Similarly a detailed analysis of the costs in each area must be undertaken to determine which costs can be measured and quantified.

The first key assumption to be tested is that of the assumed independence of costs incurred in each stage of project life. The development of the cost prediction model explicitly assumed that expenditures in one project stage would have no effect on costs incurred in any other project stage. As was argued in the introduction to this chapter, such independence is not at all assured by the operation of the "real world." On the other side of the coin, if independence of project stage costs is an invalid assumption the nature of project stage cost interdependencies must be discovered and the cost prediction model modified accordingly.

Once the project stage cost independence question has been resolved research efforts can be directed to the more straightforward (but by no means trivial) task of verifying the structure, assumptions, and hypotheses proposed in the cost prediction models for each of the three project stages. The verification of the O&M cost prediction model would presumably be the easiest since the data are most readily available and the research most complete in this area. The procurement cost prediction model would be more difficult to verify, principally because cost data may be difficult to obtain and analyze. However, since procurement cost models have been developed for other systems (notably military), some consolation can be gained from the fact that the course is not completely uncharted. Without question, the verification of the proposed RDT&E cost prediction model would be the most difficult. Fortunately, the bulk of expenditures in the transportation system area fall into the O&M and Procurement categories. Hence, a greater degree of uncertainty can be tolerated in the RDT&E cost prediction than for the other two stages.

The general problem of uncertainty was touched upon briefly in the introduction to this chapter. That cost estimates are uncertain and usually too low is something 'everyone knows'. From what we have seen in the military studies, predictions of the uncertainties of cost prediction are at least as uncertain as the cost predictions themselves. Furthermore, the military people have been in the system cost prediction business for years and have a large data base of predicted vs. actual costs to analyze where we do not. The result is that while it is good to recognize that uncertainty is inherent in any cost prediction, it would be best to postpone any detailed consideration of that problem until experience with cost prediction models in the highway transportation area is accumulated.

Even without considering the problem of uncertainty, it should be apparent that much remains to be done if the theoretical concepts considered in earlier sections of the chapter are to be translated into a useful cost prediction tool. In essence, this chapter should be considered as only providing the basic guidelines necessary for the development of hypotheses on the structure of the cost prediction models which in turn would have to be tested against actual data.

Appendix 7-A
COST COMPONENTS AND COST COMPONENT DESCRIPTORS

In the following pages a preliminary attempt will be made to indicate appropriate cost components and cost component descriptors Exhibit I (Vehicle Subsystem); Exhibit II (Highway Subsystem); and for Exhibit III (Casualty Recovery Subsystem).

Exhibit I (Vehicle Subsystem)

Cost Components and Cost Component Descriptors

A. Engine

1. Weight
2. Length
3. Height
4. Width
5. Installed Volume
6. Installed Weight
7. No. of Cylinders
8. Valve Arrangement
9. Bore and Stroke
10. Piston Displacement
11. Compression Ratio
12. Brake Horsepower at RPM
13. Foot Pounds Torque at RPM
14. Octane Requirement
15. Brake Specific Fuel Consumption
16. Idle Fuel Consumption
17. Idle Speed
18. Taxable Horsepower
19. Cylinder Block Material
20. Cylinder Liners Material
21. Engine Manifold Vacuum at Idle
22. Crankshaft Bearings Number and Type
23. Connecting Rod Bearings—Type
24. Connecting Rod Bearings—Material
25. Piston Material and Surface Treatment
26. Crankshaft Type and Material

27. Camshaft Drive
28. Number and Type of Camshaft Drive
29. Valve Lifter Type
30. Valve Spring Type
31. Oiling System Type
32. Normal Oil Pressure
33. Oil Reservoir Capacity
34. Cooling System Type
35. Water Temperature Control Type
36. Cooling System Capacity
37. Fan Diameter
38. Fan Drive

B. Fuel System

1. Fuel Type
2. Fuel Tank Capacity
3. Fuel Pump—Type
4. Fuel Pump—Drive
5. Fuel Pump Pressure - At Pump Outlet
6. Fuel Pump Pressure - At Carburetor Inlet
7. Fuel Filters - Type, Location
8. Carburetor - Type
9. Air Cleaner - Type
10. Sump Capacity
11. Manifold Heat Control Valve
12. Carburetor Heat Source

C. Exhaust System

1. Exhaust Manifold—Type
2. Exhaust Manifold—Material
3. Emission Control System—Type
4. Exhaust Pipes—Diameter, Length
5. Mufflers—Type
6. Silencers—Type
7. Tail Pipes—Diameter, Length

D. Transmission

1. Weight
2. Length
3. Height
4. Width
5. Installed Volume
6. Maximum Input Torque
7. Maximum Output Torque
8. Maximum Output Horsepower
9. Clutch Type (S-M only)
10. Clutch Pedal Pressure (do.)
11. Driver Plate Diameter (do.)
12. Driver Plate Facings (do.)
    a. Type and Number
    b. Attachment to Plate
    c. Facing Thickness
    d. Total Effective Area
13. Clutch Springs (S-M only)
    a. Type and Number
    b. Total Pressure
14. Transmission Mounting—Type
15. Oil Capacity (S-M)
16. Type of Gearing (do.)
17. Transmission Ratio (do.)
    a. Reverse
    b. First Speed
    c. Second Speed
    d. Third Speed
    e. Fourth Speed
18. Speedometer Gears—Type

19. Torque Multiplication (auto. only)
    a. Low Range and Drive Range Before Convertor Clutch Engagement (Stall)
    b. Drive Range after Convertor Clutch Engagement (Stall)
20. Convertor Clutch Engagement Control (auto. only)
21. Planetary Gearing Type (do.)
22. Number of Pinions (do.)
23. Pump Pressure Regulation (do.)
24. Fluid Capacity (do.)

E. <u>Drive Line and Axle</u>

1. Rear Axle Type
2. Drive and Torque
3. Rear Axle Oil Capacity
4. Ring and Pinion Gear Set Type
5. Reduction Ratio
6. Rear Axle Weight
7. Number of Universal Joints
8. Number of Constant Velocity Joints
9. Torque Tube
    a. Length
    b. Diameter
    c. Moment of Inertia

F. <u>Chassis Suspension System</u>

1. Wheel Type
2. Rim Type
3. Rim Size
4. Tire Size
5. Tire Type
6. Shock Absorbers
    a. Front
    b. Rear
7. Spring Type
    a. Front
    b. Rear
8. Stabilizers, Type
9. Upper Control Arms - Front
10. Lower Control Arms - Front
11. Lower Control Arms - Rear
12. Front Knuckle - Type, Mounting

G. <u>Steering Gear</u>

1. Gear Type
2. Housing Material
3. Ratio, Gear Only
4. Ratio, Overall
5. Turns of Wheel, Lt. to Rt.
6. Lubrication
7. Oil Capacity
8. Steering Wheel Diameter
9. Number and Type of Pitman Shaft Bearings
10. Effort Necessary at Wheel Rim for Initial Hydraulic Assist (Power)
11. Effort Necessary at Wheel Rim for Full Hydraulic Assist (do.)
12. Pump Capacity (do.)
13. Relief Value Opening Pressure (do.)
14. Pump Test Pressure (do.)
15. Steering Linkage Type
16. Toe-in
17. Caster

18. Camber
19. Turning Circle Diameter

H. Brake Systems

1. Operating Mechanism
    a. Service Brakes
    b. Parking Brakes
2. Percent of Total Braking Power on:
    a. Front Brakes
    b. Rear Brakes
3. Static Pressure in Hydraulic System
4. Number of Brake Shoes at Each Wheel
5. Lining Type, Thickness
6. Master Cylinder Piston Diameter
7. Wheel Cylinder Size
    a. Front
    b. Rear
8. Brake Drum I.D.
9. Power Assist Pump Capacity
10. Pump Test Pressure

I. Electrical System

1. Battery
    a. Type
    b. Capacity
    c. Voltage
    d. Case Type
    e. Dimensions
    f. Bench Changing Rate
2. Generator - Field Current Draw
3. Generator - Power Consumption at No Load
4. Generator - Power Consumption at Full Load
5. Generator Drive and Rotation
6. Current Output at Idle
7. Current Output at 1500 Engine RPM
8. Generator Regulator Type
9. Cranking Motor Type
10. Cranking Motor Shift Actuation
11. Cranking Motor Shift Operation
12. Type of Drive
13. Cranking Speed
14. No Load Test
    a. Amperes
    b. Volts
    c. RPM
15. Locked Armature Test
    a. Amperes
    b. Volts
16. Solenoid Switch, Type
17. Current Draw of Solenoid Winding
18. Ignition Coil, Type
19. Resistor, Type
20. Current Draw
    a. Engine Stopped
    b. Engine Idling
21. Coil Resistance
    a. Primary
    b. Secondary
22. Spark Plugs, Type
23. Distributor, Type
24. Condenser, Type, Capacity
25. Headlamp, Type

26. Headlamp, Lens Diameter
27. Tail, Stop, Parking, Signal Lamps; Type
28. Wiring Circuit Type
29. Circuit Protection Type
30. Number of Fuses
31. Number of Circuit Breakers
32. Horn, Type
33. Horn, Current Draw

J. Environmental Control Systems

1. Heater, Type
2. Heater Fan Motor, Type
3. Current Draw Each Speed
4. Air Conditioner, Type
5. Effective Displacement
6. Oil Content
7. Refrigerant
8. Type of Temperature Control
9. Radio, Type
10. Tape Deck, Type
11. Speaker System, Type
12. Television, Type
13. Other Control or Entertainment Systems

K. Front End Sheet Metal and Bumpers

1. Hood and Fenders, Dimensional and Structural Specifications
2. Bulkhead and Supports, Dimensional and Structural Specs.
3. Bumpers, Dimensional and Structural Specifications

L. Frame

1. Dimensions
2. Structural Specifications
3. Mounting Hardware Specifications
    a. Engine
    b. Transmission
    c. Drive Line
    d. Axle
    e. Suspension
    f. Exhaust System
    g. Miscellaneous Piping and Wiring
    h. Body Mounts

M. Body

1. Exterior Dimensions
2. Interior (Passenger Compartment)
3. Structural Properties
4. Seating Dimensions and Specifications
5. Instruments and Informations Systems
6. Controls
7. Visibility
8. Safety Equipment
9. Amenities
10. Doors and Windows
11. Stowage Compartments

Exhibit II (Highway Subsystem)

Cost Components and Cost Component Descriptors for Highway Subsystem

A. Road Surface

1. Material Type
2. Thickness
3. Lane Width
4. Number of Lanes
5. Mat Reinforcement

6. Other Surface Reinforcement
7. Coefficient of Friction
8. Roughness
9. Load Bearing Capacity
10. Surface Hardness
11. Distortion
12. Flexibility
13. Glare Characteristics
14. Color
15. Degree of Curve, Distribution
    a. Frequency Distribution
    b. Mean
    c. Variance
16. Gradient, Distribution
    a. Frequency Distribution
    b. Mean
    c. Variance
17. Superelevation on Curves, Distribution
    a. Frequency Distribution
    b. Mean
    c. Variance
18. Crown, Distribution
    a. Frequency Distribution
    b. Mean
    c. Variance
19. Separation Between Opposite Directed Lanes (Median), Distribution
    a. Frequency Distribution
    b. Mean
    c. Variance
20. Design Speed
21. Transverse Expansion Joints
22. Transverse Contraction Joints
23. Longitudinal Joint Support

B. Subgrade Structure

1. Material Types(s)
2. Thickness(es)
3. Number of Lanes
4. Effective Subgrade Lane Width
5. Load Bearing Capacity
6. Drainage Characteristics
7. Permeability
8. Flexibility
9. Undersealing

C. Shoulder

1. Material
2. Width
3. Load Bearing Capacity
4. Coefficient of Friction
5. Roughness
6. Glare Characteristics
7. Drainage Characteristics
8. Hardness
9. Geometrics
10. Curbing
11. Expansion Joints

D. Drainage

1. Gutters, Capacity
2. Paved Ditch, Capacity
3. Drainage Pipe, Type
4. Drainage Pipe, Capacity
5. Pipe Arches, Type
6. Underdrain, Type
7. Underdrain, Capacity
8. Catch Basins
9. Manholes

10. Leak-Off Chutes
11. Erosion Control Devices
12. Erosion Control Landscaping

E. Structures

1. Type
2. Number of Lanes
3. Lane Width
4. Load Bearing Capacity, Beaming
5. Load Bearing Capacity, Horizontal (Sail Loading)
6. Span
7. Height Clearance, Below
8. Abutements, Type
9. Abutements, Load Bearing Capacity
10. Number of Structures per Unit Distance
11. Surface Characteristics of Structure Roadway (see Component A)
12. Pedestrian Walkways
13. Bridge Rail and Retainers, Type
14. Open-steel-grid Floor
15. Temporary Structures and Approaches Required
16. Approach Retaining Walls
17. Lateral Stability Retaining Walls
18. Piling

F. Information and Control Devices

1. Signs, Number and Type per Unit Distance
2. Legibility Distance, Day
3. Legibility Distance, Night
4. Supporting Structures for Signs
5. Traffic Signal Lights, Number and Types per Unit Distances
6. Visibility Distance, Day
7. Visibility Distance, Night
8. Min. Height from Roadway
9. Max. Height from Roadway
10. Supporting Structures
11. Railway Crossing Signals, Number and Type per Unit Distance
12. Location w.r.t. Road-Track Intersection
13. Supporting Structures
14. Roadway Illumination, Number and Type per Unit Distance
15. Illumination per Square Unit
16. Supporting Structures
17. Power Supplies
18. Guard Rails, Number and Type per Unit Distance
19. Bearing Strength of Rail
20. Shear Strength of Supports
21. Weight of Rail per Lin. Foot
22. Height of Rail Above Road Surface
23. Traffic Pacing Systems
24. Traffic Surveillance Systems
25. Roadway Environmental Control Devices
26. Roadway Stripping and Marking, Amount per Unit Distance
27. Wear Rate of Markings
28. Visibility of Markings, Day
29. Visibility of Markings, Night

30. Toll Booths per Unit Distance

G. Right of Way

1. Width, Mean
2. Max. Distance to Roadside Obstacle
3. Earthwork (see Component D)
4. Porous Material for Tree-root Protection
5. Topsoil
6. Seeding
7. Sodding
8. Mulching
9. Sprigging
10. Trees, New Planting
11. Trees, Removal
12. Shrubs, New Planting
13. Shrubs, Removal
14. Slope Paving

H. Earthwork

1. Clearing
2. Grubbing
3. Tree Removal
4. Stump Removal
5. Removal of Existing Pavement
6. Removal of Existing Structures
7. Stripping
8. Roadway Excavation
9. Rock Excavation
10. Loam Excavation
11. Structure Excavation
12. Cofferdams and Pumping
13. Trench, Channel, and Ditch Excavation
14. Scarifying and Reshaping
15. Subbase
16. Stabilization of Road Bed
17. Borrow
18. Gravel Fill
19. Rock Fill
20. Overhaul
21. Roadside Cleanup
22. Temporary Barricade
23. Permanent Barricade
24. Maintenance and Protection of Traffic

This exhibit was compiled from various sources; one of the more complete references is Grushky, M., and Rosenbaum, B., "Contracts and Specifications, Section 15," in Woods, Kenneth B., ed., Highway Engineering Handbook, McGraw-Hill, New York, 1960, pp. 15-34—15-38.

Exhibit III (Recovery System)

Cost Components and Cost Component Descriptors

A. Casualty Recovery Vehicles

1. Main Power Plant
2. Auxiliary Power Plant(s) and Generator(s)
3. Frame
4. Body (Passenger Compartment)
5. Fuel System
6. Exhaust System
7. Transmission(s)
8. Drive Line(s)
9. Steering Gear (Controls)
10. Chassis Suspension
11. Brake System

12. Rotors
13. Front End Sheet Metal and Bumpers
14. Electrical System
15. Environmental Control Systems
16. Communication Systems
    a. Transmitter
    b. Receiver
    c. Input-Output
    d. Inter-com.
17. Reconnaissance Systems
    a. Optical
    b. Electromagnetic (e.g., Radar)
    c. Chemical (e.g., photographic)
    d. Electrostatic
    e. Infrared
    f. Other
18. Approach Warning System
    a. Audio
    b. Visual
    c. Electromagnetic
19. Navigation System
    a. Navigation Radar
    b. Computer
    c. Other
20. On-Board Medical Facilities
    a. Stretchers
    b. Inhalation Therapy Equipment
    c. Sterile Linen
    d. Bandages, Casts, etc.
    e. Medication
    f. Surgical Equipment
    g. Environmental System
    h. Anesthesia
    i. Diagnostic Equipment, Data Link
    j. Medical Equipment
    k. Medical Lighting
    l. Other
21. Surveillance System Link
    a. Data Link
    b. T.V. Display
22. Assembly Costs
23. Maintenance Shops and Garages

B. <u>Central Communication System</u>
1. Broadcast Equipment
2. Terminal (I/O)
3. Switching Gear
4. Antennas
5. Wire/Cable
6. Microwave
7. Studio
8. Installation and Construction

C. <u>Surveillance and Detection System</u>
1. Mobile Surveillance Units
    a. Vehicles
    b. Surveillance Equipment
    c. Communication Equipment
    d. Assembly
2. Remote Fixed Surveillance Units
    a. Surveillance Equipment

- b. Environmental Protection
- c. Wire/Cable
- d. Information I/O Gear
- e. Installation and Construction

3. Central Surveillance Command and Control Station
    - a. Communication Equipment
    - b. Display Devices
        - 1) T.V.
        - 2) Status Boards
        - 3) Other
    - c. Data Processing Equipment
        - 1) Data Storage
        - 2) Data Processing
        - 3) I/O
        - 4) Computational Unit
    - d. Installation and Construction

D. Treatment Facilities

1. Hospital Emergency Receiving
2. Remote Treatment Stations

E. Vehicle Recovery System

1. Vehicle Recovery Vehicles
2. Communication Network
3. Maintenance Shops and Garages

Appendix 7-B
DERIVATION OF A COST ESTIMATING RELATIONSHIP

As we have stated before, a cost estimating relationship (CER) is a statement of how cost is dependent upon other variables which describe the physical characteristics of the component parts of the system and perhaps the production characteristics of contemplated output volume and the rate index. The derivation of any CER requires these four steps:

1. Data collection
2. Identifying independent variables
3. Mathematical formulation of CER
4. Define limitations on accuracy of CER

The data collection phase is the most important since no CER can be accurate if the underlying supporting data have not been properly classified and carefully collected. We have assumed in developing the preceding cost prediction model that cost data are available by cost component and by cost element. Furthermore, we have assumed that the necessary information on component characteristics and production characteristics is available to the cost analyst. The extent to which these ideals are fulfilled in a particular situation defines the difficulty of the data collection task.

In the ideal situation, all that the cost analyst need do is pick up the necessary accounting records and move on to step 2. Usually, however, more digging will be required because the cost data was not originally collected with the cost analyst's problems in mind. The only general guideline for accurate data collection seems to be for the cost analyst to become thoroughly familiar with the production operations which generate the costs. In some cases where the operations to be estimated have not been performed before, the cost analyst will have to rely on the estimates of other people or preliminary experiments as data. The kinds of gross inaccuries which can develop in cost data are well documented for military systems. The Harvard Business School studies under the direction of Prof. Paul Cherington by Peck & Scherer (1962) give frequent examples of the difficulties inherent in data collection. Fisher's memorandum (1962) as cited earlier gives an estimate on the magnitude of the errors which can result from inaccurate CER's at 20-30%. The conclusion is then that any time spent in digging for and refining basic data is well spent.

In order to be less general in our discussion of the next three required steps, the derivation of a CER will be illustrated by way of a somewhat trivial but nonetheless useful example. Suppose we are interested in estimating the cost (to the consumer) of an automobile with some arbitrary descriptors (x_{jk}'s) such as brake horsepower, vehicle weight, etc. We will assume the vehicle whose cost we wish to estimate is a 'current generation' vehicle, i.e., basically similar to those currently in production. We will also assume (to keep things simple) that we wish to estimate the cost of a basic four door sedan. Fortunately, the data collection step is readily accomplished by reference to sales literature or trade magazines (our data is from Ward's Automotive Reports, 1965). Thus, we soon have a substantial list of costs and descriptors. The next task is to identify the significant independent variables.

Having taken a cursory look at these data we decide that a relationship exists between engine horsepower and vehicle cost, i.e., the higher the horsepower, the higher the cost. To test whether this independent variable, horsepower, serves as a good cost predictor we first plot the data on a scatter diagram as shown in Fig. 7-B-1. A trend is apparent but the points are widely scattered. In fact, when we fit a polynomial through these points (using a curve fitting program on a digital computer) we note that the curve doesn't follow the points very well and consequently cannot provide a very good estimate of vehicle cost. A measure of the inaccuracy inherent in our CER is the standard error of estimate which for the cost-BHP data is $710.10. This means if we were to make a number of cost estimates for a range of horsepowers we would, on average, expect about 1/3 of our estimates to fall beyond ±$710.10 of the true cost. Not satisfied with this kind of accuracy, we shall seek an independent variable from which a better CER can be derived.

As a second attack on the problem, let us choose vehicle weight as the independent variable. If we plot the data we obtain the scatter diagram in Fig. 7-B-2. Clearly, weight is a better cost predictor than brake horsepower. In fact the standard error of estimate has been reduced to $194.20. This may still be beyond our estimating tolerances, and so we shall make a third attempt at defining the CER.

This time, learning from our previous experience, we retain vehicle weight as a dependent variable, but also include the brake horsepower (BHP) as a factor. We have observed that both horsepower and weight affect cost and thus investigate $C = f(BHP \times Wgt)$ as a CER. Now, as shown in Fig. 7-B-3, we have obtained an even better CER with a standard error estimate of $161.50. The process of introducing progressively more x_{jl}'s may be continued until the standard error of estimate is reduced to some predetermined level or until no more improvement is possible. In our example, which utilized only two x_{jl}'s, we have reduced the standard error of estimate to about 5% of the mean cost, which may be sufficiently accurate for some applications.

The third step, that of identifying the mathematical relationship between variables, may be accomplished through a variety of statistical means. In our example we used a least squares fit of a fourth degree polynomial to fit the data points. Other techniques such as linear regression or multiple regression also are often useful in deriving the equations for the CER's. Since regression analysis is a broad and well documented area in statistics we will not attempt to

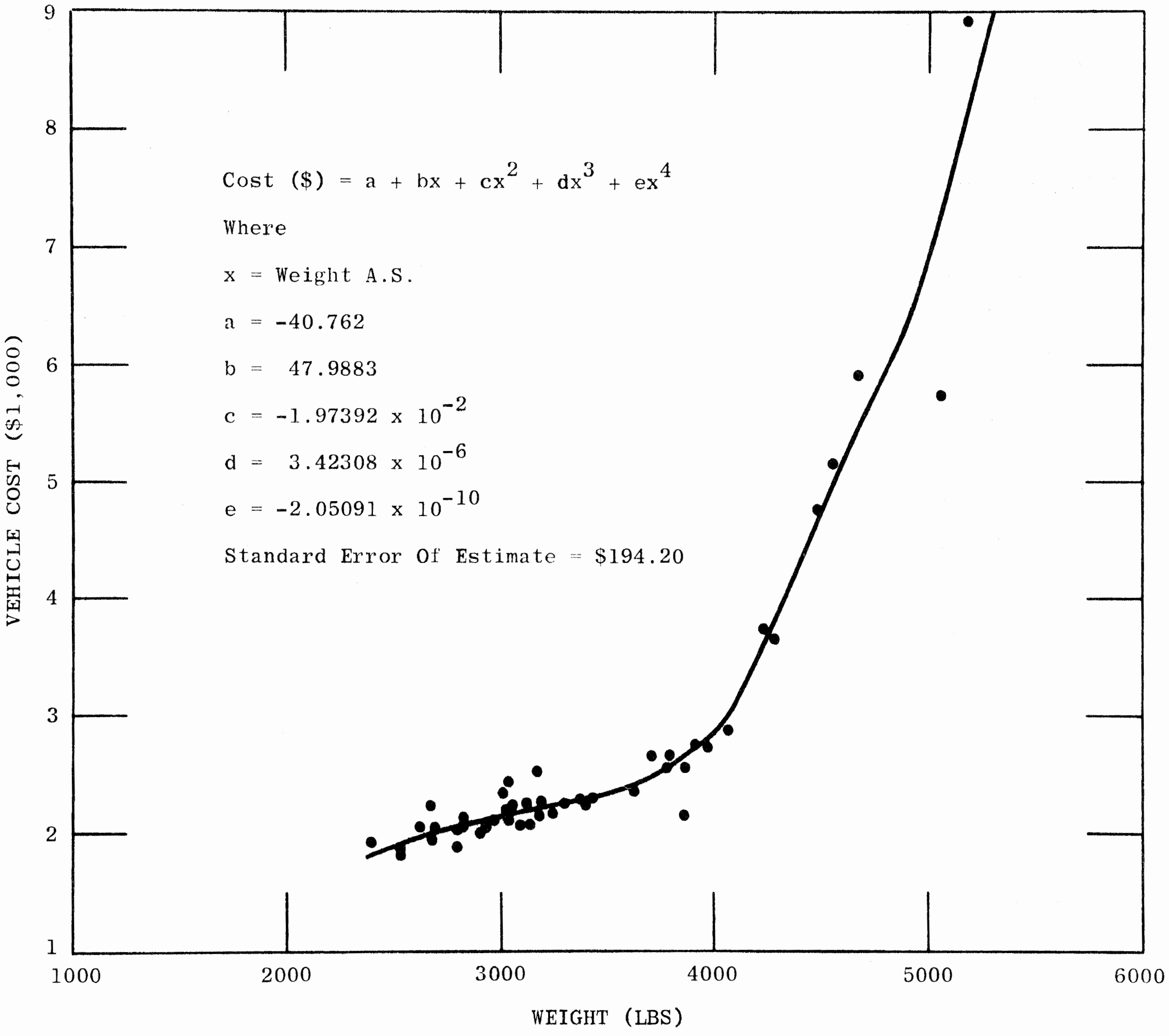


FIGURE 7-B-1. COST-ESTIMATING RELATIONSHIP -- MANUFACTURER'S SUGGESTED RETAIL PRICE VS. VEHICLE WEIGHT A.S. (Data from Ward's 1965 Automotive Yearbook, 27th Edition)

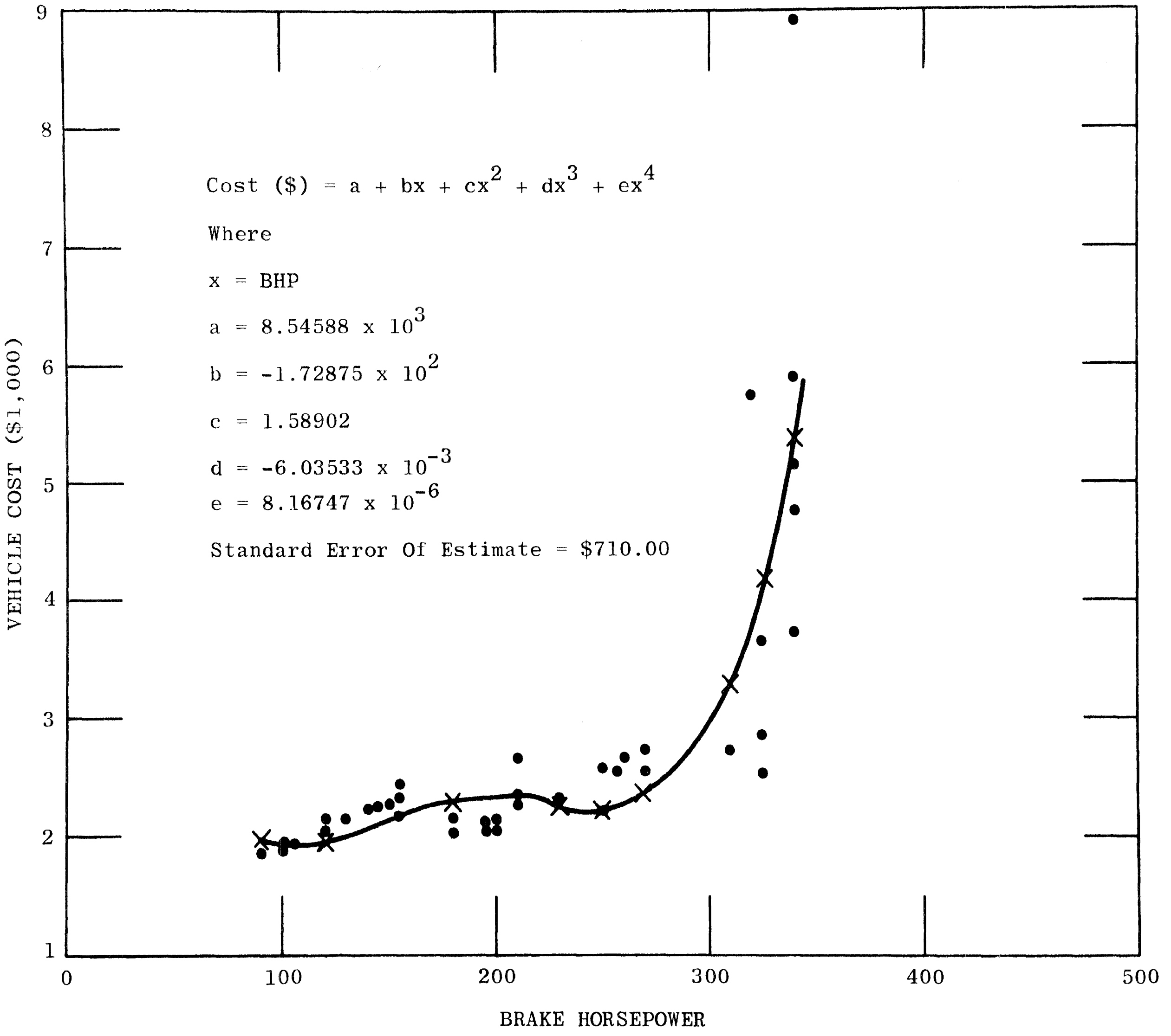


FIGURE 7-B-2. COST ESTIMATING RELATIONSHIP -- MANUFACTURER'S RETAIL PRICE VS. BHP (Data from Ward's 1965 Automotive Yearbook, 27th Edition)

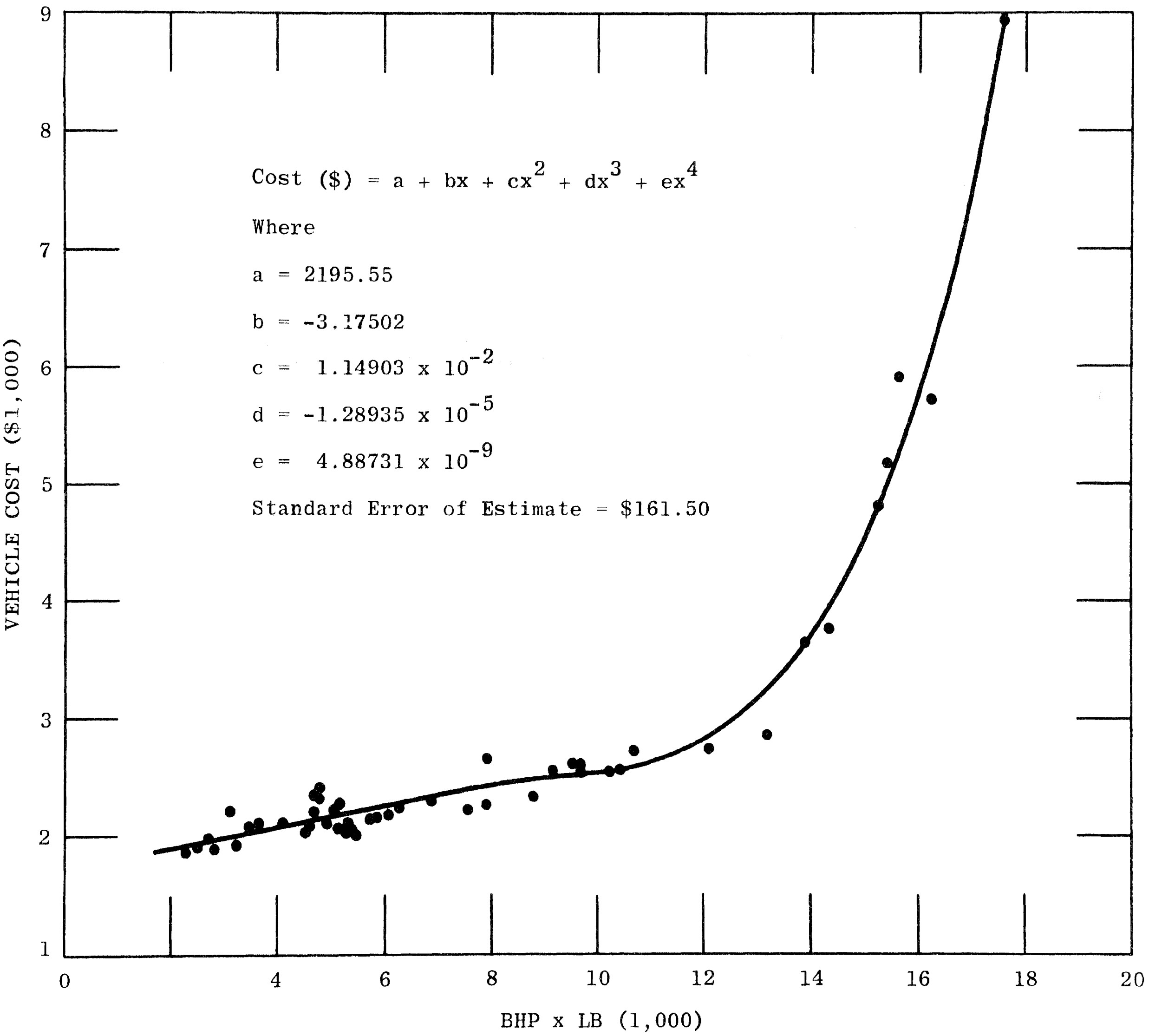


FIGURE 7-B-3. COST ESTIMATING RELATIONSHIP -- MANUFACTURER'S SUGGESTED RETAIL PRICE VS. BHP x LB (Data for basic sedan from Ward's 1965 Automotive Yearbook, 27th Edition)

go into these techniques in further detail here. The interested reader is referred to Fisher (1962), Graybill (1961), Cramer (1945), Scheffe (1959), Grenander & Rosenblatt (1959), and Johnson (1960) for additional information on statistical regression techniques.

The need for defining the limits of accuracy of CER's arises for two reasons. First, as we have noted, there may be inaccuracies in the basic data. There is no direct means of quantifying these data collection inaccuracies. Hence the analyst must use judgment in applying the CER. Essentially we are saying that while statistical techniques may help to identify and quantify the relationships between cost and other variables they do not obviate the need for experience and judgement in the application of their results. Second: there are statistical confidence limits which can be placed on the derived CER's. The width of these limits is an indication of the variability of the data points about the calculated trend line. Usually one defines a 95% confidence interval which indicates that 95% of all predictions made on the basis of the given CER will fall within the given limits. Again the reader is referred to the statistical references given above for a further discussion.

It should now be apparent in a general way how the CER is derived. We have continually referred to the component descriptors, x_{jk}'s, as the basis for the derivation of the CER. As a first step toward developing the cost prediction models for the major subsystems of the highway transportation model, lists of the cost components and the cost component descriptors, x_{jk}'s, for the Vehicle, Highway, and Recovery subsystems have been compiled. While we have attempted to include all x_{jk}'s which might affect component cost, we realize that there may still be some omissions which will become known when the actual data collection and analysis task is begun. The cost component and cost component descriptor lists appear above as exhibits I, II, and III in Appendix 7-A.

Appendix 7-C
SEVERAL PROPOSITIONS RELATING TO COST AND OUTPUT

Borrowing heavily from Alchian's work (1959) but retaining our previous notation, we will form the following propositions.

Proposition I:

$$\frac{\partial C_V}{\partial R} > 0$$

This states that for a constant volume of total production, an increase in the rate index, R, for example by increasing the mean production rate or the variability of the production rate, will cause an increase in total procurement cost. An increase in variability may occur either through planned variations according to the production schedule (deterministic variability) or through chance fluctuations in rate caused by other forces (random variability). It should be possible to test this proposition for both deterministic and random variability. The validity of Proposition I should be a direct consequence of Eq. 5, 6, and 7 above. We may further note that it appears reasonable that $\partial F_V/\partial R > 0$ also.

Proposition II

$$\frac{\partial^2 C_V}{\partial R^2} > 0$$

We thus assume C_V is convex, in R, i.e., the cost increment in R is an increasing function of R.

Proposition III

$$\frac{\partial C_R}{\partial V} > 0$$

This states that for a given rate index, R, an increase in total procurement quantity will cause an increase in total procurement cost; a direct consequence of Eq. 1. Also, certainly,

$$\frac{\partial F_R}{\partial V} \geqq 0$$

Proposition IV

$$\partial^2 C_R/\partial V^2 < 0$$

Assumes that as volume increases at a fixed R, cost increases at a decreasing rate. This follows from Eq. 5, 6, and 7, and is a result of the "economies of scale" or the "learning phenomenon" discussed previously.

Naturally these propositions, along with the four hypotheses described above, must be tested through actual experiments and they are provided here only as a structure for future research.

Appendix 7-D

FURTHER DEVELOPMENT OF RELATIONSHIPS CONCERNING INSPECTION, REPAIR, TEST, AND SCRAP COSTS

Since the nature of the inspection (destructive or nondestructive) will affect costs we may adopt a new parameter, ν, which identifies the nature of the inspection process. Also the percent of assemblies inspected, Pr_I, has a direct affect on cost. Since there may be more than one inspection during assembly the number, N, of inspections is a cost factor. Now we assume that the costs incurred at each inspection are exclusive and that an assembly coming to inspection station a has probabilities, Pr_{Ia}, of being inspected. Pr_{Iga}, of passing if inspected, Pr_{Iba}, of not passing but being repairable, and Pr_{Isa}, of being scrapped:

$$Pr_{Iga} + Pr_{Iba} + Pr_{Isa} = 1$$

If V_I assemblies enter the first inspection station the unit inspection cost of sending V_T assemblies to test is:

$$I_{V,R,M} = \frac{V_I}{V}\left(Pr_{Ia}C_{Ia} + (1 - Pr_{Isa})Pr_{Ia+1}C_{Ia+1} + \dots + \prod_{a=1}^{n-1}(1 - Pr_{Isa})P_{IN}C_{IM}\right) \tag{10}$$

where C_{Ia} is the inspection cost at station a. The inspection cost, C_{Ia}, may be predicted according to the preceding assumptions by

$$C_{Ia} = f(M, X_{jk}, \nu_a) \tag{11}$$

Here, for convenience, $X_{jk} = (X_{11}, \dots, X_{jk}, \dots, X_{MM})$. The testing cost $T_{V,R,M}$ can be computed in an analogous manner. We let Pr_{Tg}, Pr_{Tb}, Pr_{Ts}, equal (respectively) the probability of being good, defective bur repairable, and scrap when tested, where again $Pr_{Tg} + Pr_{Tb} + Pr_{Ts} = 1$). If we let $t_{V,R,M}$ be the unit testing cost we have

$$T_{V,R,M} = \frac{V_T}{V}\left(1 + \frac{Pr_{Tb}}{1 - Pr_{Tb}}\right)t_{V,R,M} \tag{12}$$

(assuming Pr_{Tb} is the same for a repaired unit as for a unit directly from inspection. The unit testing cost may be assumed to be dependent on the number of components, M, some x_{jk}'s (particularly the performance characteristics to be tested), and a factor φ describing the nature of the test, viz:

$$t_{V,R,M} = f(M, X_{jk}, \varphi) \tag{13}$$

The unit repair costs are seen to be a combination of the repair costs incurred during inspection and those incurred at test. If ρ_a and ρ_t represent the repair costs at inspection station a and at test respectively we have:

$$R_{V,R,M} = \frac{V_T}{V}\left\{Pr_{Ia}Pr_{ba}\rho_a + (1 + Pr_{Isa})Pr_{Ia+1}Pr_{Iba+1}\rho_{a+1} + \cdots + \prod_{a=1}^{N-1}(1 - Pr_{Isa})Pr_{IN}Pr_{IbN}\rho_n\right\} + \frac{V_T}{V}\left(\frac{Pr_{Tb}}{1 - Pr_{Tb}}\right)\rho_t \tag{14}$$

Similarly, for unit scrap costs if we let γ_a and γ_b be the scrap value of assemblies scrapped at station a and at test t respectively for total unit scrap cost we have:

$$S_{V,R,M} = \frac{V_I}{V}\left\{Pr_{Ia}Pr_{Isa}\gamma_a + (1 - Pr_{Isa})Pr_{Ia+1}Pr_{Isa+1}\gamma_{a+1} + \cdots + \prod_{a=1}^{N-1}(1 - Pr_{Isa})Pr_{IN}Pr_{IsN}\gamma_N\right\} + \frac{V_T}{V}\left(1 - \frac{Pr_{Tb}}{1 - Pr_{Tb}}\right)Pr_{Ts}\gamma_t \tag{15}$$

REFERENCES

Alchian, A., "Costs and Outputs," In Abramovitz, M., et al., *The Allocation of Economic Resources*, Stanford University *Press: Stanford, Calif., 1959, pp.* 23-40.

Baker, Robert F., *Highway Costs and Their Relationship to Vehicle Size*, *Engineering Experimental Station, The Ohio State University*: Columbus, 1958.

Bigelow, C. G., "Bibliography on Project Planning and Control by Network Analysis: 1959-1961," *Operations Research*, Volume 10, Number 5, 1962.

Billingsley, Charles M., and Joregenson, Dayton P., "Direct Costs and Frequencies of 1958 Illinois Motor-Vehicle Accidents," *Highway Research Record No. 12*, Highway Research Board: *Washington, 1963.*

Bonder, S., "Cost Analysis," Chapter 5 in Howland, D., and Bonder, S., *Annual Progress Report: The Tank Weapon System*, *Systems Research Group, The Ohio State University*: Columbus, 1963.

Carrier, M. J., and Smith, R. W., *Aircraft Airframe Cost Estimating Techniques*, The RAND Corporation: Santa Monica, RM-3, 375-PR, 1962.

Claffey, Paul J., *Running Cost of Motor Vehicles as Affected by Highway Design*, *Report 13, National Cooperative Highway Research Program*, Highway Research Board: Washington, 1965.

Cramer, H., *Mathematical Methods of Statistics*, Princeton University Press, 1945.

Davis, E. W., "Resource Allocation in Project Network Models—A Survey," *The Journal of Industrial Engineering*, Volume XVII, No. 4, April, 1966, p. 177.

Dean, B. V., "A Method for Selecting Research and Development Projects," unpublished report, Case Institute of Technology: Cleveland, 1959, p. 5.

DoD and NASA Guide, PERT/Cost, Office of the Secretary of Defense and National Aeronautics and Space Administration: Washington, D. C., 1962.

Final Report of the Highway Cost Allocation Study, 87th Congress, 1st Session, House Document No. 54, Washington, 1961.

Fisher, G. H., *A Discussion of Uncertainty in Cost Analysis*, The RAND Corporation: Santa Monica, RM-3071-PR, April, 1962 a.

Fisher, G. H., *Derivation of Estimating Relationships: An Illustrative Example*, The RAND Corporation: Santa Monica, RM-3366-RR, 1962 b.

Fisher, G. H., *Total Force Structure Cost Analysis*, The RAND Corporation: Santa Monica, RM-3070-PR, 1962 c.

Fulkerson, D. R., "A Network Flow Computation for Project Cost Curves," *Management Service*, Volume 7, 1961, p. 167.

Graybill, F. A., *An Introduction to Linear Statistical Models*, McGraw-Hill: New York, 1961.

Grenander, V., and Rosenblatt, M., *Statistical Analysis of Stationary Time Series*, Wiley: New York, 1959.

Hertz, D. B., *The Theory and Practice of Industrial Research*, McGraw-Hill: New York, 1950.

Hosking, Sid, "Accidents Measured in Dollars and Cents," *California Highway Patrolman*, April, 1957, p. 10.

Illinois, State of, *Cost of Motor Vehicle Accidents to Illinois Motorists 1958*, Department of Public Works and Buildings, Division of Highways, State of Illinois: Springfield, 1962.

Jacobson, L., and McGovern, P. J., Jr., "Computer Simulation of a National Economy," *Computers and Automation*, pp. 14-16.

Johnston, J., *Statistical Cost Analysis*, McGraw-Hill: New York, 1960.

Kaplan, S., "An Analysis of Manpower Requirements for Research and Development Projects," mimeographed report, Radio Corporation of America: Princeton, 1959, p. 8.

Lang, A. S., and Robbins, D. H., "A New Technique for Predicting Vehicle Operating Cost," Bulletin 308, Highway Research Board: Washington, 1962.

Levenson, G. S., and Barro, S. M., Cost-estimating Relationship for Aircraft Airframes, The RAND Corporation: Santa Monica, RM-4845-PR (Abridged), May 1966.

Mann, Lawrence, Jr., "A Procedure to Predict Maintenance Costs for Highway Systems," Bulletin 85, Division of Engineering Research, Louisiana State University and Agricultural and Mechanical College: Baton Rouge, 1965.

Norden, P. V., "Curve Fitting for a Model of Applied Research and Development Scheduling," IBM Journal of Research and Development, Volume II, No. 3, July, 1958, pp. 232-248.

Norden, P. V., "On the Anatomy of Development Projects," IRE Transactions, PGEM, Vol. EM-7, No. 1, March, 1960, p. 41.

Norden, P. V., "Resource Usage and Network Planning Techniques," Operations Research in Research and Development, Wiley: New York, 1963, pp. 149-169.

Norden, P. V., and Bakshi, A. V., "Internal Dynamics of Research and Development Projects," Management Sciences Models and Techniques, The Pergamon Press: New York, 1960, pp. 187-205.

Novick, D., System and Total Force Cost Analysis, The RAND Corporation: Santa Monica, RM-2695, 1961.

Peck, M. J., and Scherer, F. M., The Weapons Acquisition Process, Division of Business Research, Harvard University, Graduate School of Business Administration, Boston, Mass., 1962.

Petruschell, R. L., and Chester, J. M., Total System Cost Analysis - Part I, The RAND Corporation: Santa Monica, RM-3069-PR, 1963.

Poletti, G., "The Diffusion of Network Techniques Throughout Government Publications," IEEE Transactions, Volume EM-II, Number I, 1964.

Scheffe, H., The Analysis of Variance, Wiley: New York, 1959.

Stanford Research Institute, A General Framework for Analysis of the Costs of Traffic, Home, and Public Accidents, Southern California Laboratories of Stanford Research Institute: South Pasadena, 1964.

Summers, R., Cost Estimates vs. Predictors of Actual Weapon Costs: A Study of Major Hardware Articles, The RAND Corporation, RM-3061-PR, April, 1962.

Thédié, Jacques, and Abraham, Claude, "Economic Aspect of Road Accidents," Traffic Engineering and Control, February, 1961, pp. 589-595.

Twombly, Bernard B., "Economic Cost of Traffic Accidents in Relation to the Highway," Economic Cost of Traffic Accidents, Bulletin 263, Highway Research Board: Washington, 1960.

Ward's 1965 Automotive Yearbook, 27th Edition, Powers and Company, Inc., Detroit, Michigan, pp. 17-25.

8 TRANSPORTATION DEMAND

DONALD E. CLEVELAND
ALLEN COOK

The traffic generation or demand model describes the set of trips undertaken in a given highway transportation system. This description provides the input for the operational models of traffic. The demand is affected by service and safety levels, by the provision of alternative modes of transport, and by general sociological variables describing the pattern of life around the highway system. In the present system structure, the dependencies on service and safety are not explicitly modeled: the dependencies on sociological variables appear only in determining the pattern of traffic generation in time and space. Other effects are considered to be second-order effects of small magnitude.

The determination of demand depends upon knowing two things: (1) the nature of the modes of transport, and (2) the factors that generate the demand for movement; i. e., the factors that motivate people to move themselves, other people, or goods. Knowledge of the nature of the modes of transport is needed because it has frequently been demonstrated that transportation demand is affected by system capabilities and performance.

The demand model should produce the following characteristics of demand as its output:

(1) Characteristics of trips taken:
Origin, destination of users, location of parking place of vehicle, route taken, time of start for trip.

(2) Characteristics of the vehicles used for conveyance:
Vehicle type, capacity in terms of passengers and cargo.

(3) Characteristics of drivers:
Age, sex, marital status, economic status, physical characteristics, accident and traffic violation history, psychological and medical characteristics.

(4) Characteristics of passengers:
Number, age, sex, marital status, economic status, physical characteristics.

(5) Characteristics of goods moved:
Quantity, value, stowage configuration, vulnerability to damage,

potential hazards to system operation in transport.

Eventually, additional descriptors may be required by other subsystems. Some of these proposed measures may not be needed in other parts of the model.

Four fundamental conditions govern the demand for movement.

(1) A need or desire to engage in some activity necessitating the movement of people or goods from some origin. The measurement of this primary demand depends upon describing the needs or desires of the potential trip makers and shippers and the extent to which these needs can be met at the origin.

(2) The existence of destinations where demands can be satisfied.

(3) The existence of adequate transportation as measured by the resistance to movement; i.e., the mode of transport must meet standards acceptable to the decision maker. These standards include considerations of travel time, cost, safety, comfort, and convenience, and essentially constitute a problem of economics, where the costs of movement measured by resistance are compared to the advantages to be derived by such a move. For example, the survival rate of goods transported is certainly a direct cost of movement. Less easy to measure are standards of human comfort or convenience.

(4) The availability of transportation facilities serving the point of trip origin. Availability may itself create a desire for movement.

We may consider the problem in a manner similar to supply-and-demand analysis in economics. There exist individual and social propensities to travel or transport goods, the amount of such travel being related to the "price" determined on the supply curve. In our case the price includes both the monetary expense and the cost in time and inconvenience to the traveler or transport user. A general term for all these items is resistance to travel. It is typically measured in such dimensions as:

(1) Travel cost: in terms of time, direct costs of transportation, and investment costs in the system.

(2) Safety: in terms of the expected loss, damage, injury, or delay incurred by accidents, fires, or vibration or jerk in transit.

(3) Comfort: in terms of temperature, noise, vibration, jerk,

acceleration, deceleration, seating arrangements, and internal vehicle environment.

(4) Convenience: in terms of travel time, frequency, reliability of service, location of transportation origins and terminals in relation to person origins and destinations, flexibility of service, capacity, and facilities for baggage handling.

When considering resistance to travel and the related trip-generation process, it is necessary to note that, in addition to considerations of price versus propensity to travel, some of the inherent properties of both passengers and goods will establish certain minimum satisfactory service levels for the transportation system. People demand a certain degree of comfort of the travel environment (including such factors as temperature and humidity in the travel vehicle) and protection against unpleasant acceleration, jerk, engine heat, and other elements inherent in the mode of transport. It is the function of the vehicle performance model to provide information about these considerations.

Goods movement does not involve those amenities associated with human travel demands, but does require a measure of protection which depends upon the nature of the goods moved. Such diverse cargos as dry bulk, liquid bulk, live animals and plant matter, refrigerated goods, long-length or heavy-lift cargo, and mobile cargo, all require special facilities for handling. Terminal facilities must also be provided that both handle and protect goods. The identification of needs and desires for movement include determining the specifications required to protect people and goods.

The predicted rate of trip generation will be determined from the resistance to transportation and the availability of transportation and its level of service (in terms of time, cost, etc.). The prediction of the resistance to transportation will be made from social, environmental, and economic data. Demographic, social and psychological descriptors of the population, and economic descriptors of its business enterprises, are necessary in order to determine their propensity to travel.

In summary, trips are generated when there exists a need or desire at an origin, one or more destinations or transportation facilities capable of satisfying this desire, and a mode of transport between the origin and destination capable of accomplishing this movement at a cost acceptable to the decision-maker. One objective of the demand model is to relate this demand to existing or potential transport systems, subject to constraints imposed by other models. For example, Tanner (1952) has studied the effects of weather (as defined by temperature, hours of sunshine, and rainfall recorded) upon possible changes in demand in Great Britain. He cites four possible reactions to bad weather in trip planning:

(1) Plans are unchanged.

(2) Mode of transport is changed, as from bicycle to automobile, or automobile to train.

(3) Trip is postponed either because travel resistance is increased or the destination is made undesirable.

(4) Trip is cancelled.

He found, for example, that fog caused a 1/3 to 1/2 reduction in traffic volume. Rain also reduces traffic volume, but is less likely to interfere with essential trips.

Given as input an array of potential decision-makers and goods or people to be moved, and the potential performance characteristics of modes of transport, parameters should be developed relating demand to transport service.

In recent years, the interest of governmental agencies at federal, state, and municipal levels in projecting future demand has greatly increased. Stemming from this interest has been considerable research and formulations seeking to establish meaningful relationships in which to describe and predict the following:

(1) Trip generation or amount of trip making
(2) Trip distribution or linking origin with suitable destinations.
(3) Model choice
(4) Route selection
(5) Interactions between demand and costs and services provided.

Origin and destination studies have related trip generation to the following variables: trip purpose, family income, vehicle ownership, land use at the origin, distance from the central business district, length of trip, mode of travel, automobile occupancy, land use at destination, and time of day. To a large extent the interrelationships among those variables have been charted for urban transportation. The independent variables directly correlated with trips per person or vehicles per day are determined and combined into a multiple regression model of the following form:

$$T_p = a + b_1X_1 + b_2X_2 + \ldots$$

where T_p represents trips per person or vehicle, per day; a is the intercept on the T_p axis; the b terms are multiple regression coefficients for individual variables; and the X terms are appropriate variables such as family income and vehicle ownership. Estimation for the future is contingent upon predicting future changes in the variables, and upon the multiple regression equation coefficient stability. The four significant factors which have been employed to predict trends for the future are land use patterns, average vehicle ownership, average family income, and zonal centroidal distance from the central business district.

The prediction of the mode of travel (i.e., auto, bus, rail transit, walk, etc.) in urban areas depends upon the degree to which future use may be based on past trends and the way in which changes in public policy might alter the use of mass transit. Mass transportation primarily serves two basic trip categories: (1) Trips to and from the central business district for work-related purposes, and (2) school trips.

The amount of travel by mass transit in an urban area has been shown to depend on the extent of automobile ownership, the intensity of land use, and the relative "costs" of automobile versus public transportation. Two methods are used to predict mass-transit trips:

(1) Prediction of modal use solely from socio-economic information on the trip maker and his trip purpose.

(2) Predictions of modal use based on comparing transit and auto for a specified trip with the usage of one mode relative to others dependent on relative travel time or cost.

Adams has developed the following equation for the relative use of transit and the automobile in an urban area:

$$y = a + b_1 \log P + b_2 \log E + b_3 \log T + b_4 \log U + b_5 \log M$$

where y is the percentage of all person trips made by transit
P is the fraction of population over five years of age
T is a measure of transit service
E is an economic factor considering automobile ownership and dwelling unit characteristics

U is a land utilization factor

M is the size of the urbanized area

and where a and the b_1 terms are empirically derived constants.

The linking of origins and destinations has been the subject of considerable research and several models have been developed:

(1) Extrapolation methods: uniform factor, average factor, Detroit, and Fratar

(2) Travel determinant methods: gravity model, interactance model, opportunity model

As of now these models have not been statistically evaluated and critically compared. All are based upon one of two systems of logic, either future projections based upon trends in current interzonal transfers, or synthetic patterns derived from analysis of current origin-destination patterns applied to future land-use estimates.

The growth factor methods assume that present travel patterns can be projected into the future through a process based upon anticipated zonal growth rates. The uniform factor method is simplest, expressed as follows:

$$T_{ij} = t_{ij}F \qquad F = \sum_{i=1}^{n} \sum_{j=1}^{n} T_{ij} / \sum_{i=1}^{n} \sum_{j=1}^{n} t_{ij}$$

where T_{ij} = future trips from i to j

t_{ij} = present trips

F = a growth factor for the entire urban area

T = the total future vehicle trips in the urban area

t = the total present vehicle trips in the urban area

n = the number of zones in the urban area

This procedure ignores differential growth rates in parts of the urban area and can lead to considerable error. The average factor method modifies this by introducing individual growth factors F_i and F_j for different zones where $F_i = \sum_j T_{ij} / \sum_j t_{ij}$.

In the Fratar method, the distribution of future vehicle trips from zone i to zone j is proportional to the present trips from i to j, modified by the growth factor of the two zones. The volume of trips from zone i to zone j is expressed by

$$T_{ij} = t_{ij}F_iF_j \frac{\sum_{k=1}^{n} t_{ik}}{\sum_{k=1}^{n} (F_k t_{ik})}$$

An iteration process is used. The Detroit method is a modification of the Fratar method.

Martin, Memmott, and Bone (1961) cite simplicity of application and applicability to updating origin and destination survey data as two advantages of the growth factor methods. The disadvantages include the required comprehensive input data on origins and destinations, the lack of sensitivity of the method to substantial changes in land use patterns or transit facilities, and its weakness in handling small interchange volumes.

The gravity model is based upon the attraction of a destination zone, the magnitude of which is directly associated with size of land use development and inversely related to the separation between the origin and attracting zone. For destination zone j this ratio is expressed as

$$S_j/f(X_{ij})$$

where S_j is the attraction at j

X_{ij} is the travel time or distance between k and j.

T_{ij}, the number of vehicle trips between i and j, is the number of present trips originating from t_i, 0_i, times the attraction ratio for j divided by the sum of the attraction ratios for all the zones as expressed by

$$T_{ij} = 0_i \frac{S_j/f(X_{ij})}{\Sigma S_j/f(X_{ij})}$$

As in the other models an iteration process is needed to guarantee the correct number of arrivals in the destination zones. For projection, values of S_j and X_{ij} are determined and applied to the equation.

The gravity model accounts for competition among different zones for trips. It requires little input data (compared to other models) and is sensitive to the resistance to movement between zones. However, it is an arbitrary attempt to describe trip distribution with an "alien" physical law. The difficulty of expressing resistance to travel as a constant over time questions the use of a single D value and the model requires considerable manipulation of proportionality factors to produce realistic results correlating with actual traffic patterns.

The interactance model combines the gravity model with origin-destination survey data by establishing relationships between the attractive force and distances for different trip purposes developed from survey data.

The opportunity model is based on the probability that total travel time from a point is minimized subject to the condition that every destination point has a probability of being accepted. The resulting relationship is

$$T_{ij} = O_i \quad \exp\,(-LD_{j-1}) - \exp(-LD_j)$$

where L is the probability that a destination will be accepted

D_{j-1} is the number of destinations closer in time to zone i than zone j is to zone i

O_i is the number of trip origins in zone i

This model has demonstrated accuracy similar to that of the gravity model. It requires fewer parameters and is independent of zonal boundaries. The L factors are difficult to determine.

Since no model is satisfactory in all respects, research is needed to compare different models with the objective of developing standards with which to construct improved models.

In selecting a travel route the driver is generally confronted by several routes, and the choice he makes reflects personal criteria for travel resistance, driver effort, safety, and distance. Relationships have been developed in safety which relate travel resistance in terms of the above factors to the use of the routes being compared. In Toronto these curves have been applied to transit assignments. Much work remains to be done in this field, as resistance is an intangible concept not easily defined by such factors as travel time and driver effort.

The prediction of goods movement demand is inherently simpler, although relatively little modeling effort has been described in the open literature. Research in this significant area is currently under way at a number of places.

FURTHER RESEARCH

Considerable research is currently being done in predicting demand for highway travel. There remain, however, no adequate quantitative methods to define resistance to travel as a function of the socio-economic environment and the transport available. Expressions are needed to relate level of service to demand; current demand models are weakened by the lack of such expressions. However, it is believed that research in this area by the Highway Safety Research Institute should not be undertaken until the requirements of other subsystems indicate a clear need for research of this type.

REFERENCES

Martin, Brian V.; Memmott, Frederick W., III; and Bone, Alexander J.; *Principles and Techniques of Predicting Future Demand for Urban Area Transportation*, Massachusetts Institute of Technology Press: Cambridge, Massachusetts, 1961.

Tanner, J. C., "Weather and Road Traffic Flow," *Weather*, Vol. 7, No. 9, September 1952, pp. 270-275.

BIBLIOGRAPHY

GENERAL

Martin, Brian V., Memmott, Frederick W. III, and Bone, Alexander J., *Principles and Techniques of Predicting Future Demand for Urban Area Transportation*, Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1961.
(A survey of existing techniques of predicting travel demand.)

SPECIFIC

Adams, Warren T., "Factors Influencing Transit and Automobile Use in Urban Areas," *Bulletin 230: Trip Generation and Urban Freeway Planning*, Highway Research Board, 1959, pp. 101-110.
(Transit modal split.)

Brokke, Glen E., "Evaluating Trip Forecasting Methods with an Electronic Computer," *Bulletin 203: Travel Characteristics in Urban Areas*, Highway Research Board, 1958, pp. 52-75.
(All growth factor methods.)

Chicago Area Transportation Study, *Final Report*, Vol. II, *Data Projections*, July, 1960, p. 111.
(Opportunity model.)

Mertz, William L., and Hammer, Lamelle B., "A Study of Factors Related to Urban Travel," *Public Roads*, Vol. 29, No. 7, April 1957, pp. 171-74.
(Trip generation.)

Voorhees, Alan M., "A General Theory of Traffic Movement," *Proceedings of the Institute of Traffic Engineers*, 1955.
(Gravity model.)

Wynn, F. Houston, "Studies of Trip Generation in the Nation's Capitol, 1956-1958," *Bulletin 230: Trip Generation and Urban Freeway Planning*, Highway Research Board, 1959, pp. 18-21.
(Trip generation.)

Wynn, F. Houston, and Linder, Eric E., "Tests of Interactance Formulas Derived from O-D Data," *Bulletin 253: Traffic Origin-and-Destination Studies: Appraisal of Methods*, Highway Research Board, 1960, pp. 62-85.
(Interactance models.)

9 THE ENVIRONMENT

Allen Cook

INTRODUCTION

The least controllable aspect of the highway transportation system is the dynamic environment external to the vehicle. By dynamic environment, we mean the various transient states or conditions that comprise the environment as it affects the driver, the vehicle, and the roadway. More specifically, these are conditions to which the system generally must adapt; they function independent of the system, and generally do not permit practical modification. Their significance is readily apparent in terms of system disruption or even failure. No model of a highway system would be analogous to the real-world structure if it failed to take environmental variables into account.

Of fundamental interest in such an analysis are predictions of the frequency of occurrence of relevant aspects of the environment, and of their effects upon the operation of the highway system. At this stage of the development of the model frequency of occurrence is of primary concern. As this model is interrelated with models describing other aspects of the highway safety problem, the operational effects will become more significant and pertinent. A description of these operational effects requires first a description of the nature of the environmental stimulus itself, in terms of both physical and dynamic characteristics, and second, a description which relates it to a predictable or identifiable response. Fundamental here is the determination of the impact of the environment upon operational efficiency and utilization of the system.

In terms of general system operation, some measure of the rate of occurrence is essential. The degree of expectation on the part of the driver and his consequent response and adaptation will be functions of this rate and the event durations.

Available data for the parameter of interest can be employed to compute frequency distributions, mean values, standard deviations, and other possible methods of interpreting the data in a useful context. Pertinent mathematical statistics procedures are used to estimate the probability of encountering the various environmental situations of interest. Thus the availability of useful and appropriate data becomes of fundamental interest, as well as the applicability of existing data to predictive analysis. Emphasis should be upon using available data as efficiently as possible.

It is apparent, however, that at this preliminary stage the appropriate concern is to delimit and define potentially significant environmental stimuli and their organization so that they may be more readily analyzed. Table 9-I lists obvious environmental factors and estimates their controllability.

Until they are more carefully analyzed, the relative significance of any of these stimuli cannot be firmly established. In fact, the validity of including such stimuli as sun mirages or distractions remains in some doubt. Weather and pavement conditions in the area of an accident are observed and included in accident reports, but accident-related information about distractions such as noise or objects along the roadside can only be highly speculative or at best taken from subjective accounts.

Measurement of certain stimuli also present problems. That live animals do in fact obstruct traffic from time to time is not in question. However, any attempt to predict such occurrences with any assurance would require one to measure the effectiveness of right-of-way fences against the lack of same, measure the numbers and distribution of the animal population, and acquire insight into animal motivation for intruding upon the highway system. Obviously, the existence and identification of any specific environmental stimulus does not in itself constitute justification for detailed study of that stimulus. One must separate the trivial from the significant, and not only provide proper emphasis, but proper context, for studying these stimuli. For example, many of those which are too trivial to study within the context of the environmental model may be considered more efficiently in the context of the vehicle controller model under driver perception and information processing.

TABLE 9-I. ENVIRONMENTAL FACTORS NEGATIVELY AFFECTING THE VEHICLE OPERATOR, VEHICLE OPERATION, AND THE EFFICIENCY OF THE PHYSICAL SYSTEM

I. CLIMATE (The weather conditions local to the system region)
   (a) Temperature (affecting both operator and vehicle) (occurrence not controllable)
   (b) Wind: effect on vehicle handling and driver response; also, effect on precipitation, relative velocities, and angle of incidence (occurrence not controllable)
      (1) Distribution of short period
      (2) Mean velocity
      (3) Regularity, amount of gustiness
   (c) Weather-related atmospheric contaminants within driving vision range
      (1) Precipitation: snow, sleet, rain, mist (occurrence not controllable)
      (2) Fog (occurrence not controllable)
      (3) Other particles suspended in the atmosphere, man- or wind-induced: smoke, dust storms, smog, flying insects, etc. (occurrence partially controllable)
   (d) Weather-related surface conditions affecting physical system efficiency (such as pavement wetness, snow accumulation, mud, ice) (occurrence partially controllable)

II. LIGHT CONDITIONS
   (a) Sun glare (direct or reflected) (occurrence not controllable)
   (b) Sun shadows (as detrimental to visibility) (occurrence partially controllable)
   (c) Mirages, ephemeral effects of the sun (occurrence not controllable)
   (d) Night conditions (absence of sun) (occurrence not controllable)

III. ERGONOMICS NOISE (occurrence partially controllable)
IV. OLFACTORY DISTRACTIONS (occurrence partially controllable)
V. TRANSIENT DISTRACTIONS (Off the roadway and not physically obstructive per se)

- (a) Attractive nuisances: scenic elements, advertising, other attractive activity including off-road repair and maintenance operations, accidents, etc. (occurrence partially controllable)
- (b) Potential physical obstructions: animals, pedestrians, vehicles off the roadway but which might move onto the roadway (occurrence partially controllable)

VI. TRANSIENT PHYSICAL OBSTRUCTIONS (On the roadway)

- (a) Random debris: litter, vehicle-generated debris, dead animals, rocks, sand, and gravel (occurrence partially controllable)
- (b) Pedestrians or livestock on the roadway (occurrence partially controllable)
- (c) Maintenance and repair operations on the roadway (occurrence controllable)

VII. NATURAL DISASTERS (Floods, landslides, earthquakes, forest fires) (occurrence not controllable)

RELEVANT SUBMODEL VARIABLES

The significant submodel variables to be isolated at this point are those of fundamental importance to the vehicle operator, the vehicle, and the highway system, i.e., those relevant stimuli which can be shown to have definite bearing upon the mechanics of system operation, operational efficiency, and safety. The chief goal of this submodel should be to present these environmental stimuli in such form as to facilitate their ready integration into the response models; that is, driver response and limitation, vehicle response, and system response. For example, these might include the driver's ability to perform the necessary operations under a disruptive sequence of visual distractions; the handling characteristics of a vehicle under irregular wind loading; or the detrimental effects of wet pavements upon regular vehicle operation. In order to properly analyze environmental stimuli, then, the origin and nature of the particular factor must be known and a predictive model developed giving probable intensities, spatial distributions, and durations of the particular stimulus.

To facilitate analysis, the factors listed in Table I have been rearranged and categorized under four fundamental topics of concern: vision through the atmosphere, wind force, operationally disruptive stimuli, and operationally distractive stimuli.

VISION THROUGH THE ATMOSPHERE (θ_1)

The visibility problem can be expressed in terms of four fundamental factors:

(1) The optical properties of the atmosphere itself

(2) The amount and distribution of natural and artificial light

(3) The characteristics of the objects to be perceived, particularly as to position, shape, color, optical properties, and as light sources in their own right

(4) The properties of the vehicle's glass system and the controller's eyes or optical system

In this chapter only the first two factors are of concern.

The following aspects of visibility through the atmosphere are appropriate to this study:

(1) Optical properties of the atmosphere in its natural state

(2) Modification of these optical properties through the presence of suspensions in the atmosphere that serve to impair visibility by scattering and obstructing light

    (a) Meteorologically induced: precipitation, fog, clouds
    (b) Otherwise induced: dust, smog, insect swarms

(3) The presence of light sources (artificial and natural), their intensity and configuration

WIND FORCE (θ_2)

The relevant properties of wind are:

(1) Time distribution pattern

(2) Short-period fluctuations

(3) Spatial and directional characteristics

Also relevant are the properties of wind as related to the following:

(1) Vehicle operator response

(2) Influence of wind on alteration of visual properties of the atmosphere

OPERATIONALLY DISRUPTIVE STIMULI (θ_3)

Those environmental factors (excluding the above) which serve to impair normal and efficient operation of the total highway system are:

(1) Meteorologically induced: precipitation to cause wet or slippery pavements, snow accumulation, ice accumulation

(2) Natural disasters: floods, landslides, earthquakes, fires

(3) Transient obstructions: litter, live animals, or maintenance operations on roadway

OPERATIONALLY DISTRACTIVE STIMULI (θ_4)

Those environmental factors which impair normal operation of the highway system only by distracting the vehicle operator are:

(1) Temperature

(2) Noise

(3) Olfactory distractions

(4) Distractions off the roadway (accident sites, disabled vehicles, pedestrians, maintenance work)

Thus, any variable is determined by various and appropriate subfactors, any of which may vary in significance depending on the input conditions. But it can be seen that all the subfactors under a given θ_k produce a similar effect upon some aspect of the highway system. Although all the stimuli may be said to be either distractive or disruptive, it is believed that meteorologically caused pavement effects, atmospheric visual capacity and the impact of wind force are far and away the most important factors in the environmental model. Consequently, the last two categories—distractive and disruptive—serve to identify in a general way those other odd assorted potential stimuli of interest as well as the meteorologically induced disruptive stimuli which are vital to pavement performance and related responses. In this study the objective is to provide a predictive analysis for these stimuli. It is obvious that this analysis is virtually identical to that for meteorologically induced impediments to visibility, that is, the prediction of precipitation. The problem will not be discussed separately. Ultimately, it is probable that no attempt will be made to separate disruptive from distractive stimuli when analyzing their probability. The value in so classifying them is that it permits detailed analyses of their effects.

SPECIFICATION OF REQUISITE RELATIONSHIPS

θ_1: VISION THROUGH THE ATMOSPHERE

Middleton (1958) states the basic objective of visual analysis is "to establish usable theoretical relationships between light, eye, target, and atmosphere that will permit the calculation of the visual range at any time; and to provide means of measuring the necessary parameters quickly and accurately enough." Here visual range means the distance at which something can be seen. It is a function of atmospheric conditions affecting the visual area between the stimulus (e.g., the object) and the receptor (e.g., the eye), and properties of the object and its surroundings which determine how well the object can be distinguished from its surroundings.

Fundamental to this study is the interaction between light and the atmosphere. Manifestations of this interaction may be crudely described as either daylight or night darkness, but it should be noted that they may include some extremely subtle effects, as noted by Middleton (op. cit., p. 18): "the luminance and color of the sky play a major part in determining our sub-conscious feelings about the weather." Without question, this topic is extremely complex, and theoretical analysis becomes very involved. This report does not attempt to present the specifics of such analysis, and the interested reader should consult Middleton's work directly. Though he concedes that theory does not yet explain all aspects of the visibility problem, his treatment is extremely thorough and rigorous.

Extinction of light in the atmosphere is fundamentally the result of interference from the air itself and particles in the air. The source of the earth's light and radiant energy is the sun. Although the measure of this energy, the solar constant, varies according to the energy output of the sun and its distance from the earth (slight seasonal variation, 91.5 to 94.5 million miles), the actual range of values (from 1.88 to 2.01 gram calories) has little effect on the earth. Rather, it is the "transparency" of the atmosphere to these rays of radiant energy that determines the heat energy received on the surface, and the intensity and nature of light. On the average, over the surface of the planet, 34% of this energy is transmitted back into space (9% is scattered, 2% is reflected by the earth's surface and 23% by clouds); of the remainder, 19% is absorbed by the atmosphere, 23% is absorbed by the earth in diffuse and scattered form, and 24% is absorbed directly by the earth. It is very difficult to trace a single ray because of multiple scattering induced by the atmosphere, suspended particles, and the earth.

Visible light may be considered as "radiant energy evaluated in proportion to its ability to stimulate our sense of sight" (Middleton, p. 7). The scattering coefficient b for pure dry air is defined as:

$$b = \frac{8\pi^3}{3\lambda^4} \frac{(u^2 - 1)^2}{n} \cdot \frac{6 + 3p}{6 - 7p}$$

where u is the index of refraction of air of wavelength λ

n is the number of molecules per unit volume

$p (= 0.042)$ is the factor of depolarization for air

However, such theoretical considerations are rather academic because of the presence of the atmospheric aerosol (air, plus the particles suspended in it as a single system of medium plus suspension). These suspended particles take many forms; smoke, dust, living organisms of microscopic size, salt particles from the sea, water vapor, water droplets, ice crystals, and snow flakes. Water can exist in the atmosphere in all three of its phases, as an invisible vapor, as droplets of water, or in solid form as particles of ice. Knowledge of the size and form of the individual water particles is necessary for application to light extinction theory, and has been the subject of much research with fog, humidity, and the like. Water droplets form only when there are "nuclei," minute particles on which vapor can condense and which are the objects of considerable speculation. Even today the exact nature of the ions inducing cloud formation are not known.

Given the size and concentration of suspended particles, some means must be devised for predicting extinction coefficients, or predicting the nature of the particles given their optical properties. According to Middleton (op. cit., p. 26), work done in this field of research takes essentially three directions: theoretical investigations based on the electromagnetic theory of light, theoretical work on the lines of geometrical and physical optics, and experimental investigations devoted to the discovery of empirical relationships, especially between the extinction coefficient and wavelength.

The problem encountered in any theoretical analysis is that of correlation with conditions as they actually are in the atmosphere. There is confusion about the precise definition of many terms (such as "fog") and descriptions of precipitation and obscurity in general are vague. The common practice of measuring visibility in terms of distances over which an object remains visible does not begin to take into consideration the nature of the interfering atmospheric particles, which could be of many different forms and produce the same visual effects. There is thus a gap between theory and actual practice in visibility studies, particularly in the interpretation of existing weather records to deduce probabilities of certain levels of visibility.

The subject of visibility is by no means completed by analyzing media through which light must be transmitted. Contrast, an important measure of visibility, is a function of the object and background, as well as the atmosphere. Middleton (op. cit., p. 60) defines contrast as:

$$C = \frac{B - B'}{B'}$$

where B is the luminance of the object, and B' that of the background. Contrast (C) can range from -1 to infinity, with higher values where the object is more luminous than the background, as an object in the headlights of a car at night. A general expression for the law of contrast reduction by the atmosphere is as follows:

$$C_R = C_o(B'_o/B'_R)e^{-\sigma_o \bar{R}}$$

where B'_o is the luminance of an object close at hand, B'_R is its luminance at distance R, and σ_o is the extinction coefficient. The following formula relates visibility distance d to similar brightness factors:

$$d = \frac{1}{\sigma} \ell n \frac{\frac{B_L}{B_B} - 1}{E}$$

where B_L denotes the average brightness of the light being seen, B_B denotes the average brightness of the atmosphere, E the threshold constant, and σ the fog attenuation coefficient. The above results show the significance of the optical properties of mediums of light transmission and the importance of contrast in visibility. Although artificial light will not be considered here, these fundamental principles are directly applicable to lighting design. Finally, analysis of glare is closely related to the matter of contrast, in addition to the adaptability of the human eye to rapid variations in light intensity.

In general, a proper background in knowledge and an understanding of the principles of light transmission through varying mediums, and of their interplay with properties of visibility (e.g., absorptance, reflectance, wavelength, etc.) of physical objects, are necessary to any competent study of the diverse environmental elements which contribute to visibility through the atmosphere.

The fundamental characteristics of the environmental elements have also to be considered. The first is the presence or absence of the sun, both as a light and heat source. In order to describe adequately its effect upon the system at any given time, one must determine from the season and time of day the sun's position in the sky and consequently the angle at which the sun's rays are penetrating the atmosphere. The percentage and expected duration of cloud cover must also be known.

A predictive model for forecasting weather conditions appropriate to this study is the proper domain of the agencies assigned this task, such as the U. S. Weather Bureau and other military and private organizations. The primary reason for this is their access to their own extensive records of past weather conditions; without such records attempts to forecast are generally inadequate. The easiest mistake to make in forecasting is to depend solely on a prognosis developed from abstract relationships. A realistic forecast is to some extent an estimate of probability based upon weather known to have occurred in the past under similar conditions.

The proper objective of the weather submodel in the highway's safety system is to predict the probability of occurrence of weather conditions relevant to system operation. In terms of visibility, the model must provide some information on diurnal and seasonal variations for the following climatological factors: humidity, precipitation, fog, cloud cover, and the interrelated factors such as pressure, wind, and temperature which alter the former factors. The following comprehensive description of the prevailing weather in a given area are measures of the following eight interrelated variables (after Critchfield, 1960):

(1) Temperature
(2) Sunshine
(3) Pressure
(4) Winds
(5) Humidity
(6) Cloudiness
(7) Precipitation
(8) Visibility

The principal climatic controls producing a given state are the following:

(1) Latitude
(2) Altitude
(3) Land and water surfaces
(4) Mountain barriers
(5) Prevailing winds
(6) Air masses
(7) Storm and pressure centers
(8) Ocean currents
(9) Local relief

and, of course, the sun. It can be seen that it is the interaction among these factors that produce diurnal and annual variations, spatial distribution, and intensity.

As an example of such analysis consider fog, the most well known of the visibility detractors on the highway. There are four positive factors, one of which must be present in order to produce fog (George, 1960, p. 298):

(1) Existence of fog areas the night before (in that even though there might be no other positive factors present, the fact that fog is being developed under existing conditions is sufficient to be a positive factor itself)

(2) Precipitation areas

(3) Climatology for season and locality

(4) Upslope or downslope air flow

The following negative factors operate to prevent fog:

(5) State of the earth's surface

(6) Cloud cover

(7) Stability of atmosphere (for lowest 50-150 mb.)

(8) State of hydrolapse (vertical gradient of specific humidity)

(9) Temperature

(10) Wind direction

(11) Wind velocity

(12) Wind shear

(13) Wind accelerations

Favorable conditions for fog include low wind velocity, a cloudless sky, depression of the dewpoint, the onset of night cooling, direction of prevailing winds, rate of dew point change, and nature of airflows (fog is more likely at points of divergent flow). A predictive model for fog on a general scale would produce a set of maximum likelihood factors for a given season of year, and with regard to specific nights or weeks, a probability estimation. Given a fog bank, the spatial distribution is desired and obtained from such modifying factors as topography and air masses. In addition, a model for the prediction of fog clearing is desirable, usually related to time of sunrise, but also to the onset of any disruptive factor. Available records upon which analysis may be based are data on fog occurrence compiled by location and season.

Precipitation likewise has its set of factors, generally related to temperature and pressure changes. In a given rainstorm, intensity varies according to the duration of the storm and its hydrological classification. Fitted to a rainfall of specified frequency, an intensity-duration curve is expressed by the following equation:

$$p = \frac{A}{(t + B)^n}$$

where p equals the average rain intensity in inches per hour, t is the duration in minutes, and A, B, and n are empirical constants. The pattern expressed is one of high intensity of rainfall in the beginning moments of a storm and a gradual tapering off.

Particle size, in addition to the accumulation of particles per given volume, is fundamental to the visibility model. It has already been mentioned that knowledge in this area is sketchy, but in any event particle size must be related to precipitation definitions to accord with the available records. Fortunately, the U. S. Weather Bureau provides a standard of definition of precipitation type based upon the nature of the particles. In fact, there are some 23 identifiable classifications of precipitation, the most important of which are as follows:

(1) Rain (drops greater than 0.5 mm diameter)
    Includes: freezing rain (freezing on impact)

(2) Drizzle (drops less than 0.5 mm diameter, particles uniform and closely spaced)
    Includes: freezing drizzle (freezing on impact)

(3) Snow

(4) Snow pellets (diameter 2 to 5 mm, granular)

(5) Snow grams (diameter less than 1 mm, flat or elongated)

(6) Ice pellets (less than 5 mm in diameter, translucent)

(7) Hail (5 to 50 mm in diameter)

(8) Ice prisms (needles, plates)

(9) Fog

(10) Ice fog

(11) Mist (reduces visibility to not less than 1 km)

For theoretical visibility calculations this list is probably not very satisfying, but it must be remembered that this study is being based upon past records. Further research can perhaps remedy this, but not over any very short period of time.

Descriptions of cloud cover are based upon the percentage of sky coverage and the classification of the cloud, particularly by its absorption and light-scattering properties. Records are available that give the number of days per year during which cloud cover (overcast) is present, and probabilities of the extent of cloud cover by time of day and season of year. For any given day afternoons are most often the cloudiest, with night and early morning the least cloud-covered.

Practical weather forecasting is based primarily on observations of air mass systems: prediction of the movement of the air masses and storm systems (which generate all weather), and knowledge of the weather associated with them (singly or in combination). And, up to the recent advent of weather satellites, it was generally difficult to piece together a very localized description of weather conditions or of the spatial distribution of pressure systems, precipitation extent, and wind patterns because of the limitations inherent in scattered, isolated weather observation stations. Because long-term records of comprehensive observations are not yet available, a predictive model is still unable to give pinpoint probabilities (at one spot) or a completely satisfying overall picture.

In visibility analysis primary emphasis is upon relating basic optical concepts to available meteorological data so that a predictive model can be constructed to give probabilities of occurrence of meteorological conditions leading to impaired visibility (and would not exclude impaired visibility caused by "normal" environmental conditions). These data must then be analyzed in terms of the human response model (capacity of the human eye) and visual characteristics of the physical highway system. In terms of vehicle performance, the only relevant factor would be the optical efficiency of the windshield, and the ability of the vehicle's equipment to keep it clear during precipitation.

θ_2: WIND FORCE

The force of the wind has a significant effect upon vehicle operation and response, but is very difficult to determine. We need some means of expressing the effect of wind in terms of both velocity and dynamic considerations. The problem is made exceedingly complex by erratic local wind currents, eddies, and gusts generally resulting from surface irregularities, vehicular motion and proximity to other moving vehicles, particularly large trucks and busses. Consequently, no meaningful analytical expression can be derived to characterize total wind motion. In structural analysis the common alternative has been to emphasize the average effects of wind behavior, a solution which unfortunately works adequately only for objects much larger than motor vehicles.

Gustiness is most often referred to in terms of a "gust factor," or the ratio of the peak wind speed to the mean wind speed for a 5-minute period. At wind speeds greater than 30 knots there is no real variation in the gust factor, but at lesser velocities one can obtain average gust factors and gust durations. These values, related to average speed, can be found in the Handbook of Geophysics.

There are several possible ways to model gustiness. One might be to assume wind as flowing with periodic fluctuation; that is (Davenport, 1961; p. 454):

$$V_t = \overline{V}(1 + v \sin 2\pi nt/\overline{V})$$

where $\overline{V}$ is the mean velocity, with a superimposed fluctuation of amplitude v, small compared to $\overline{V}$, and frequency n.

Another way is to consider flow as a stationary random process. The statement implies an unpredictable velocity from one moment to the next except as governed by certain stationary laws of probability, i.e., independent of the origin of time and duration of the record. Davenport finds this to be characteristic of most turbulent flows and wind in general. In structural analysis, the response of the structure cannot be discussed in terms of amplitude of input and output as with the former formulation, but must be analyzed with relation to stress analysis. A general wind equation is devised as follows (Davenport, op. cit.; p. 455):

$$\overline{V^j} = \frac{1}{T}\int_o^T V_t^j \, dt$$

where T is the duration of the record. It has been found that the normal, or Gaussian distribution can be applied to velocity distributions of wind, thus requiring only a knowledge of the mean velocity $\overline{V}$, and the mean square velocity, $\overline{V^2}$ (or the variance, $\sigma^2(v)$). A Fourier analysis can break down records of irregular wind velocity, such as into sine terms involving a phase angle:

$$V_t = \overline{V}\left[1 + \sum_{j=1}^{\infty}(v_j/\overline{V}) \sin\left(2\pi j\frac{t}{T} + \theta_j\right)\right]$$

If T is increased to make the relative interval between consecutive frequencies indefinitely small, then the summation can be replaced with an integral (Davenport, op. cit.; p. 456):

$$V_t = \overline{V}\left[1 + \int_o^{\infty} a(n) \sin(2\pi nt + \theta_n) \, dn\right]$$

where n is the frequency, and a(n) is a coefficient (velocity per unit frequency).

The mean-square velocity is:

$$\overline{V_t^2} = \overline{V}^2 + \int_o^{\infty} \frac{a^2(n)}{2} \, dn$$

The term $a^2(n)/2$, denoted as S(n), is important, the spectrum of gustiness (power spectral density), and defines the variation of the mean-square intensity of the velocity fluctuation with frequency (Davenport, op. cit.; p. 457). Thus, the variance of the fluctuations is given by:

$$\sigma^2(v) = \int_o^{\infty} S(n) \, dn$$

At high average wind speeds the fluctuations comprise a smaller average percentage of the total and simplifying assumptions can be made.

Other studies, directly related to vehicle response, have been published by the Society of Automotive Engineers (see bibliography); these treat wind velocity as a variation of fluid flow and employ hydrodynamic relationships and wind tunnel tests almost exclusively.

Analysis of driver response to wind fluctuations is primarily concerned with the extent of fluctuation and the periodicity of gusts. Wind is produced by large-scale pressure gradients, but surface winds are modified by the effect of surface roughness. The gradient wind velocity is obtained as the wind's velocity at an elevation of 1000 to 2000 feet, i.e., its velocity when unobstructed by frictional forces from the ground. Davenport relates surface wind velocities to the gradient wind velocity by means of a power function related to elevation and surface roughness. Surface roughness is created by topography, vegetation, and "culture" (such as buildings). For relatively strong winds, gusts of a few seconds duration arise primarily from surface frictional effects on the wind flow, while gusts of a minute's duration generally arise from internal perturbations in the air flow. The wind gradient may be constant over a great area but it is apparent that the surface wind velocity will not be, making spatial distributions of wind conditions difficult to determine unless they can be related to the prevailing gradient wind. Davenport has employed statistical approaches to this problem. At this time it is doubtful that any meaningful results have been obtained to describe the gusts that buffet a car. The variability of surface roughness renders a model difficult to develop; further research may be of value here.

It appears, then, that the most meaningful analysis of wind force at this time is its impact on vehicle response. It has already been said that it is the contrast between peak and average velocities and the frequency distribution of gusts that should be of concern. Vehicle operation in wind is greatly affected by the driver's response, which is affected both by the "unexpectedness" as described by the gust factor, and by the response of the vehicle itself. Thus, the gustiness model should be structured to correlate with the vehicle response model.

θ_3: OPERATIONALLY DISRUPTIVE STIMULI

The meteorological model requires only a shift in emphasis to precipitation that dampens or freezes over the roadway surface, impairing optimal utilization. Snow and ice accumulation must be added to the model, and related to system capacity. Accumulation is, of course, highly controllable, and thus is closely linked with maintenance efficiency.

There are other ways in which weather conditions disrupt normal highway system operation. Julian cites the failure of drivers to adjust to driving hazards in wet weather, and particularly rain, as outranking all other traffic hazards combined. Apparently the surprise and novelty of snow offset its hazards somewhat by increasing the driver's attention to his environment.

The submodel shown includes provisions for predicting transient obstructions and natural disasters, but attention should also be given to their significance. It is possible that in general their occurrence should merely be accepted; descriptions of such events would probably be limited to statistical descriptions relating occurrence to environment. By the same token, control of these events would essentially be limited to considerations of efficient maintenance and clean-up operations. Vehicle-generated debris (fallen exhaust pipes, treads, cargo, etc.) could possibly be considered more properly the realm of the vehicle performance model.

θ_4: OPERATIONALLY DISTRACTIVE STIMULI

Statistical methods would be applied here similar to the ones for disruptive factors mentioned above. Included among these stimuli would be noise and fumes from exogenous sources such as factories, construction or repair work, farms (using or producing large quantities of organic material), and even dead skunks.

SUMMARY

The recommended priority of research emphasis is as follows:

(1) Vision Through the Atmosphere. Apply basic concepts to available meteorological data to develop a highway system visibility model. Accompany this with the development of a predictive model which will give the general probabilities of the occurrence of detrimental meteorological conditions. These data should be interrelated

with the human response model (limitations of the human eye) and visual characteristics of the physical highway system.

(2) Wind Forces. Perform research toward developing a descriptive model of wind force, with emphasis on spatial and time characteristics. The wind model should be scaled to interrelate with the vehicle response model.

(3) Operationally Disruptive Stimuli. The meteorological problem is similar to that in the Vision section above. Research should be performed on those aspects of driver response to the dynamic environment that have been shown to be significantly related to highway safety. Where appropriate, study such dynamic characteristics as onset of weather conditions or generation of debris over time.

(4) Operationally Distractive Stimuli. Lowest priority. Reserve study unless analysis of traffic accidents and safety considerations warrant statistical analysis.

REFERENCES

Critchfield, General Climatology, 1960.

Davenport, A. G., "Application of Statistical Concepts to Wind Loading of Structures," Proc. Inst. Civil Engin. (U.K.), Volume 20, August 1961, pp. 449-472.

George, J. J., Weather Forecasting for Aeronautics, Academic Press, New York, 1960.

Middleton, W. E. Knowles, Vision Through the Atmosphere, University of Toronto Press, 1958; 1963 (revised).

SUBJECT BIBLIOGRAPHY

GENERAL METEOROLOGY

Critchfield, General Climatology, 1960; or any other good text on climatology, for background reference.

George, Joseph J., Weather Forecasting for Aeronautics (New York: Academic Press, 1960). (Weather forecasting from an operational point of view.)

VISION THROUGH THE ATMOSPHERE

Middleton, W. E. Knowles, Vision Through the Atmosphere (University of Toronto Press, Toronto, 1958, 1963r.). (A standard work, thorough and comprehensive.)

NIGHT VISIBILITY

(Topic given great attention by highway and traffic engineers. Consult annual reports by *Highway Research Board*, and magazines in the field such as *Traffic Engineering* and *Traffic Quarterly*.)

WIND FORCE

Barth, "Effect of Unsymmetric Wind Incidence on Aerodynamic Forces Acting on Vehicle Models and Similar Bodies," SAE Preprint 948 E, 1965. (Little on wind force (treated as fluid flow), but relates to vehicle response.)

Davenport, A. G., "Application of Statistical Concepts to Wind Loading of Structures," *Proc. Inst. Civil Engin. (U.K.)*, Vol. 20, paper 6480, August 1961, pp. 449-472. (Several theories on dynamic wind flow, unusually lucid treatment.)

Shiotani, M., "Dynamic Wind Pressure to Structures with Reference to Turbulent Properties in Wind Near Ground," *Railway Technical Research Institute, Quarterly Report*, Tokyo (English translation), June 1963, Vol. 4, No. 2, pp. 1-4.

Toti, G., "Aerodynamic Effects on Vehicle Moving in Stationary Air and Their Influence on Stability and Steering Control," SAE Preprint 948 D, 1965. (Vehicle response model.)

Atmospheric Turbulence and its Relation to Aircraft, Symposium, Royal Aircraft Establishment, Farnborough, 16 November 1961. (Her Majesty's Stationary Office, London.) (Not entirely appropriate, but exhaustive treatment of all aspects.)

National Symposium on Winds for Aerospace Vehicle Design, Air Force Surveys in Geophysics No. 140, March 1962. (U. S. Air Force, Bedford, Mass.) (How the U. S. treats the subject of wind.)

METEOROLOGICALLY-INDUCED DISRUPTIVE STIMULI

Julian, Dan W., and Arwine, Carline M., "Weather in Frequency of Traffic Casualties," *Traffic Safety*, December 1957, pp. 94-97. (Driver response to changing conditions.)

Tanner, J. C., "Weather and Road Traffic Flow," *Weather*, Volume 7, No. 9, September 1952, pp. 270-275. (Relation of bad weather to traffic demand and flow.)

10 TRAFFIC FLOW

ROBERT FARRELL

INTRODUCTION

The general form of the highway systems model comprises two complementary classes of submodel: those which describe the system's physical components and their performance capabilities, and those which describe the operating process of traffic movement. The latter include models of traffic movement, traffic accidents, and traffic generation. This chapter is concerned with models of traffic movement, or flow.

Traffic for the purposes of this chapter, is the movement of a number of vehicles through a transportation network. In normal operations, the vehicles are constrained to enter and exit the network at known sources and sinks, and to have nonintersecting world-lines* which must not intersect the world lines of any other objects. Any violations of these constraints are accidents (see Chapter 11). The network may vary with time (as, for instance, when roads are built or closed), but we will be little concerned with temporary changes in traffic caused by the actual action of change and much more concerned with predicting the stable behavior of traffic in the modified network. Thus, in all future discussion, we will consider only an invariant network.

We think of each vehicle in traffic as both traveling along a route conceived in advance by some human agency and as commencing travel at a time determined in advance by some human agency. The times and routes considered together constitute the demand. (Thus, in general, we do not consider route changes decided upon in traffic; the theory may have to be extended to include these.) The demand process drives the traffic: if there were no demand, there would be no traffic.

In traffic, a vehicle which has entered a transportation network with a predetermined route moves in a manner dependent on the local

---

*The world-line of an object is the space-time volume it occupies; it is the trace of the object through space as time elapses.

traffic situation. Specifically, the acceleration vector of the automobile or other vehicle in the highway-road-street network is a function of the immediate environment, including the nearby traffic. The mechanism which underlies and determines this function is made up of the actions of:

(1) A surveillance system, which actively elicits and passively receives data from the environment. (Generally embodied in the human visual, aural, and kinesthetic perception systems.)

(2) A guidance system, which accepts the data from the surveillance system and operates on it to arrive at control settings for the vehicle control system. (Generally embodied in the human central nervous system.)

(3) A control system, which alters the vehicle's force and energy relations with the environment in accordance with the control settings. (Generally embodied in mechanical systems within the vehicle.)

A systematic theory of traffic, then, must be a system which accepts statements describing

(1) The vehicles (and possibly the cargoes)
(2) The transportation network
(3) The demand
(4) The surveillance-guidance-control systems

and produces statements about the character of traffic. Specifically, the theory might produce statements about:

(1) Traffic Flow (the number of vehicles passing a given point in a given time period)

(2) Traffic Density (the number of vehicles occupying a given section of network at a given time)

(3) Traffic Velocity (the harmonic average of the vehicle velocities, which is the flow divided by the density, is generally used. This is useful in predicting travel times.)

(4) Trip Time for location A to location B

(5) Traffic Delay Time (defined as the actual trip time from A to B less the time that would be required in the absence of traffic.)

(6) Number of stops on trip

(7) Time stopped on trip

(8) Cargo Flow (passengers or freight tons per hour passing a given point)

(9) Cargo Density

(10) Cargo Velocity

The output of any model of traffic flow will provide data for the accident prediction model and for the calculation of service (efficiency) measures (see Appendix 1-A). As the accident prediction model is defined and improved, the traffic flow model must be extended to provide the data needed as input to the next stage.

The mechanisms behind the dependence of the output variables upon the system variables (above) are extremely complex, and have thus far proven unsatisfactory for detailed investigation. Instead, four different approaches have been used and have yielded results which describe, very approximately, certain of the dependencies when all else in the situation is held fixed at "normal" values. These approaches are outlined below.

CAPACITY THEORIES

METHOD

In these theories, invented by and for those who design traffic facilities, the independent variable is the facility design and the dependent variable is the facility capacity—the maximum traffic flow that it will accomodate under prescribed conditions. Generally neglected are the dependencies on control strategies, vehicle characteristics, etc. The facility design is principally treated in terms of traffic lane geometry, and secondarily in terms of major control devices (stop signs, traffic lights, etc.). (The control devices may be seen as mechanisms for correlating control strategies, and insofar as these correlations affect the capacity, the dependencies on control strategies are partly included in the theory.) The capacity of a facility is determined from empirically discovered relationships with the geometry.

The capacity theories have a natural extension to considerations of larger networks. The total capacity of facilities leading into (or out of) a region may be defined as the sum of the capacities of the individual facilities. This extension, however, may lead to capacities which exceed maximum obtainable flows because of interrelations of the facilities inside the region, or because of the location of demand (see Fig. 10-1).

If a large outbound demand originates in region A, and no additional outbound demand originates in region B, the net flow across the boundary of B and C is limited by the capacity of the A-B boundary, not the capacity of the B-C boundary (unless the B-C boundary has the lesser capacity).

LITERATURE

The principal results of capacity theory are available in the Highway Capacity Manual (U. S. Bureau of Public Roads, 1965). The results presented in this document are too numerous to reproduce here. They are generally presented in tabular or graphic form, and sometimes (rarely) in formulas. The relations are generally between operating

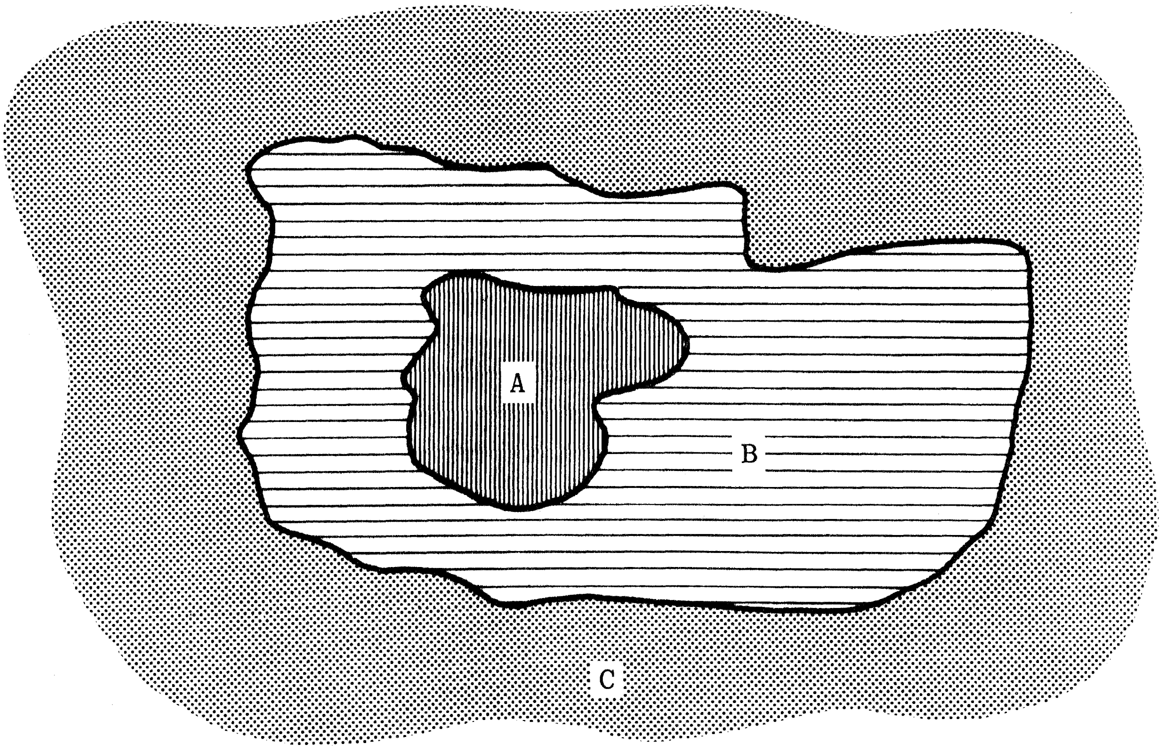


FIGURE 10-1. EXAMPLE

speeds, flow rates, and descriptions of the roadway or facility (e.g., lane width and distance to lateral obstructions in open road; and traffic-signal timing, provision of turn lanes, and percentage of left and right turns for intersections).

FLOW-DENSITY RELATION THEORIES

METHOD

These theories begin with the assumption that the interaction of the control strategies and the highway configuration lead to a functional relationship between flow and density. This relationship is assumed to (1) pass through (0, 0), (2) have increasing flow up to some maximum flow and decreasing flow thereafter, and (3) have no flow beyond some maximum density. The density at maximum flow is termed the optimal density. The theory is generally assumed valid only for single roadways, and the exact functional relationship is determined empirically. Little work has been done on the relationship of the function to the highway characteristics or the control strategies. Some attempt has been made to derive a flow-density relation from control strategies in very general form. This is discussed in the section on car-following theories.

Several papers consider the effects of this assumption on traffic behavior. (These are the "hydrodynamic" theories.) They generally discuss the space-time variation in flow and density (wave behavior) under certain demand assumptions. They remain concerned with single roadways.

There are several problems with these theories: extensions to networks of roads are not immediate. The curves proposed to fit the flow-density relation, though they fit the case in which density is less than optimal density very well, are a poor fit to the case of higher density. This is not due to the particular form of the curve proposed, but to the fact that the data for high-density flows have a large dispersion. It would be well to consider here also the fact that the high-density flows are actually caused by queueing effects behind downstream merges or constrictions, and that the assumptions of the flow-density theory may well not hold in such areas: the flow and density may be mutually dependent upon the character of the merge or constriction, rather than upon the immediate local traffic.

The assumption that flow and density, as instantaneous derivatives, are related by a particular function does not lead to any obvious conclusions about the average flow and density over even moderate periods and distances. If flow and density are varying in either time or space, the averages will not, in general, be on the function curve. (This may help to explain the dispersion of the high-density average data.)

LITERATURE

The earliest proposal for a model of the flow-density relationship was made by Greenshields (no date). He utilized a linear relation between speed and density, or, equivalently, a parabolic flow-density curve. Only one aberrant data point was available for densities greater than the "optimal" density. More recently, Greenberg (no date) used the relation

$$v = c \ln (K_j/K)$$

equivalently

$$q = cK \ln (K_j/K)$$

where v = traffic velocity
K = traffic density
q = traffic flow

and c and K_j are parameters varying with the road and the vehicle and driver populations. Guerin (no date) proposed that curves of the form

$$V_e = \frac{A_\theta K^2}{(B_\theta - K)^{1/2}} + c_\theta$$

be fitted to the relation between V_θ, the θ-percentile value of traffic velocity, and K, the traffic density. Underwood (no date) has suggested that the relation

$$V = Ae^{-K/B}$$

may be adequate for predicting flows on nonexpressway facilities.

Several researchers, beginning probably with the late O.K. Normann of the Bureau of Public Roads, have proposed that analytically distinct curves be fitted to the flow-density relation at densities less than the "optimal" density and at densities greater than the "optimal" density. A natural extension of this idea, taken with a conclusion of Edie's (undated)* is to consider the flow-density relation—whatever curve is fitted to it—valid only up to the "optimal" density of adjacent downstream sections of road. Beyond this, any flow-density pattern has much less accuracy; that is, constant flow lines generally fit the empirical data as well as or better than the curves generally used, but all curves have large standard errors. At these greater densities, theories should concentrate on the physical behavior of traffic engaged in queueing and leaving queues, and should not rely on the assumption that the flow and density are functionally related.

The extensions of flow-density theory involve the analyses of the dependence of the traffic variables q, K, and v and their time and space variations. Central to this discussion are the results of Richards (no date) and Lighthill and Whitham (no date). The principal results of the analyses are

$$v = q/K$$

and wave velocity $= Df(K)$ (where $f(K) = q$ is the flow-density relation)

$= \Delta q/\Delta K$ (if a shock wave)

(Wave velocity is the speed of a density pattern's motion, not the velocity of any vehicle or vehicles. Mathematically, it is the space-time slope of an equidensity [contour] line.) These relations enable the engineer to predict the effects of time-varying input flows. The accuracy is limited in the most useful applications (high-density congestion effects) by the inaccuracy of the assumption of flow-density dependence.

QUEUEING THEORIES

METHOD

By considering the traffic at any point as determining a statistical arrival process, one may consider the queueing that results at certain types of blockages—stop signs, stop lights, merges, etc.

---

*Edie suggested, and all data not distorted by pre-summarizing (either by classification or long-term averaging techniques) confirms, that densities greater than the critical density occur only in queueing situations upstream from restricted flow.

The control strategy is incorporated into the service time distribution: generally, the problems treated are those in which the control strategy is principally concerned with moving the vehicle into an acceptable gap in a traffic stream, making the gap distribution determine the service time distribution of the queue.

The principal weaknesses in this approach to traffic are:

(1) The interaction of several queueing problems in a network is beyond present theory.

(2) The actual physical characteristics of traffic in a queueing situation are inadequately understood (e.g., traffic density and its rate of change).

LITERATURE

The major queueing studies have considered three situations:

(1) "Stop" or "yield" merges or crossings in which traffic on one (major) facility is undisturbed by that on the other (minor) road

(2) Traffic-light controlled intersections

(3) Two-lane, two-way road traffic

In the first two cases, one first finds the gap (headway; time between successive vehicles) distribution in each traffic stream involved. It is assumed that this distribution is not affected by the system state (queue size, etc.) or by the preceding gap value. This is false-to-fact, neglecting both the effects of queue length (physical) and of dependencies between successive gaps. It is almost always assumed that these gap distributions are exponential, or Pearson Type III.

Then, in the first case, assuming that a particular vehicle A will merge if and only if the major traffic stream has a gap greater than or equal to A's minimum acceptance criterion, one has a standard queueing problem, with the arrivals distributed in accordance with the minor stream process and the service distributed as large gaps in the major stream. In the second case, counting distribution rather than gap distribution is used to specify the arrival process: one considers the queue size at the beginning of a green-light cycle as a renewal process or queueing process. Again, standard methods give average waiting time, average queue size, etc. The formulas for these results may be found in Chapter 3 of Introduction to Traffic Flow Theory (HRB) and in Haight's book (1963).

For the third case, results are at present entirely unsatisfactory; those that exist will be found in the same references as above.

CAR-FOLLOWING THEORIES

METHOD

In these theories, the control strategy is assumed to depend only upon the general configuration of the highway and upon both the absolute and relative acceleration, velocity, and position of two vehicles in the traffic stream, the one being controlled and the one immediately preceding. No work except the most qualitative has indicated anything about dependence upon highway design. Various assumptions about the precise form of the control strategy function have been made, and some attention has been devoted to two problems, (1) the effects of the control strategy on the relation between traffic density and traffic flow (generally as instantaneous, local quantities, not as averages); and (2) the transmission of oscillatory acceleration down a line of traffic (the studies of traffic "stability").

LITERATURE

Principal results in this area have been obtained by R. Herman and his associates, and are discussed in leading papers by Herman and Rothery (1965), Herman and Potts (1960), and Gazis, Herman and Rothery (no date).

The basic relation given in these papers is

$$\mathrm{acc}_f(t + T) = \gamma(\mathrm{vel}_f(t) - \mathrm{vel}_\ell(t))$$

where γ may take one of the following or certain other similar forms:

$$\gamma = \alpha/(\mathrm{pos}_\ell\ (t) - \mathrm{pos}_f\ (t)) \qquad \text{(Case 1)}$$

$$\gamma = \alpha \qquad \text{(Case 2)}$$

$$\gamma = \alpha \mathrm{vel}_f(t)/(\mathrm{pos}_\ell(t) - \mathrm{pos}_f(t) \qquad \text{(Case 3)}$$

and where pos is position
vel is velocity
acc is acceleration
f denotes the following car
ℓ denotes the leading car
T is time lag
α is a constant
t is the time coordinate

For case 2, it has been shown that:

If and only if $\alpha T < 1/2$, a disturbance due to acceleration by the lead car of a long line decreases asymptotically.

If and only if $\alpha T < 1/e$, the spacing between any pair of cars is nonoscillatory and damped.

Analysis indicates that the stability limits are slightly greater in case 1.

Case 1 has also been analyzed to show that if a pair of cars proceeds for a long time after the last acceleration by the lead car, the asymptotic headway-velocity relation is equivalent to Greenberg's flow-density relation.

Other results have indicated that, under normal conditions, the accuracy of the results is slightly improved if γ is assumed to have different values for positive and for negative relative velocities; and that there is almost no improvement if the formula is extended to include a dependency on the car preceding the lead car.

It should be noted that these stability studies might be considered under accident-prediction models, rather than flow models, inasmuch as they are actually concerned with the conditions of rear-end collisions in single-lane traffic.

SUMMARY

The examples of research results in each of the four areas of traffic flow theory given above are not complete in any sense, but explore the general results of the areas. General sources from which a fairly complete bibliography of the area is readily obtained are:

Highway Capacity Manual, 1965, Bureau of Public Roads

Introduction to Traffic Flow Theory, Highway Research Board

F. Haight, Mathematical Theory of Traffic Flow, Academic Press, 1963

F. Haight, "Annotated Bibliography of Scientific Research in Road Traffic and Safety," Operations Research, Vol. 12, pp. 976-1039.

It is obvious that of the four approaches only two, the car-following and queueing approaches, present a composite model in which the structure is analogous in any way to the structure of the real-world process. Better traffic theories might best be obtained by extending these approaches in the direction of improved correspondence to reality.

In addition to the extensions, alluded to above, queueing theories need improved descriptions of the physical characteristics of the traffic queue. Studies are needed to describe the dependence of the flow and density of the traffic on the traffic queue parameters. Development of these studies will be aided by the data already obtained in experiments concerned with all four types of traffic models: these data will be usable in many model verification tests without requiring completely new experiments.

Car-following theories should be extended in essentially two ways:

(1) Continue and expand the analysis of the effects of car-following behavior on gross traffic parameters (flow, density, etc.). Present asymptotic results are at best only moderately useful.

(2) Extend the car-following model to include nonfollowing behavior. The control system behavior actually depends on perceptions other than those of the immediately preceding vehicle; these dependencies (on visible highway geometry, visible traffic control devices, visible lateral obstructions, etc.) must be explicitly modeled.

These extensions of present theory will require major, long-term studies. But there is a fairly easy, short-term project which offers immediate rewards in increased understanding of traffic and would contribute to the extensions of car-following theory called for above. This is the unification of the present capacity and flow-density theories, and the study of their union with the queueing theories. Since the capacity of a facility is a single point on its flow-density curve, the union is readily accomplished. The unification will give results indicating the dependence of traffic behavior on both control-system and geometric factors (the control system being reflected in the form of the flow-density curve, the geometry in the capacity point and perhaps also in the form of the curve). The increased understanding of this dependence will aid in modeling the actual process through which the geometry influences the behavior, just as the flow-density relation studies aided and led to the car-following results.

REFERENCES

Bureau of Public Roads, Highway Capacity Manual, 1965.

Edie, in HRB Proceedings, No. 37, p. 334.

Gazis, Herman, and Rothery, Operations Research, Volume 9, p. 545.

Greenberg, Operations Research, Volume 7, p. 79.

Greenshields, HRB Proceedings, No. 14, p. 448.

Guerin, in Quality and Theory of Traffic Flow, p. 71.

Haight, F., Mathematical Theory of Traffic Flow, Academic Press, 1963.

Herman, R., and Potts, Theory of Traffic Flow, Elsevier, Amsterdam, 1960.

Herman, R., and Rothery, in Proc. 2nd Int'l. Symp. on the Theory of Traffic Flow, Organization for Economic Cooperation and Development, Paris, 1965.

Highway Research Board, Introduction to Traffic Flow Theory.

Lighthill and Whitham in Proc. Roy Soc. A 229, pp. 281 and 317.

Richards, Operations Research, Volume 4, p. 42.

Underwood, in Quality and Theory of Traffic Flow, p. 141.

11 ACCIDENT PREDICTION

Dean Wilson

The accident prediction model describes the occurrence of accident events in the highway system. The necessary inputs to such a model are:

1. The joint performance capabilities of the vehicle, controller, and highway

2. A description of control strategies for the vehicles involved in travel

3. Summary statements concerning the nature of the traffic flow in the system (outputs of the traffic flow model)

4. Environmental specifications

The outputs are descriptions of the accidents arising in the system, giving:

1. Locations
2. Types and numbers of vehicles involved
3. Distribution of times of accidents
4. Distribution of impact vectors and objects impacted against

The literature on models for predicting collisions is sparse. All of the literature on collision prediction was concerned with models of a type referred to in Chapter 1 as correlational models. Suchman (1961) labels this type as the "epidemiological model" and suggests that this approach is limited in its utility. Jacobs (1961) puts it more strongly:

> It has always been believed that one of the most powerful approaches to the problem of identifying accident causes is that of exploring the nature of these differences (among accidents) and of attempting to determine the characteristics that account for them. This research formulation, unfortunately, has encountered some of the most profound and difficult conceptual and methodological problems in accident research. Equally unfortunately, these problems have rarely been recognized by research investigators working in the area, *with the result that most of the findings that have been reported are meaningless*.

The literature search did uncover two research projects (no publications) whose objectives appear to be directed toward the development of composite collision prediction models, i.e., models for which the internal structure corresponds to the real-world process in terms of the sequence of events which lead to the collision. Descriptions of these projects follow.

Cornell Aeronautical Laboratory

Project:	Urban Intersection Accidents
Sponsor:	Bureau of Public Roads, Office of Research and Development
Purpose:	To observe the actual dynamics of accidents and normal traffic, especially as they relate to the geometry of the intersection and the response limitations of the driver, and to develop a simulation model of the accident producing process.

Harvard University

> A project underway at Harvard University involves the use of interdisciplinary teams of physicians, lawyers, engineers, and behavioral scientists who proceed to the scene of a traffic accident as soon as possible and attempt to make an on-the-spot evaluation.
>
> Some efforts have also been made to simulate the accident sequence under experimental conditions. This type of systems analysis attempts to coordinate all three factors of host, agent and environment into a single human-vehicle environment system including a series of controls affecting driver vehicle and environment simultaneously. It is too early to evaluate the results of this approach, but this kind of systems model is a far cry from the simple classification of host, agent and environmental factors associated with traditional epidemiological research on accidents. — Suchman, 1961.

Aside from these two as yet uncompleted projects, the typical approach is illustrated at its best by Head (1959), in which accident frequency (overall) is correlated with six variables describing the roadway and environment:

(1) Number of commercial units per mile adjacent to roadway

(2) Number of traffic signals per mile

(3) Number of grade intersections per mile

(4) Amount of traffic (ADT)

(5) Indicated speed (accidents increased as the speed indicated on speed signs was lowered)

(6) Pavement width

The study indicated that accident frequency increased with each parameter except for (5), to which it was inversely related. The most significant correlation was with (1). But it is clear that commercial developments near roadsides do not actually of themselves cause accidents; they may distract the attention of the drivers, or the drivers may be concentrating on their shopping rather than their driving, or they may be associated with a larger number of entrance and exit motions, etc. The mere fact of correlation tells very little. Evidence of the dangers inherent in approaches which fail to consider the overall structure of the accident process is reported in a study (Automotive Safety Foundation and U. S. Bureau of Public Roads, 1963) of median dividers in New Jersey which had been found to reduce accidents by 40%. When dividers of the same type were placed on a similar (with respect to the parameters involved) stretch of roadway in California, they were associated with a 75% increase in the accident rate and a 116% increase in the injury rate.

In summarizing the literature search it is clear that there is no danger of needless duplication of effort on the part of The University of Michigan's Highway Safety Research Institute if the systems approach is followed in developing accident prediction models. It is also clear that the literature does not contain even a first-order approximation to the composite model of the accident-producing process.

One approach to the development of such a model might be based upon the assumption that collisions occur when the maximum performance capabilities of the vehicle-driver combination for the particular environment are less than the performance capabilities required to avoid a collision. The inputs for such a model would be the range of vehicle performance capabilities (listed in Chapter 2) such as acceleration, deceleration, speed, etc. as constrained by the driver's performance capabilities (reaction time, senses, etc.) to provide a range of performance capabilities for vehicle-driver combinations which are further constrained by the roadway factors (width, information system, visibility, etc.). Additional inputs are the demand for service (number of cars) on this particular stretch of roadway, and those driving strategies (decisions to pass, slow down, stop, etc.) which are independent of the strategies of the other drivers.

The model would then compute the dynamic trajectories for each vehicle, and for any two trajectories on a collision course, if the performance capability required to avoid the collision exceeds the maximum performance capabilities of the pair, the probability of collision is 1. When the maximum performance capabilities of the vehicles are not exceeded, the likelihood of a collision is the likelihood that the vehicle-driver combination will operate at the required level or greater. The output of such a model would be the location, time, impact magnitude, and angle for each collision. These should be a sufficient set of inputs to the second collision (passenger and vehicle) model. Other outputs such as the number of near-collisions, total number of accidents by type, frequency of accidents per mile of road, etc., could, in principle, also be computed.

The difficulty with the above approach is that it would seem to be impossible to do the computations even if given the values for all of the input variables. Some statistical method would be required to treat this problem. Almost certainly, original attempts to structure such a model would have to simply treat one or more separate types of road situation, and quite possibly one or more separate types of accident (e.g., rear-end collisions on expressway facilities, two-car accidents at urban intersections, or one-car accidents on rural roads).

ACCIDENT PREDICTION MODELS: SELECTED LIST FROM THE LITERATURE

A. RELATED TO DESIGN VARIABLES

Highway:

1. Head, J. A., "Predicting Traffic Accidents From Roadway Elements on Urban Extensions of State Highways," H.R.B. Bulletin No. 208, Highway Research Board: Washington, 1959.

Driver:

None.

Vehicle:

None.

Highway, Driver, and Vehicle:

1. Oklahoma State Highway Department, Special Accident Studies, Oklahoma State Highway Department, Current Planning Branch, U. S. Bureau of Public Roads, 1964.
2. Schoppert, D. W., "Predicting Traffic Accidents From Roadway Elements of Rural Two-Lane Highways with Gravel Shoulders," H.R.B. Bulletin No. 158, Highway Research Board: Washington, 1957.
3. Versace, J., "Factor Analysis of Roadway and Accident Data," H.R.B. Bulletin No. 240, Highway Research Board: Washington, 1960.
4. Malo, A. F., and Mika, H. S., "Accident Analysis of an Urban Expressway System," H.R.B. Bulletin No. 240, Highway Research Board: Washington, 1960.

B. RELATED TO NON-DESIGN VARIABLES

1. Allgaier, E., and Wood, K., "Predicting Traffic Death Rates," H.R.B. Proceedings No. 23, Highway Research Board: Washington, 1943.
2. Case, H. W., "Prediction of Recorded Accidents and Violations Using Non-Driving Predictors," H.R.B. Record No. 4, Highway Research Board: Washington, 1963.
3. Crossette, J., Interstate System Accident Study, Parts I and II, Arizona Highway Department, 1964.

GENERAL

1. Platt, F. N., Operations Analysis of Traffic Safety, Traffic Safety and Highway Improvement Department, Ford Motor Company: Dearborn, Michigan, 1959.
2. Lundy, R. A., "Effect of Traffic Volume and Number of Lanes on Freeway Accident Rates," H.R.B. Record No. 99, Highway Research Board: Washington, 1965.

REFERENCES

Head, J. A., "Predicting Traffic Accidents From Roadway Elements on Urban Extensions of State Highways," H.R.B. Bulletin No. 208, Highway Research Board: Washington, 1959.

Jacobs, H. H., "Conceptual and Methodological Problems in Accident Research," Behavioral Approaches to Accident Research, Ganis and Harris: New York, 1961.

Suchman, E. A., "A Conceptual Analysis of the Accident Phenomenon," Behavioral Approaches to Accident Research, Ganis and Harris: New York, 1961.

"Traffic Control and Roadway Elements," Automotive Safety Foundation and U. S. Bureau of Public Roads, 1963, p. 22.