MODEL-BASED EVALUATIONS FOR CRIMINAL JUSTICE PROGRAMS

by

EDWARD HARRIS KAPLAN

B.A., McGill University
(1977)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREES OF

MASTER IN CITY PLANNING

and

MASTER OF SCIENCE IN OPERATIONS RESEARCH

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

(MAY 11, 1979)

Signature of Author..
Department of Urban Studies and Planning, May 11, 1979

Certified by..
Richard C. Larson, Thesis Supervisor

Accepted by..
Ralph Gakenheimer, Chairman, MCP Committee
Department of Urban Studies and Planning

MODEL-BASED EVALUATIONS FOR CRIMINAL JUSTICE PROGRAMS

by

EDWARD HARRIS KAPLAN

Submitted to the Department of Urban Studies and Planning

on May 11, 1979 in partial fulfillment of the requirements

for the Degrees of Master in City Planning and

Master of Science in Operations Research.

ABSTRACT

This thesis addresses the use of quantitative modeling techniques in the evaluation of criminal justice programs. An overview of evaluation and evaluation methodology is presented, followed by a review of the limitations of these methods. Model-Based Evaluation (MBE) is introduced as a means of resolving some of the difficulties discussed. The major advantages of MBE are presented with examples drawn from the criminal justice evaluation literature.

Some recent evaluations performed in the area of police patrol are presented. The statistical shortcomings of these studies are illustrated, and reasons for the existence of such errors are suggested.

Two concrete applications of MBE are discussed in detail. The first example presents a MBE of one- versus two-officer patrol staffing. Postulated arguments for and against each strategy are outlined as they appear in the literature. Performance measures are elicited from this discussion. Several models are constructed which allow for a comparative analysis using these performance measures; equal cost staffing options are considered.

The second example presents a discussion of methods available for conducting MBEs of treatment-release corrections programs. A general model of rearrest patterns over time is described along with a numerical example illustrating model behavior under alternative assumptions. Classical and Bayesian methods for the estimation of model parameters are reviewed, as are complementary MBE procedures.

Richard C. Larson, Thesis Supervisor
Associate Professor of Electrical Engineering and Urban Studies
Co-Director, Operations Research Center

ACKNOWLEDGEMENTS

This thesis could not have been written and produced were it not for the kind assistance received from several individuals. I would like to thank Dick Larson for two years of guidance and support as both a teacher and a friend. Dick critically reviewed the entirety of this thesis; the finished product has certainly benefited from his comments and suggestions. Keith Stevenson also contributed his thoughts and advice towards the material now contained in Chapters 3 and 4; his encouragement remains greatly appreciated.

Those members of the Operations Research Center and the Department of Urban Studies and Planning who informally comprise the Evaluation Research Cluster have provided a useful sounding board for ideas of all sorts during the past two years. Past and present co-conspirators include: Arnie Barnett, Leni Berliner, Vicki Bier, Tim Eckels, Dorothy Green, Cheryl Mattingly, Nancy Reichman, and Tom Willemain.

The typing of this thesis required a good deal of time and effort. I wish to thank Joan Kanavich for typing Chapter 1, Cheri Johnson for Chapter 4, and Beverly Brown for Chapters 2, 3, and coordinating the production of the final document.

This research was supported in part by Grant number 78NI-AX-0007 from the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice. Points of view or opinions stated in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Finally, I would like to thank my friends and fellow dancers at MIT Israeli Folk Dancing for making this whole process bearable.

Ed Kaplan
MIT, Cambridge, Mass.
May 11, 1979

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

CHAPTER 1

MODEL-BASED EVALUATIONS

I. INTRODUCTION

Each year, millions of dollars are spent on the design and delivery of social service programs. With such massive expenditures in mind, the need for assessing the worth of these programs has become imperative. Increasing attention must be paid to the evaluation of public programs.

However, there are several ways in which evaluation may be approached, ranging from audit-oriented input studies to rigorous experimental designs. To understand the strengths and weaknesses of program evaluation, one needs to understand the strengths and weaknesses of the tools that evaluators employ. If evaluations are to be more effective, the development of more effective evaluation methods is in order.

This thesis addresses the problems which accompany current methods used in the evaluation of criminal justice programs. In particular, a different approach--Model-Based Evaluation--is proposed for application to evaluation problems. The intent of the thesis, then, is to define and discuss Model-Based Evaluation, and illustrate its applicability to situations in criminal justice evaluation.

This chapter serves to set the pace by presenting an overview of evaluation and evaluation methodology. A review of the limitations of these methods follows, and Model-Based Evaluation is introduced as a means of resolving some of the difficulties discussed. The major advantages of Model-Based Evaluation are presented with examples drawn from the criminal justice evaluation literature. Finally, an example of a recently performed

Model-Based Evaluation is discussed. The chapter concludes with a guide to Chapters 2, 3 and 4 of this thesis.

II. INTRODUCTION TO PROGRAM EVALUATION

Perhaps the best way to begin a discussion of program evaluation is to present some general definitions of evaluation as stated in the literature.

1. "Evaluation is the appraisal of the extent to which a program realizes certain goals." (Weiss and Rein, 1972: 247)

2. "'Program evaluation' is the systematic examination of specific government activities to provide information on the full range of the program's short and long term effects on citizens." (Hatry _et al_., 1973: 8)

3. "...evaluation research is the application of social science methodologies to the assessment of human resource programs, so that it is possible to determine, empirically and with the confidence that results from employing scientific procedures, whether or not they are useful." (Freeman, 1976: 14)

4. "...evaluation is a pronouncement concerning the effectiveness of some treatment or plan that has been tried or put into effect." (Deming, 1975: 53)

5. "The starting point of any evaluation study is the question of whether or not a particular social policy 'works'." (Rossi and Wright, 1977: 7)

6. "Evaluation reserach is a rational enterprise. It examines the effects of policies and programs on their targets (individuals, groups, institutions, communities) in terms of the goals they are meant to achieve. By objective and systematic methods, evaluation research assesses the extent to which goals are realized and looks at the factors associated with successful or unsuccessful outcomes." (Weiss, 1975: 13)

All of these definitions are in agreement that the function of evaluation is to assess the performance of _particular_ programs or policies. Rossi and Wright state that:

> "In all cases, the basic assumption of evaluation research is that the program itself, its goals, and

> the criteria for its success are sufficiently well-defined so as to allow an appropriate research plan to be designed." (Rossi and Wright, 1977: 7)

This does not imply that issues such as program goal selection are not problematic; it does imply that such issues fall elsewhere within the broad realm of policy analysis, and are not considered here as problems of program evaluation.

III. EVALUATION METHODOLOGY: NORMATIVE APPROACHES

A generic social program may be conceived as consisting of input, process and outcome components as illustrated in Figure 1, hence it is useful to classify evaluations according to the program element being evaluated. We will now proceed to discuss three approches to program evaluation and the methodological issues that accompany each:

1. Input Evaluation
2. Process Evaluation
3. Outcome Evaluation

i. Input Evaluation

Input evaluations examine the resources which constitute a given social program; as such, input evaluations are not able to answer the question

"Did the program meet its goals?"

without the assistance of other evaluation approaches. However, input evaluations are important in that they serve to summarize the resource constraints under which a program is expected to operate, and whether or not the program did in fact function within these constraints.

In some ways, input evaluations are similar to feasibility studies. However, while input evaluations might specify levels of program operation given fixed resources, feasibility studies attempt to discover which resource mixes are capable of achieving requisite levels of program operation. Feasibility studies are normally associated with the program design process and tend to be somewhat exploratory; input evaluations are less exploratory in nature.

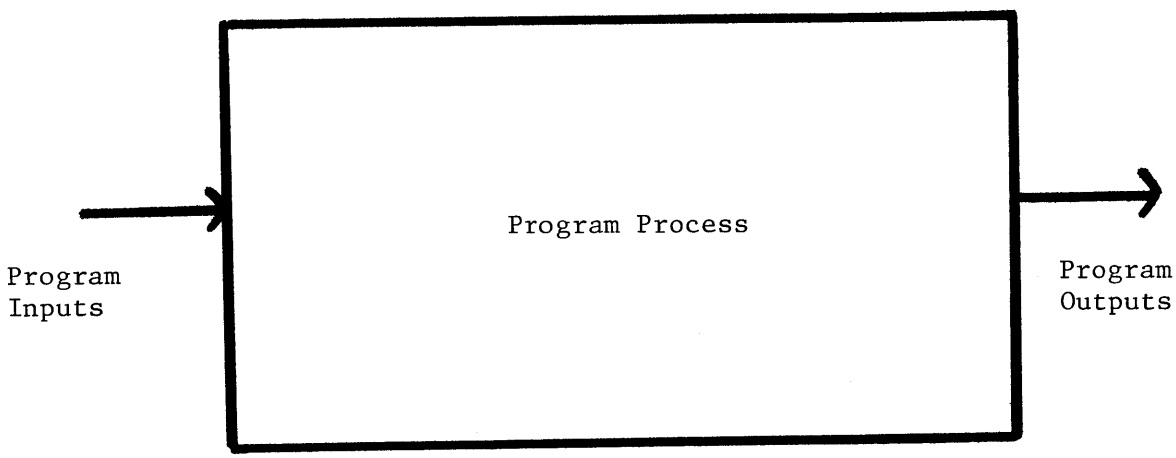


Figure 1

A Generic Social Program

Program auditing is perhaps the most common form of input evaluation regularly practiced. The typical program audit matches program-espoused funding allocations to staff, facilities and equipment with actual program expenditures to see if the program was initially implemented as designed. While such audits cannot be expected to completely answer this question, they do provide useful information to both evaluators and decision makers.

ii. Process Evaluation

While input evaluations identify the resources available to a social program, process evaluations examine the actual utilization of these resources by studying what the program physically does (as opposed to what the program accomplishes--this is an issue of outcome evaluation to be discussed later). As Weiss and Rein (1972) suggest, the essence of process evaluation revolves around the question

> "When such a program is introduced, what then happens?"

Richard Larson (1977) has offered the following definition:

> "A process evaluation of a program seeks to understand the causal mechanism that translates program inputs into program outputs. Or, if program outputs are unobtainable, it seeks to understand the mechanism whereby program inputs are translated into action.
>
> "It utilizes a mixture of qualitative and quantitative techniques to understand the usually multifaceted nature of the process. These include analysis of process-related performance measures, and utilization of participant observers, interviews, questionnaires, and other methods for understanding the total environment of the program. The outlook is 'Bayesian' rather than statistical, meaning that impressionistic information can play a role equally as important as statistical information."

Howard Freeman writes that process evaluation attempts to answer two questions: (Freeman, 1976: 15)

1. Has the program been directed at the appropriate and specified target population or target area?

2. Were the various practices and intervention efforts undertaken as specified in the program design or derived from the principles explicated in that design?

Freeman's conception of process evaluation is somewhat similar to our earlier presentation of input evaluation. However, while in input evaluation we were concerned that resource allocations as described in the program design and observed in the program implementation be congruent, process evaluation searches for congruence in the manner that the resources were utilized. As an example, consider a crime-control program where ten officers were to be utilized for special saturation patrols. Whether or not ten officers were available would be determined via input evaluation; whether or not ten officers actually engaged in saturation patrol would be determined via process evaluation. While this example is admittedly contrived, it does serve to illustrate a conceptual difference between input evaluation and process evaluation.

Several of the techniques associated with process evaluation are included in Larson's statement presented earlier. In order to determine "what happens" after a program is introduced, process evaluators may utilize a number of social science survey and interview methods to obtain information from program officials, staff and clients. Or, evaluators might choose to "live in the system" to see if their "case" is handled as it should be according to the program design. Another monitoring method includes the review of program statistics to see if they match the figures anticipated in the program design (e.g., are service times as anticipated; are caseloads as anticipated).

Process evaluation can yield information which helps to establish the degree of program influence on observed outcomes (this point is especially pertinent to experimental designs; we will return to it later on). Process evaluation may also identify unintended side effects of the program through the methods of inquiry already discussed. What process evaluation does not attempt to determine is whether or not the program achieved its stated goals. This question is addressed via outcome evaluation.

iii. Outcome Evaluation

Outcome evaluations, or impact evaluations, specifically address the question of whether or not social programs have accomplished what they were designed to accomplish. In short, outcome evaluations ask the question

"Did the program meet its goals?"

There are several methodologies available for performing an outcome evaluation. However, from the evaluation literature it is clear that experimental/quasi-experimental designs currently constitute the preferred approach to outcome evaluation. Since many of the ideas of this thesis refer back to experimental situations, this method of evaluation will be reviewed in some detail.

Experimental Design

As mentioned, experimental design (Campbell and Stanley, 1966; Campbell, 1975; Lamar, 1978) has become the dominant approach suggested for application to the evaluation of public programs. In general, experiments may be thought of as consisting of the following five steps:

1. Selection of hypotheses to be tested
2. Selection of performance measures by which to test the stated hypotheses
3. Design of an experimental mechanism for testing purposes
4. Execution of the experimental mechanism
5. Analysis of the experimental results

It is advantageous to briefly discuss each of these steps.

1. Selection of hypotheses to be tested. In evaluation, the basic hypothesis to be tested often has the perverse form

"The program had no effect"

This assertion is termed the "null hypothesis." The substantive form taken by experimental evaluation hypotheses, is, in principle, determined by the stated goals of the program. For example, if the purpose of a saturation patrol program is to reduce the crime rate in an area, then the null hypothesis

"Saturation patrol had no effect on the crime rate"

would be employed.

2. Selection of performance measures. The evaluator's problem is not so much one of identifying hypotheses; rather the evaluator must be able to operationalize conceptual hypotheses in testable terms. The testing of specific hypotheses requires their redefinition in quantitative terms via the use of performance measures. Selected performance measures should satisfy certain criteria:

a. Deterministic criterion - performance measures should be able to gauge the desired phenomena under known conditions

b. Probabilistic criterion - performance measures should behave in predictable fashion under conditions of uncertainty

c. Observability criterion - performance measures should be based on available or collectable data

The first of these points implies that the conditional values taken on by the performance measures are known given different levels of program performance (i.e., the deterministic behavior of the performance measures is well understood). The second point implies that the random behavior of the phenomena under study can be accounted for (i.e., the probabilistic behavior of the performance measures is well understood). This point is important, as it allows one to determine whether or not observed fluctuations in a performance measure may be plausibly attributed to chance (as opposed to the presence of an experimental program). The third point implies the obvious--a performance measure is of no use if it cannot be observed and recorded.

In social evaluation research, it is difficult to satisfy the first two points mentioned above. Thus, it is important to utilize a family of measures for any given study rather than rely solely on one or two variables. According to Hatry et al.:

> "Rarely are a single objective and a single evaluation criterion sufficient to describe the impacts of a program. Inevitably, a program involves numerous objectives, and numerous evaluation criteria will be needed to measure their effects." (Hatry et al., 1973: 31)

The performance measures which are chosen for any given evaluation study are of course dependent upon the hypotheses being examined. However, for a given class of studies, the same measures almost always appear to surface. Some examples from the evaluation of police patrol are presented in Table I.

3. Designing an Experiment. The key concern with the design of social experiments rests with the formation of "experimental" and "control" groups (Campbell and Stanley, 1966). Theoretically, both experimental and control groups should be identical in character. However, while control groups function exactly as before the initiation of the experiment, experimental groups partake in new experimental programs. The presence of the experimental program should be the only difference between the experimental and control groups.

For example, an experiment in the corrections area might consist of two equivalent groups of offenders who have been randomly allocated to control or experimental groups. The control group would be released via conventional parole, while the experimental group would participate in a new program (e.g., work release) upon release from prison. The program being evaluated is the work release program, with standard parole providing a basis for comparison and assessment.

Over the duration of the experimental period (for example, typical experiments in police patrol have had durations of about one year), the levels of predetermined performance measures are continuously monitored. At the end of the experimental period, the control and experimental groups are compared statistically (this will be further discussed under the analysis of experimental results).

In many evaluations, it has not proved possible to obtain a controlled situation of the type described above. Several quasi-experimental

TABLE I

Common Performance Measures Found in Police Patrol Evaluation

Type of Study	Associated Measures
Crime Prevention	UCR Index Locally, reported crime rates Victimization rates (survey) Probability of crime interception Patrol visibility Citizen-perceived fear of crime Citizen-perceived level of safety
Police Response Time	Travel time (with/without dispatch time) Travel distance Apprehension probability Citizen satisfaction with response time
Patrol Productivity/ Manpower Allocation	Frequency of patrol passings Patrol officer workload Patrol officer safety (injuries) Crime/Victimization rates Travel time Citizen complaints Officer complaints

designs are popular in the evaluation field for examining such situations (Campbell and Stanley, 1966; Campbell, 1975; Hatry *et al.*, 1973). These quasi-experiments rely on less powerful comparison groups and/or before-and-after observations. Some quasi-experimental designs incorporate statistical models[1] of the relevant performance measures into the evaluation design (e.g., time-series models, regression models). Such models attempt to predict levels of the performance measures that would have occurred in the absence of the experimental treatment, and hence serve the same function as that of a control group. Of course, the strength of these designs is largely dependent upon the accuracy of the statistical models involved.

4. Maintaining an Experimental Design. In the previous section on the design of experiments, the importance of obtaining good experimental and control groups was stressed. While actually conducting an experiment, the emphasis shifts toward the *maintenance* of good experimental and control groups.

While the laboratories of the physicist or chemist allow desired conditions to be prolonged almost indefinitely, the designs of social experiments are only approximations of these laboratories. Hence, there is good reason to be concerned about problems such as the contamination of the experimental and/or control groups. The conduct of a successful experiment requires that the general conditions in the control and experimental groups remain the same throughout the course of the experiment.

To achieve this, it is advisable to monitor the experiment for the entirety of its duration. The ideas of process evaluation discussed earlier are useful here. The potential payoffs of process evaluation are more pronounced when the evaluation is of the quasi-experimental variety. Without the controls that normally accompany a true experimental design, one has an accountability problem with respect to the true determinants of program outcomes. Was it the program or some other environmental condition that caused the observed results? The collection of process information will not guarantee an answer to this, but such information may surely provide clues not available elsewhere.

5. Analysis of Experimental Results. The evaluation of experimental results in program evaluation has relied heavily upon the use of classical statistical procedures. The use of such procedures is predicated on the control group/experimental group design discussed earlier. If it is in fact true that the sole difference between experimental and control groups rests with the presence of an experimental program (or "treatment") in the experimental group, then observed differences in the levels of performance measures between the two groups may be attributed to one of two sources:

(a) chance, or

(b) the experimental treatment

Statistical procedures along the lines of hypothesis testing check to see if observed differences can be plausibly attributed to chance. If plausible attribution to chance cannot be established, then the experimental treatment is assumed to be responsible for the observed differences through the logic of elimination.

Similarly, quasi-experimental procedures utilize statistical routines to examine program performance in the light of comparison groups, "before-and-after" periods, or a statistical model as discussed earlier. However, as the design of a test program deviates from that of the classic experiment, the rationale behind the use of statistical evaluation devices is weakened. In such cases (which constitute the majority of social evaluation efforts), the collection of process data is extremely important, as such information aids in determining whether or not it was the experimental innovation or some other combination of factors which was responsible for observed outcomes.

* * * * * * *

This description of experimentation/quasi-experimentation has been the preferred approach to program evaluation within recent years. Rossi and Wright claim that:

> "There is almost universal agreement among evaluation researchers that the randomized controlled experiment is the ideal model for evaluating the effectiveness of a public policy." (Rossi and Wright, 1977: 13)

However, experimental evaluations are not problem free; indeed <u>most</u> evaluation efforts, whether input, process or outcome, have been fraught with difficulties. Some of these difficulties stem from misapplications of evaluation techniques (see Chapter 2 for several examples of this sort), while other difficulties seem more attached to the choice of evaluation approach. The limitations of the three approaches to evaluation reviewed here are important to recognize, and it is toward the problematic aspects of popular evaluation methodology that our attention is now directed.

IV. LIMITATIONS OF CURRENT APPROACHES

To understand the difficulties associated with evaluation methodology, it is useful to first ask the question

> "What type of information should be produced by the ideal evaluation under ideal circumstances?"

By focusing on the type of information desired from an evaluation effort, it is possible to consider whether or not current evaluation designs are capable of producing such information.

At this point, the reader may protest that we have already discussed this matter; surely the information desired from an evaluation is whether or not the relevant social program achieved its goals. However, one might be inclined to argue that the situation is more complex.

Let us reconsider Figure 1. The inputs to a social program may be viewed as an initial allocation of resources. The manner in which these resources are allocated is controllable. If, for example, one wishes to use saturation patrol methods to reduce crime rates, one may specify whether two, five, or fifteen officers are to participate.

Similarly, the manner in which these resources are utilized in the program itself is controllable. Are the officers to be deployed simultaneously in "burst" patrols? Are the officers to patrol overtly or covertly? These questions may be answered by program officials.

When a program is evaluated, it is not sufficient to merely state "Yes, it works" or "No, it doesn't work." In the ideal situation, decision makers will want to know:

> "What are program outcomes as a function of program inputs?"

In our patrol example, decision makers will want to know not only the crime rate as related to a patricular patrol configuration; rather, decision makers will want some idea of how the crime rate varies under alternative allocations and utilizations of patrol resources.

Consider Figure 2. Points X_1 and X_2 denote the observed crime rate as a function of number of officers on patrol in a hypothetical patrol program. If X_1 and X_2 are the only pieces of information presented to decision makers, then one could hypothesize relationships A, B and C, all of which are consistent with X_1 and X_2, and all of which have radically different implications toward the deployment of patrol resources.

Thus, an ideal evaluation would provide information explaining how a particular program process (or theory) translates program inputs into program outcomes, and whether or not the resultant program outcomes are commensurate with stated program goals. According to this reasoning, any evaluation which fails to examine program inputs and program process and program outcomes will not produce the type of information which is expected of an evaluation. Input evaluations alone do not examine how program resources are translated into observable impacts. Process evaluations alone do not determine whether or not programs are meeting their goals. Outcome evaluations alone do not examine impacts in the light of program inputs and process.

In particular, it is important to examine experimental evaluation designs. If experimental designs could be perfectly implemented, then through replication, statistical relations between inputs and outcomes could be established. Rather than attempting to examine program structure

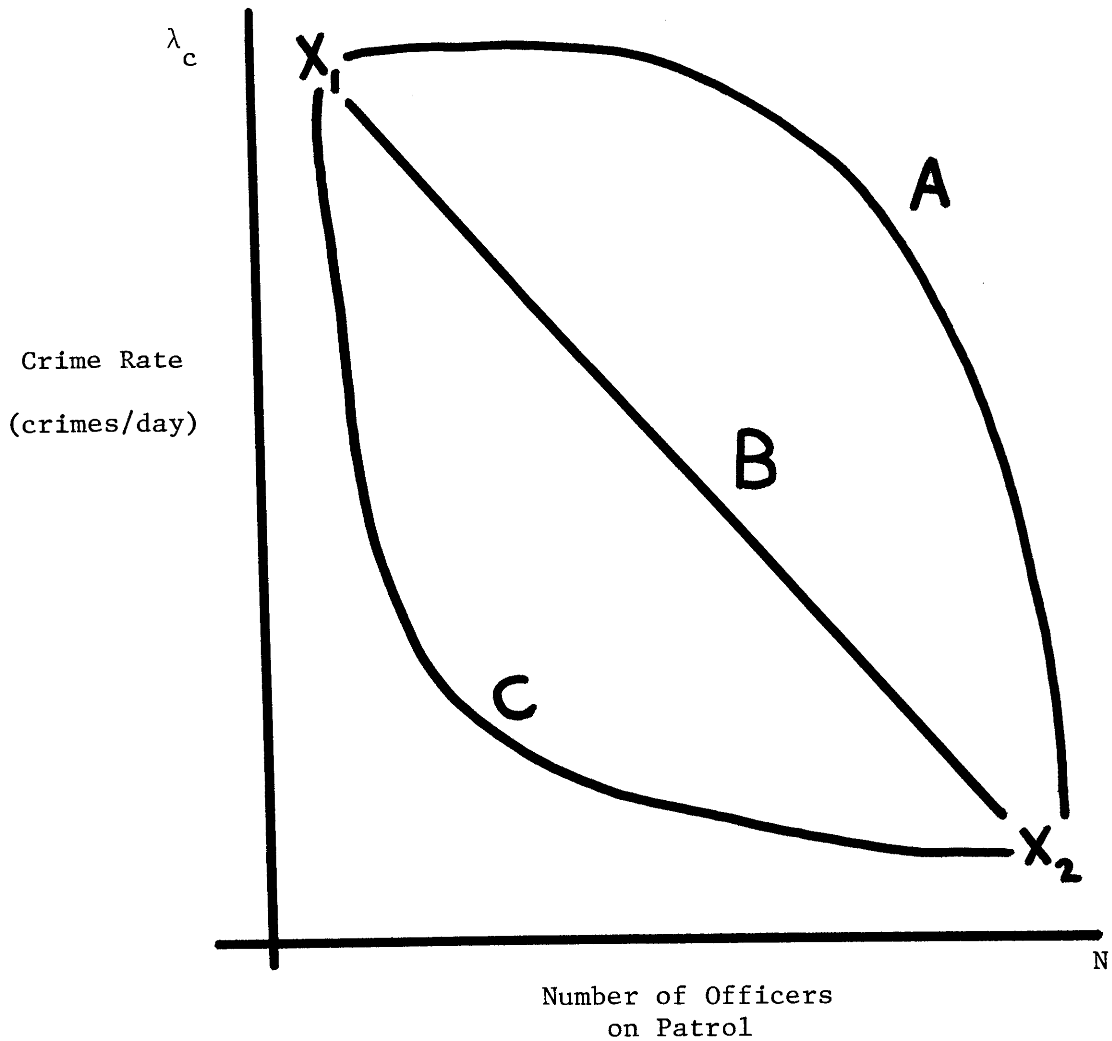


Figure 2

Possible Relations between the
Crime Rate and Patrol Intensity

analytically, experimental methods would build descriptive relationships between inputs and outcomes.

However, in the area of social policy in general, and the field of criminal justice in particular, experimentation as a means of evaluation is difficult to implement in a credible fashion. The most serious problem with experimentation relates to the nearly impossible task of establishing experimental controls. As discussed, the presence of controls is essentially intended to reduce the number of explanatory factors for a particular outcome from infinity to two (chance and the experimental program). In social settings, this type of rigidity is often not obtainable.

For example, a recent LEAA-funded study[2] of two hundred evaluations in criminal justice asked the question

> "Can outcomes directly be attributed to program activities?"

The answer to this question was "no" for 72% of those studies where this question was answered! (Larson et al., 1979: 49) The most frequent reason stated for this finding was that "...the study was essentially uncontrolled (in the sense of research design)." (Larson et al., 1979: 49)

Another problem with experimental-style evaluations has to do with a built-in bias toward the "null hypothesis" (i.e., the no-change option) present in such studies. Rather than attempting to determine how well an innovation or intervention functions, the emphasis is on a double negative: disprove the hypothesis that the program does not work (see Fienberg, Larntz and Reiss, 1976). It is simple to construct hypothetical cases where a perfectly healthy program with a high probability of success would be declared a failure according to a classical statistical test.

There is every reason to suspect that such cases occur in the evaluation of social programs.

Finally, it should be noted that experimental evaluations are very expensive research designs to implement (Hatry *et al*. (1973) refer to experimental design as the "Cadillac of program evaluations"). Given this cost, the problems accompanying social experimentation cannot be ignored.

It would be useful if a different approach could be developed with a sensitivity toward the comprehensive linkage of program inputs, process and outcomes, and a recognition of structural weaknesses in current experimental methodology. One such approach is proposed in the next section.

V. MODEL-BASED EVALUATION: CONCEPTS

The approach presented here and illustrated in Chapters 3 and 4 of this thesis is referred to as Model-Based Evaluation (MBE). As is suggested by its name, MBE involves the use of one or more models in the evaluation process. With reference to MBE, Richard Larson states that

> "The term model here suggests axiomatically derived model in contrast to statistically 'discovered' model." (Larson, 1978: 4)

Thus, the types of models found in MBE are not statistical models such as regression equations. We also do not intend to address mental or other non-mathematical modeling methods within the context of MBE (though such conceptual modeling is important to any evaluation). Rather, the models we are interested in for evaluation purposes are formal models. According to Drake:

> "...a formal model is an explicit abstraction that represents the more significant features of an issue to be considered, identifies the variables of primary interest, and (usually) proposes measures of effectiveness for the comparison of alternative policies." (Drake, 1972: 76)

We may now combine the notions of formal models and program evaluation in an attempt to define MBE. One such definition has already been suggested by Larson (1979: 12):

> "...a model-based evaluation is one in which a comprehensive evaluation (i.e., one studying inputs, process and outcome) is aided by the use of conjectured models that predict one or more process or outcome variables as a function of one or more program inputs."

This definition, it will be shown, is adequate for the most powerful forms of MBE. However, in order to allow less powerful but frequently occurring scenarios to be included, we will modify Larson's definition as follows:

> A model-based evaluation is one in which an evaluation is aided by the use of conjectured models that predict one or more process or outcome measures as
>
> (i) a function of one or more program inputs, and/or
>
> (ii) a result of systematically examining the implications of alternative hypotheses governing program performance, without necessarily making reference to program inputs.

We will refer to MBEs following (i) and (ii) as Full MBEs, while evaluations which are MBEs by virtue of (ii) alone will be termed Partial MBEs. If (i) is satisfied, then it is always possible to explore alternative hypotheses of program performance via sensitivity analysis (changing the parameters of fixed-structure models), structural analysis (altering the form of fixed-parameter models), or both. However, it is sometimes not possible to examine program performance explicitly including program inputs in the analysis, yet models may still prove useful in such instances.

The ensuing paragraphs present the major issues which MBEs are expected to address. It is useful to categorize these issues under the following five headings:

1. Linkage of Program Inputs, Process and Outcomes
2. Systematic Generation and Examination of Alternative Hypotheses of Program Performance
3. Determination of the Time Frame Necessary for Evaluation
4. Identification and Understanding of Alternative Performance Measures

5. Determination of the Cost-Effectivess of Alternative Programs

Each of these issues will now be discussed.

i. Linking Program Inputs, Process and Outcomes

This is the primary incentive for pursuing MBEs. By examining how program inputs are translated into program outputs, decision makers may be provided with information which causally links program activities to observed results. Perhaps an example will make matters clearer.

Consider a saturation patrol program which operates with the objective of reducing observable street crime (i.e., crimes which can be detected by the police patrol). For this example, the only input we consider is the number of patrol units allocated to the program, N. To determine program effectiveness, it would appear appropriate to utilize the probability of on-scene apprehension, P_A, as a performance measure.

The question asked by MBE is, can we build a model which will present P_A (outcome) as a function of N (input)? Without deriving a model here, we may quickly state some assumptions and a result:

(i) Let λ be the rate at which a single unit patrols the experimental area, and assume that all N units patrol independently with identical rates λ.

(ii) Let b be the fraction of time a single unit would be busy (or out of service) if a single unit patrolled the entire experimental area alone.

(iii) Let γ be the rate at which observable street crimes are committed in the experimental area (i.e., the length of the average street crime is $1/\gamma$--see Chapter 3, Section VII).

(iv) Assume that the locations of street crime are uniformly distributed throughout the experimental area.

(v) Let P_I be the probability of intercepting a randomly occurring crime in progress (i.e., the probability of a patrol unit being in the vicinity of a crime while the crime is being committed).

(vi) Let $P_{A|I}$ be the probability of apprehending a randomly occurring crime given interception.

Using probabilistic reasoning coupled with assumptions (i) through (vi) yields a "back-of-the-envelope" formulation for P_A in terms of N:

$$P_A(N|b, \lambda, \gamma, P_{A|I}) = P_{A|I} \cdot P_I$$

$$= P_{A|I} \cdot \frac{N\lambda-\lambda b}{N\lambda-\lambda b+\gamma} \quad (1)$$

Of the variables listed on the right hand side of equation (1) above, N, λ, and to a degree b and $P_{A|I}$ are controllable. While N is strictly an input measure, λ, b and $P_{A|I}$ are process measures. By raising patrol speeds, for example, λ may be increased; by changing reporting practices, b may be reduced, and through alternative search procedures, $P_{A|I}$ may be increased. All of these changes would serve to increase P_A.

Indeed, equation (1) yields a method by which prior expectations of program performance may be computed. These expectations are logically derived through a model which takes the program's input and process explicitly into account. A model similar to that proposed in equation (1) has actually been used in conjunction with an evaluation in the field of police patrol (Elliott and Sardino, 1971).

The application of Full MBEs enhances the comprehensiveness of evaluations by definition. Data collection will be required at all three stages of program functioning in order to implement Full MBEs. A Full

MBE may even be viewed as a framework for comprehensive evaluation in general--by explicitly stating input, process and outcome variables within the model structure, Full MBEs dictate what types of data need to be collected for evaluation purposes, and indicate how such data should be related.

ii. Systematic Examination of Alternative Hypotheses

Sometimes it is very difficult to explicitly state functional relationships between program inputs and program outputs. Consider a corrections program such as work-release which provides an alternative to traditional parole. It is not known to this author if a relationship between the number of counsellors and ultimate recidivism rates can be established, nor is it clear how time spent by a certain number of program workers and the same time spent by the same number of parole officers differentially affect the length of time until clients commit another offense.

In such cases where the existence of causal relationships is not at all obvious, it may be worthwhile to pursue a range of conceivable hypotheses governing program performance. Some of these hypotheses would be rooted in theories which are pro-program success, while others would be rooted in theories which are anti-program success. The emphasis would then be on the determination of which hypotheses are most consistent with observed outcomes.

Note that this approach lies in contrast to the pure experimental approach where one hypothesis, the null hypothesis of program failure, is scrutinized to the exclusion of possible competing hypotheses exhibiting varying degrees of program success or failure. Although the approach of systematic hypothesis generation and analysis is consistent with the notion of controlled field experiments, the emphasis has shifted to the identification of plausible hypotheses for what is observed, rather than the establishment of credibility for or against the null hypothesis.

Models provide a method for systematically generating and studying hypotheses. In the corrections example discussed earlier, alternative theories governing the recidivism patterns of released clients may be expressed as models, and these models may be tested against actual data for their relative credibilities. Chapter 4 of this thesis develops Partial MBE methods for this class of problems.

A good example of the use of models as hypothesis generators/ analyzers stems from recent work in the deterrence area performed by Barnett (1979). In attempting to determine the effect of capital punishment on homicide levels, Barnett examined several models of homicide rates including some models which assume a deterrent effect, and some models which assume no deterrent effect of capital punishment. The predictions of these models were compared with observed homicide levels to determine the relative credibilities of the models. Using the stringent "Region I" admissibility criterion,[3] Barnett reported that the group of models consistent with observed data was "...composed predominantly of models that attribute a noticeable but not gigantic deterrent effect to capital punishment. But--and it is a crucial but--certain models assuming no deterrence at all...are also present in Region I." (Barnett, 1979: 25) Our interest here is not in the actual results stated by Barnett, but rather in the creative fashion that he has employed models to evaluate the effect of capital punishment on homicide levels. The logic of Barnett's approach is the logic to be employed when conducting Partial MBEs.

iii. Determining the Time Frame for Evaluation Research

Many evaluations involve some type of pre/post-program comparison, where dates of program implementation and completion, and lengths of follow-up data collection periods are fixed. However, it is possible that by the workings of the program process, major points in the life of the evaluation may not at all coincide with complementary points in the dynamic manifestation of program effects. If this is true, the evaluation is not likely to produce accurate information.

An example of this is discussed in detail in Chapter 4, but we will briefly discuss this same example now. Consider a corrections program where clients are released following a period of treatment. An evaluation is to take place after twenty-four months have gone by. It has been pre-specified that if 40% or less of program clients have been rearrested after twenty-four months, then the program is a success. Figure 3 presents rearrest patterns over time for three models which all are consistent with the first six months of data collection.

As can be seen from Figure 3, Model C demonstrates a plausible behavior in which less than 40% of program clients have been rearrested by twenty-four months, but more than 40% of all clients are rearrested in the long run. In this case, assuming that 40% rearrested is still the

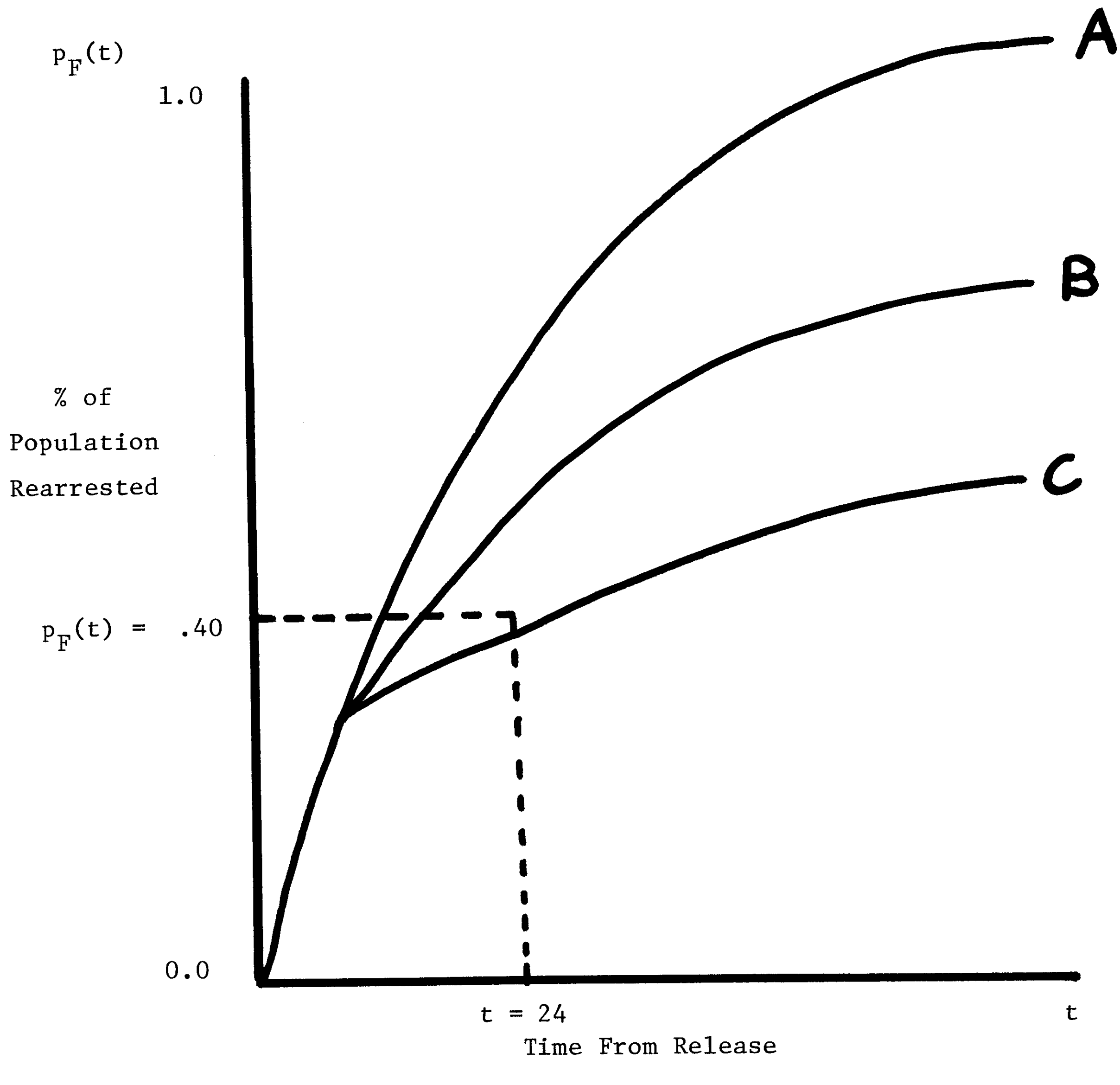


Figure 3

Rearrest Patterns over Time: Three Models desired standard, the evaluation should have continued at least until the time that Model C predicted 40% of all program clients would be rearrested.

Models also allow one to project the long-term consequences of programs. In the same example just discussed, Model A actually predicts the rearrest of <u>all</u> program clients in the long run, while Models B and C predict an eventual 50% rearrest rate. Predictions of this sort are valuable to decision makers who are interested in the ultimate impacts of the programs and policies they choose to adopt.

iv. Identification and Understanding of Alternative Performance Measures

In Section III of this chapter, we identified three criteria for performance measures: the deterministic, probabilistic and observability criteria. Models are an invaluable aid toward the satisfaction of the deterministic and probabilistic criteria. In some cases, models even help to satisfy the observability criterion by manipulating easily gathered information to produce performance measures whose direct observation would be rather difficult. An example of the latter phenomenon is the use of modeling methods to predict queueing delays and other statistics pertaining to the police emergency response system from simply gathered information like call-for-service rates and service times. Some examples of this are shown in Section IV of Chapter 3.

Returning to the deterministic and probabilistic criteria, and the identification of alternative performance measures, models of the performance measures indicate what behavior to expect under alternative assumptions. A good example involves a model built by Blumstein and Larson (1971) of the recidivism process. This model identifies a new performance measure $\bar{n}_c$, the average number of future crimes committed, as a function of r, the recidivism probability. The simple relation discovered was (Blumstein and Larson, 1971)

$$\bar{n}_c(r) = \frac{1}{1 - r} \tag{2}$$

This measure is interesting. Equation (2) tells us that as the recidivism rate decreases from .9 to .8, a difference of 11%, the number of future crimes committed decreases from 10 to 5, a decrease of 50% (Blumstein and Larson, 1971: 128). The performance measure $\bar{n}_c(r)$ is sensitive to small changes in r when r is near 1.0, and is perhaps a better performance measure for this reason.

Equation (2) satisfies the deterministic criterion; if we know that $r = .5$, then we know that $\bar{n}_c(r) = 2$. Suppose, however, that the recidivism probability varies from individual to individual according to some distribution f(r). For example, if the assumption is that individuals recidivate independently of each other, then the beta distribution[4]

$$f(r) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1 - r)^{\beta-1} \tag{3}$$

$$0 < r < 1, \ \alpha, \ \beta > 0$$

could well describe the probabilistic behavior of r. How about the probabilistic behavior of $\bar{n}_c(r)$? One could derive the probability distribution of $\bar{n}_c(r)$ using change of variable methods. In particular, the expected value of $\bar{n}_c(r)$ may be computed as

$$E[\bar{n}_c(r)] = \frac{\alpha + \beta - 1}{\beta - 1} \tag{4}$$

Thus, the probabilistic criterion may also be satisfied for $\overline{n}_c(r)$ through the use of a mathematical model.

Research into the behavior of performance measures is clearly an important task for any evaluation. One of the major outcomes of such research relates back to the notion of prior expectations of program performance as discussed earlier. It is obviously to the advantage of the evaluator to know that observed results are in line with certain situations made explicit via MBE.

v. Determination of Program Cost-Effectiveness

One last use of MBE which should be mentioned--models of social programs can be used to determine their costs, and program effectiveness can be explored by subjecting models of alternative programs to equal cost budget constraints. Although techniques of cost benefit analysis have been developed separately from MBE, the existence of a model provides a convenient framework for costing out a program. It is not the purpose of this thesis to discuss model-based methods of cost-benefit or cost-effectiveness analysis, but a simple example is provided in Section IX of Chapter 3.

* * * * * * *

This tour of MBE has suggested that a different style of evaluation may help to resolve some, but not all, of the problems discussed in Section IV of this chapter. While we have presented scattered examples throughout, it would be useful to examine an actual MBE that was performed and disseminated. Such an example is presented in the next section.

VI. MODEL-BASED EVALUATION: FORMAL MODELS AND THE KANSAS CITY PREVENTIVE PATROL EXPERIMENT

The Kansas City Preventive Patrol Experiment (KCPPE) is perhaps the most well-known social experiment performed in the area of criminal justice. The intent of this study was to determine the effect of varying levels of preventive patrol on outcome measures such as the crime rate. Basically, the experiment attempted to implement three levels of patrol activity:

(i) Reactive Beats - no preventive patrol was to be performed in these areas

(ii) Control Beats - preventive patrol was to be carried out as usual

(iii) Proactive Beats - two to three times the normal level of patrol was to be implemented

The major finding of the study was that variations in the level of preventive patrol had no effect on the relevant outcome measures (crime rate, citizen's feelings of safety, etc.). The details of this experiment may be found in Kelling *et al*. (1974).

The implications of these results could be far reaching. Indeed, if it really is true that patrol has little influence on the incidence of crime, then perhaps the amount of resources allocated to the patrol function should be seriously questioned. However, before any such steps are taken, it is important to question the validity of the experimental results from Kansas City.

In 1975, an article by Richard Larson appeared in the *Journal of Criminal Justice* entitled "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment." This article constituted a reanalysis of the KCPPE. Using data generated by the experiment, Larson probed the likely results one could have expected if the experimental design was properly implemented. He also examined the maintenance of the experimental design, and determined several serious threats to the experimental set-up. More interesting to us here, though, is the method by which Larson carried out his analysis. Throughout his paper, Larson employed probabilistic models to aid him in his investigation. Although the idea of MBE was not fully developed at the time Larson wrote his paper, his work provides a shining example of Full MBE as applied to the evaluation of an evaluation.

To see what it was that Larson was able to show with his models, we will quote from the abstract to his paper (Larson, 1975: 267):

> "Where appropriate, simple probabilistic models are employed to estimate frequencies of preventive patrols and response times in each of the experimental areas. These models, together with experimental data, demonstrate that (1) typical patrol intensities in Kansas City are not large enough to encompass the range of patrol intensities experienced in other cities, and (2) patrol visibility in the depleted areas (the reactive beats) due to responding calls for service is relatively quite large, perhaps even equalling the pre-experimental levels during high workload periods. Such models also demonstrate that travel distances into the reactive beats should not be markedly increased, as the researchers had expected.
>
> "Based on models and experimental data, the analysis indicates that the particular experimental design used in Kansas City resulted in a significant continued patrol presence in the depleted areas, with little increase in travel times in those areas."

Rather than repeat all of Larson's analysis here, the reader is invited to read Larson's paper itself. However, it should be noted that Larson's work does respond to some of the issues discussed in the last section. By examining process-oriented models which utilized input variables as predictive factors, Larson was able to establish a comprehensive link between program inputs and outputs. In some of his models, Larson examined alternative hypotheses governing program performance, establishing a plausible range of likely outcomes (for example, the use of "liberal" and "conservative" estimates in his model of patrol frequency--Larson (1975: 278)). Of course, by constructing models of various performance measures, the behavior of these measures becomes clearer. Larson's models enable his performance measures to satisfy the deterministic and probabilistic criteria as discussed in Section III of this chapter.

The appearance of Larson's work has sparked a controversy in the evaluation field. The evaluators of the KCPPE, following the traditional norms of social program evaluation, devoted much energy to the collection and analysis of empirical data. Larson, a theoretician, attempted to introduce explanatory models to the debate. The two procedures are not mutually exclusive, but they should be recognized as different approaches to problems of evaluation research.

The overall impact of Larson's research appears to have reduced the credibility of the KCPPE results. This is perhaps best evidenced by the Law Enforcement Assistance Administration's (LEAA) recent Request for Proposals to design an experiment to effectively determine the deterrent effect of preventive patrol (see The Effectiveness of Police Preventive Patrol, 1979). There are those who still support the KCPPE results in spite of Larson's work. Nonetheless, Larson has demonstrated the feasibility of using models in evaluation; his work remains the best example of MBE to date.

VII. SUMMARY AND GUIDE TO THE THESIS

This chapter introduced the reader to the field of evaluation research, and to the methodologies currently employed in the evaluation of public programs. After discussing the limitations of these methodologies, we presented an alternative approach in the form of Model-Based Evaluation. In review, we argued that Full and Partial MBEs could be expected to address five issues:

1. Linkage of Program Inputs, Process and Outcomes
2. Systematic Generation and Examination of Alternative Hypotheses of Program Performance
3. Determination of the Time Frame Necessary for Evaluation
4. Identification and Understanding of Alternative Performance Measures
5. Determination of the Cost-Effectiveness of Alternative Programs

We ended by discussing an example of MBE that was performed in response to a large-scale social experiment.

The discussion in this chapter was somewhat abstract. It is the purpose of the following three chapters to motivate, illustrate and apply the ideas we have presented here in specific instances of criminal justice evaluation. These three chapters taken in conjunction with this introductory chapter should give the reader a strong sense of what it means to develop and utilize models within the context of criminal justice evaluation.

Chapter 2 presents some recent evaluation research performed in the police patrol area. The statistical shortcomings of these studies are illustrated, and reasons for the existence of such errors are suggested.

The studies examined in Chapter 2 could well have benefited from the ideas of MBE, and it is intended that the problems observed in that chapter serve as motivating examples of the need for a more structured approach to evaluation. Such an approach is available through MBE.

Chapter 3 presents an in-depth Full MBE of one- versus two-officer patrol units. Utilizing actual data generated by a recent social experiment in San Diego, the models of that chapter examine the implications of staffing policies on performance measures such as expected area covered by patrol, response time, patrol frequency, patrol visibility, probability of crime interception, and injury probability. In addition, equal-cost staffing options are examined. From this MBE, the advantages of one-officer patrol become strikingly apparent.

Finally, Chapter 4 develops and discusses a class of models applicable to the evaluation of treatment-release programs. The use of these models is appropriate in situations involving experimental designs, as the models are valuable generators and analyzers of alternative theories of program performance. Statistical procedures derived from the models are presented, and examples of their use in program evaluation are discussed. That chapter concludes with a brief discussion of possible extensions to the models described.

FOOTNOTES

[1]These statistical models are not to be confused with axiomatically derived "formal models." Indeed, most of this thesis will center around the role of formal models in evaluation.

[2]The study is "An Empirical Study of Methods Used in Criminal Justice Evaluations," Richard C. Larson, Principal Investigator.

[3]A model is deemed statistically credible if it is admitted to Region I. "A homicide model should be included in Region I if its observed systematic error is smaller than the effect it attributes to capital punishment." (Barnett, 1979: 21)

[4]See Freund (1971), page 114.

REFERENCES

1. Barnett, A. (1979). "The Deterrent Effect of Capital Punishment: Yet Another Study." Technical Report #159, Operations Research Center, MIT.

2. Blumstein, A., and Larson, R.C. (1971). "Problems in Modeling and Measuring Recidivism." Journal of Research on Crime and Delinquency, 8: 124-132.

3. Campbell, D.T. (1975). "Reforms as Experiments." In Marcia Guttentag and Elmer L. Struening, eds., Handbook of Evaluation Research, Volume 1. Beverly Hills: Sage Publications, Inc., pp. 71-100.

4. Campbell, D.T. and Stanley, J.C. (1966). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand-McNally.

5. Deming, W.E. (1975). "The Logic of Evaluation." In Marcia Guttentag and Elmer L. Struening, eds., Handbook of Evaluation Research, Volume 1. Beverly Hills: Sage Publications, Inc., pp. 53-68.

6. Drake, A.W. (1972). "Quantitative Models in Public Administration: Some Educational Needs." In Alvin W. Drake, Ralph L. Keeney and Philip M. Morse (eds.), Analysis of Public Systems. Cambridge, Massachusetts: MIT Press, pp. 75-93.

7. Elliott, J.F. and Sardino, T.J. (1971). Crime Control Team: An Experiment in Municipal Police Department Management and Operations. Springfield: Charles C. Thomas.

8. Fienberg, S.E., Larntz, K. and Reiss, A.J. (1976). "Redesigning the Kansas City Preventive Patrol Experiment." Evaluation, Volume 3, Numbers 1 and 2: 124-131.

9. Freeman, H.E. (1976). "The Present Status of Evaluation Research." Paper presented at the 1976 UNESCO Evaluation Research Conference, Washington, D.C.

10. Freund, J.E. (1971). Mathematical Statistics. Englewood Cliffs: Prentice-Hall, Inc.

11. Hatry, H.P., Winnie, R.E. and Fisk, D.M. (1973). Practical Program Evaluation for State and Local Government Officials. Washington: The Urban Institute.

12. Kelling, G.L., Pate, T., Dieckman, D. and Brown, C.E. (1974). The Kansas City Preventive Patrol Experiment: A Technical Report. Washington: Police Foundation.

13. Lamar, B.W. (1978). "What the Textbooks Say about the Design of Experiments." Working Paper OR 073-78, Operations Research Center, MIT.

14. Larson, R.C. (1975). "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment." Journal of Criminal Justice, 3: 267-297.

15. Larson, R.C. (1977). Class Lecture, 11.523 - Evaluation of Public Programs. Department of Urban Studies and Planning, MIT.

16. Larson, R.C. (1978). Third Quarterly Progress Report of An Empirical Study of Methods Used in Criminal Justice Evaluations. Operations Research Center, MIT.

17. Larson, R.C. (1979). "New Tools for Comprehensive Evaluations in Criminal Justice." Grant proposal submitted to the Office of Research and Evaluation Methods, Law Enforcement Assistance Administration, The National Institute of Law Enforcement and Criminal Justice.

18. Larson, R.C., Bier, V.M., Kaplan, E.H., Mattingly, C., Eckels, T.J., Reichman, N., Berliner, L.S. (1979). "Interim Analysis of 200 Evaluations on Criminal Justice." (Draft Copy) Operations Research Center, MIT.

19. Rossi, P.H. and Wright, S.R. (1976). "Evaluation Research: An Assessment of Theory, Practice and Politics." Evaluation Quarterly, Volume 1, Number 1: 5-52.

20. The Effectiveness of Police Preventive Patrol (1979). Criminal Justice Research Solicitation, National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice.

21. Weiss, C.H. (1975). "Evaluation Research in the Political Context." In Marcia Guttentag and Elmer L. Struening, eds. Handbook of Evaluation Research, Volume 1. Beverly Hills: Sage Publications, Inc., pp. 13-26.

22. Weiss, R.S. and Rein, M. (1974). "The Evaluation of Broad-Aim Programs: Difficulties in Experimental Design and an Alternative." In Carol H. Weiss, ed., Evaluating Action Programs: Readings in Social Action and Education. Boston: Allyn and Bacon, Inc., pp. 236-249.

CHAPTER 2

RECURRENT TECHNICAL DIFFICULTIES IN PATROL EVALUATION: MOTIVATING EXAMPLES

I. INTRODUCTION

The world of criminal justice evaluation is characterized by constrained studies with limited budgets attempting to analyze the complex consequences of broad reforms. In the areas of the courts, corrections and policing, studies can be found ranging from the effects of jury selection on sentencing, to the effects of community treatment programs on rearrest probabilities, to the effects of team policing on crime rates. To determine the effectiveness of these programs, evaluators are often placed in the position of having to examine data via some form of quantitative analysis.

Of the three subcomponents (police, courts, corrections) of the criminal justice system, quantitative methods are most easily applied to the logistical problems of the police. Police patrol evaluation in particular has been pursued from a quantitative point of view.

A current LEAA-funded study of 200 criminal justice evaluations included 46 studies in the "police logistics" area; many of these 46 studies involved well-known evaluations of alternative patrol strategies.[1] Of these 46 studies, 42 relied heavily on numerical data, and 32 invoked some form of quantitative method (e.g., hypothesis testing, regression) beyond the tabular display of raw data.[2] When these studies were subjected to methodological scrutiny, some disappointing findings emerged.

Many studies examined exhibited low technical quality. Inappropriate methods were used in some cases, while reasonable techniques were poorly applied in other instances. Since conclusions regarding program effectiveness are reached through such analysis, the credibility of these studies becomes questionable.

This chapter presents four examples of misapplied methodology in police patrol. Since patrol studies tend to be more sophisticated (in a technical sense) than other criminal justice evaluations, the difficulties described in these examples are not felt to be peculiar to patrol evlauations; the problem of misapplied evaluation methodology is of broad concern to criminal justice system evaluators and decision-makers. We conclude by attempting to account for these technical shortcomings, and formulating the case for a different approach to quantitative evaluation.

II. EXAMPLES OF RECENT PATROL EVALUATION ANALYSES

The four examples which follow are in many ways typical of the analytical difficulties found in patrol evaluation. Though the tone of this section may sound highly critical, it must be remembered that these examples stem from state-of-the-art patrol evaluations; in fact, the studies mentioned here are without a doubt among the most rigorous patrol evaluations performed to date. Thus, the comments which follow should not be interpreted as premeditated assaults against specific studies; rather these examples are meant to illustrate a general problem which currently troubles the entire patrol (and criminal justice!) evaluation fields.

i. Injuries in San Diego

The San Diego One-Officer/Two-Officer Patrol Experiment was conducted to compare the performance of one- versus two-officer patrol (Boydstun *et al.*, 1977). While there were many issues tackled in this experiment (most of which are reviewed in Chapter 3), one major reported finding surrounded the issue of safety. Patrol officers were not in favor of one-officer patrol primarily due to the feeling that one-officer patrol is simply more dangerous than two-officer patrol, yet the final evaluation report *Patrol Staffing in San Diego: One- or Two-Officer Units* stated that in general, one-officer patrol was as safe as, if not safer than, two-officer patrol (Boydstun *et al.*, 1977: 69-70). The basis for such a statement stems from a number of statistical tests performed on measures such as assaults, resisting arrests, and officer injuries (Boydstun *et al.*, 1977: 62). For this example, we will only consider injury levels as the other measures may be approached in a similar fashion.

The San Diego evaluators examined the average number of injuries per unit that occurred during the experimental period for both one- and two-officer patrol. These two figures were compared statistically using the well-known Student's t-test (Freund, 1971: 319), and it was concluded that at a significance level of $\alpha = 0.05$, there was no significant difference between these figures, and hence no difference between the likelihoods of injury resulting from one-officer or two-officer patrol (Boydstun et al., 1977: 62). Ignoring the question of whether or not the t-test is appropriate for the data generated by this study, this finding appears to have been reasonably derived. Or was it?

The average number of injuries per unit is not as interesting a measure as the average number of injuries per officer. Assume for the moment that a priori, an officer in a one-officer unit has a fixed probability of injury (over time, type of call and particular officer) given involvement in an incident (this assumption should not be troublesome to supporters of the San Diego study; a comparable assumption is necessary to invoke the t-test). If each of the officers in a two-officer unit has this same officer specific probability of injury, then our intuition would lead us to expect twice as many injuries from two-officer units as from one-officer units. In Section VIII of Chapter 3, we show that this intuitive notion is in fact true.

What we have learned is that a test of the null hypothesis

$$H_o: \frac{\mu_{\text{Injuries}}}{\text{One-Officer unit}} = \frac{\mu_{\text{Injuries}}}{\text{Two-Officer Unit}} \qquad (1)$$

is not equivalent to testing the null hypothesis

$$H_o: \Pr\{\text{officer specific injury} | \text{one-officer unit}\} = \Pr\{\text{officer specific injury} | \text{two-officer unit}\}. \qquad (2)$$

Rather, to test (2), one would examine

$$H_o: \frac{2 \cdot \mu_{\text{Injuries}}}{\text{One-Officer Unit}} = \frac{\mu_{\text{Injuries}}}{\text{Two-Officer Unit}} \quad . \tag{3}$$

If $\frac{2 \cdot \mu_{\text{Injuries}}}{\text{One-Officer Unit}} > \frac{\mu_{\text{Injuries}}}{\text{Two-Officer Unit}}$, then in fact one is more likely to be injured when patrolling in a one-officer unit than when patrolling in a two-officer unit. From the San Diego study, we see that 19 injuries were incurred by officers in 22 one-officer units, as opposed to 31 injuries in the same number of two-officer units for equivalent exposures to dangerous incidents (Boydstun et al., 1977: 63). Since 38 > 31, we see the evidence suggesting that two-officer patrol is safer than one-officer patrol, if only slightly so (this is addressed in Section VIII of Chapter 3).

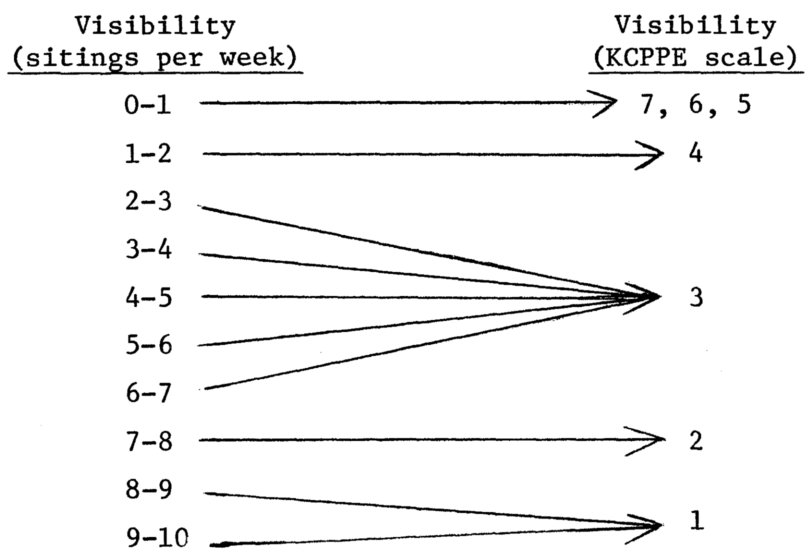
ii. Aggregating Data: Patrol Visibility in Kansas City

The Kansas City Preventive Patrol Experiment (KCPPE) is probably the most well-known of all research efforts in the criminal justice field. In this experiment, 15 Kansas City police beats were divided into three groups of five beats. The first of these groups consisted of five "proactive" beats where the amount of preventive patrol was doubled to tripled. The second group of five was made up of "control" beats where the preventive patrol effort remained at normal operating levels. The remaining five beats were termed "reactive"; in these beats, preventive patrol was presumably terminated completely (Kelling et al., 1974: 28).

To determine whether or not police visibility decreased markedly in reactive areas, surveys were administered which, among other things, asked respondents how often they observed police cars (Kelling et al., 1974: 39). Respondents were able to choose one of the following answers to this question:

1. more than once a day
2. once a day
3. more than once a week (but less than once a day)
4. once a week
5. less than once a week (but more than once a month)
6. irregularly
7. never.

As Larson (1975) has noted, these intervals are not of equal size. In fact, if we compare this data aggregation scheme to one based on an equal-interval scale, the following correspondence is noted (Carter and Kaplan, 1977: 12).



The following example is meant to illustrate one type of result that could be completely determined by the way that the Kansas City evaluators classified their data. While the framework used in the example is analogous to the Kansas City design, we stress that in this example, the data are hypothetical.

Consider the data shown in Table I. One way to check for the existence of a statistically significant relationship between Police Visibility and Type of Beat is to test the null hypothesis

H_o: Police Visibility and Type of Beat are <u>independent</u>

using the statistic (Freund, 1971, 334-337)

$$S = \sum_{i=1}^{5} \sum_{j=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{4}$$

where i,j = row, column subscripts

O_{ij} = observed frequency in cell ij

E_{ij} = expected frequency in cell ij

$$= \left(\sum_{k=1}^{5} O_{kj} \cdot \sum_{\ell=1}^{3} O_{i\ell} \right) \Big/ \sum_{k=1}^{5} \sum_{\ell=1}^{3} O_{k\ell}$$

S = a chi-square random variable with (5-1)(3-1)=8 degrees of freedom.

Applying equation (4) to Table I, we compute $S = 18.79$. At a significance level of $\alpha = .05$, the critical value of a chi-square random variable with 8 degrees of freedom is $S^* = 15.507$ (Freund, 1971: 438). Since $S = 18.79 > 15.507 = S^*$, we may <u>reject</u> the null hypothesis of independence at $\alpha = 0.05$, and conclude that Police Visibility and Type of Beat are related (the desired result from the viewpoint of the Kansas City evaluators).

TABLE I

		BEAT			
		Reactive	Control	Proactive	TOTAL
	I	15	12	10	37
POLICE	II	30	14	10	54
VISIBILITY	III	40	55	60	155
(Sitings)	IV	5	6	5	16
	V	10	13	15	38
	TOTAL	100	100	100	300

where I = less than once/week, but more than once/month

II = once/week

III = more than once/week (but less than once/day)

IV = once/day

V = more than once/day

Now, if we turn our attention to Table II, one notices that if these figures were classified using the aggregation scheme of Table I, Table I would be identically reproduced. Using the equal-interval scale of Table II, we may again compute a chi-square statistic to test for the independence of Police Visibility and Type of Beat. In this case, we examine

$$S = \sum_{i=1}^{10} \sum_{j=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{5}$$

where all symbols are defined as before (but with reference to Table II), and S now has (10-1)(3-1) = 18 degrees of freedom.

Equation (5) produces the value S = 21.32. Again checking against a significance level of $\alpha = .05$, the critical value of a chi-square random variable with 18 degrees of freedom is S* = 28.869 (Freund, 1971: 438). Since $S = 21.32 < 28.869 = S^*$, we cannot reject the null hypothesis of independence at $\alpha = .05$. It could be that there is no relationship between Police Visibility and Type of Beat.

As mentioned before, Tables I and II may be viewed as containing the same data under alternative aggregation schemes. From a statistical viewpoint, the equal-interval data aggregation scheme of Table II is preferable to the scheme utilized in Table I. Yet in the actual experiment, visibility data were aggregated along the lines of Table I (Kelling *et al.*, 1974: 39). In the Kansas City Experiment, relationships between Police Visibility and Beat Type may have be found solely due to the manner in which survey data were classified.[3]

TABLE II

		BEAT			
		Reactive	Control	Proactive	TOTAL
	0 - 1	15	12	10	37
	1 - 2	30	14	10	54
	2 - 3	8	11	12	31
POLICE	3 - 4	11	11	13	35
VISIBILITY	4 - 5	10	11	11	32
(Sitings	5 - 6	6	10	10	26
per Week)	6 - 7	5	12	14	31
	7 - 8	5	6	5	16
	8 - 9	5	6	7	18
	9 - 10	5	7	8	20
	TOTAL	100	100	100	300

iii. The Determinants of Police Response Time

As part of a study related to the KCPPE entitled Police Response Time: Its Determinants and Effects, researchers attempted to determine statistically those factors thought to affect police response time. Response time was defined as "...the difference between the time an officer received a call and the time the officer contacted the citizen." (Pate et al., 1976: 22) Those variables chosen to estimate response time included the distance to the call, the time taken to start the call after the officer had received the call, driving speed, and the time elapsed before the arrival of an assisting officer (where applicable) (Pate et al., 1976: 22-23). The analysis consisted of computing Pearson's product-moment correlation coefficients (r) for response time versus each of the four explanatory variables, and testing to see if this coefficient was significantly different from zero (Pate et al., 1976: 63).

Applying such a technique to a problem of this sort is erroneous. First, let us define the equation that these evaluators implicitly suggested through their framing of the research question:

$$t = D_1 + \frac{d}{s} + D_2 \qquad (6)$$

where t = response time

D_1 = time taken to start call

D_2 = time elapsed before the arrival of an assisting officer

d = distance to the call

s = driving speed associated with the call

As is immediately evident from equation (6), there are four variables which affect response time, hence correlating response time with any one variable demands that the other three be held constant!

A few illustrations will help to establish this point. Allow us for the moment to combine D_1 and D_2 via summation (i.e., define $D \equiv D_1 + D_2$). Suppose we wish to correlate response time with distance. Figure 1 illustrates one set of relations between response time and distance where speed and non-travel delays (D) are held constant; such sets of relations exist for all values of s and D. It is apparent from Figure 1 that for any fixed pair (s, D), response time and distance are linearly related, and the correlation coefficient between these two variables would equal 1.0. However, if s and D are not controlled, then no definite relation is guaranteed to emerge (hence the reported correlations between response time and distance of r = .567, r = .532, r = .475 and r = .550 (Pate et al., 1976: 24-25)).

Similarly, examining response time versus speed controlling for distance and non-travel delays yields Figure 2. Patterns like those shown in Figure 2 exist for all combinations of D and d; aggregating all collected data without structuring it as in Figure 2 is bound to weaken the appearance of this conditional relationship between response time and travel speed.

However, Figure 2 is instructive in another sense. The conditional relationship that exists between response time and travel speed is not linear; this is also obvious from equation (6).

In the actual study, true response speeds were not measured. Rather, the following subjective rating scale was used (Pate et al., 1976: 28).[4]

1. very fast
2. moderately fast
3. slightly fast

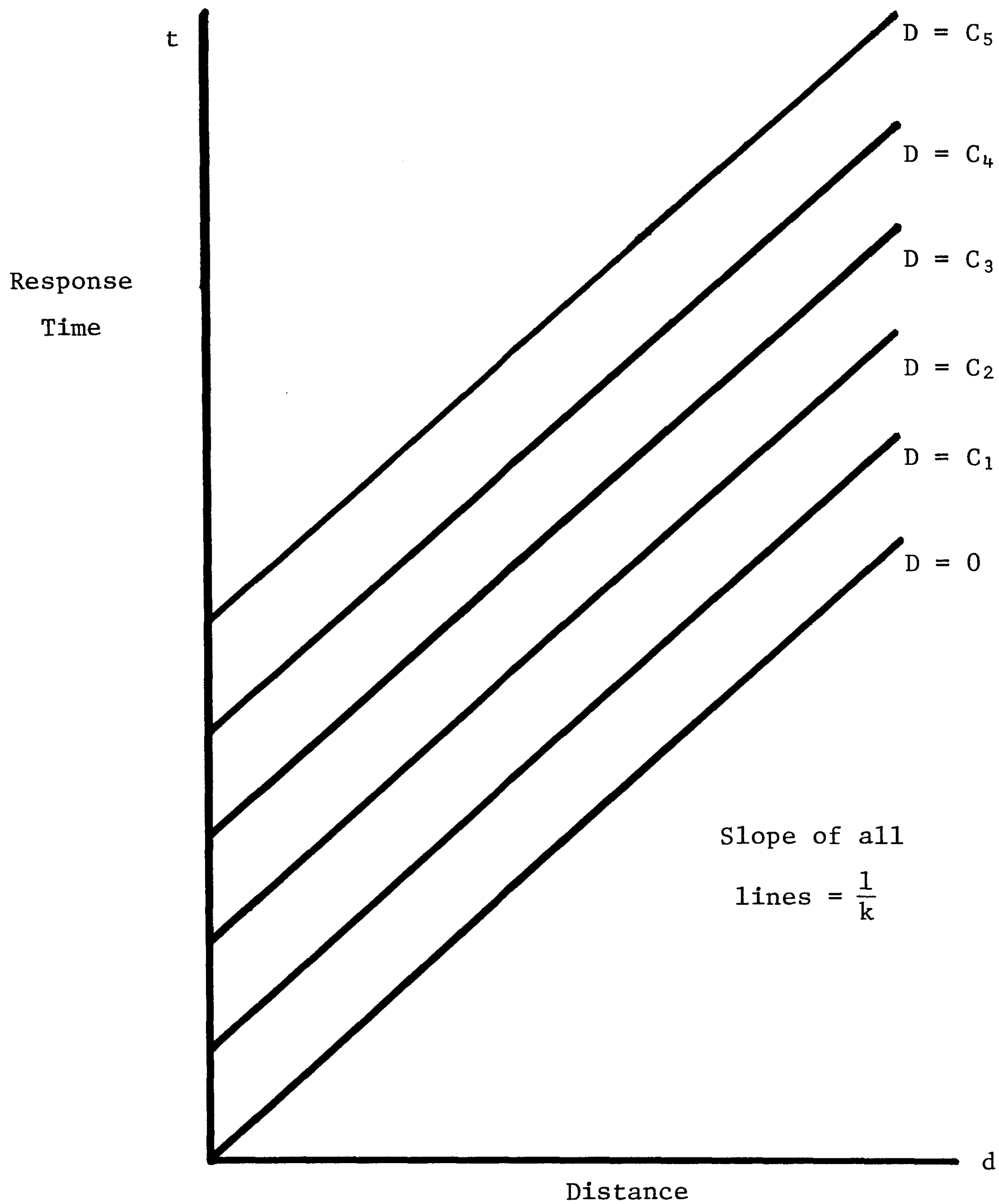


Figure 1

Response Time Versus Distance (s=k)

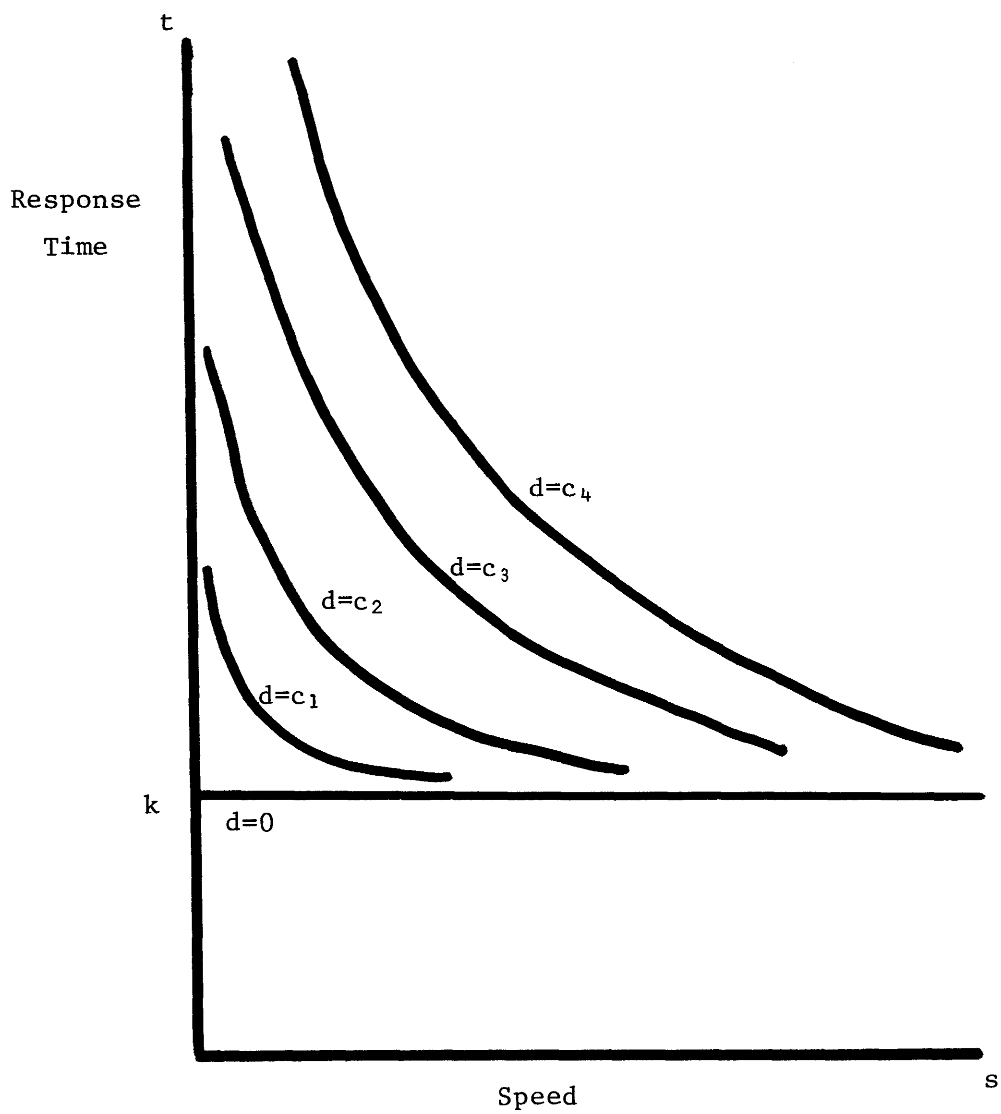


Figure 2

Response Time Versus Speed (D=k)

4. speed limit

5. slightly slow

6. moderately slow

7. very slow.

This scale undoubtedly suffers from problems similar to those of the KCPPE's visibility scale of the last example. In particular, for linear correlation methods to be used, the intervals of this subjective scale need to correspond to equal intervals of 1/s, the inverse of speed for any given call, as the scale increases from 1 to 7. Otherwise, the conditional relation between response time and "subjective speed" as scaled in this report cannot be linear, and the use of linear correlation analysis becomes even more incorrect.

It is little wonder that the reported correlations between response time and driving speed are so low (r = .016, r = .059 (Pate et al., 1976: 28-29)). The researchers attribute this finding to the narrow range of driving speeds that is found in Kansas City:

> "... the speed limits generally ranged only from 20 to 35 miles per hour. With so little variation among reported driving speeds, the low statistical correlations between driving speed and response time should not be construed to indicate that driving speeds do not affect response time. The method of estimating driving speeds may not be sufficiently reliable, and the actual range of speeds may be too narrow to permit the accuracy required for correlation analysis." (Pate et al., 1976: 28)

This writer is inclined to attribute this result to the overall method of analysis employed rather than to the particulars of observed data.

If we return to equation (6), we realize that this equation defines response time. The term d/s corresponds to travel time, while the sum

$D_1 + D_2 = D$ corresponds to non-travel delay time. The intent of the researchers translates as an attempt to show statistically that equation (6) is correct by definition! This example has presented the analytical errors committed by one study which attempted to verify the obvious.

iv. Citizen Satisfaction with Response Time

In a different response time study, a system of regression models was proposed to determine the effects of various factors on citizen satisfaction with response time. The five equations analyzed were (Response Time Analysis, 1977: 120)

$$TT = a + b_1 \; TOC + e \tag{7}$$

$$DT = a + b_2 \; TOC + e \tag{8}$$

$$IRT = a + b_3 \; SC + b_4 \; TOC + b_5 \; TT + b_6 \; DT + e \tag{9}$$

$$(P-E)/E = a + b_7 \; SC + b_8 \; TOC + b_9 \; TT + b_{10} \; DT + b_{11} \; IRT + e \tag{10}$$

$$CS = a + b_{12} \; SC + b_{13} \; TOC + b_{14} \; TT + b_{15} \; DT + b_{16} \; IRT + b_{17} \; (P-E)/E + e \tag{11}$$

where SC = social characteristics of the involved citizen

TOC = type of crime

TT = travel time

DT = dispatch time

IRT = importance of response time

(P-E)/E = perceptions and expectations index

CS = citizen satisfaction

a's,b's = constants to be estimated

e's = residual variation

According to the evaluation report,

> "This model was analyzed through successive multiple regression analysis of each equation listed above. By examining the path coefficients (b's), it was possible to obtain the total effect that an independent variable had on citizen satisfaction by examining both its direct effects and its indirect effects through other variables." (Response Time Analysis, 1977: 120)

The "model" that is presented here has some peculiar implications. Suppose that equations (7) through (11) are true. It is a simple algebraic exercise to show via the substitution of equations (7) through (10) into equation (11) that the following result is also true,

$$CS = A + B \cdot TOC + C \cdot SC + \varepsilon \qquad (12)$$

where A, B and C are constants, ε is a random error term, and CS, TOC and SC are as defined previously. Equation (12) states that citizen satisfaction with response time is determined solely by the type of crime committed, the social characteristics of the involved citizen, and random fluctuations.

Somehow, this result is troublesome. Aside from the methodological flaw of including "independent" variables which are postulated to be dependent upon each other in the regressions of equations (9) through (11), the best result that could be produced by this model is one that dictates police ineffectiveness. If the proposed model is true, all attempts by the police to increase the level of citizen satisfaction with response time are doomed to fail.

Intuitively, one senses that citizen satisfaction with response time should be a monotonically decreasing function of response time. A model exhibiting this behavior could provide a useful framework for analysis.

While the model depicted by equations (7) - (11) does include some of the factors which may be thought of as affecting citizen satisfaction with response time, the way these factors are related in that model is neither useful nor meaningful.

III. PROCEDURE AND THEORY IN PROGRAM EVALUATION

As mentioned at the beginning of Section II, the examples we have discussed typify the kinds of errors commonly committed in patrol evaluation. Though one might react by seriously questioning the integrity of patrol evaluation research when confronted with these examples, such a reaction is not constructive. In this section, we will attempt to account for the technical difficulties apparent in patrol evaluation.

Initially, it is useful to distinguish between two classes of methodological errors in evaluation. The first class of errors we will label "procedural." The second class consists of errors that are more "conceptual" in nature.

Procedural errors refer to flaws that arise during the mechanics of analysis. Improper data aggregation (as in the visibility example), algebraic errors and computational miscalculations may all be viewed as procedural problems. Procedural errors may be committed within the framework of an analytical design which is in itself technically sound, and some of these errors, if detected, may be corrected without too much difficulty. It is our guess (and hope) that these errors do not occur in isolation with sufficient frequency to warrant further attention.

Those errors we have termed as conceptual comprise a more serious group of technical problems. Given a specific evaluation concern, conceptual errors may arise from the evaluator's understanding (or misunderstanding) of the process by which program inputs are transformed into program outputs, and from the evaluator's insertion of this process into a methodological framework for purposes of analysis. In short, conceptual errors result when the relationship linking program inputs to program outputs is not well-structured by the evaluator.

Figure 3 presents the familiar schematic of a generic program, including the components normally associated with program evaluation. Aside from program goals which we will take as "given," and evaluation conclusions which we will take as "implied," each of these components will be briefly scanned for their potential contributions to conceptual error in evaluation.

i. Hypothesis Formulation

This step in many ways sets the stage for subsequent evaluation effort. An ill-conceived hypothesis can dictate the selection of meaningless performance measures, the collection of irrelevant data, and the application of poor analytical techniques (or the unnecessary application of sophisticated methods). In the injuries example of last section, the hypothesis

> "If one-officer and two-officer patrol are equally safe, then the number of injuries resulting from one-officer patrol should equal the number of injuries resulting from two-officer patrol."

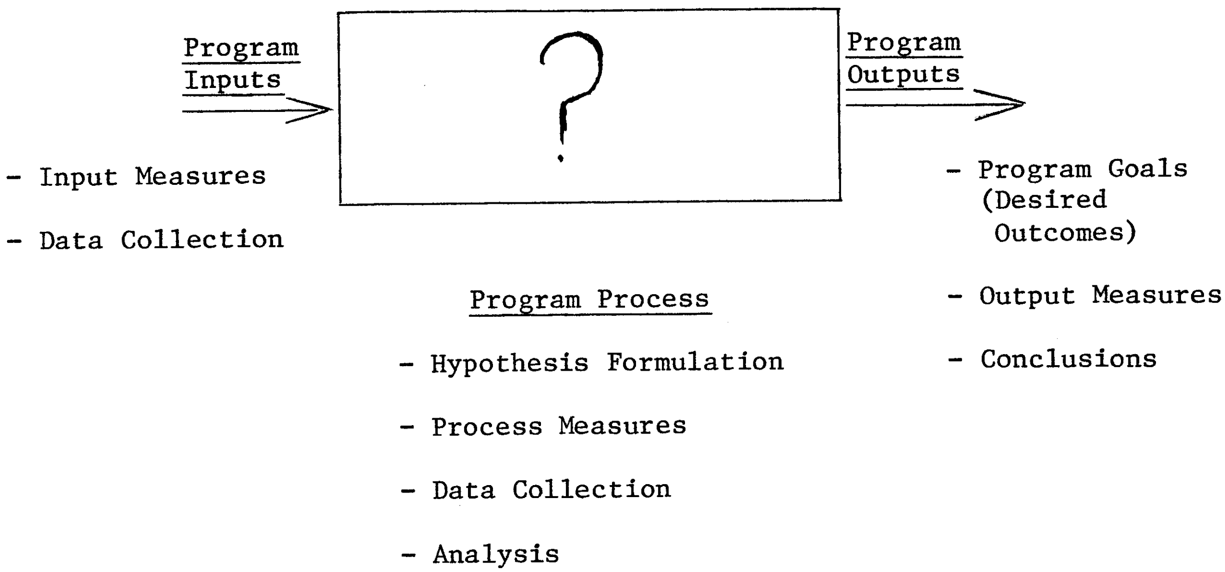
led to the specification of injuries per unit (as opposed to injuries per officer) as the key performance measure. This in turn led to the t-test which acted to support the hypothesis in question. Luckily, this example was easy to correct, and luckier still, the correct result does not differ sharply enough from the incorrect result to affect policy significance. However, it should be obvious that this will not always be the case.

ii. Selection of Performance Measures

Performance measures have always caused problems for evaluators. In the attempt to specify measures for program inputs, process and outcomes,

FIGURE 3

A GENERIC PROGRAM



it is easy to confuse the purpose of any given measure. For example, if the aim of a program is to increase police productivity in a given area, then response time is not a proper outcome measure to use. Rather, response time is a process measure (along with measures like patrol workload), which when combined with certain input measures (like officer salaries, number of available personnel) may be used to compute prespecified outcome measures (such as efficiency/effectiveness ratios). To evaluate such a program as if it were intended to minimize response time is conceptually erroneous.

iii. Data Collection

While errors committed during the manual recording of data may be seen as procedural, there are some conceptual problems that also arise. Foremost among these are problems associated with sampling. The means by which data are collected has direct implications towards the means by which data may be analyzed. For example, Larson (1975: 274) has noted that the use of participant-observers in high activity beats as data collectors for the KCPPE generated a highly nonrandom sample. To use standard statistical procedures on a sample such as this is simply not valid, as such procedures are predicated on the notion of random sampling. Interestingly enough, the data for Police Response Time: Its Determinants and Effects were generated by the observers Larson mentions in his article (Pate et al., 1976: 9).

iv. Analysis

This is the step which appears to be associated with the most serious conceptual errors. The two response time examples of last section are clear instances of this; they need not be repeated here.

What is symptomatic of conceptual errors in patrol evaluation is the reliance on a narrow range of analytical tools. Two of the most popular (and frequently misused) methods are conventional hypothesis testing and regression analysis. Due to their omnipresence in the evaluation literature, it is useful to speak briefly about each.

Hypothesis testing (e.g., z, t, F, X^2) is a useful technique if it is administered correctly. Unfortunately, within the context of program evaluation, the requisite environment for the proper application of these methods is often unachievable (as discussed at length in Chapter 1).

With reference to the relationship between program inputs and outputs, hypothesis testing provides a framework for the examination of one theory; this will be the level of outcome measures if the program has no effect. This framework is narrow indeed for a situation as complex as the evaluation of a social intervention program.

Regression analysis is perhaps the most abused technique of all. Due to its ability to fit multi-factor equations to data, many feel this method to be invaluable. Yet, the use of this technique highly restricts the evaluation to the assumption of intrinsically linear relationships between variables. As the measures associated with social programs exhibit high degrees of uncertainty, the use of such linear analysis to examine nonlinear stochastic relationships is too simplistic an approach.

Clearly, another approach is necessary. If such an approach could capture the major features of the mechanisms by which program inputs are translated into program outputs, then much of the conceptual confusion related to hypothesis formulation, selection of performance measures, data collection and analysis could be avoided.

IV. LINKING THEORY TO PROCEDURE VIA MODEL-BASED EVALUATION

In Chapter 1, we outlined a theory of model-based evaluation. It is evident that process-oriented models could have aided the evaluations cited in Section II. The two response-time studies would particularly have benefited from the use of modeling procedures.

Having probed the field of evaluation methodology and proposed an alternative to current methods of analysis, it is time to attempt some applications of our ideas. The remaining two chapters of this thesis are both applications of modeling techniques. Chapter 3 presents a model-based evaluation of one- versus two-officer police patrol; it represents a modeling effort aimed at a specific topic. Our focus becomes more general in Chapter 4 where we develop a class of models for the evaluation of treatment-release corrections programs.

FOOTNOTES

[1]The study referred to is "An Empirical Study of Methods Used in Criminal Justice Evaluations," currently underway at MIT under the direction of Richard C. Larson.

[2]These figures are derived from the MIT study.

[3]Of course, the reverse situation could also be true--significant differences resulting from the aggregation scheme of Table II could well be masked by the use of the aggregation scheme of Table I.

[4]A modified five-point scale was also used, but the criticisms which follow apply equally to both scales.

REFERENCES

1. Boydstun, J.E., Sherry, M.E., and Moelter, N.P., (1977). Patrol Staffing in San Diego: One- or Two-Officer Units. Washington: Police Foundation.

2. Carter, C.L. and Kaplan, E.H., (1977). "Did Anything Happen in Kansas City?" Unpublished paper, Department of Urban Studies and Planning, MIT.

3. Freund, J.E. (1971). Mathematical Statistics. Englewood Cliffs. Prentice-Hall.

4. Kelling. G.L., Pate T., Dickman, D. and Brown, C.E. (1974). The Kansas City Preventive Patrol Experiment: A Technical Report. Washington: Police Foundation.

5. Larson, R.C. (1975). "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment." Journal of Criminal Justice, 3: 267-297.

6. Pate, T., Ferrara, A., Bowers, R., and Lorence, J. (1976). Police Response Time: Its Determinants and Effects. Washington: Police Foundation.

7. Response Time Analysis: Volume I, Methodology; Volume II, Analysis (1977). Kansas City: Kansas City Police Department.

CHAPTER 3

EVALUATING THE EFFECTIVENESS OF ONE- VERSUS TWO-OFFICER PATROL UNITS

I. INTRODUCTION

The issue of one- versus two-officer patrol units has been a subject of controversial debate among police researchers for over twenty years. While many decry the postulated merits of one-officer patrol, timely surveys indicate that an increasing number of U.S. metropolitan police departments are adopting this staffing strategy (Governmental Research Institute (1957); Boydstun *et al.*, 1977: A1-A8, Clawson and Tarr, 1977). Despite this major change in manpower allocation, the impact such a change would have on police patrol has not been widely researched. Most statements made on the subject have been speculative in nature; it is not known what the implications of using one- versus two-officer patrol units are, let alone which of the two methods of patrol should be adopted under different circumstances.

The little evaluative research that has been carried out in this area has been experimental in nature. While a carefully monitored and controlled social experiment can provide valuable information, such experiments are very difficult and expensive to perform in the field of police patrol. Also, such experimental data emphasizes only the presence or absence of correlations among variables; structural, indeed casual relationships between variables are not defined through such empirical reasoning.

On the other hand, model-based evaluations provide an attractive alternative to the approach of social experimentation. Through the construction of simple mathematical models, structural relationships between variables are explicitly defined. These relationships are invoked by the evaluator to determine numerical values for a wide range of performance measures relevant to the problem under consideration.

A model-based evaluation of one- versus two-officer patrol staffing would examine the operational aspects of the two staffing strategies. Such an evaluation attempts to formalize expectations of comparative strategic performance under varying circumstances. These expectations may then be used by police administrators to aid in their policy decisions with respect to one- versus two-officer patrol staffing issues.

This chapter presents a model-based evaluation of one- versus two-officer patrol. The pros and cons of the two strategies are discussed in the next section. Emerging from this discussion is a set of performance measures. These performance measures are modeled for both staffing strategies, allowing for a comparative analysis. Important results are summarized at the end of the paper, as are directions for further work.

II. ISSUES AND PERFORMANCE MEASURES

The pros and cons of each staffing strategy as debated in the literature are remarkably consistent (Chicago Police Department, 1963; Governmental Research Institute, 1957; Boydstun et al., 1977: F-8, A1-A8). Briefly, the arguments are as follows.

For a given manpower level, the use of one-officer units enables the fielding of twice the number of patrol cars as the use of two-officer units. This doubling of units allows for an increase in police visibility; time spent on preventive patrol, hence increasing the probability of detecting a crime in progress; average area covered by patrol; and a decrease in response time. To achieve this same level of patrol utilizing two-officer units would involve doubling the number of patrol officers, a very expensive proposition. Alternatively, for a given cost constraint, more units may be fielded using one- as opposed to two-officer staffing, again allowing for improved patrol performance. Thus the use of one-officer patrol units provides a more efficient, cost-effective service system than could be realized through the use of two-officer units.

On the other hand, two-officer cars are preferred to one-officer units for reasons of safety. The presence of a second officer provides a "built-in cover" for the first officer. Two-officer units do not take as long to service calls as one-officer units due to the additional manpower available on the scene. One two-officer unit is less expensive than two one-officer units, and since many calls cannot be handled by single one-officer units, the cost of servicing such calls is higher for one-officer staffing than for two-officer staffing. Thus, the use of two-officer patrol units provides a safer and better quality service than the use of one-officer units.

It is interesting to note the type of reasoning which is being employed in these arguments. For example, the arguments in favor of one-officer patrol are products of "linear" thinking (Chicago Police Department, 1963: 213).

> "It is obvious that patrol coverage can be twice as intensive with 2 one-man cars as it can be with 1 two-car."

As will be shown later, situations arise where patrol "intensity" due to two one-officer cars may be less than twice, twice, or greater than twice that of one two-officer car.

From the arguments stated above, one may elicit relevant performance measures by which to evaluate the effectiveness of one- versus two-officer units. Those measures chosen for comparative analysis include:

1. expected area covered by patrol
2. response time from the nearest vehicle to a randomly occuring incident
3. expected frequency of patrol
4. visibility of patrol
5. probability of intercepting a randomly occurring crime in progress
6. probability of officer injury
7. comparative costs.

Where appropriate, simple probabilistic models will be constructed in order to compare the two staffing strategies. In some instances, numerical examples using actual experimental data from the Police Foundation's recent study Patrol Staffing in San Diego: One- or Two-Officer Units will be presented to illustrate the models used.

III. COVERAGE

Upon contemplating a staffing change from two officers to one officer per unit, the police administrator might wish to consider expected benefits to be gained from an increase in patrol coverage. As mentioned in the last section, supporters of one-officer patrol have expressed linear expectations of increases in coverage; that is, twice the number of patrol cars will double the expected area covered. The model presented in this section demonstrates that, assuming full availabilities of patrol units, a doubling of the number of patrol cars in an area increases the expected area covered by less than twice; we will examine cases where units are unavailable for service later on in this chapter.

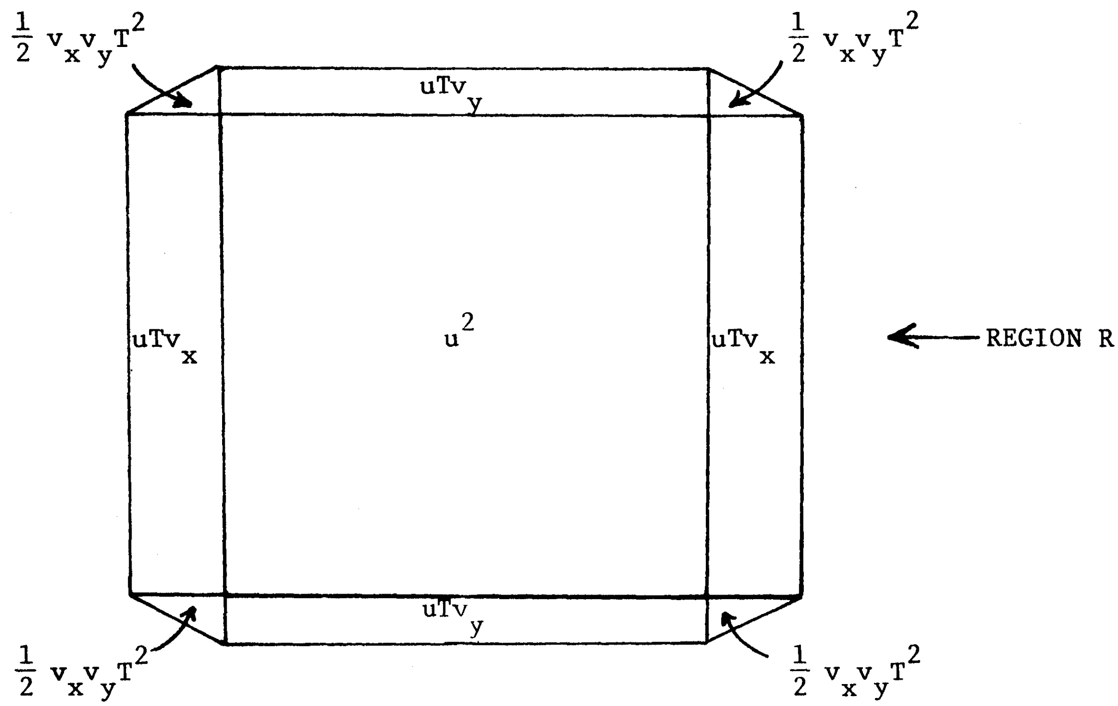
Consider the square beat with area u^2 shown in Figure 1. A point in this beat is said to be covered if a police car is within T time units of that point (Larson, 1972: 263). Thus, if T=2 minutes, a point Q is covered if an available police car is no further than 2 minutes travel time form Q. The expected (or average) area covered is simply the average area formed by collecting all covered points. Here, it is assumed that travel time can be closely approximated by the right angle metric

$$t_r = \frac{d_x}{v_x} + \frac{d_y}{v_y} \qquad (1)$$

where t_r = travel time between patrol car and the point

$d_x(d_y)$ = East-West (North-South) distances between the patrol car and the point

$v_x(v_y)$ = East-West (North-South) speed of travel.



$$\text{AREA (R)} = u^2 + 2uT(v_x + v_y) + 2v_x v_y T^2$$

assume $v_x > v_y$

Figure 1

Geometry of the Coverage Model

To eliminate boundary effects, it is assumed that patrol cars are uniformly and independently distributed over the entire region R shown in Figure 1. Under these conditions, a simple result from geometrical probability shows that the probability that any point in the beat is covered given that there are N patrol cars uniformly and independently distributed over R is

$$P_{C|N} \equiv \Pr \text{ (point is covered } | \text{ N cars in R)}$$

$$= 1 - \left[\frac{u^2 + 2uT(v_x + v_y)}{u^2 + 2uT(v_x + v_y) + 2v_x v_y T^2} \right]^N \quad (2)$$

and that the expected area covered in the beat is given by

$$E \text{ (area covered)} = P_{C|N} \cdot u^2 \quad (3)$$

Figure 2 shows a graph of the expected area covered versus N, the number of units in the beat, for the special case u = 1 mile, $v_x = v_y = 20$ m.p.h., and T = 3 minutes. Note that the expected area covered does not increase linearly with N; rather the additional expected area covered is marginally decreasing with N. In this example, the increase in expected area covered due to a switch from two-officer patrol unit staffing (N = 1) to one-officer patrol unit staffing (N = 2) is on the order of 71%.

Although the expected area covered is not doubled from its previous value due to the presence of a second unit, the increase in expected area covered is substantial. In general, such increases will be more significant to the police administrator when considering large beats rather than small beats, when working with low values of v_x and v_y as opposed to high values, and when coverage is defined for small values of T rather

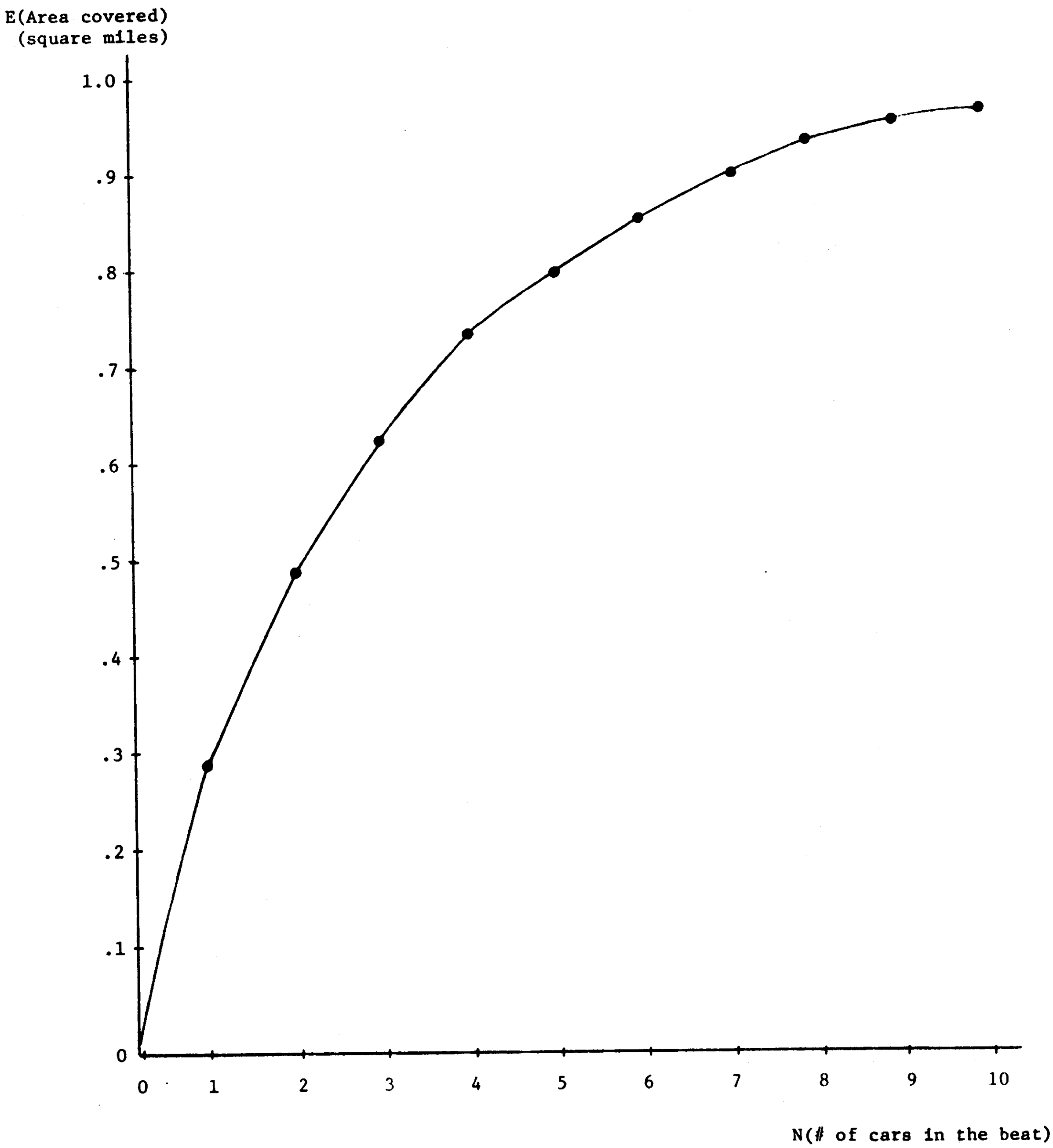


Figure 2

Expected Area Covered is Marginally Decreasing with N
(A smooth curve has been placed through these points to aid the reader.)

than large values of T. In any event, the percentage increase in expected area covered due to patrol by two cars as opposed to one car per beat will always equal $100 \cdot (1 - P_{C|1})$ under the conditions of the model presented here.

IV. RESPONSE TIME

Mean response time is an important variable which pertains to the quality of police service in an emergency setting. One of the postulated benefits of one-officer unit staffing is that police response time may be considerably reduced. This assertion is now made the subject of analytical investigation.

For simplicity, consider the square beat of area u^2 shown in Figure 3. It is assumed that travel speed is held constant at v throughout the beat; that is, $v_x = v_y = v$. Also, travel time between patrol units and incidents is assumed to be given by the right angle metric stated earlier. Patrol units and incident locations are distributed uniformly and independently over the beat.

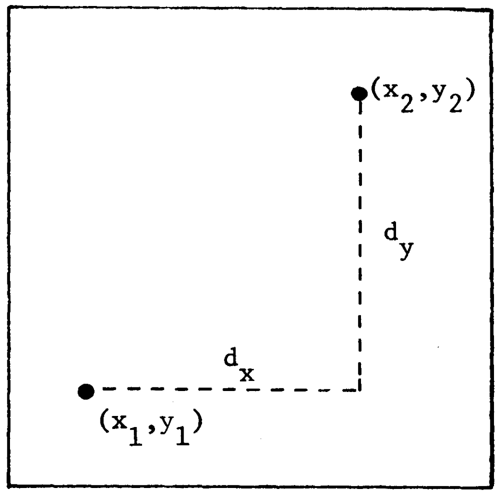
Initially, we may consider again the case where all units are available. Consider the case of two-officer staffing (i.e., one car in the beat). Under the conditions of this model, it is well known that the expected travel time from the patrol unit to the incident is (Larson, 1972: 79)

$$E(t_r) = \frac{2u}{3v} \approx .67 \frac{u}{v} \quad . \tag{4}$$

Now consider the case of one-officer staffing (i.e., two cars in the beat). If we assume that the responding unit is the nearest unit to the scene of the incident, we show in the Appendix that the expected travel time is:

$$E(t_r^*) = \frac{1354}{2835} \frac{u}{v} \approx .48 \frac{u}{v} \ . \tag{5}$$

The reduction in expected travel time gained by switching from 2-officer units (1 car per beat) to 1-officer units (2 cars per beat is only around 28%.



$$t_r = \frac{d_x + d_y}{v}$$

$$= \frac{|x_1 - x_2| + |y_1 - y_2|}{v}$$

Figure 3

Illustration of the Right-Angle Metric in a Square Beat

The analysis performed so far has not considered the case where patrol units might not be able to respond due to possible commitments to other incidents. This case is now examined within the context of a spatially distributed queueing model. For simplicity, we will only consider the case where exactly one car is dispatched to an incident.

Consider the square city shown in Figure 4. The city is divided into four beats; each beat is a square with area u^2. Calls for police service arrive in a Poisson manner at a rate λ_c per hour for the entire city; each beat receives on average $\lambda_c/4$ calls per hour. Assume that the time to service a call is an exponentially distributed random variable with mean service time (including travel time) equal to $\frac{1}{\mu}$. Travel time is closely approximated by the right angle metric. People in this city are patient in that they will wait any length of time until their call is eventually serviced; thus no calls are lost.

Now imagine that there is one two-officer unit patrolling each of the four beats, thus there are four police cars at work. Each car is uniformly located over its home beat, and all units are located independently of one another. Consider the following four mutually exclusive events conditioned on beat 1:

$A_{1|1} \Leftrightarrow$ a call originates in beat 1 and the unit patrolling beat 1 is available;

$A_{2|1} \Leftrightarrow$ a call originates in beat 1, the unit patrolling beat 1 is not available, but either or both of the units patrolling beats 2 and 3 are available;

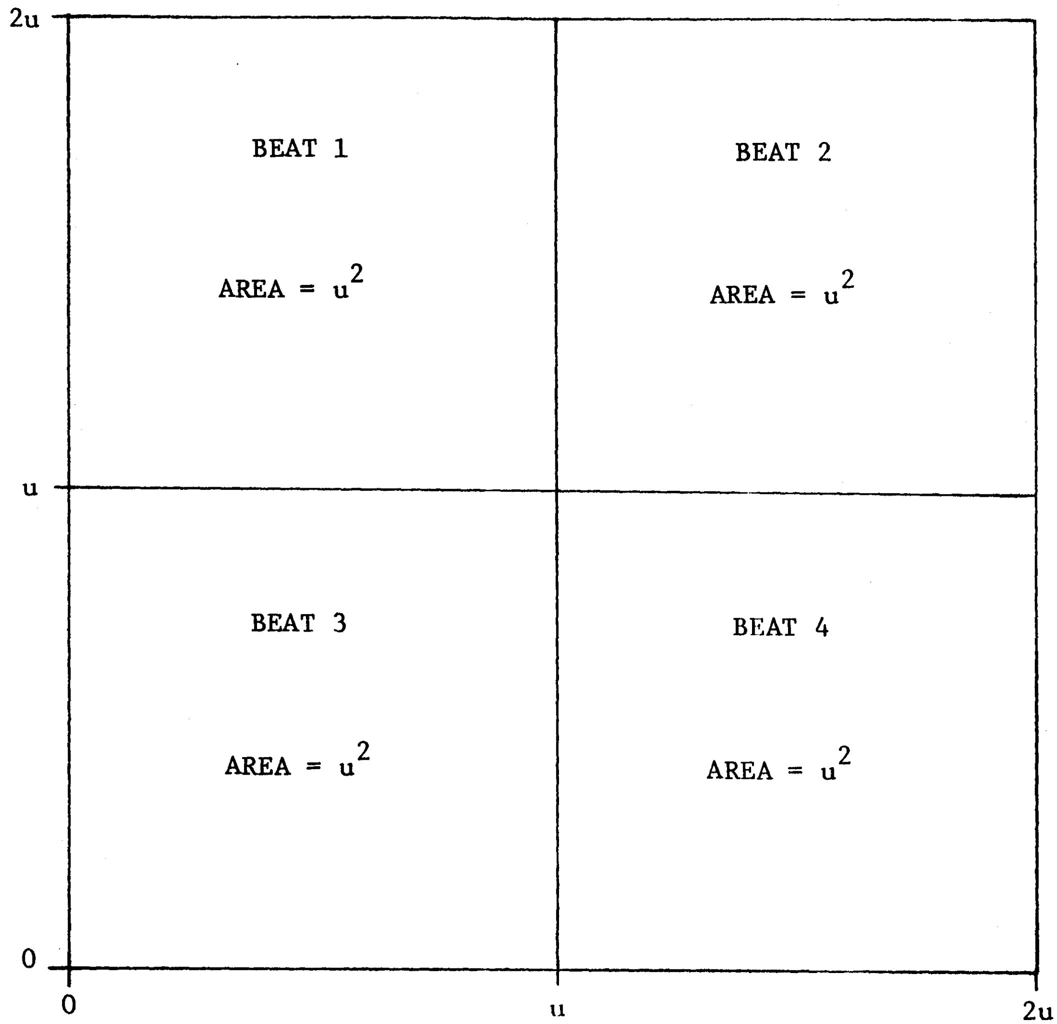


Figure 4

The Four Beat Model

$A_{3|1}$ <=> a call originates in beat 1, neither of the units patrolling beats 1, 2 or 3 are available, but the unit patrolling beat 4 is available;

$A_{4|1}$ <=> a call originates in beat 1, no units are available, hence the call is placed in queue where it will be serviced as soon as a unit becomes available. (We assume that the queue discipline is First-Come, First-Serve.)

In the case of $A_{1|1}$, the beat unit responds and the expected response time is given by $\frac{2u}{3v}$. In the case of $A_{2|1}$, one unit from an adjacent beat (2 or 3) responds, and it is easy to show that the expected response time in this case is $\frac{4u}{3v}$. Once the service has been completed, it is assumed that the responding unit returns to its home beat before accepting another call for service. $A_{3|1}$ implies that the beat 4 unit responds, and it is again easy to show that the expected travel time in this case is equal to $\frac{2u}{v}$. Analogous to $A_{2|1}$, it is assumed that the beat 4 unit returns to its home beat upon service completion before accepting a new call for service. $A_{4|1}$ implies a queue. In such an event, the time it will take for a vehicle to respond is equal to the time spent waiting for a unit to become available, plus the travel time to the incident. It is easy to show that the expected travel time to the incident for this case equals $\frac{4u}{3v}$; we will determine the average queueing delay later on.

Now the unconditioning of these events from beat 1 is remarkably easy. First, since the beat layout is perfectly symmetric, we could define the events discussed conditional on each of beats 2, 3 and 4 in a manner analogous to beat 1. Second, since the probability that a call originates in any one of the 4 beats is equal to $\frac{1}{4}$, it is clearly true that

$$\Pr(A_i) = \Pr(A_{i|1}) \quad i = 1, 2, 3, 4$$

where A_1 <=> unit in beat where call for service originated is available for service;

A_2 <=> unit in beat where call for service originated is not available for service, but a unit in an adjacent beat is available;

A_3 <=> the only unit available for service is located in the beat diagonal to the beat where the call for service originated;

A_4 <=> no units are available.

The point of all this is that once $\Pr\{A_i\}$, $i = 1,\ldots,4$ are known, then the expected travel time is:

$$E(t_r) = \sum_{i=1}^{4} E(t_r|A_i) \cdot \Pr\{A_i\} \tag{6}$$

$$= \frac{2u}{3v} \Pr\{A_1\} + \frac{4u}{3v} \Pr\{A_2\} + \frac{2u}{v} \Pr\{A_3\} + \frac{4u}{3v} \Pr\{A_4\},$$

and the expected response time is given by

$$E(RT) = E(t_r) + W_q \tag{7}$$

where W_q is the average queueing delay incurred due to unavailable units. In the appendix, we determine expressions for $\Pr\{A_i\}$ $i = 1,\ldots,4$ and W_q which enables us to use equation (7) for values of λ_c and μ.

It is straightforward to extend this model to the case of one-officer staffing (2 units per beat). We assume for simplicity that exactly one unit is dispatched to any call. The following events may be defined:

B_1 <=> both units in beat where call for service originated are available for service;

B_2 <=> only one unit in beat where call for service originated is available for service;

B_3 <=> both units in beat where call for service originated are busy, but a unit in an adjacent beat is available;

B_4 <=> the only unit(s) available for service is located in the beat diagonal to the beat where the call for service originated;

B_5 <=> no units are available, hence the call is placed in queue where it will be serviced as soon as a unit becomes available.

In the case of B_1, it is assumed that the nearest response unit answers the call, hence the expected travel time is given by $\frac{1354u}{2835v}$. The conditional expected travel time given the other events are straightforward. The expected travel time for the case of two cars per beat is given by

$$E(t_r^*) = \sum_{i=1}^{5} E(t_r^* | B_i) \Pr\{B_i\}, \qquad (8)$$

and the expected response time is given by

$$E(RT^*) = E(t_r^*) + W_q^* \qquad (9)$$

where W_q^* is the expected queueing delay. Expressions for $\Pr\{B_i\}$ and W_q^* are derived in the Appendix.

In Figure 5, $E(t_r)$, $E(t_r^*)$ and W_q are plotted as functions of $\rho = \frac{\lambda_c}{\mu}$; W_q^* is effectively equal to zero for this example, though exact calculations may be made using equation (A25) from the Appendix. If it takes

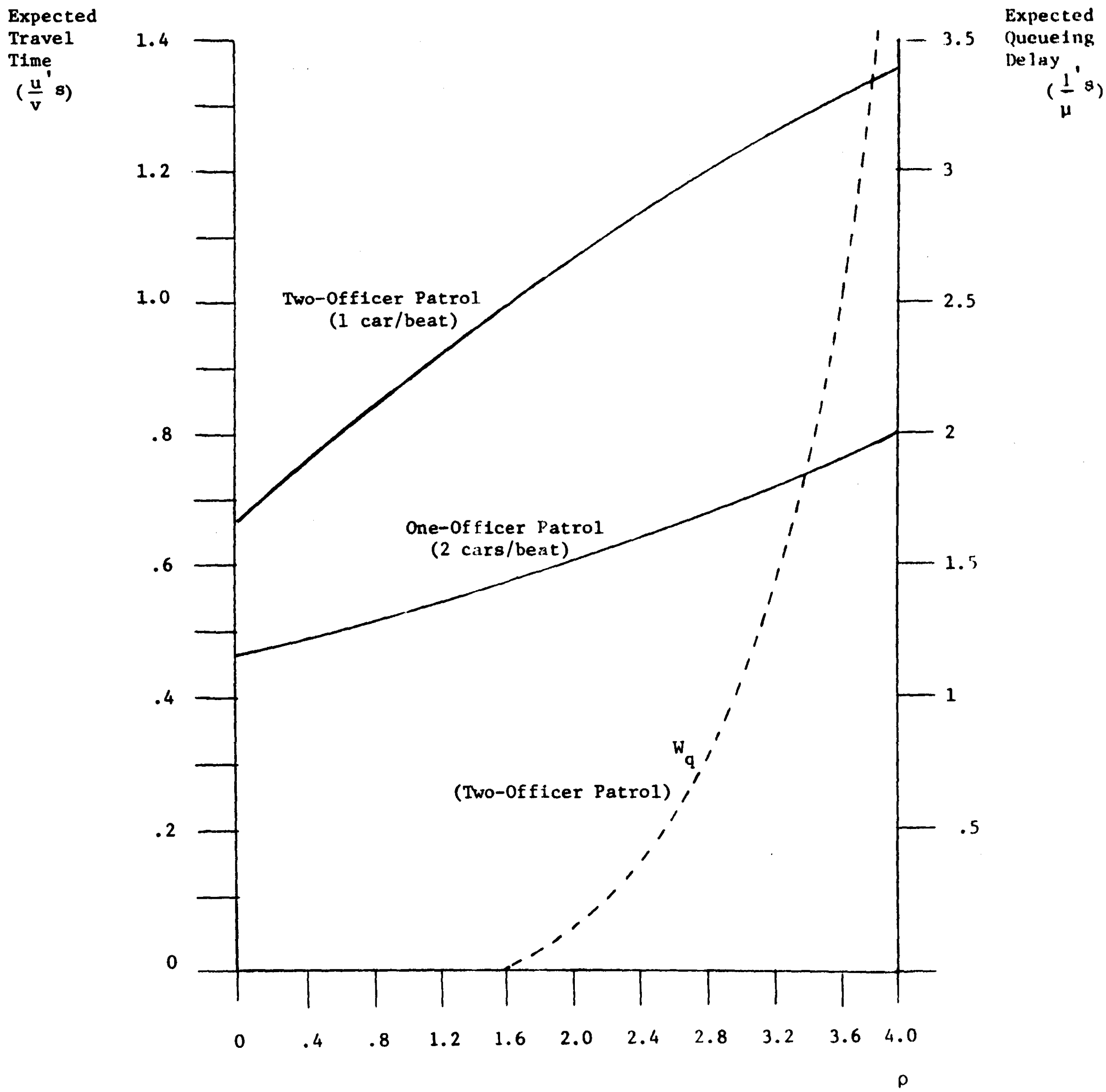


Figure 5

Response Time Statistics for the Four-Beat Model the same amount of time for a one-officer unit as a two-officer unit to service a call, then the expected travel times resulting from the two staffing strategies may be identified by choosing the appropriate value of ρ and reading the values for $E(t_r)$ and $E(t_r^*)$ off the graph. $E(RT)$ may be found by reading the value for W_q off of the graph, multiplying it by $1/\mu$, and adding it to $E(t_r)$; $E(RT^*) \approx E(t_r^*)$ since W_q is effectively zero, though exact values for $E(RT^*)$ may be found using equation (A25) from the Appendix. Note that the percentage difference in expected travel time does not exceed 40% when service times are equal. If the service times are not equivalent for one-officer and two-officer units, then one needs to consider ρ(1 car, 2 officers) and ρ(2 cars, 1 officer) separately in order to determine $E(t_r)$ and $E(t_r^*)$ before a comparison can be made. However, the percentage difference is expected <u>response time</u> can be enormous, as this depends on the magnitude of $1/\mu$. Hence, from the standpoint of response time the major advantage of fielding two one-officer units as opposed to one two-officer unit is that the expected amount of time waiting for a unit to become available is drastically reduced.

<u>Example</u>: Patrol Staffing in San Diego

From a crude map study, it has been determined that the average beat size in the San Diego experiment lies somewhere between four and six square miles, so for the sake of discussion, u^2 is set equal to 5 with $u = \sqrt{5} \approx 2.24$ miles. If we assume an average travel speed of 15 miles per hour, then the expected travel time to a random incident in a one-car beat given that the randomly located unit is available is given by

$$T_r = \frac{2 \cdot 2.24}{3 \cdot 15} \approx 6 \text{ minutes.}$$

If in a 2-car beat, both units are available, then the expected travel time from the nearest unit to the incident is

$$T_r^* = \frac{1354}{2835} \cdot \frac{2.24}{15} \approx 4.3 \text{ minutes.}$$

The response time reported for two-officer units in the experiment is 7.5 minutes (Boydstun et al., 1977: 50), which is within 25% of the modeled value for T_r. Since the San Diego experiment did not double the number of units in beats where one-officer cars were used, it is hard to compare T_r^* to an empirical value. However, the response time for two one-officer units responding to the same call was recorded, the value reported being the maximum of the two individual response times. The figure given is 5.9 minutes which is not that much larger than the modeled value for T_r^*. However, the maximum of the two response times should be larger than T_r^*; it should also be larger than T_r. It is indeed puzzling (as admitted by the San Diego researchers) why this reported maximum value is so low.

Thus far, units have been assumed to be available. Within the context of the four-beat queueing model, the effects of busy units may be taken into account using San Diego data. During the course of the experiment, two-officer units received an average of 6.68 calls per 8-hour tour as compared with 6.28 calls per 8-hour tour for one-officer units (Boydstun et al., 1977: 23). On the average then, there were 6.48 calls per 8-hour tour per car. For the four-beat model, λ_c is calculated as

$$\lambda_c = \frac{6.48 \text{ calls/car} \cdot 4 \text{ cars}}{8 \text{ hours}} = 3.24 \text{ calls/hour.}$$

Two-officer units spent an average of 37.3 minutes/call, while one-officer units spent an average of 48.8 minutes/call. Thus:

$$\rho_{2 \text{ officers}} = \frac{\lambda_c}{\mu_{2 \text{ officers}}} = 3.24 \text{ calls/hour} \cdot 37.3 \frac{\text{min.}}{\text{call}}$$

$$\cdot \frac{1 \text{ hr.}}{60 \text{ min.}} = 2.02.$$

$$\rho_{1 \text{ officer}} = \frac{\lambda_c}{\mu_{1 \text{ officer}}} = 3.24 \text{ calls/hour} \cdot 48.8 \frac{\text{min.}}{\text{call}}$$

$$\cdot \frac{1 \text{ hr.}}{60 \text{ min.}} = 2.63.$$

From Figure 5, and equations (A13), (A16), A23) and (A25) in the Appendix, we see that

$$E(t_r) = 1.029 \frac{(2.24)}{15} 60 = 9.22 \text{ minutes;}$$

$$E(t^*_r) = .661 \frac{(2.24)}{15} 60 = 5.92 \text{ minutes;}$$

$$Pr\{A_4\} \approx .18;$$

$$Pr\{B_5\} \approx .01;$$

$$E(RT) = 9.22 + 37.3 \cdot .087 = 12.47 \text{ minutes;}$$

$$E(RT^*) = 5.92 + 48.8 \cdot .0012 = 5.98 \text{ minutes.}$$

The decrease in response time due to the presence of an additional unit is about 52%. Perhaps more significantly, the probability of a call not being serviced immediately due to the presence of a queue has been reduced by 94%.

From this analysis, it appears that while actual travel times are not reduced that substantially, the probability of having to wait for a free unit is reduced considerably by raising the number of units per beat from one to two. In this San Diego example, the above statement holds even though it does take one-officer units around 1.3 times longer than two-officer units to service a call. Thus, it is better from a queueing standpoint to have many slow servers than a few fast servers.

V. PATROL FREQUENCY

Another of the postulated benefits that one-officer staffing is supposed to deliver is an increase in the level of preventive patrol. Preventive patrol is felt to be an important police function for a number of reasons. In particular, police visibility is felt to be directly proportional to the time spent on prevent patrol. Preventive patrol is also intended to provide the threat of interception and apprehension to the criminal. The next three sections of this chapter concentrate on issues of preventive patrol.

Before one can study police visibility and interception probabilities, one must establish the stochastic process by which patrol takes place. For this reason, we now direct our attention to patrol frequency.

Larson (1972: 135-142) has developed a patrol frequency model, a simplified version of which will be reported here. Let

D = number of street miles in a beat,

s = average patrol speed.

Suppose that a single two-officer unit is allocated in a uniform manner over the streets of the beat. Under the condition that the unit is free, the mean rate of patrol (in passes/hour, say) will be equal to

$$\lambda_p = s/D \qquad (10)$$

If the Poisson assumptions are made, then

$$p_n(t) = \frac{(\lambda_p t)^n e^{-\lambda_p t}}{n!} \qquad \begin{array}{l} n = 0, 1, 2, \ldots \\ t > 0 \end{array}$$

where $P_n(t) \equiv$ probability that a random point has been passed n times in a period of length t.

The time between patrol passings at a random point is an exponentially distributed random variable with parameter λ_p; i.e.,

$$f_t(t_0) = \lambda_p e^{-\lambda_p t_0} \qquad t_0 > 0 \tag{11}$$

where t_0 = time between patrol passings. The expected time between passings is simply

$$1/\lambda_p = D/s.$$

Now, consider the case where the unit is busy a fraction b of the time. Under this condition, the mean rate of patrol is given by

$$\tilde{\lambda}_{2\ \text{officers}} = \lambda_p\ (1-b), \tag{12}$$

and the time between patrols at a random point is assumed to be exponentially distributed but with parameter $\tilde{\lambda}_{2\ \text{officers}}$.

As usual, the effect of one-officer patrol is introduced via a doubling of the number of patrol units. If the service times for one-officer and two-officer units are identical, then on average, either of the two one-officer units will be busy a fraction $\frac{b}{2}$ of the time due to a halving of the call for service workload per unit. Thus, the average patrol frequency given that there are now two one-officer units in the beat is given by

$$\tilde{\lambda}_{1\ \text{officer}} = 2\lambda_p(1 - b/2), \tag{13}$$

and the pdf for the time between patrol passings at a random point is again assumed to be an exponential distribution with parameter $\tilde{\lambda}_{1\ \text{officer}}$.

To see the increase in patrol frequency due to the use of 2 one-officer units as opposed to one two-officer unit, look at the ratio

$$\frac{\tilde{\lambda}_{1\ \text{officer}}}{\tilde{\lambda}_{2\ \text{officers}}} = \frac{2\lambda_p(1 - b/2)}{\lambda_p(1 - b)} \tag{14}$$

$$= \frac{2\lambda_p - b\lambda_p}{\lambda_p - b\lambda_p} \geq 2 \ . \quad 0 \leq b < 1$$

Thus, the increase in patrol frequency is <u>at least</u> 100% when changing from 1 two-officer unit to 2 one-officer units if service times are the same for one- and two-officer units.

For the case where service times are not equal for both staffing strategies, $\tilde{\lambda}_{1\ \text{officer}}$ must be redefined as follows:

$$\tilde{\lambda}_{1\ \text{officer}} = 2\lambda_p\ (1 - b_1/2) \tag{15}$$

where

$$b_1 = \text{avg. \# calls/service/hr.} \cdot \begin{pmatrix} \text{service time (hrs.)/call} \\ \text{for a one-beat unit} \end{pmatrix}$$

$$0 \leq b_1 < 1.$$

The ratio

$$\frac{\tilde{\lambda}_{1\ \text{officer}}}{\tilde{\lambda}_{2\ \text{officers}}} = \frac{2\lambda_p(1 - b_{1/2})}{\lambda_p(1 - b)} \tag{16}$$

is: greater than 2 if $b_1 < 2b$,

equal to 2 if $b_1 = 2b$,

less than 2 if $b_1 > 2b$.

Since b_1 is nearly always less than 2b, the switch to one-officer staffing is likely to imply increasing returns to scale with respect to patrol frequency.

Example: Patrol Staffing in San Diego

In San Diego, as previously mentioned, two-officer units averaged a 37.3 minutes/call service time, while one-officer units averaged a 48.8 minutes/call service time. On average, units received $\frac{6.48}{8}$ = .81 calls/hour. Thus, two-officer units were busy on average

$$b = .81 \frac{37.3}{60} \approx .50 \text{ of the time,}$$

while one-officer units were busy on average

$$b_1 = .81 \frac{48.8}{60} \approx .66 \text{ of the time.}$$

From the experimental report, it may be derived that patrol units drove an average of 60 miles per 8-hour tour which implies that s = 7.5 m.p.h.[1] Map analysis has yielded a rough estimate of the number of street miles in an average beat to be in the neighborhood of 30 miles. Thus,

$$\lambda_p = s/D = 7.5/30 = .25 \text{ patrols/hour;}$$

$$\tilde{\lambda}_{2 \text{ officers}} = \lambda_p(1 - b) = .25(1 - .5) = .125 \text{ patrols/hour;}$$

and the expected time between two-officer patrol passings of a random point is

$$1/\tilde{\lambda}_{2 \text{ officers}} = \frac{1}{.125} = 8 \text{ hours between patrols.}$$

Also,

$$\tilde{\lambda}_{1\ \text{officer}} = 2\lambda_p(1 - b_{1/2}) = 2(.25)(1 - .66/_2) = .335 \text{ patrols per hour;}$$

and the expected time between patrol passings of a random point using two one-officer units is

$$1/\tilde{\lambda}_{1\ \text{officer}} \approx 3 \text{ hours between patrols.}$$

The ratio of mean one-officer to two-officer patrol rates is

$$\frac{\tilde{\lambda}_{1\ \text{officer}}}{\tilde{\lambda}_{2\ \text{officer}}} = \frac{.335}{.125} = 2.68\ .$$

Of course, we knew from equation (16) that this ratio would be greater than 2 since

$$.66 = b_1 < 1.0 = 2b.$$

VI. VISIBILITY

The notion that an increase in the level of patrol via a switch from two-officer to one-officer patrol staffing will substantially increase police visibility is questionable. Although we have shown that patrol frequencies will increase by a factor greater than two in most cases, the verm "visibility" requests the participation of "viewers." People do not, by and large, spend a large fraction of their day observing the street. Thus, although percentage increases in visibility may be substantial, absolute increases in visibility may be too small to be perceived by the public.

Consider a person located at a random point in a beat. Imagine that this person observes the street, say, 5% of the time (i.e., 72 minutes per day). How often will our $\frac{1}{20}$th time street watcher see a police car? On the average, he will observe

$$S = .05\tilde{\lambda} \text{ sightings/hour.} \tag{17}$$

For example, in San Diego where $\tilde{\lambda}_{2 \text{ officers}} = .125$,

$$S = .05 \cdot .125 = .00625 \text{ sightings per hour}$$

$$= 1.05 \text{ sightings per week.}$$

On average, our street watcher will see a patrol car once every 160 hours.

If we now introduce two one-officer units, our San Diego example would have $\tilde{\lambda}_{1 \text{ officer}} = .335$. This implies

$$S = .05 \cdot .335 = .01675 \text{ sightings per hour}$$

$$= 2.8 \text{ sightings per week.}$$

On average, the street watcher would see a car about once every 60 hours.

If one- and two-officer units had equal service times, then using two-officer service times as a base, $\tilde{\lambda}_{1 \text{ officer}}$ would be reset to $2\lambda_p(1 - b/2) = 2(.25)(.75) = .379$. This would put

$$S = .05 \cdot .375 = .01875 \text{ sightings per hour}$$

$$= 3.2 \text{ sightings per week,}$$

or one sighting every 53.3 hours on average.

The percentage changes in these numbers are substantial; the numerical differences may in fact be statistically significant (Carter and Kaplan, 1977: 14-15). However, the magnitudes of these figures are too small to make much of a difference. People are not likely to notice changes in visibility due to changes in patrol effort of the type mentioned here.

Larson (1975) has suggested a similar consideration with respect to the Kansas City Preventive Patrol Experiment. In this study, patrol levels were supposedly tripled in some areas, yet the change in the level of patrol was difficult to perceive.

However, this need not always be the case. If the initial level of patrol is high (for example, $\tilde{\lambda}_p \gtrsim .8$ patrols per hour), then the percentage increases due to a switch from two-officer to one-officer staffing may become perceived increases as well.

For example, if $\tilde{\lambda}_{2 \text{ officers}} = 1$ patrol per hour, and $\tilde{\lambda}_{1 \text{ officer}} =$ 2.2 patrols per hour, this would imply (using $\frac{1}{20}$th time street watcher).

S = .05 sightings per hour = 8.4 sightings per week for two-officer cars, and

S = .05 • 2.2 = .11 sightings per hour = 18.5 sightings per week for one-officer units.

In this example, the percentage increase is less than in our previous examples. However, one might expect that the absolute increase would be a noticeable one.

Thus, perceived increases in police "visibility" due to the adoption of one-officer patrol units are only likely to set in after some initial, "critical" level of one-unit patrol has been reached. For cities with patrol frequencies such as those found in San Diego or Kansas City, the increase in perceived patrol visibility will likely occur only on paper.

VII. CRIME INTERCEPTION

One of the postulated major functions of preventive patrol is that of intercepting crimes in progress. It is argued that the use of two one-officer units will provide for a higher probability of interception than the use of one two-officer unit. This assertion will now be investigated in some detail.

Suppose that patrol is occurring according to the Poisson process described in Section V. Assume that crime is also uniformly and independently distributed over the beat in question; that is to say, a crime is equally likely to occur at any particular location in the beat independent of the locations of other crimes. A crime is said to be intercepted if the point where the crime is occurring is passed by a patrol car while the crime is still in progress. Note that there is no attempt being made to distinguish between apprehension and interception; equivalently, the tacit assumption being made is

$$\Pr(\text{apprehension} \mid \text{interception}) = 1.0$$

Now the question becomes one of determining the probability of intercepting a crime of duration τ. This is a problem of random incidence. The probability of intercepting a crime of duration τ is simply the probability that a patrol car arrives at the point where the crime is being committed before the crime has been completed, given that the crime started at some random time.

For Poisson patrol, the pdf for the time between patrol passings at at random point is the exponential distribution with parameter $\tilde{\lambda}$

$$f_t(t_0) = \tilde{\lambda} e^{-\tilde{\lambda} t_0} \qquad t_0 > 0. \tag{18}$$

The cdf for time between patrol passings at random point is given by

$$F_t(t_0) = 1 - e^{-\tilde{\lambda} t_0} \qquad t_0 > 0 \tag{19}$$

If we define z as the time from the beginning of a randomly occurring crime (in time and space) until a patrol car arrives at the same random point, then using random incidence arguments (Drake, 1967: 149-153), the probability of intercepting a crime of duration τ is given by

$$\begin{aligned} \psi(\tau) &\equiv \Pr \{\text{intercepting a crime of duration } \tau\} \\ &= \Pr \{z < \tau\} \\ &= \int_0^{\tau} \frac{1 - F_t(z)}{E(t)} \, dz \\ &= 1 - e^{-\lambda\tau} \cdot \tau \text{ is deterministic.} \end{aligned} \tag{20}$$

Models of this sort (e.g., Elliott (1968), Larson (1972)) where crime duration is treated deterministically are reviewed by Riccio (1974).

However, it is more realistic to treat crime as a random variable. As a first cut, it seems reasonable to hypothesize that there will be relatively many crimes of short duration and few crimes of long duration (e.g., purse snatches vs. bank robberies). Hence, it is assumed that crime duration is exponentially distributed with parameter γ, i.e.,

$$f_\tau(\tau_0) = \gamma e^{-\gamma\tau_0} \qquad \tau_0 > 0, \tag{21}$$

and

$$E(\tau) = \frac{1}{\gamma} .$$

Hence, if $\gamma = 20$ crimes per hour, then the expected length of a crime would be 3 minutes.[2]

Of course, if τ is a random variable, then the probability of intercepting a crime of duration τ, $\psi(\tau)$, is itself a random variable. Noting that $\psi(\tau)$ is monotonically increasing with τ, it is straightforward to obtain the pdf for ψ (Freund, 1971: 119-126):

$$f_\psi(p) = f_\tau(\psi^{-1}(p)) \cdot \frac{1}{\frac{d\psi}{d\tau}} \Big|_{\tau = \psi^{-1}(p)}$$

$$= f_\tau(-\ell n\,(1 - p)/\tilde{\lambda}) \cdot \frac{1}{\tilde{\lambda} e^{-\tilde{\lambda}(-\ell n(1 - p)/\lambda)}}$$

$$= \omega(1 - p)^{\omega-1} \quad 0 \leq p \leq 1 \text{ where } \omega \equiv \frac{\gamma}{\tilde{\lambda}}, \ \omega > 0 \ . \tag{22}$$

This new result is remarkable in that the derived $f_\psi(p)$ is a beta distribution (Freund, 1971; 114), with $\alpha = 1$ and $\beta = \omega$. The expected value is given by

$$E(\psi) = \frac{1}{1 + \omega} = \frac{\tilde{\lambda}}{\tilde{\lambda} + \gamma} \ . \tag{23}$$

Before proceeding to compare one-officer versus two-officer staffing using this result, an important sampling property of this distribution should be mentioned. If the police administrator is not sure of the value of γ ($\tilde{\lambda}$ may be calculated using the models presented earlier), he will not be able to state outright what the expected intercept probability is. However, if he assumes a value for γ which seems reasonable to him, then his estimates for $E(\psi)$ may be updated in the following manner.

Suppose after choosing a value for γ, the administrator discovers that of N potentially "observable" crimes committed in a given time period, n crimes were actually intercepted. Under the crime independence assumption postulated earlier, the above information corresponds to n "successes" in N independent Bernoulli trials. It is a property of beta distributions that the pdf for a beta distributed random variable conditioned on Bernoulli events is also a beta distribution. In the case presented here;

$$f_{\psi|N,n}(p|N,n) = \frac{\Gamma(N+\omega+1)}{\Gamma(n+1)\ \Gamma(N-n+\omega)}\ p^{n}(1-p)^{N-n+\omega-1} \quad (24)$$

$0 \leq p \leq 1,$

$n \leq N$, n & N interger.

and

$$E(\psi|N,n) = \frac{n+1}{N+\omega+1} \quad . \quad (25)$$

Since N and n may be treated as cumulative figures, the police administrator's estimates of the probability of intercepting a randomly occurring crime may be periodically updated.

To compare $E(\psi|1$ 2-officer unit) to $E(\psi|2$ 1-officer units), it is obvious that changes are going to occur due to different values of $\tilde{\lambda}$. As in Section V, the cases where service times are equal and not equal for one- and two-officer units will be considered separately.

For the case where service times are equal, from Section V we have

$$E(\psi | 1 \text{ 2-officer car}) = \frac{\tilde{\lambda}_{2 \text{ officers}}}{\tilde{\lambda}_{2 \text{ officers}} + \gamma}$$

$$= \frac{\lambda_p (1 - b)}{\lambda_p (1 - b) + \gamma}, \quad 0 \leq b < 1, \tag{26}$$

where b is the fraction of time spent out of service by a single two-officer unit. Also,

$$E(\psi | 1 \text{ 2-officer car}) = \frac{\tilde{\lambda}_{1 \text{ officer}}}{\tilde{\lambda}_{1 \text{ officer}} + \gamma}$$

$$= \frac{2\lambda_p (1 - b/2)}{2\lambda_p (1 - b/2) + \gamma}, \quad 0 \leq b < 1. \tag{27}$$

The percentage increase in expected probability of interception due to using two one-officer cars depends on b. In particular, there is a critical value b* such that

% increase in $E(\psi)$ > 100% if b > b*, (increasing returns to scale)

% increase in $E(\psi)$ = 100% if b = b*, (constant returns to scale)

% increase in $E(\psi)$ < 100% if b < b*. (decreasing returns to scale)

It is easy to obtain the value of b*:

$$\frac{\lambda_p (3\lambda_p + \gamma) - \sqrt{[\lambda_p (3\lambda_p + \gamma)]^2 - 8\lambda_p^4}}{2\lambda_p^2} \tag{28}$$

Figure 6 shows percentage increases in $E(\psi)$ due to a change to one-officer staffing as a function of b for selected values of λ_p and γ.

Where service times are not equal, $E(\psi|2$ 1 officer cars) is changed to:

$$E(\psi|2 \text{ 1-officer cars}) = \frac{2\lambda_p(1 - b_{1/2})}{2\lambda_p(1 - b_{1/2}) + \gamma} \quad , \quad 0 \leq b_1 < 1 \qquad (29)$$

where b_1 is as defined in Section V. The critical value b_1^* has the properties:

% increase in $E(\psi) > 100\%$ if $b_1 < b_1^*$, (increasing returns to scale)

% increase in $E(\psi) = 100\%$ if $b_1 = b_1^*$, (constant returns to scale)

% increase in $E(\psi) < 100\%$ if $b_1 > b_1^*$. (decreasing returns to scale)

Again, it is easy to show that

$$b_1^* = 2\left[\frac{\lambda_p(1 - b) - b\gamma}{\lambda_p(1 - b) - \gamma}\right] \quad , \quad 0 \leq b < 1 \qquad (30)$$

Note that b_1^* may be greater than 1, but b_1 cannot be greater than 1. A value of $b_1^* > 1$ simply means that increasing returns to scale are guaranteed by a switch to two one-officer units from one two-officer unit.

Figure 7 shows percentage increases in $E(\psi)$ as a function of b_1 for selected values of λ_p, γ, and b. The models developed here will now be illustrated using San Diego data.

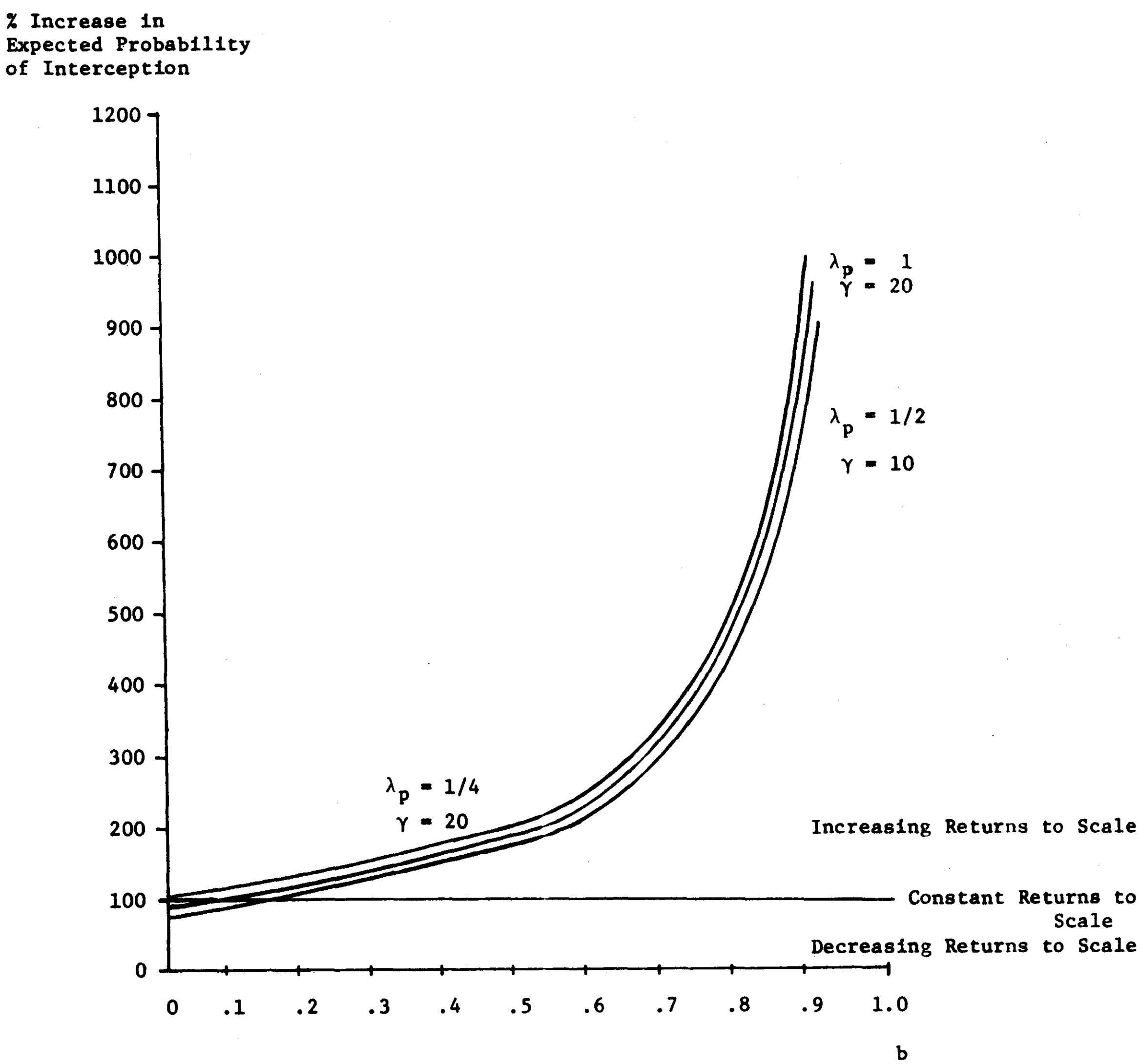


Figure 6

Percentage Increases in $E(\psi)$ for Equal Service Times

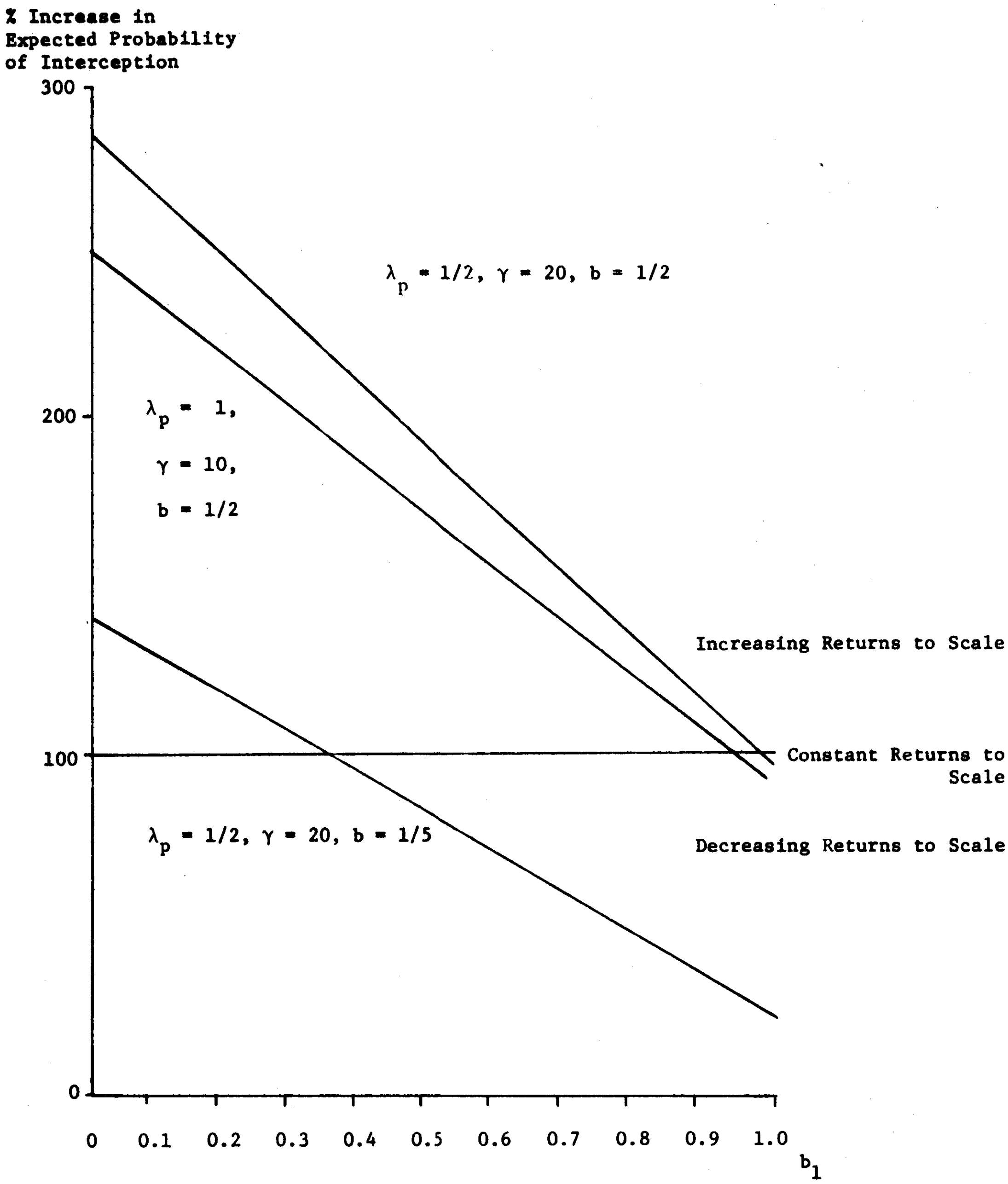


Figure 7

Percentage Increases in $E(\psi)$ for Unequal Service Times

Example: Patrol Staffing in San Diego

Recall from Section V that the values for λ_p, b, b_1 in San Diego were determined as

λ_p = .25 patrols per hour,

b = 0.50,

b_1 = 0.66.

If we assume that the average crime takes 3 minutes to occur (i.e., γ = 20) then:

$$E(\psi|1 \text{ 2-officer unit}) = \frac{.25(1 - .50)}{.25(1 - .50) + 20}$$

$$\approx .006,$$

$$E(\psi|2 \text{ 1-officer unit} = \frac{(2).25(1 - .66/2)}{(2).25(1 - .66/2) + 20}$$

$$\approx .016.$$

The % increase is around 167%; we knew that this increase would be greater than 100% since from equation (30),

$b_1 = .66 < .994 = b_1^*$.

If service times were the same for one-officer and two-officer units and using two-officer service times as a base measure, we compute using equation (28)

$b^* = .024$,

and since $b = .5 > b^* = .024$, one would still expect increasing returns to scale.

To see how these figures compare with what actually happened in San Diego, an approximation of the number of arrests made as a result of interception was made. Of all crimes which occurred in San Diego in the experimental period, Table I shows those which are felt by this analyst to be "observable," along with arrests made.[3] Table I represents about 93% of all crimes reported in the experimental period. There were a total of 101 criminal arrests made due to calls for service (Boydstun et al., 1977 29). Thus about 93% of these were arrests made on observable crimes, i.e., $.93 \cdot 101 \approx 94$.

Hence, given that 118 observable arrests were made in total, a maximum of 118 - 94 = 24 arrests could have been made due to interception by preventive patrol.

Using our estimate for $E(\psi | 1\ \text{2-officer car}) = .006$, one would expect $2301 \cdot .006 \approx 14$ arrests due to preventive patrol. Thus, the model is predicting within the feasible upper bound, and the results are certainly correct within an order of magnitude.

One thing that is interesting is the small numerical values of $E(\psi)$. These figures indicate that preventive patrol does not generate a high probability of intercepting a crime. This finding has also been discussed by Larson (1972: 147).

However, depending upon the crime level in the area in question, the switch from two-officer to one-officer patrol could result in significant increases in the number of intercepted crimes. The results of this section show that the most significant increases will occur (increasing returns to scale) for large values of b. This is consistent with the results generated from the queueing analysis of Section V. Again it has been shown that the switch to two one-officer units

TABLE I

Selected Crime Statistics from Patrol Staffing In San Diego, Table B-2

Observable Crimes	# Crimes Committed	# Arrests
Robbery	249	20
Burglary	1251	64
Grand Theft	55	2
Petty Theft	240	9
Auto Theft	208	6
Tamper with Vehicle	107	11
Malicious Mischief	191	6
Total	2301	118

This table accounts for roughly 93% of all crimes reported in the experimental period.

is more justified if the present two-officer unit is busy for a large fraction of the time than if it is busy for a small fraction of the time.

VIII. OFFICER INJURY

The major argument in favor of two-officer units hinges around the claim that two-officer patrol is safer than one-officer patrol. One way to investigate this claim is to examine officer injury rates. While no attempt will be made to develop detailed models here, some simple probabilistic arguments applied to empirical data will act to investigate the claim mentioned above.

Suppose there is a probability θ that an officer could be injured on any given incident. Then,

$$Pr\{\text{exactly one injury} | \text{one one-officer unit at incident}\} = \theta$$

$$Pr\{\text{exactly one injury} | \text{one 2-0 unit at incident}\} = 2\,\theta\, Pr\{\text{no 2nd injury} | \text{1 injury}\}.$$

$$Pr\{\text{exactly two injuries} | \text{one 2-0 unit at incident}\} = \theta\, Pr\{\text{2nd injury} | \text{1 injury}\}.$$

The expected number of injuries per incident for 1-0 units is given simply by θ. For two-officer units, the expected number of injuries is given by

$$1 \cdot 2\,\theta\, Pr\{\text{no 2nd injury} | \text{1 injury}\} + 2 \cdot \theta\, Pr\{\text{2nd injury} | \text{1 injury}\} = 2\,\theta\,.$$

Hence, one would expect, on average, twice as many injuries from two-officer units than from one-officer units. This is the case when the probability of injury is the same for officers in both one-and two-officer units.

In the San Diego experiment, it was concluded that one-officer patrol was safer than two-officer patrol. To substantiate this claim, the evaluators argued that given roughly the same number of exposures to potentially hazardous incidents, two-officer units became entangled in more "critical" incidents than one-officer units (117 v.s. 62), and as a result incurred more injuries (31 v.s. 18) (Boydstun et al., 1977: 61-70). However, this argument does not recognize the fact that more injuries are to be expected from two-officer units due to the presence of the second officer.

If the number of exposures to potentially hazardous incidents is K, and the number of critical incidents which actually occur is Q, then

$$\Pr\{\text{critical incident}|\text{exposure}\} = \begin{cases} \dfrac{Q_{1-0}}{K_{1-0}} & \text{for 1-0 units,} \\ \\ \dfrac{Q_{2-0}}{K_{2-0}} & \text{for 2-0 units.} \end{cases}$$

Also, if I injuries are incurred over the Q critical incidents, then

$$\Pr\{\text{injury}|\text{critical incident}\} = \begin{cases} \dfrac{I_{1-0}}{Q_{1-0}} & \text{for 1-0 units,} \\ \\ \dfrac{I_{2-0}}{2Q_{2-0}} & \text{for 2-0 units.} \end{cases}$$

If we assume that

$$\Pr\{\text{injury}|\text{incident not critical}\} = 0,$$

then the ultimate probability of injury given a potentially hazardous situation is

$$\Pr\{\text{injury}|\text{hazardous situation}\} = \begin{cases} \dfrac{I_{1-0}}{Q_{1-0}} \dfrac{Q_{1-0}}{K_{1-0}} = \dfrac{I_{1-0}}{K_{1-0}} \text{ for 1-0 units,} \\ \\ \dfrac{I_{2-0}}{2Q_{2-0}} \dfrac{Q_{2-0}}{K_{2-0}} = \dfrac{I_{2-0}}{2K_{2-0}} \text{ for 2-0 units.} \end{cases} \quad (31)$$

Note that this result does not depend on Q, the number of critical incidents which actually occur.

If both one- and two-officer units are exposed to the same number of potentially hazardous units (i.e., $K_{1-0} = K_{2-0}$ as is suggested by the San Diego study), then

$$\Pr\{\text{injury to an officer in a 1-0 unit}|\text{hazardous situation}\}$$
$$< \Pr\{\text{injury to an officer in a 2-0 unit}|\text{hazardous situation}\}$$
$$\text{only if } I_{1-0} < \frac{I_{2-0}}{2}$$

In San Diego, $I_{1-0} = 18$, while $I_{2-0} = 31$. Since $\frac{I_{2-0}}{2} = 15.5 < 18 = I_{1-0}$, we conclude that one-officer staffing is in fact more dangerous than two-officer staffing, if only slightly so. In figures,

$$\Pr\{\text{injury to 1-0 officer}|\text{hazardous situation}\} \approx \frac{18}{225} = .08.$$

$$\Pr\{\text{injury to 2-0 officer}|\text{hazardous situation}\} \approx \frac{31}{2 \cdot 225} = .07.$$

These figures are not that different. It would appear that, at least in San Diego, the chances of getting injured are roughly equivalent for officers in both one- and two-officer units.

IX. COSTS

If one studies the cost aspect of the one- versus two-officer staffing question, it is obvious that one one-officer unit will be less expensive than one two-officer unit; the difference being the additional salary required by the second officer in a two-officer car. Also, two one-officer units are more expensive to field than one two-officer unit, the difference being the additional cost of obtaining and maintaining the second patrol car. In the context of this paper, the relevant cost consideration is that which compares two one-officer units to one two-officer unit.

Let

C_o = cost per officer,

C_c = cost per car,

N_{1-0} = # of 1-0 units,

N_{2-0} = # of 2-0 units.

Then the cost associated with any mix of 1-0 and 2-0 units is given by

$$C = N_{1-0}(C_o + C_c) + N_{2-0}(2C_o + C_c). \tag{32}$$

For a fixed cost C, equal cost alternatives may be visualized as those pairs (N_{1-0}, N_{2-0}) which fall on the line graphed in Figure 8.

In this paper, we have considered the two cases of all two-officer staffing, or all one-officer staffing. If initially there were N two-officer cars, than an equal cost option would allow for

$$N\left[\frac{2C_o + C_c}{C_o + C_c}\right] \text{ one-officer cars.}$$

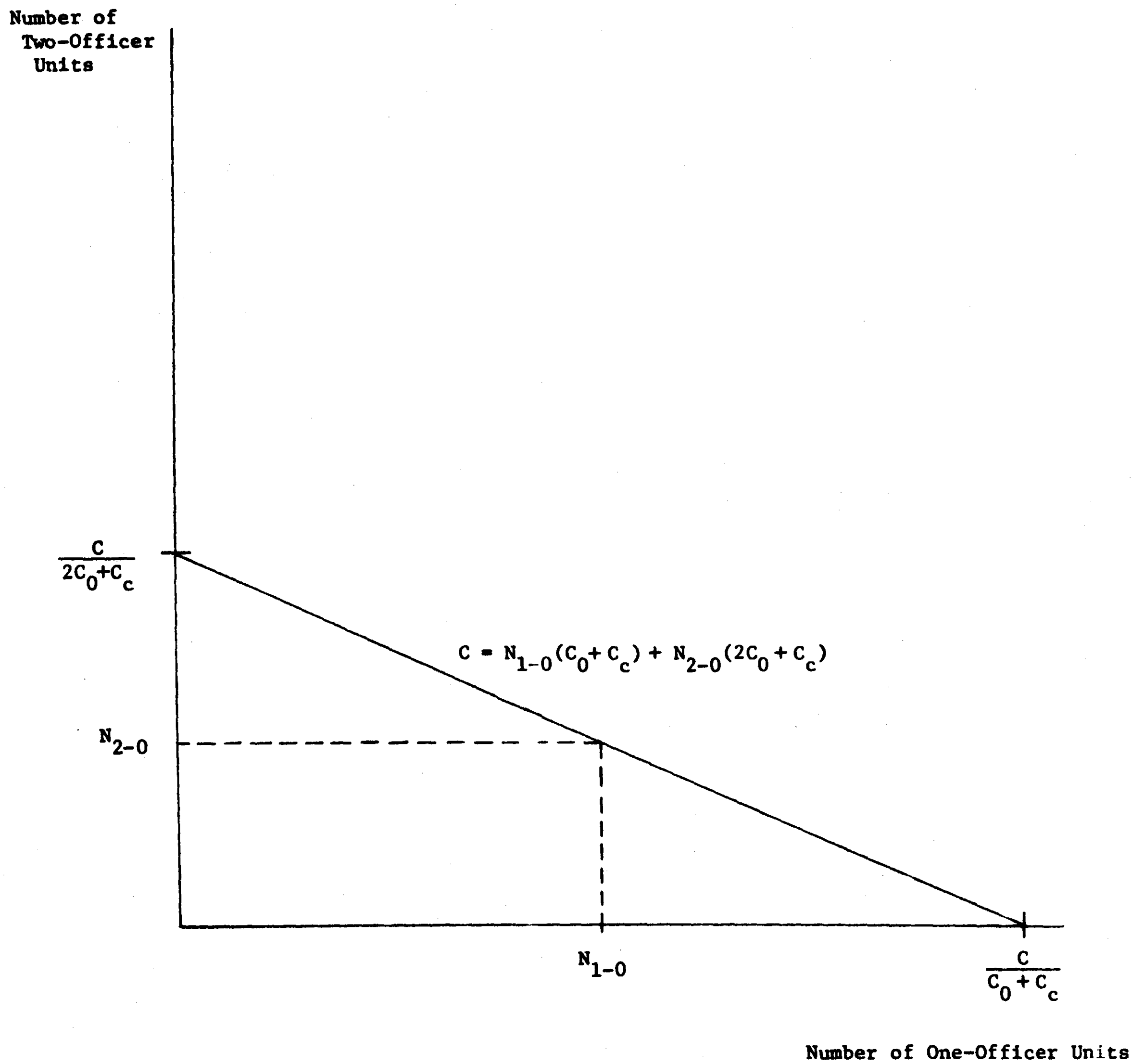


Figure 8

Equal Cost Staffing Options
(Total Cost = C)

It stands to reason that from the models presented in this paper, one would like to maximize the usage of one-officer cars. Using criteria such as expected area covered, response time, patrol frequency/visibility, and the probability of intercepting a randomly occurring crime in progress, one-officer patrol is the preferred staffing mode as long as two one-officer units are allocated where one two-officer unit was assigned before. Also, from the last section, it does not appear that a policy utilizing 1-0 cars needs to be compromised for reasons of safety.

In San Diego, the following figures were reported (Boydstun et al., 1977: 54-55)

C_o = \$12.79/hr.,

C_c = \$ 2.51/hr.

Thus, in San Diego, one could staff N two-officer units or

$$\left[\frac{2 \cdot 12.79 + 2.51}{12.79 + 2.51}\right] N = 1.84N$$

one-officer units at the same cost. How should these 1.84N units be allocated? This is by no means a trivial problem, but some illustrative simple-minded solutions do come to mind.

For example, a reasonable allocation scheme for a city with N two-officer units and no one-officer units might be to use

$\frac{N}{6}$ units for the first 8-hour tour,

$\frac{2N}{6}$ units for the second 8-hour tour,

$\frac{3N}{6}$ units for the third 8-hour tour.

This, of course, assumes that a car can't be used on more than one tour per day; this assumption could be relaxed without too much difficulty. Now, if 1.84N one-officer units were available, one could allocate

$\frac{4N}{6}$ units for the second tour,

$\frac{6N}{6}$ units for the third tour, leaving

$1.84N - \frac{4N}{6} - \frac{6N}{6} = .173N$ units for the first tour.

Tours two and three would contain exactly twice the number of units as before, hence the models of this paper would apply. Tour one would contain only a few more cars than before, and less manpower. However, since this shift was not heavily manned before relative to other shifts, the degradation in service would not be a serious one.

X. SUMMARY AND CONCLUSIONS

This chapter began with a discussion of the postulated advantages of one- versus two-officer patrol. Relevant performance measures were elicited from this discussion, and models were constructed in order to formalize expectations with respect to the levels of these performance measures under the two staffing strategies. The results include:

1. Doubling the number of available units in a beat increases the expected area covered by less than twice.
2. Increases in coverage gained by a switch to two one-officer units per beat are most significant when:
   (i) beat sizes are large,
   (ii) travel speeds are low,
   (iii) coverage is defined for low travel times.
3. Given that units are free, the reduction in expected travel time due to a switch to two one-officer units is only around 28%.
4. If one-officer and two-officer units have equal service times, and allowing for the possibility of busy units, the reduction in expected travel time due to a switch to two one-officer units does not exceed 40%.
5. Response delays are less frequent with many "slow" servers (one-officer units) than with few "fast" servers (two-officer units). Hence, the reduction in response times (including queueing delay) due to a switch to one-officer staffing may be quite large.

6. If one-officer and two-officer units have equal service times, then a switch to two one-officer units per beat will increase patrol frequency by at least 100%. If one- and two-officer units have different service times, patrol frequency is likely to increase by more than 100%, though it may increase by less than 100%.
7. Percentage increases in absolute visibility are the same as percentage increases in patrol frequency. Noticeable increases in police visibility are only likely to set in after some initial, "critical" level of one unit patrol has been reached.
8. The increase in the probability of intercepting a randomly occurring crime in progress due to a switch to two one-officer units per beat may be less than, equal to, or greater than 100%, depending upon the fraction of time units are out of service.
9. The magnitudes of interception probabilities are small.
10. Increases in interception probabilities due to the adoption of of one-officer per unit staffing are most significant for beats with heavy workloads.
11. In San Diego, the probability of officer specific injury is roughly the same for one- and two-officer staffing.
12. The additional cost due to switching from one two-officer unit per beat to two one-officer units per beat is the cost of maintaining a second vehicle, which in San Diego is $2.51/hour/unit.
13. Equal-cost alternatives favor maximum usage of one-officer units. In San Diego, 1.84N one-officer units may be fielded for the same cost as N two-officer units.

This research has not attempted to be exhaustive; there are other relevant areas which could be subjected to analysis. One important item is the need for backup support when using one-officer vehicles. Both the frequency of backup requests, and the logistics of backup assignments may be investigated analytically. The patrol frequency and crime interception models may be improved by relaxing the uniformity assumptions: patrol may be spatially distributed according to a coverage function (Larson, 1972: 135-137) and crime may take on a spatial as well as temporal distribution. Also, the distinction between interception and apprehension may be investigated. In the area of officer safety, more examination of injury statistics from different cities is required. Finally, an assessment of on-scene performance of one- versus two-officer units would be useful; the models presented here do not begin to address this this question.

FOOTNOTES

[1]On page 36 of Patrol Staffing in San Diego, we are told that over 12 one-week periods, one-officer units drove on average 4856 miles/unit while two-officer units drove on average 5026 miles/unit. Thus, one-officer units averaged

$$\frac{4856 \text{ miles}}{12 \text{ weeks}} \cdot \frac{1 \text{ week}}{7 \text{ days}} = 58 \text{ miles/day/unit}$$

while two-officer units drove an average of

$$\frac{5026 \text{ miles}}{12 \text{ weeks}} \cdot \frac{1 \text{ week}}{7 \text{ days}} = 60 \text{ miles/day/unit.}$$

For any given unit, there are only 8 hours in a day (i.e., one tour) as opposed to 24. Rounding off to the nearest mile would yield an average of 59 miles/8-hour tour/unit for both one- and two-officer units. For convenience, the number 60 was chosen instead of 59.

[2]By saying $\gamma = 20$ crimes/hour, it is meant that on average, 20 crimes could occur in an hour, not that on average, 20 crimes do occur in an hour.

[3]These figures are taken from Table B-2 in Patrol Staffing in San Diego. Number of crimes committed is set equal to the number of reported incidents minus the number of unfounded calls for each category.

REFERENCES

1. Boydstun, J.E., Sherry, M.E., and Moelter, N.P. (1977). Patrol Staffing in San Diego: One- or Two-Officer Units. Washington: Police Foundation.

2. Carter, C.L., and Kaplan, E.H. (1977). "Did Anything Happen in Kansas City?" Unpublished paper, Department of Urban Studies and Planning, M.I.T.

3. Chicago Police Department (1963). "One-Man Patrol Cars." In Samuel G. Chapman, ed., Police Patrol Readings. Springfield: Charles C. Thomas, pp. 212-220.

4. Clawson, C., and Tarr, D. (1977). Summary of One- and Two-Officer Patrol Car Study. Seattle: Seattle Police Department.

5. Drake, A.W. (1967). Fundamentals of Applied Probability Theory. New York: McGraw-Hill.

6. Elliott, J.F. (1968). "Random Patrol." In Samuel G. Chapman, ed., Police Patrol Readings. Springfield: Charles C. Thomas, pp. 357-365.

7. Freund, J.E. (1971). Mathematical Statistics. Englewood Cliffs: Prentice-Hall.

8. Governmental Research Institute (1957). "One-Man Police Patrol Car Operation." In Samuel G. Chapman, ed., Police Patrol Readings. Springfield: Charles C. Thomas, pp. 220-244.

9. Larson, R.C. (1972). Urban Police Patrol Analysis. Cambridge: M.I.T. Press.

10. Larson, R.C. (1975). "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment." Journal of Criminal Justice, 3: 267-297.

11. Riccio, L.J. (1974). "Direct Deterrence - An Analysis of the Effectiveness of Police Patrol and other Crime Prevention Technologies." Journal of Criminal Justice, 2: 207-217.

12. White, J.A., Schmidt, J.W., and Bennett, G.K. (1975). Analysis of Queueing Systems. New York: Academic Press.

APPENDIX

1. Derivation of expected travel time from the nearest of two units to an incident.

Under the conditions of the model, it is relatively easy to show that the pdf for travel time from a randomly located patrol car to the randomly located incident is:

$$f_{t_r}(t_o) = \begin{cases} \frac{4v}{u^4}(u^2vt_o - uv^2t_o^{\,2} + \frac{1}{6}v^3t_o^{\,3}) & 0 \le t_o \le \frac{u}{v} \\ \frac{4v}{u^4}(\frac{4}{3}u^3 - 2u^2vt_o + uv^2t_o^{\,2} - \frac{v^3t_o^{\,3}}{6}) & \frac{u}{v} < t_o \le \frac{2u}{v} \end{cases} \tag{A1}$$

If we consider the case of two units in the beat, then assuming that the responding unit is the nearest unit to the scene of the incident, the travel time is given by:

$$t_r^* = \min\,(t_{r_1}, t_{r_2}) \tag{A2}$$

where t_{r_1} = travel time from 1st unit to incident.

t_{r_2} = travel time from 2nd unit to incident.

To find the pdf for t_r^*, one can use the following rule:

$$f_x(x_o) = \sum_{i=1}^{n} f_{x|A_i}(x_o|A_i) \cdot \Pr(A_i)$$

where the A_i are mutually exclusive, collectively exhaustive events in x_o sample space. Returning to the example at hand, let

$$A_1 \iff 0 \le t_{r_1} \le \frac{u}{v}, \quad 0 \le t_{r_2} \le \frac{u}{v};$$

$$A_2 \iff 0 \le t_{r_1} \le \frac{u}{v}, \quad \frac{u}{v} < t_{r_2} \le \frac{2u}{v};$$

$$A_3 \iff \frac{u}{v} < t_{r_1} \le \frac{2u}{v}, \quad 0 \le t_{r_2} \le \frac{u}{v};$$

$$A_4 \iff \frac{u}{v} < t_{r_1} \le \frac{2u}{v}, \quad \frac{u}{v} < t_{r_2} \le \frac{2u}{v}.$$

Using the pdf $f_{t_r}(t_o)$ and the fact that t_{r_1} and t_{r_2} are independent, one can verify that:

$$Pr(A_1) = \frac{25}{36}, \quad Pr(A_2) = Pr(A_3) = \frac{5}{36}, \quad Pr(A_4) = \frac{1}{36}.$$

Immediately, one can state that:

$$f_{t_r^*|A_2}(t_o|A_2) = \frac{f_{t_r}(t_o)}{Pr(A_2)} \qquad 0 \le t_o \le \frac{u}{v}$$

$$= \frac{6}{5} \cdot \frac{4v}{u^4}\left(u^2 v t_o - u v^2 t_o^2 + \frac{1}{6} v^3 t_o^3\right) \qquad 0 \le t_o \le \frac{u}{v} \tag{A3}$$

By symmetry:

$$f_{t_r^*|A_3}(t_o|A_3) = f_{t_r^*|A_2}(t_o|A_2) \qquad 0 \le t_o \le \frac{u}{v}$$

The method for obtaining $f_{t_r^*|A_1}(t_o|A_1)$ and $f_{t_r^*|A_4}(t_o|A_4)$ is tedious and will not be discussed here. The results are:

$$f_{t_r^*|A_1}(t_o|A_1) = \frac{36}{25} \cdot \frac{32v}{u^8}\left[\frac{5}{24}u^6vt_o - \frac{5}{24}u^5v^2t_o^{\,2} - \frac{67}{144}u^4v^3t_o^{\,3}\right. \tag{A4}$$

$$\left. + \frac{5}{6}u^3v^4t_o^{\,4} - \frac{11}{24}u^2v^5t_o^{\,5} + \frac{7}{72}uv^6t_o^{\,6} - \frac{1}{144}v^7t_o^{\,7}\right]$$

$$0 \le t_o \le \frac{u}{v}$$

$$f_{t_r^*|A_4}(t_o|A_4) = 36 \cdot \frac{32v}{u^8}\left[\frac{8}{9}u^7 - \frac{28}{9}u^6vt_o + \frac{14}{3}u^5v^2t_o^{\,2}\right. \tag{A5}$$

$$- \frac{35}{9}u^4v^3t_o^{\,3} + \frac{35}{18}u^3v^4t_o^{\,4} - \frac{7}{12}u^2v^5t_o^{\,5} + \frac{7}{72}uv^6t_o^{\,6}$$

$$\left. - \frac{1}{144}v^7t_o^{\,7}\right] \frac{u}{v} < t_o \le \frac{2u}{v}$$

Thus, the unconditional pdf $f_{t_r^*}(t_o)$ is given by:

$$f_{t_r^*}(t_o) = \begin{cases} \dfrac{16v}{u^8}\left[\dfrac{1}{2}u^6vt_o - \dfrac{1}{2}u^5v^2t_o^{\,2} - \dfrac{11}{12}u^4v^3t_o^{\,3} + \dfrac{5}{3}u^3v^4t_o^{\,4}\right. \\ \qquad \left. - \dfrac{11}{12}u^2v^5t_o^{\,5} + \dfrac{7}{36}uv^6t_o^{\,6} - \dfrac{1}{72}v^7t_o^{\,7}\right] \quad 0 \le t_o \le \dfrac{u}{v} \\ \dfrac{32v}{3u^8}\left[\dfrac{8}{3}u^7 - \dfrac{28}{3}u^6vt_o + 14u^5v^2t_o^{\,2} - \dfrac{35}{3}u^4v^3t_o^{\,3}\right. \\ \qquad \left. + \dfrac{35}{6}u^3v^4t_o^{\,4} - \dfrac{7}{4}u^2v^5t_o^{\,5} + \dfrac{7}{24}uv^6t_o^{\,6} - \dfrac{1}{48}v^7t_o^{\,7}\right] \\ \dfrac{u}{v} < t_o \le \dfrac{2u}{v} \end{cases} \tag{A6}$$

The expected travel time from the nearest unit to the incident is:

$$E(t_r^*) = \int_o^{\frac{2u}{v}} t_o f_{t_r^*}(t_o)dt_o$$

$$= \frac{1354u}{2835v} \qquad \text{(A7)}$$

2. Determination of $Pr\{A_i\}$ and W_q for the four-beat model: Two-officer staffing.

To determine $Pr\{A_i\}$, i = 1,..4, it is useful to distinguish between a macrostate and a microstate. In the example considered here, a macrostate reports how many patrol cars are busy. A microstate reports a particular combination of busy and free cars by beat. It should be obvious that several microstates may give rise to the same macrostate.

Consider the macrostate S_1 corresponding to the event "one patrol car is busy." There are four microstates that could give rise to S_1:

M_{11}: only beat 1 car is busy,

M_{12}: only beat 2 car is busy,

M_{13}: only beat 3 car is busy,

M_{14}: only beat 4 car is busy.

The events A_i i = 1,...,4 can be expressed in terns of microstates. Since for a given macrostate, the associated microstates are equally likely, if we know the probability of the macrostates, then we can find $Pr(A_i)$ $i = 1,\ldots,4$ through the enumeration of the relevant microstates.

Since we know that $Pr(A_i) = Pr(A_{i|1})$ $i = 1,\ldots,4$, let's examine the case discussed earlier where calls originate in beat 1. Consider Table AI. Here, all the events $A_{i|1}$ $i = 1,\ldots,4$ have been enumerated in terms of the relevant microstates. For example, given that 2 cars are busy there are:

3 microstates corresponding to $A_{1|1}$ (busy cars in beats 2-3, 2-4, 3-4);

3 microstates corresponding to $A_{2|1}$ (busy cars in beats 1-2, 1-3, 1-4).

Since all microstates are equally likely for a given macrostate, we have:

$$Pr\{A_{1|1}\} = Pr\{S_0\} + \frac{3}{4} Pr\{S_1\} + \frac{3}{6} Pr\{S_2\} + \frac{1}{4} Pr\{S_3\} \tag{A8}$$

$$Pr\{A_{2|1}\} = \frac{1}{4} Pr\{S_1\} + \frac{3}{6} Pr\{S_2\} + \frac{2}{4} Pr\{S_3\} \tag{A9}$$

$$Pr\{A_{3|1}\} = \frac{1}{4} Pr\{S_3\} \tag{A10}$$

$$Pr\{A_{4|1}\} = Pr\{S_4\} \tag{A11}$$

It has already been noted that $Pr(A_i) = Pr(A_{i|1})$ $i = 1,\ldots,4$, so the stated equations hold for the general case. If the macrostate probabilities $Pr(S_i)$ $i = 0,\ldots,4$ can be found, the model will be complete.

TABLE AI

STATE STRUCTURE FOR FOUR-BEAT MODEL, TWO-OFFICER PATROL

Macrostate (# of busy cars)	# Microstates/ Macrostate	# Microstates per Event			
		$A_{1\|1}$	$A_{2\|1}$	$A_{3\|1}$	$A_{4\|1}$
S_0	1	1	0	0	0
S_1	4	3	1	0	0
S_2	6	3	3	0	0
S_3	4	1	2	1	0
S_4	1	0	0	0	1

Given any macrostate, all microstates are equally likely.

From the way this model has been constructed, the macrostate probabilities are given by the M/M/4/∞ queueing model, i.e.:

$$\Pr\{S_i\} = \Pr\{S_0\}\ \rho^i/i! \qquad i = 0, 1, 2, 3 \tag{A12}$$

$$\Pr\{S_4\} = 1 - \sum_{i=0}^{3} \Pr\{S_i\} \tag{A13}$$

where

$$\Pr\{S_0\} = \left[\sum_{i=0}^{3} \rho^i/i! + \frac{1}{4!}\rho^4 \cdot \frac{1}{1 - \rho/4}\right]^{-1} \tag{A14}$$

W_q is simply the expected length of time spent waiting in queue for the M/M/4/∞ queueing model. It is easy to show that for the general M/M/N/∞ model (White et al., 1975: 103)

$$W_q = \frac{\rho^N}{N \cdot N!\ \mu} \cdot \frac{\Pr\{S_o\}}{(1 - \rho/N)^2} \tag{A15}$$

so for this case,

$$W_q = \frac{\rho^4}{4 \cdot 4!\ \mu} \cdot \frac{\Pr\{S_o\}}{(1 - \rho/4)^2}$$

3. Determination of $\Pr\{B_i\}$ and W_q^* for the four-beat model: One-officer staffing.

Table AII presents the macrostate/microstate/event breakdown for the one officer per unit case. Since all microstates are equally likely for any macrostate, the events $B_1,\ldots,B_5$ take on the following probabilities.

TABLE AII

STATE STRUCTURE FOR FOUR-BEAT MODEL, ONE-OFFICER PATROL

Macrostates (# of busy cars)	# Microstates/ Macrostate	# Microstates per Event				
		B_1	B_2	B_3	B_4	B_5
S_0	1	1	0	0	0	0
S_1	8	6	2	0	0	0
S_2	28	15	12	1	0	0
S_3	56	20	30	6	0	0
S_4	70	15	40	15	0	0
S_5	56	6	30	20	0	0
S_6	28	1	12	14	1	0
S_7	8	0	2	4	2	0
S_8	1	0	0	0	0	1

Given any macrostate, all microstates are equally likely.

$$\Pr\{B_1\} = \Pr\{S_0\} + \frac{6}{8}\Pr\{S_1\} + \frac{15}{28}\Pr\{S_2\} + \frac{20}{56}\Pr\{S_3\} + \frac{15}{70}\Pr\{S_4\} + \frac{6}{56}\Pr\{S_5\} + \frac{1}{28}\Pr\{S_6\} \tag{A17}$$

$$\Pr\{B_2\} = \frac{2}{8}\Pr\{S_1\} + \frac{12}{28}\Pr\{S_2\} + \frac{30}{56}\Pr\{S_3\} + \frac{40}{70}\Pr\{S_4\} + \frac{30}{56}\Pr\{S_5\} + \frac{12}{28}\Pr\{S_6\} + \frac{2}{8}\Pr\{S_7\} \tag{A18}$$

$$\Pr\{B_3\} = \frac{1}{28}\Pr\{S_2\} + \frac{6}{56}\Pr\{S_3\} + \frac{15}{70}\Pr\{S_4\} + \frac{20}{56}\Pr\{S_5\} + \frac{14}{28}\Pr\{S_6\} + \frac{4}{8}\Pr\{S_7\} \tag{A19}$$

$$\Pr\{B_4\} = \frac{1}{28}\Pr\{S_6\} + \frac{2}{8}\Pr\{S_7\} \tag{A20}$$

$$\Pr\{B_5\} = \Pr\{S_8\} \tag{A21}$$

The macrostate probabilities in this case are generated by the M/M/8/∞ queueing model:

$$\Pr\{S_i\} = \Pr\{S_0\}\ \rho^i/i! \qquad i = 0, .,\ldots,7 \tag{A22}$$

$$\Pr\{S_8\} = 1 - \sum_{i=0}^{7} \Pr\{S_i\} \tag{A23}$$

where

$$\Pr\{S_0\} = \left[\sum_{i=0}^{7} \rho^i/i! + \frac{1}{8!}\rho^8 \cdot \frac{1}{1-\rho/8}\right]^{-1} \tag{A24}$$

and

$$\rho = \frac{\lambda_c}{\mu}\ , \ \rho/8 < 1\ .$$

For this case, W_q^* is calculated from equation (A14) substituting $N = 8$, i.e.,

$$W_q^* = \frac{\rho^8}{8 \cdot 8!\ \mu} \frac{Pr\{S_o\}}{(1 - \rho/8)^2} \tag{A25}$$

CHAPTER 4

MODELS FOR THE EVALUATION OF TREATMENT - RELEASE CORRECTIONS PROGRAMS

I. INTRODUCTION

Within the corrections component of the criminal justice system, a range of programs aimed at the rehabilitation of selected individuals has been established. These programs (including prison, parole, residential centers for drug offenders, alcohol abuse counselling, etc.) all share the following feature in common: individuals committed to a program are subjected to a period of "treatment"; upon satisfactory completion of the treatment period, these individuals are "released" (hence the term "treatment-release program"). Of interest to the officials of such programs is the event that a randomly chosen program client commits an offense after release; the likelihood of this event is termed the "recidivism probability."

While the concept of a recidivism probability poses no immediate difficulty, the measurement of recidivism is not an easy task. In a clever paper, Blumstein and Larson (1969) discussed measurement problems which arise from alternative definitions of recidivism, and from improper interpretation of sample statistics. If we allow recidivism to refer solely to the event where a program client commits an offense after release, then recidivism cannot be measured directly.

When an individual commits a crime, there is no guarantee that (s)he will be apprehended. Of all individuals who commit crimes, some fraction will in fact be arrested by the police. Placed into the context of a treatment-release corrections program, only those recidivists who are rearrested are in fact observed as having recidivated; indeed, the difference between rearrest levels and true recidivism levels may be greater than one might expect (Barnett and Stabile, 1979). Whether or not an offender is apprehended depends upon police performance as well as upon the nature of the offense. Thus, the methods of this chapter will be presented as applied to rearrest patterns over time, as this is the type of data which is frequently available.

What is found in the remainder of this chapter is a discussion of methodology for conducting model-based evaluations of treatment-release corrections programs. Model-based techniques have proved useful in evaluating police patrol programs (Larson, 1975; Kaplan, 1978a), and it is felt that the advantages provided by the modeling approach can carry over to the corrections area. We begin with the description of a general model for rearrest patterns over time; the behavior of this model is examined under alternative assumptions in a numerical example. Classical and Bayesian estimation methods are presented, followed by a discussion of model-based evaluation procedures. The chapter concludes with a brief discussion of possible extensions to the work reported here.

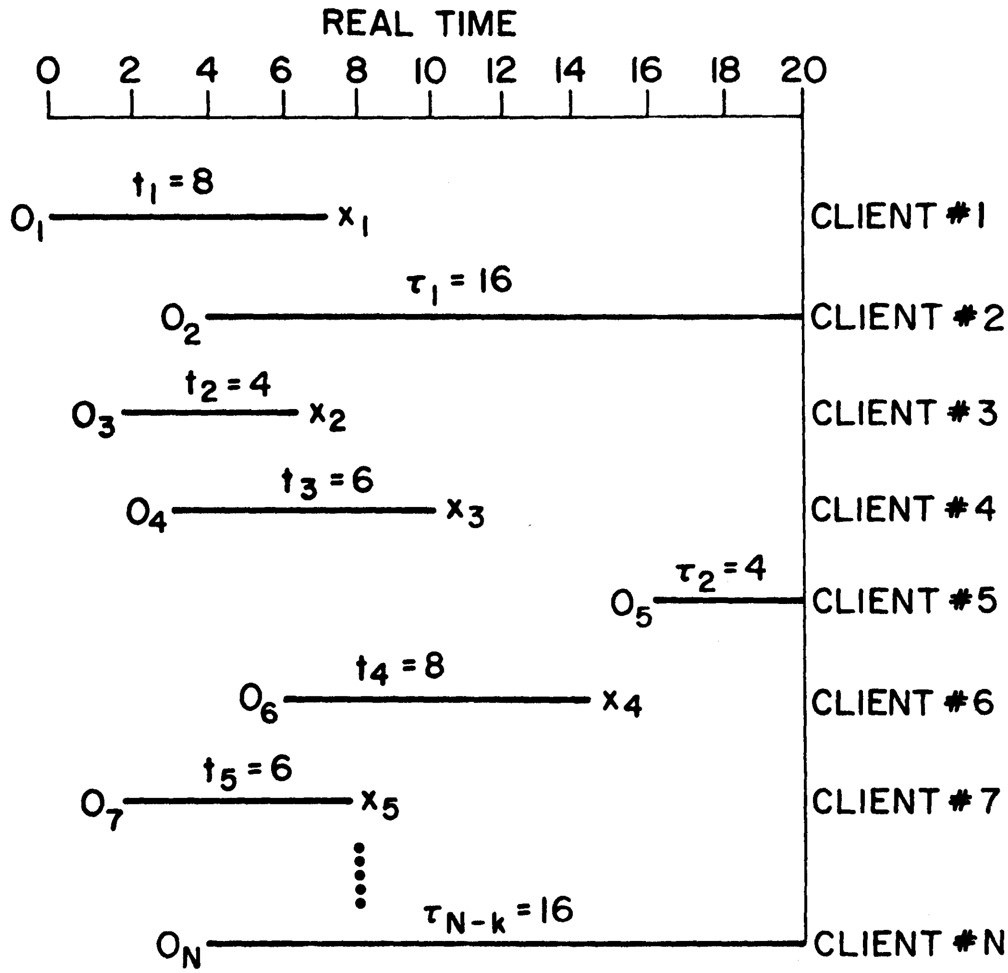
II. A GENERAL MODEL OF REARREST PATTERNS

Consider any corrections program which subjects its clients to a treatment period after which they are released. Ideally, clients released from treatment return to the community as law - abiding citizens. Realistically, sizeable fractions of program populations are known to recidivate. Of those who do recidivate, some are apprehended and rearrested.

In the model to be presented, we exploit the similarities between the rearrest process and a branch of probabilistic reasoning known as reliability theory (see Chapter 4 in Tsokos (1972), Chapter 13 in Hillier and Lieberman (1974).[1] Figure 1 depicts our observation of a corrections program which releases clients at different points in time; in total, N clients are released. Of these N individuals, some are rearrested, while the others are not rearrested during the period of time allowed for observation.

For all clients in the program, our model will measure time according to time from individual release, hence all of our arguments are conditioned on release occurring at time zero. An observed rearrest is referred to as a "failure", and the length of time that elapses between an individual's release and failure is denoted as the "time until failure". Hence, the statement "Five failures occurred by the eighth day after release" is interpreted to mean that five (of N) individuals were rearrested within eight days of the particular day on which they were each individually released.

Consider an individual who is released from treatment at time 0($t_R = 0$). We are interested in the probability that this same individual will be rearrested at some future time t given release at $t_R = 0$. Let



O_n = TIME OF n th RELEASE $n = 1, 2, \ldots, N$
x_i = TIME OF i th REARREST $i = 1, 2, \ldots, k$
t_i = TIME FROM RELEASE UNTIL i th FAILURE $i = 1, 2, \ldots, k$
τ_j = OBSERVED TIME FROM RELEASE OF j th SUCCESS, $j = 1, 2, \ldots, N-k$

Figure 1

A Rearrest Pattern Over Time Measured According To Time From Release this <u>probability of failure</u> be denoted by $p_F(t)$. In order to derive expressions for $p_F(t)$ and other performance measures, we need to make a few simplifying assumptions.

(i) All individuals fail independently of each other.

(ii) Pr {any client fails in (t,t + dt)|ultimate failure, but not in (0,t), $t_R = 0$} $= \phi(t|F)dt$.

(iii) Pr {any client does not fail in (t,t + dt)|ultimate failure, but not in (0,t), $t_R = 0$} $= 1 - \phi(t|F)dt$.

(iv) The fraction of the population that will ultimately fail is given by r; $0 \leq r \leq 1$.

From reliability theory, it is well known that these assumptions determine the conditional probability of failure by time t given $t_R = 0$ and ultimate failure to be (Hillier and Lieberman, (1974))

$$p_F(t|F) = 1 - e^{-\int_0^t \phi(x|F)dx}, \qquad t > 0 \qquad (1)$$

subject to:

(i) $\phi(t|F) \geq 0 \ \forall \ t > 0.$

(ii) $\int_0^\infty \phi(t|F)dt = \infty$

The unconditional probability of failure by time t after release, $p_F(t)$, is then given by

$$p_F(t) = r(1 - e^{-\int_0^t \phi(x|F)dx}). \qquad t > 0 \qquad (2)$$

$$0 \leq r \leq 1$$

It is apparent that the behavior of this model is completely dependent on the nature of the function $\phi(t|F)$. This function, which is referred to as the hazard function, dictates the probability of failure in the next time instant for those clients who have yet to fail but will ultimately fail. Several functions come to mind for $\phi(t|F)$. These are shown in Figure 2. Curves of Type I are of the form $\phi(t|F) = \lambda$; such an assumption implies that $p_F(t|F)$ will take on the simple negative exponential distribution. Stollmack and Harris (1974) studied this model, they also assumed that the fraction of utlimate failures was equal to one, a rather restrictive assumption. Maltz and McCleary (1977) studied this model without the r = 1 assumption. Both of these models will be examined later on in this paper. Type II and Type III curves involve increasing or decreasing propensities to be rearrested over time. Some of these curves can be formulated as $\phi(t|F) = \alpha\beta t^{\beta-1}$, which implies that $p_F(t|F)$ takes on the Weibull distribution (Freund, 1971:117); a model of this sort is illustrated later on. A more complicated model involving an exponential - type hazard function is discussed by Bloom (1978). In the appendix, we show that Bloom's model performs equivalently to the Maltz-McCleary model. Hence, Bloom's model will not be reviewed. Of course, plausible arguments for more complicated curves such as Types IV and V can be made; these would lead to still more complex forms for $p_F(t|F)$. The choice of an appropriate function for $\phi(t|F)$ is a data analysis question not pursued in this paper; since there is no one correct function $\phi(t|F)$ which works for all situations, the results which follow will be notated for the general case.

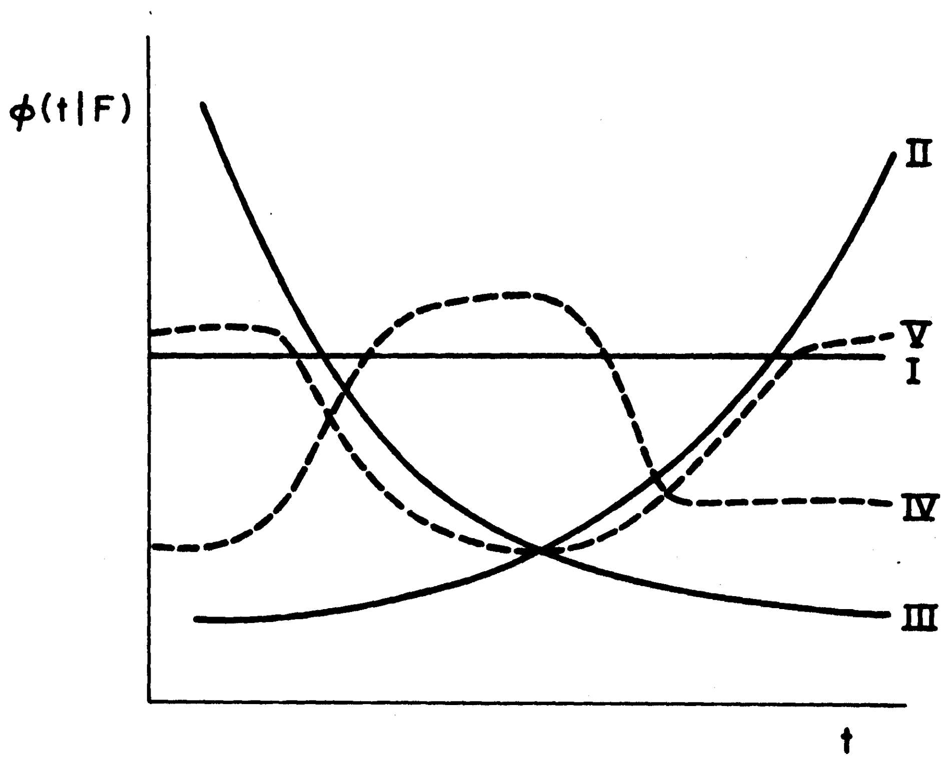


Figure 2

Possible Curves for $\phi(t|F)$

Performance Measures for the General Model

(i) Ultimate Probability of Failure: r

The ultimate probability of failure, $r = p_F(\infty)$, is directly estimated along with whatever parameters accompany the hazard function $\phi(t|F)$; this estimation problem is discussed in Section IV.

(ii) Expected Time Until Failure: $\bar{t}_F$

We have already computed the unconditional probability of failure by time t to be

$$p_F(t) = r\left(1 - e^{-\int_0^t \phi(x|F)\,dx}\right). \qquad t > 0$$

Differentiating this expression with respect to time from release provides us with the pseudo - pdf[2] for clients' individual times until failure

$$f_{t_F}(t) = r\phi(t|F)e^{-\int_0^t \phi(x|F)\,dx} . \qquad t > 0 \qquad (3)$$

The expected time until failures for ultimate failures is thus formulated as

$$\bar{t}_F = \frac{1}{r}\int_0^\infty t\ f_{t_F}(t)\,dt. \qquad (4)$$

Note that unless r = 1, the unconditional expected time until failure is infinite, since

E(time to failure) = E(time to failure|ultimate failure)·Pr {ultimate failure}
+ E(time to failure|ultimate success)·Pr {ultimate success}

$$= \bar{t}_F \cdot r + \infty \cdot (1-r)$$

$$= \infty$$

(iii) Median Time Until Failure: $t_{F(.5)}$

The median time until failure for ultimate failures is given by the equation

$$t_{F(.5)} = p_F^{-1}(.5r). \tag{5}$$

This is the time by which 50% of all eventual rearrests will have occurred.

(iv) Safety Time: $t^*(\varepsilon)$

The performance measure $t^*(\varepsilon)$ satisfies the equation

$$\Pr\ \{\text{individual fails in } (t^*,\infty)\,|\,\text{didn't fail in } (0,t^*) \text{ and } t_R{=}0\} = \varepsilon;$$

and hence defines the safety time at risk level ε (see Bloom (1978)). Intuitively, if one wishes to observe a client after release until the rearrest probability of that client is less than or equal to ε, then one must observe that client for at least $t^*(\varepsilon)$ time units after release at $t_R = 0$. The safety time is found by solving

$$t^*(\varepsilon) = p_F^{-1}\left(\frac{r - \varepsilon}{1 - \varepsilon}\right), \quad 0 < \varepsilon < r. \tag{6}$$

(v) Probability Mass Function of the Number of Rearrests: $P_n(t)$

Recall that by assumption, all individuals fail independently of one another. Since the probability of failure by time t given release at $t_R = 0$, $p_F(t)$, applies to any client, the probability that exactly n individuals have failed by time t is given by the well known binomial pmf

$$P_n(t) = \binom{N}{n} [p_F(t)]^n [1 - p_F(t)]^{N-n} \qquad (7)$$

$$n = 0,1,..,N$$

$$t \geq 0 \ .$$

Note that the probabilities

$$P_n(\infty) = \binom{N}{n} r^n (1 - r)^{N-n}$$

$$n = 0,1,..,N \qquad (8)$$

$$t \geq 0$$

may be interpreted as long run probabilities of failure, in that equation (8) determines the probability distribution of the ultimate failure population.

(vi) <u>Expected Number of Failures: $\bar{n}(t)$</u>

The expected number of failures that have occurred by time t is simply

$$\bar{n}(t) = Np_F(t). \qquad (9)$$

since n(t) is a binomially distributed random variable.

(vii) <u>Variance of the Number of Failures: $\sigma_n^2(t)$</u>

This measure is also easily obtained; it is given by

$$\sigma_n^2(t) = Np_F(t)\ (1 - p_F(t)). \tag{10}$$

It is interesting to note that since $0 \leq p_F(t) \leq 1$, $\sigma_n^2(t)$ will achieve a maximum when $p_F(t) = 1/2$. This implies that regardless of the functional form of $p_F(t)$, the maximum achievable variance of this model is equal to N/4. Under the conditions of this model, at worst one can be 95% certain that the observed number of failures by any time t is within $\pm\sqrt{N}$ of $\bar{n}(t)$.

(viii) Gausian Approximation to $P_n(t)$: $f_{n(t)}(x)$

Most correctional programs involve a large number of clients. In such situations where N is large, the calculations of $P_n(t)$ are both tedious and perhaps unrewarding. However, if N is sufficiently large (i.e., if $Np_F(t)$ and $N(1 - p_F(t))$, are both greater than 5 (Freund, 1971:177)), one may approximate the discrete binomial distribution of n(t) by a Gaussian distribution with mean $\bar{n}(t)$ and variance $\sigma_n^2(t)$. If we allow the interval $(n - 1/2, n + 1/2]$ to represent the integer number of failures n, then the number of failures which occur by a given time t may be approximated as a continuous random variable with pdf

$$f_{n(t)}(x) = \frac{1}{\sigma_n(t)\sqrt{2\pi}} e^{-1/2\left[\frac{x - \bar{n}(t)}{\sigma_n(t)}\right]^2}. \tag{11}$$

$$-\infty < x < \infty$$

The corresponding steady state pdf is obtained by setting $t = \infty$ in the above formulation. For large values of N, these distributions $f_{n(t)}(x)$ will be quite accurate.

(ix) <u>Time by Which the Probability of Failure Equals k/N: T_k</u>

One may also be interested in the time until the k^{th} failure. While there is no theoretical problem formulating the expected time until the k^{th} failure given that at least k failures ultimately occur, the computations involved are prohibitively difficult. To gain some indication of the timing of rearrests, it seems reasonable to examine

$$T_k \equiv \bar{n}^{-1}(k). \tag{12}$$

This statistic reports the time by which the probability of failure equals k/N; we will use T_k in Section V to compute <u>fractile</u> times until failure for ultimate failures.

In summary, this section has presented a general model which can provide a framework for analyzing rearrest patterns over time. Having discussed this model, it is useful to examine the differences in model performance that result from the choice of alternative hazard functions. Such an example is presented in the next section.

III. AN ILLUSTRATIVE EXAMPLE

Consider a treatment-release corrections program which serves a client population of N = 100. Suppose that a preliminary review of rearrest data has revealed that 25 of the program's 100 clients were rearrested within 6 months after release (hence $p_F(6)$ is estimated to equal .25). Program officials are committed to evaluate the program after 24 months of exposure data have been collected. The program will be considered a success if $p_F(24) < .40$. In the meantime, the program staff would like some indications of the range of rearrest patterns that could occur over time under alternative assumptions governing the rearrest process.

Three conjectures are of particular interest to the program staff; each may be formulated as a model consisted with the data point $p_F(6) = .25$.

Model 1 (Stollmack-Harris)

Let $\phi(t|F) = \lambda$, $\lambda > 0$, and assume $r = 1.0$.

Model 2 (Maltz-McCleary)

Let $\phi(t|F) = a$, $a > 0$, and assume $r = 0.5$.

Model 3 (Weibull)

Let $\phi(t|F) = \alpha\ \beta\ t^{\beta-1}$, α, $\beta > 0$, and assume $r = 0.5$.

The implications of these postulates may be examined in some detail. Table I presents the formulas used to compute the measures associated with the models, while Table II reports numerical values for selected measures.

If we direct our attention to Figure 3, we notice that the three models do represent quite different rearrest patterns over time.

TABLE I

FORMULAS FOR PERFORMANCE MEASURES

PERFORMANCE MEASURES	EQUATION	STOLLMACK-HARRIS	MALTZ-MCCLEARY	WEIBULL
$p_F(t)$	(2)	$1 - e^{-\lambda t}$	$r(1 - e^{-at})$	$r(1 - e^{-\alpha t^{\beta}})$
$\bar{t}_F$	(4)	$\frac{1}{\lambda}$	$\frac{1}{a}$	$(\frac{1}{\alpha})^{\frac{1}{\beta}}\Gamma(1 + \frac{1}{\beta})$
$t_F(.5)$	(5)	$-\frac{1}{\lambda}\ln(.5)$	$-\frac{1}{a}\ln(.5)$	$\left[-\frac{1}{\alpha}\ln(.5)\right]^{\frac{1}{\beta}}$
$t^*(\varepsilon)$	(6)	N.A.	$-\frac{1}{a}\ln(\frac{\varepsilon(1-r)}{r(1-\varepsilon)})$	$\left[-\frac{1}{\alpha}\ln(\frac{\varepsilon(1-r)}{r(1-\varepsilon)})\right]^{\frac{1}{\beta}}$
$\bar{n}(t)$	(9)	$N(1 - e^{-\lambda t})$	$Nr(1 - e^{-at})$	$Nr(1 - e^{-\alpha t^{\beta}})$
$\sigma_n^2(t)$	(10)	$N(1 - e^{-\lambda t})e^{-\lambda t}$	$Nr(1 - e^{-at}) \cdot (1-r + re^{-at})$	$Nr(1 - e^{-\alpha t^{\beta}}) \cdot (1-r + re^{-\alpha t\beta})$
T_k	(12)	$-\frac{1}{\lambda}\ln(1 - \frac{k}{N})$	$-\frac{1}{a}\ln(1 - \frac{k}{Nr})$	$\left[-\frac{1}{\alpha}\ln(1 - \frac{k}{Nr})\right]^{\frac{1}{\beta}}$

TABLE II

NUMERICAL RESULTS

PERFORMANCE MEASURE		STOLLMACK-HARRIS ($\lambda = .05$)	MALTZ-MCCLEARY ($a = .12$)	WEIBULL ($\alpha = .28$, $\beta = .50$)
r		1.0	0.5	0.5
$t_{F(.5)}$	(months)	13.9	6.0	6.0
$\bar{t}_F$		20.0	8.3	25.5
$t^*(.1)$	(months)	N.A.	18.3	61.6
$t^*(.2)$		N.A.	11.6	24.5
$t^*(.3)$		N.A.	7.1	9.2
$t^*(.4)$		N.A.	3.4	2.1
T_1	(months)	2.1	0.2	0.0
T_{20}		4.5	4.3	3.3
T_{40}		10.2	13.4	33.0

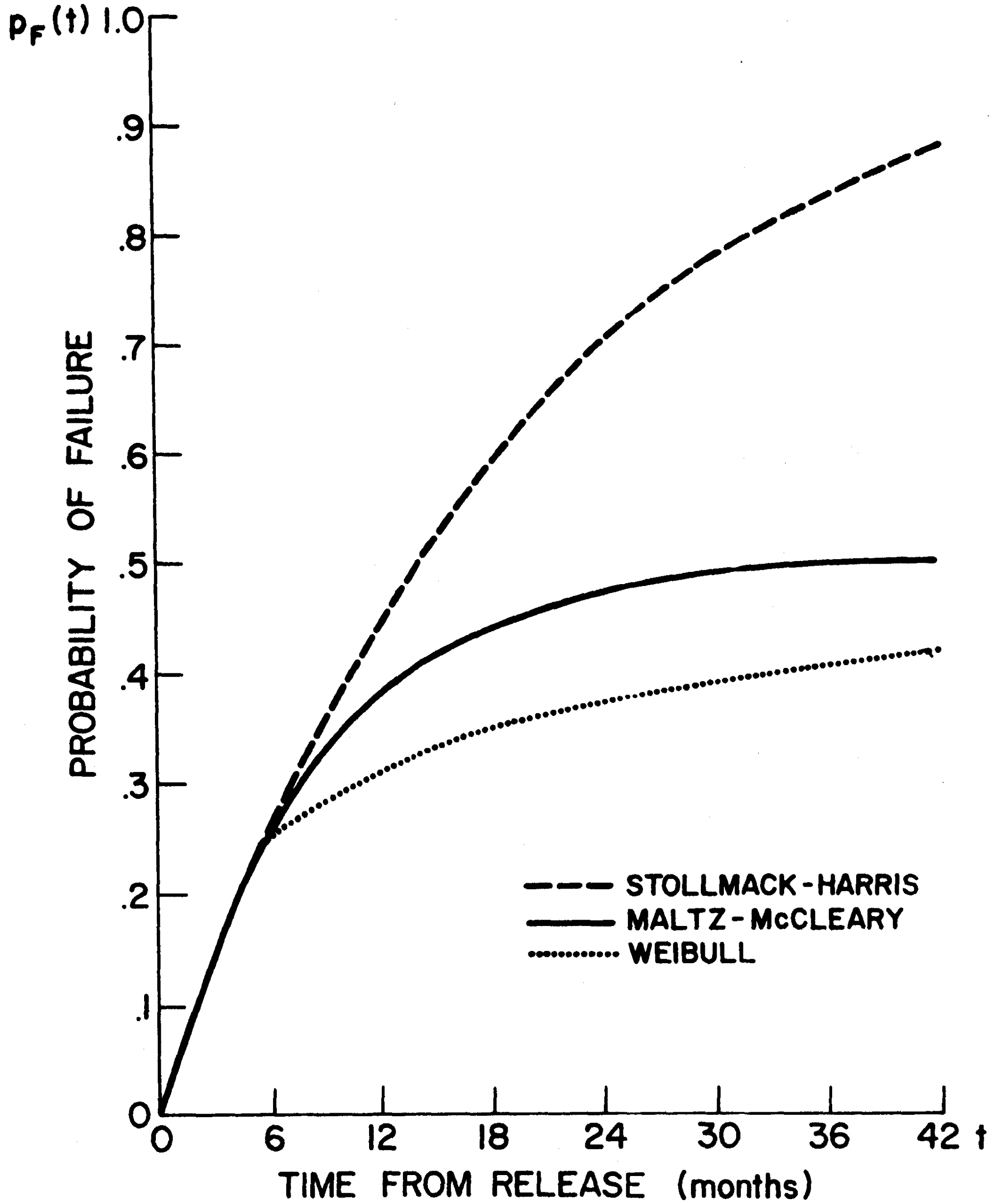


Figure 3

Failure Probability Over Time

Recalling that an evaluation is to be performed on the basis of 24 months of exposure data, the model-based values of $p_F(24)$ are useful indices for shaping prior expectations of program performance. As an example, our Weibull model demonstrates that even if $p_F(24) = .37 < .40$, the criterion set _a priori_ by program administrators, the _long run_ probability of re-arrest can equal .50. Thus, program success in the short run is not inconsistent with long run program failure; evaluations of such programs should take this possibility into account.

If we now consider the timing of rearrests, some differences in model behavior are noteworthy. Although the Maltz-McCleary and Weibull models produce equivalent _median_ times until failure, the Weibull expected time until failure is more than three times that of the Maltz-McCleary model. Of the three models considered, the Maltz-McCleary model clearly exhibits the most rapid failure process over time for ultimate failures; this is best reflected by Figure 3.

Figure 4 presents a graph of the safety time $t^*(\varepsilon)$ versus ε for the Maltz-McCleary and Weibull models. The Weibull model is clearly conservative in its implications. Only after an arrest-free release of 61.6 months can one be 90% certain that a client will not fail according to the Weibull model. At an equivalent 90% confidence level, the Maltz-McCleary model requires 18.3 months of arrest free releases.

To examine the uncertainty associated with these models, the variance of the number of failures is plotted as a function of time in Figure 5 for each of our three sets of assumptions. All three models reach the maximum achievable variance of N/4, since $p_F(t)$ approaches 1/2 as t approaches infinity for the Maltz-McCleary and Weibull models,

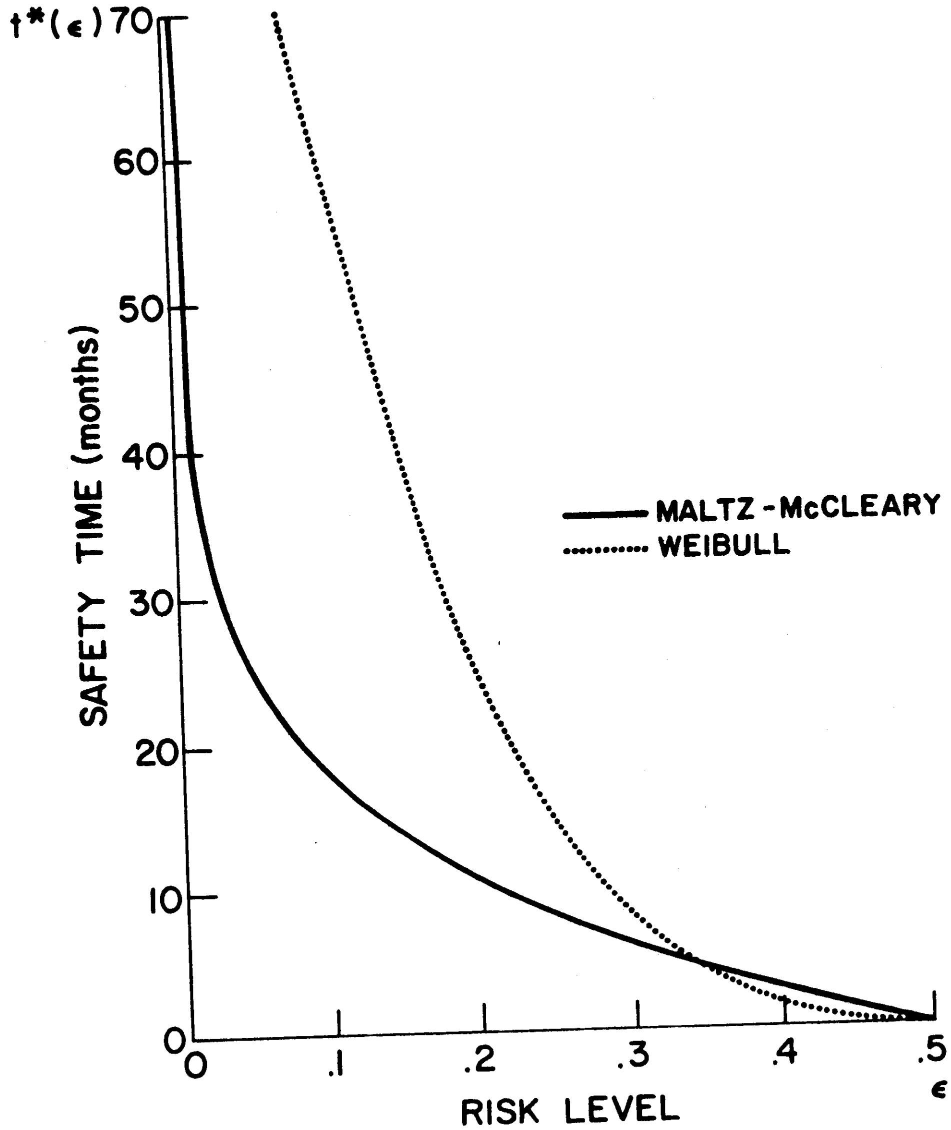


Figure 4

Safety Time Versus Risk Level

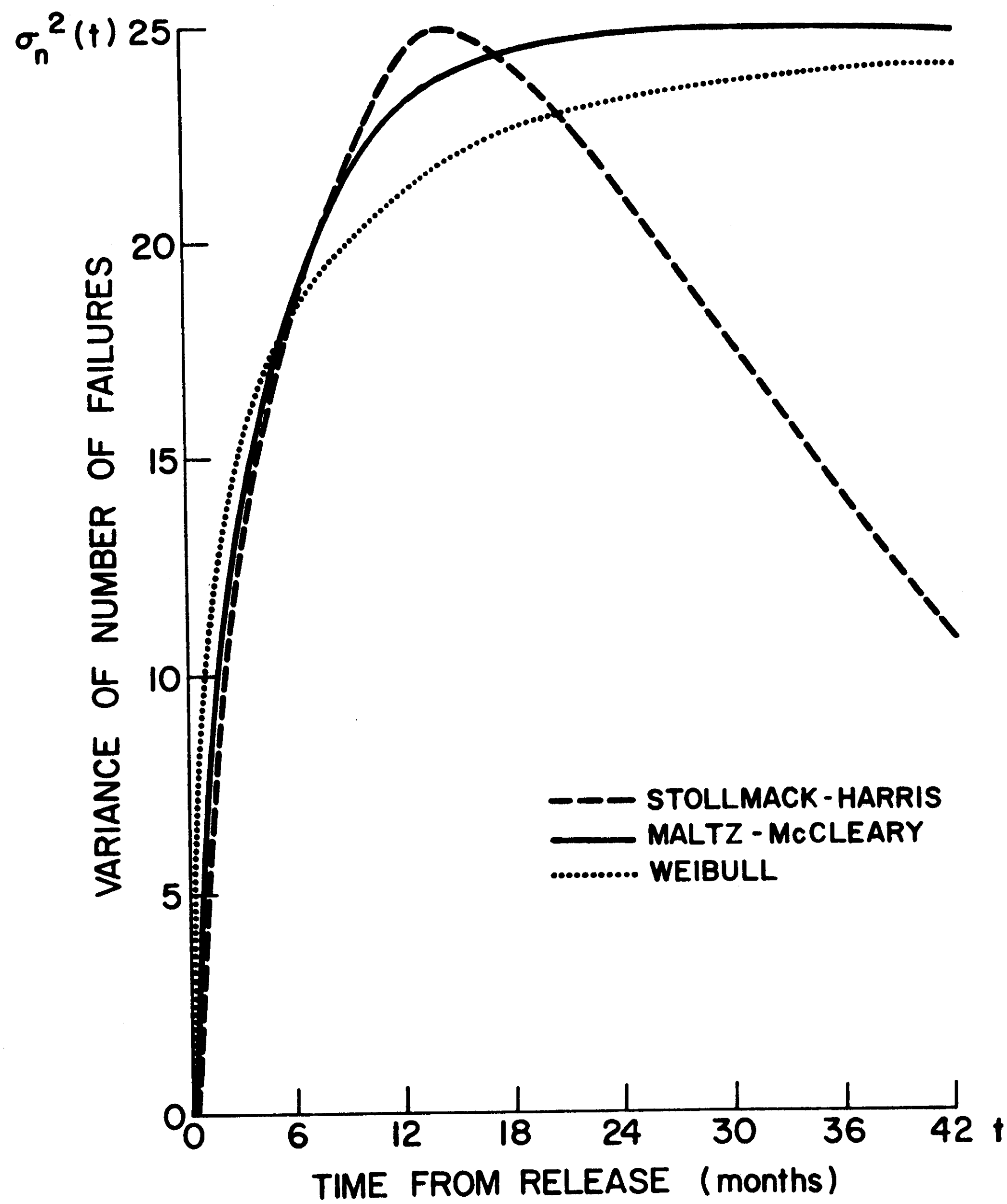


Figure 5

Variance of Failure Population Over Time while $p_F(13.9) = 1/2$ for the Stollmack-Harris model.

Finally, Figures 6, 7 and 8 present the Gaussian approximations to $P_n(t)$ for $t = 6$, 24 and ∞. At $t = 6$, all three models are identical. The distributions are quite distinct at $t = 24$, with the Stollmack-Harris model translated the furthest to the right, and the Weibull model the furthest to the left. As t approaches infinity, the Stollmack-Harris model produces an infinite spike at $n(\infty) = N$. Of course, the Maltz-McCleary and Weibull models reproduce each other for this case.

It is apparent that the behavior of the models can be drastically different at various points in time. This stems from the alternative formulations of $p_F(t)$ which in turn depend upon the assumptions governing the behavior of $\phi(t|F)$, the conditional hazard function. While the models demonstrated here do not by any means exhaust the world of possible models, they do illustrate the different types of model behavior achievable via the specification of alternative hazard functions.

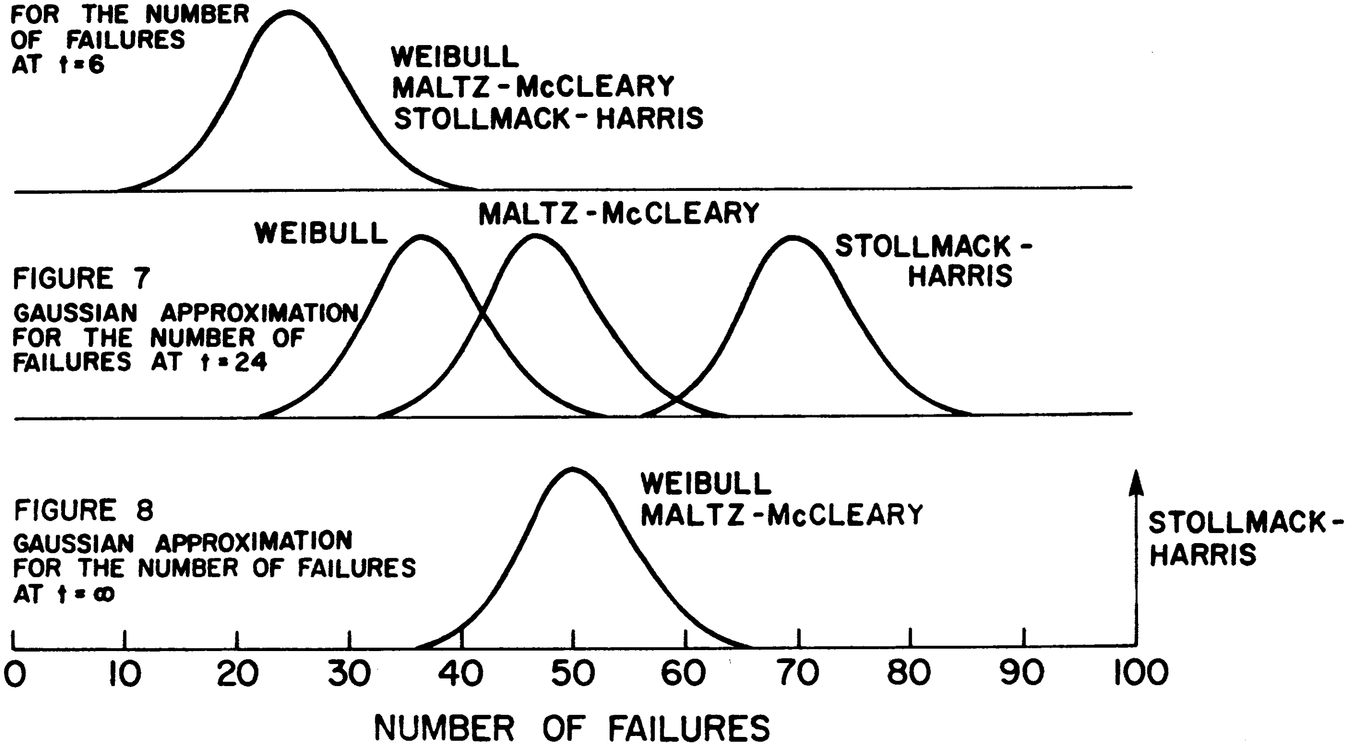


FIGURE 6
GAUSSIAN APPROXIMATION FOR THE NUMBER OF FAILURES AT t=6

FIGURE 7
GAUSSIAN APPROXIMATION FOR THE NUMBER OF FAILURES AT t=24

FIGURE 8
GAUSSIAN APPROXIMATION FOR THE NUMBER OF FAILURES AT t=∞

IV. ESTIMATION OF PARAMETERS FOR THE REARREST MODEL

In order to utilize in practice the general formulation developed thus far, we need to consider some reasonable techniques for estimating r and the parameters of $\phi(t|F)$. To this end, both maximum likelihood and Bayesian methods are suggested. The data aggregation scheme presented in conjunction with this discussion is attributable to Stollmack and Harris (1974).

Recall our discussion of Figure 1 from Section II. Suppose that of the N individuals who were released, k have been rearrested by the time we begin our analysis. If we let t_i correspond to the time from release until the i^{th} failure (i = 1,2,..,k), then the likelihood of observing these k failures at the times they occurred under the conditions of our model is given by

$$L\{k, \underset{\sim}{t}\} = \prod_{i=1}^{k} r\, \phi(t_i|F) e^{-\int_0^{t_i} \phi(x|F)\,dx} . \qquad (13)$$

Similarly, let τ_j represent the time from release that the j^{th} client (j = 1,2,..,N-k) has been observed <u>to remain unarrested.</u> The probability of observing this combination of the N-k success times is given by

$$L\{N\text{-}k, \underset{\sim}{\tau}\} = \prod_{j=1}^{N-k} (1 - r + re^{-\int_0^{\tau_j} \phi(x|F)\,dx}). \qquad (14)$$

(i) Maximum Likelihood Estimation

Let $\phi(t;\underset{\sim}{\psi}|F)$ denote the conditional hazard function where $\underset{\sim}{\psi}$ is the set of parameters contained in this function (e.g. for the Weibull model, $\underset{\sim}{\psi} = \{\alpha, \beta\}$). The overall likelihood of observing a particular pattern of k failures over time according to our model is given by

$$L\{k, N, \underset{\sim}{t}, \underset{\sim}{\tau} | r, \underset{\sim}{\psi}\} = \left[\prod_{i=1}^{k} r\phi(t_i; \underset{\sim}{\psi}|F) e^{-\int_0^{t_i} \phi(x; \underset{\sim}{\psi}|F)dx} \right] \cdot \left[\prod_{j=1}^{N-k} (1-r + re^{-\int_0^{\tau_j} \phi(x; \underset{\sim}{\psi}|F)dx}) \right]. \qquad (15)$$

To find the maximum likelihood estimates for r and $\underset{\sim}{\psi}$, one must solve the optimization problem

$$\max_{r, \underset{\sim}{\psi}} L\{k, N, \underset{\sim}{t}, \underset{\sim}{\tau} | r, \underset{\sim}{\psi}\} \qquad (16)$$

subject to $0 < r < 1$, constraints on $\underset{\sim}{\psi}$

In general, this is not an easy problem to solve. Maximum likelihood estimates have been obtained analytically for the Stollmack-Harris model (Stollmack-Harris, 1974), numerically for Bloom's model (Bloom, 1978) and numerically for the Maltz-McCleary model under the special condition that $\tau_j = \tau$, $j = 1, 2,..,N-k$ (Maltz and McCleary, 1977). To obtain maximum likelihood estimates for more complicated forms of $\phi(t|F)$ may require the use of non-linear programming routines.

(ii) <u>Bayesian Estimation</u>

In Bayesian analysis, we allow both subjective and objective information to play a role in our model (Freund, 1971: 280-281). The parameters of interest, r and $\underset{\sim}{\psi}$, are not viewed as being fixed and unchanging as is the case with classical techniques such as maximum likelihood estimation. Rather, r and $\underset{\sim}{\psi}$ are assumed to behave as random variables with *a priori* probability distributions. These distributions may be objectively or subjectively derived.

As an example, assume for the moment that r and ψ are independent within the Bayesian scheme of things. A corrections program official might make a series of statements of the form:

"The probability that some fraction not exceeding r of our clients will fail is given by F(r)."

These statements may be based on both the past experience of other programs and personal convictions regarding the likelihood of program success. This subjective cumulative distribution F(r) is then converted to a density function by examining successive differences (e.g., F(.1) - F(.05), F(.15) - F(.1), ... F(1.0) - F(.95)) and fitting a curve to the resultant histogram. Such a prior distribution f(r) is interpreted as the probability distribution of ultimate failure likelihoods across the population of programs similar to the one in question. The true value of r that will actually be observed is treated as a random selection from this population.

Suppose that a program administrator specified the following values for r and F(r):

r	F(r)
.2	.10
.4	.50
.6	.80
.8	.95
1.0	1.00

The prior distribution graphed in Figure 9 would result. This distribution formally represents prior expectations of ultimate rearrest probabilities.

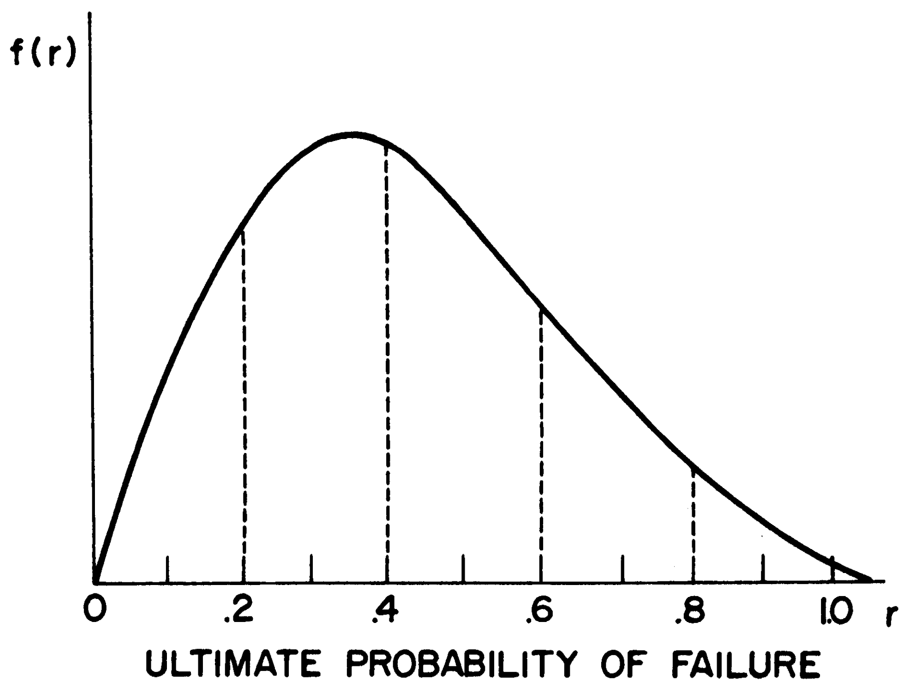


Figure 9

Prior Distribution of the Ultimate
Probability of Failure

Let $f(r, \underset{\sim}{\psi})$ represent the joint prior distribution of r and $\underset{\sim}{\psi}$. Since r and $\underset{\sim}{\psi}$ are now being viewed as random variables, the likelihood function of equation (15) is nothing more than the conditional probability of having observed a particular rearrest pattern given specific values of r and $\underset{\sim}{\psi}$. What we wish to compute is the joint conditional distribution of r and $\underset{\sim}{\psi}$ having observed a particular rearrest pattern characterized by k failures out of N released clients, and times to failure $\underset{\sim}{t}$, observed success times $\underset{\sim}{\tau}$.

This posterior distribution of r and $\underset{\sim}{\psi}$ is formulated via the use of Bayes' Rule (Freund, 1971: 280-281)

$$g(r, \underset{\sim}{\psi}|k, N, \underset{\sim}{t}, \underset{\sim}{\tau}) = \frac{f(r, \underset{\sim}{\psi})L\{k, N, \underset{\sim}{t}, \underset{\sim}{\tau}|r, \underset{\sim}{\psi}\}}{R(k, N, \underset{\sim}{t}, \underset{\sim}{\tau})} \tag{17}$$

where $R(k, N, \underset{\sim}{t}, \underset{\sim}{\tau})$ is a normalizing constant. The posterior distribution of r may be found by integrating out over $\underset{\sim}{\psi}$

$$h(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau}) = \int_{\psi_1 \varepsilon \underset{\sim}{\psi}} \cdots\cdots \int_{\psi_n \varepsilon \underset{\sim}{\psi}} g(r, \underset{\sim}{\psi}|k, N, \underset{\sim}{t}, \underset{\sim}{\tau})d\psi_1 \ldots d\psi_n . \tag{18}$$

The posterior expected value of r is then found by computing

$$E(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau}) = \int_{r=0}^{1} r\, h(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau})dr. \tag{19}$$

Similarly, the posterior expected values of each of the performance measures discussed in Section II can be found. Let a generic performance measure be denoted by $M(r, \underset{\sim}{\psi})$; the posterior expected value of this measure is given by

$$E[M(r, \underset{\sim}{\psi}) | k, N, \underset{\sim}{t}, \underset{\sim}{\tau}] = \int_{r=0}^{1} \int_{\psi_1 \varepsilon \underset{\sim}{\psi}} \cdots \int_{\psi_n \varepsilon \underset{\sim}{\psi}} M(r,\psi) g(r,\underset{\sim}{\psi} | k,N,\underset{\sim}{t},\underset{\sim}{\tau}) d\psi_1 \ldots d\psi_n \, dr. \quad (20)$$

Now, since the calculations in the equations of this Bayesian analysis require nothing more complex than integration, the Bayesian estimates discussed can be obtained numerically. It would be possible to write a computer program to perform these calculations for any given function $\phi(t|F)$, though the design of such a program has not yet been attempted.

V. EVALUATION ISSUES

In the introduction to this paper, we stressed the difference that exists between observed rearrest rates and true recidivism rates. Since we have not attempted to account for the relative influence the police have on the rearrest process, the models which have been discussed should be used in conjunction with controlled evaluation designs (see Campbell and Stanley, 1966) if the comparison of two corrections programs is being pursued.

The manner in which our model is used for evaluation purposes depends upon whether the estimation approach chosen is classical or Bayesian. The differences resulting from these alternative approaches are illustrated throughout.

The first measure of evaluative interest is the ultimate rearrest probability r. In general, program success is seen to vary inversely with the value of r.[3] It may be established a priori by program officials that one program target is the achievement of an r value less than some desired tolerance level r^*. If r was estimated via maximum likelihood techniques, then for large N, the null hypothesis H_o: $r = r^*$ may be tested using

$$Z = \frac{\hat{r} - r^*}{\sqrt{\frac{\hat{r}(1-\hat{r})}{N}}} \tag{21}$$

where $\hat{r}$ is the maximum likelihood estimate of r, and Z is distributed as a standardized Gaussian random variable (Freund, 1971:329).[4] The rationale for this test stems from the Gaussian approximation $f_{n(t)}(x)$ discussed in Section II.

The Bayesian version of this test would consist of computing the probability that r is truly less than r^*. This is computed as

$$\Pr\{r < r^* | k, N, \underset{\sim}{t}, \underset{\sim}{\tau}\} = \int_0^{r^*} h(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau})\, dr \qquad (22)$$

where $h(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau})$ is as defined in Section IV. What constitutes an acceptable likelihood of program success in this instance is a decision problem for program officials.

One would also be interested in the average rearrest probability. To this end, $E(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau})$ may be found through use of equation (19). A computed value of $E(r|k, N, \underset{\sim}{t}, \underset{\sim}{\tau}) < r^*$ is indicative of program success.

The procedure just presented may also prove useful as indicators of whether or not additional data collection is necessary during the life of a program. Suppose that after some initial fixed period of data collection, k failures out of N releases have occurred. For the classical procedure, compute the maximum likelihood estimate $\hat{r}$, and substitute k/n for r^* in equation (21). If there appears to be no significant difference between $\hat{r}$ and k/N, then perhaps there is no need to continue collecting data, and resources available for this segment of the evaluation may be channeled to other evaluation tasks (e.g. interviews with program clients). The Bayesian analogy consists of substituting k/N for r^* in equation (22); if $\Pr\{r < k/N|k, N, \underset{\sim}{t}, \underset{\sim}{\tau}\}$ is relatively large, then this may also be an appropriate signal to end data collection.

Conversely, if there is a strong disagreement between the observed fraction of failures and the estimated ultimate fraction of failures, maybe more data should be collected, even if the time by which k rearrests have occurred corresponds to the scheduled completion date for the data collection effort. For an example of this sort, data extracted from Pre-Trial Intervention: A Program Evaluation of Nine Manpower-based Pre-Trial Intervention Projects revealed that after one year of release time, 18.3% of all released clients had been rearrested, yet application of the Maltz-McCleary model to this same data yielded a maximum likelihood estimate of $\hat{r} = .43$ (Kaplan, 1978b:23). Since $\hat{r}$ is almost 2 1/2 times as large as k/N in this instance, it might have been a good idea to sustain the data collection effort for this evaluation; if models like those presented here had been available to these evaluators, this finding could have been discovered during the data collection phase.

If we now consider the case where two programs are being compared in a controlled environment, a number of our model-based performance measures may be utilized. Again focusing our attention on the ultimate probability of rearrest r, the null hypothesis $H_o: r_1 = r_2$ may be tested using

$$Z = \frac{\hat{r}_1 - \hat{r}_2}{\sqrt{\frac{\hat{r}_1(1 - \hat{r}_1)}{N_1} + \frac{\hat{r}_2(1 - \hat{r}_2)}{N_2}}} \qquad (23)$$

where:

N_1, N_2 = client populations of the programs;

$\hat{r}_1$, $\hat{r}_2$ = maximum likelihood estimates of the ultimate failure probabilities;

Z = a standardized Gausian random variable.

As with all of the procedures we have discussed which are dependent upon the Gausian approximation $f_{n(t)}(x)$, this test is valid only if N_1 and N_2 are large (Freund, 1971: 332).

The comparison of two programs from a Bayesian perspective using r commends a more graphic analysis. Essentially, the posterior distributions of r for each program may be plotted on the same figure; such a presentation provides a visual method for comparing program performance. An example of such a plot is shown in Figure 10.

The Bayesian approach does allow for numerical comparisons as well. The simplest of such comparisons would be to compute the expected posterior probabilities of ultimate rearrest using equation (19), and check to see which program produced the lower value. A more meaningful comparison involves finding the likelihood that one program produced a lower probability of rearrest than the other program. If we let

$$\underset{\sim}{A}_1 \equiv \{k_1, N_1, \underset{\sim}{t}_1, \underset{\sim}{T}_1\}$$
$$\underset{\sim}{A}_2 \equiv \{k_2, N_2, \underset{\sim}{t}_2, \underset{\sim}{T}_2\}$$

where the subscripts denote program one and two, then the expected probability that $r_1 < r_2$ is computed as

$$\Pr\{r_1 < r_2 | \underset{\sim}{A}_1, \underset{\sim}{A}_2\} = \int_{r_2=0}^{1} \int_{r_1=0}^{r_2} h_1(r_1|\underset{\sim}{A}_1) h_2(r_2|\underset{\sim}{A}_2) dr_1 dr_2 \tag{24}$$

where $h(r|\underset{\sim}{A})$ is as defined in equation (18). Conversely, $\Pr\{r_1 > r_2|\underset{\sim}{A}_1, \underset{\sim}{A}_2\} = 1 - \Pr\{r_1 < r_2|\underset{\sim}{A}_1, \underset{\sim}{A}_2\}$. If the result of equation (24) is greater than 1/2, then it would appear that program one has outperformed program two using r as a performance measure.

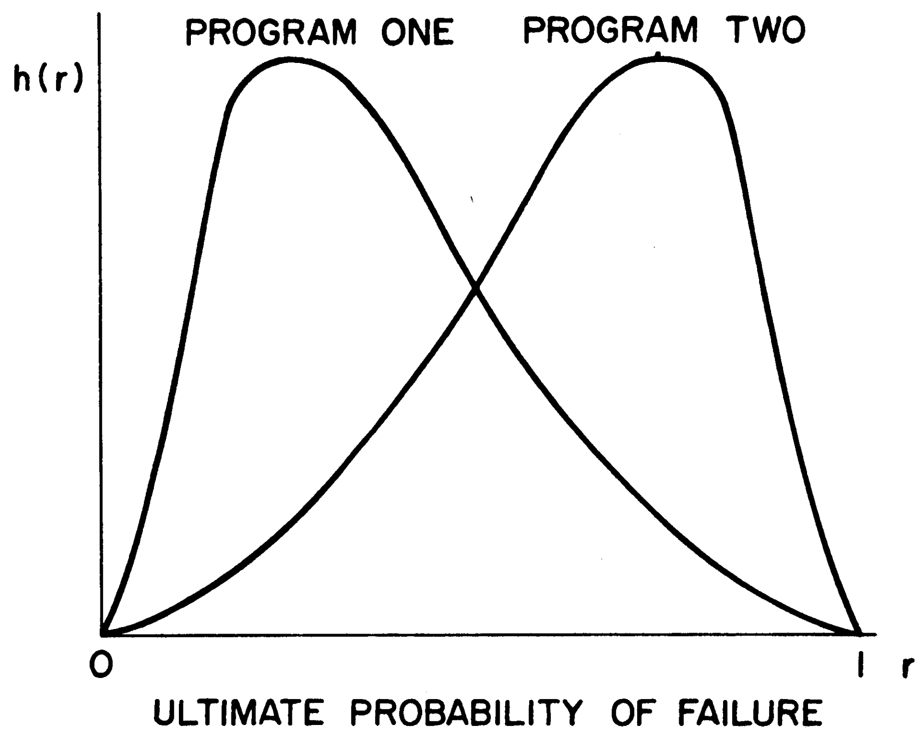


Figure 10

Comparing the Posterior Distributions
Of Ultimate Failure Probabilities

When comparing two programs, the timing of rearrests becomes important. For example, if two programs produced equivalent values for r, one could argue that the program with the lesser value for $\bar{t}_F$ (or $E(\bar{t}_F)$ if you're a Bayesian) was the more successful since such a program quickly distinguishes ultimate failures from the rest of the client population. Indeed, Maltz and McCleary recognized this possibility when they wrote that "Knowledge of a program's failure rate can help in matching programs to participants" (Maltz and McCleary, 1977: 432); in the example presented here, clients who are felt likely to fail _a priori_ by program officials could be assigned to the "quick failure given ultimate failure" program to the benefit of the other program participants (Maltz and McCleary, 1977).

The use of time until failure measures has process implications as well. In the example concerning the evaluation of pre-trial intervention projects presented earlier, it was found through application of the Maltz-McCleary model that the median time until failure for ultimate rearrests was equal to 462 days; thus the one year data collection effort terminated before 50% of all ultimate rearrests had occurred! Again, had this calculation been performed, it could have been seen as a signal to prolong the data collection phase of the evaluation (Kaplan, 1978b:24).

To compare the timing of rearrests resulting from two programs with different client populations, the measure T_k (or $E(T_k)$ for Bayesians) is useful. One can examine the _fractile_ times until failure to perform a relative comparison. Suppose we are interested in the time it takes until a fraction q of the population of ultimate failures has failed. For both of the programs being compared, compute

$$T_{k_i}|q \equiv \bar{n}^{-1}(N_i \cdot r_i \cdot q) \qquad i = 1, 2, \quad 0 < q < 1 \qquad (25)$$

for various values of q between 0 and 1. The values of $T_{k_i}|q$ (or $E(T_{k_i}|q)$) are then comparable for similar values of q (note that for q = 1/2, $T_{k_i}|q = t_{F(.5)}$, the median time until failure).

The final evaluation measure we will consider for the comparison of two programs is the safety time $t^*(\varepsilon)$. The values of t^* (or $E(t^*)$) as a function of ε may be plotted for each program on the same graph. It is then possible to check for dominance. Consider Figure 11 where $t^*(\varepsilon)$ has been plotted for two hypothetical programs. Here it is clear that Program A dominates Program B, since for any value of ε, $t^*_A < t^*_B$. To assert with confidence (1-ε) that an individual will not be arrested given that (s)he hasn't failed by t^* will always require a longer time from release for individuals in Program B than for individuals in Program A. Of course, it is possible for partial dominance to occur; A could dominate B for low values of ε, while B could dominate A for high values of ε. For evaluation purposes, dominance over low values of ε characterizes a successful program.

It should be noted that in our discussion of the methods of this section, no specific form was assumed for $\phi(t|F)$. In fact, different hazard functions could be engaged for different programs, and the comparative procedures discussed here could still be invoked. Also, as mentioned by Maltz and McCleary (1977: 432), it is not necessary for programs to exist for the same length of time in order to use these techniques.

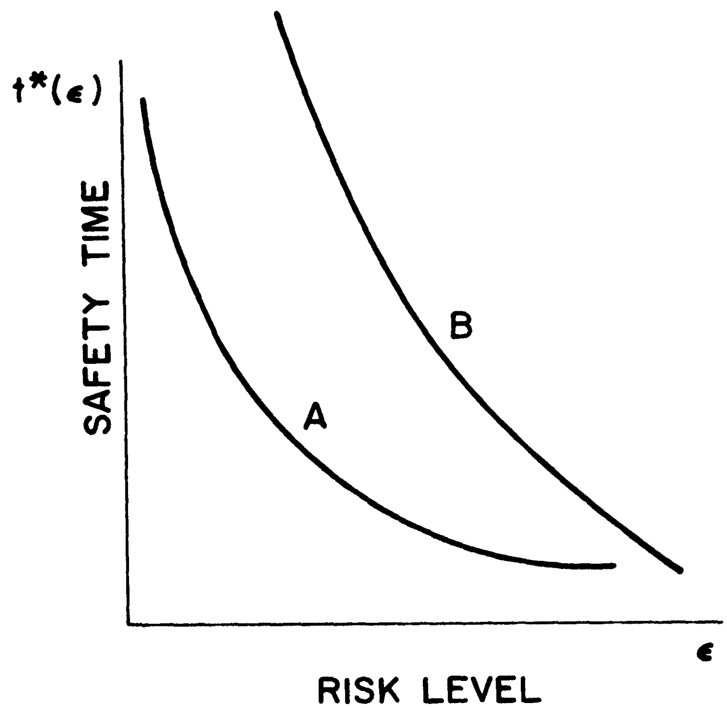


Figure 11

Program A Dominates Program B

VI. SUMMARY AND EXTENSIONS

This paper has analyzed the common structure shared by treatment-release corrections programs within the framework of reliability theory. Having presented a general model of the rearrest process, we examined the performance of this model under alternative assumptions, and illustrated appropriate techniques for estimating model parameters. We then discussed classical and Bayesian model-based evaluation procedures for use in both process and outcome situations.

While the substantive focus of this paper has been on models for rearrest patterns, it should be noted that the mathematics involved are appropriate for generic failure problems. Thus, if one was interested in performance measures based on alternative failure patterns over time, the reliability models of this paper could prove useful. For example, suppose one wished to judge a program participant as a failure only if that client was reconvicted. The time until failure would then correspond to the time until reconviction. Thus, the models presented here are responsive to the notion that different types of programs may require different definitions of client failure for evaluation purposes.

FOOTNOTES

[1]Reliability theory addresses problems associated with the failure of systems (or system components) over time.

[2]The pdf is a pseudo-pdf since $\int_0^\infty f_{t_F}(t)dt = r$, and r is in general not equal to one.

[3]While r is often considered to be a fundamental performance measure, the model presented by Blumstein and Larson (1971) suggests that 1/1-r is a more readily interpretable performance measure. The expression 1/1-r represents the number of future crimes committed per individual after release from treatment in the Blumstein/Larson model; this expression is very sensitive to changes in the value of r when r is close to one.

[4]It is assumed that the reader is familiar with the procedures of hypothesis testing; a good discussion of hypothesis testing is found in Chapters 10 through 12 of Freund (1971).

REFERENCES

1. Barnett, A. and Stabile, D., (1979). Estimating the "Virtual" Incapacitation Time for those Released from Corrections Programs. (Forthcoming).

2. Bloom, H.S., (1978). Evaluating Human Service and Criminal Justice Programs by Modeling the Probability and Timing of Recidivism. Discussion Paper D78-9, Department of City and Regional Planning, Harvard University.

3. Blumstein, A. and Larson, R.C., (1971). Problems in Modeling and Measuring Recidivism. Journal of Research on Crime and Delinquency 8: 124-132.

4. Campbell, D.T. and Stanley, J.C., (1966). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand-McNally.

5. Freund, J.E., (1971). Mathematical Statistics. Englewood Cliffs: Prentice-Hall, Inc.

6. Hillier, F.S. and Lieberman, G.J., (1974). Operations Research. San Francisco: Holden-Day, Inc.

7. Kaplan, E.H., (1978a). Evaluating the Effectiveness of One- versus Two-Officer Patrol Units. To appear in the Journal of Criminal Justice.

8. Kaplan, E.H., (1978b). Nine Pre-Trial Intervention Programs: A Re-Analysis. Unpublished paper, Department of Urban Studies and Planning, M.I.T.

9. Larson, R.C., (1975). What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment. Journal of Criminal Justice, 3: 267-297.

10. Maltz, M. and McCleary, R. (1977). The Mathematics of Behavioral Change: Recidivism and Construct Validity. Evaluation Quarterly, 1: 421-438.

11. Mullen, J., (Project Director) (1974). Pre-Trial Intervention: A Program Evaluation of Nine Manpower-based Pre-Trial Intervention Projects. Cambridge, Massachusetts: Abt Associates, Inc.

12. Purcell, E.J., (1972). Calculus with Analytic Geometry. Englewood Cliffs: Prentice-Hall, Inc.

13. Tsokos, C.P., (1972). Probability Distributions: An Introduction to Probability Theory with Applications. Belmont, California: Duxbury Press.

Appendix: Assymptotic Equivalence of the Bloom and Maltz-McCleary Models

As mentioned in Section II, Bloom (1978) has proposed a model involving an exponential hazard function. We also stated that Bloom's model was operationally equivalent to the Maltz-McCleary model. In this appendix, we will explain why this is so.

Bloom did not rely upon the notion of a conditional hazard function when formulating his model, as he rejected the assumption that some fraction r of the population released could be conceived a priori as consisting of ultimate failures (Bloom, 1978: 4). Rather, Bloom defined an unconditional hazard function for his model of the form (Bloom, 1978: 6)

$$\phi(t) = be^{-ct}, \; b, \; c > 0 \qquad \text{(A1)}$$

In Bloom's model, $\phi(t)$ represents the likelihood that an individual will fail in the next time instant given release at $t_R = 0$. Note that Bloom does not explicitly restrict the application of his hazard function to ultimate failures.

To obtain an expression for $p_F(t)$ given the hazard function of equation (A1), the well-known reliability result of equation (1) may again be invoked yielding (Bloom, 1978: 6)

$$p_F(t) = 1 - e^{-\frac{b}{c}} \, e^{\frac{b}{c} e^{-ct}} \qquad b, \; c > 0, \; t > 0 \qquad \text{(A2)}$$

This model has some interesting properties; foremost among these is the fact that by setting $t = \infty$ in equation (A2), one arrives at the expression (Bloom, 1978: 7)

$$p_F(\infty) = 1 - e^{-\frac{b}{c}} \quad b,\ c > 0 \tag{A3}$$

Thus, Bloom's model implies that some fraction $1 - e^{-\frac{b}{c}}$ of the population released will ultimately fail. Bloom does interpret equation (A3) as an expression for the "ultimate probability of failure" (Bloom, 1978: 7), yet he rejects the notion that one may assume the a priori existence of an ultimate rearrest probability r. This distinction is at best artificial, as r and $1 - e^{-\frac{b}{c}}$ are both constants.

Recall that for the Maltz-McCleary model, $\phi(t|F) = a$, a positive constant. To see how the Bloom model asymptotically approaches the Maltz-McCleary model, we will formulate the conditional hazard function $\phi(t|F)$ for Bloom's model, and show that as t approaches infinity, $\phi(t|F)$ approaches a positive constant. If we define $p_F(t|F)$ as the probability of failure by time t given release at $t_R = 0$ and ultimate failure, then for Bloom's model,

$$p_F(t|F) = \frac{1 - e^{-\frac{b}{c}} e^{\frac{b}{c} e^{-ct}}}{1 - e^{-\frac{b}{c}}} \quad b,\ c > 0,\ t > 0 \tag{A4}$$

Differentiating (A4) with respect to time from release to obtain $f_{t_F}(t|F)$, the conditional pdf for time until failure given ultimate failure, yields

$$f_{t_F}(t|F) = \frac{be^{-ct}\, e^{-\frac{b}{c}} e^{\frac{b}{c} e^{-ct}}}{1 - e^{-\frac{b}{c}}} \quad b,\ c > 0,\ t \geq 0 \tag{A5}$$

Now, the probability that an <u>ultimate failure</u> will be rearrested in the next time instant conditioned on the event that (s)he has not failed by time t is given by

$$\phi(t|F)dt = \frac{\dfrac{be^{-ct}\, e^{-\frac{b}{c}}\, e^{\frac{b}{c} e^{-ct}}\, dt}{1 - e^{-\frac{b}{c}}}}{1 - \dfrac{1 - e^{-\frac{b}{c}}\, e^{\frac{b}{c} e^{-ct}}}{1 - e^{-\frac{b}{c}}}},$$

which algebraically reduces to

$$\phi(t|F) = \frac{be^{-ct}\, e^{\frac{b}{c} e^{-ct}}}{e^{\frac{b}{c} e^{-ct}} - 1} \qquad b,\ c > 0,\ t > 0 \tag{A6}$$

If we examine the limit of $\phi(t|F)$ as t approaches infinity, we realize that we cannot evaluate this limit directly since both the numerator and denominator of (A6) approach 0 as t approaches infinity. Applying L'Hôpital's rule (Purcell, 1972: 562), we obtain

$$\lim_{t \to \infty} \phi(t|F) = \lim_{t \to \infty} \frac{\frac{d}{dt}[be^{-ct}\, e^{\frac{b}{c} e^{-ct}}]}{\frac{d}{dt}[e^{\frac{b}{c} e^{-ct}} - 1]} \tag{A7}$$

$$= \lim\ [c + be^{-ct}]$$

$$= c \qquad\qquad c > 0.$$

Thus, the conditional hazard function for Bloom's model does approach a positive constant over time.

To illustrate the operational similarities between Bloom's model and the Maltz-McCleary model, we will return to our example of Section III. First, we impose the same restrictions on Bloom's model as those that were imposed on the Maltz-McCleary model:

(i) $p_F(6) = .2$

(ii) $p_F(\infty) = .50$.

To satisfy restriction (ii), we use (A3) to obtain

$$b = -c \ln (.50). \qquad \text{(A8)}$$

Substitution of (A8) into Bloom's expression for $p_F(t)$ given in (A2) combines with restriction (i) to produce the result c = .089. Placing this value for c in equation (A8) yields b = .062. These values of b and c may be used with equation (A2) to produce the results shown in Table AI. It is evident from Table Al that these models behave in equivalent fashions.

Hence, it is not surprising that Bloom found the performance of his model and the Maltz-McCleary model to be operationally equivalent, despite their mathematical differences (Bloom, 1978: 16). When applied to the same data set, these two models will produce comparable results.

TABLE AI

FAILURE PROBABILITIES FOR THE MALTZ-MCCLEARY AND BLOOM MODELS:
THE EXAMPLE OF SECTION III

Time from Release (months)	Maltz-McCleary ($r = .5$, $a = .12$)	Bloom ($b = .062$, $c = .089$)
0	0.00	0.00
6	0.25	0.25
12	0.38	0.36
18	0.44	0.42
24	0.47	0.46
30	0.49	0.48
36	0.49	0.49
42	0.50	0.49