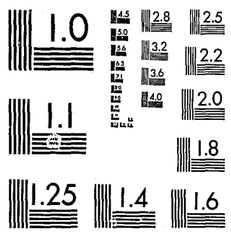
If you have issues viewing or accessing this file contact us at NCJRS.gov.

National Criminal Justice Reference Service

ncjrs

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504.

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U. S. Department of Justice.

National Institute of Justice
United States Department of Justice
Washington, D.C. 20531

DATE FILMED

June 15, 1981

DJC DECISION SCIENCE CONSORTIUM, INC.

THE IRRELEVANCE OF EVALUATION RESEARCH FOR DECISION MAKING: CASE STUDIES FROM THE COMMUNITY ANTI-CRIME PROGRAM

Kurt J. Snapper
and
David A. Seaver

7644

October 1980

Technical Report 80-12

U.S. Department of Justice
National Institute of Justice

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this copyrighted material has been granted by
Public Domain
U.S. Dept. of Justice

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the copyright owner.

Technical Report 80-12

THE IRRELEVANCE OF EVALUATION RESEARCH FOR DECISION MAKING: CASE STUDIES FROM THE COMMUNITY ANTI-CRIME PROGRAM

Kurt J. Snapper
and
David A. Seaver

October 1980

The research was supported under Grant No. 79NI-AX-0107 awarded to the Decision Science Consortium, Inc. by the Law Enforcement Assistance Administration, U. S. Department of Justice, pursuant to the Omnibus Crime Control and Safe Streets Act of 1968, as amended. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U. S. Department of Justice.

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, Virginia 22043
(703) 790-0510

NCJRS
APR 9 1981
ACQUISITIONS

ACKNOWLEDGEMENTS

The work described in this report involved the participation of more people and groups than we can properly acknowledge. These include personnel from several community anti-crime projects, the Office of Community Anti-Crime Programs, the National Institute of Justice, several members of House and Senate committee staffs, and colleagues at Research for Social Change, Inc. We will specially note Dr. Rex V. Brown, of DSC, who headed the decision analysis in Section 6.0 of this report; and Drs. Larry Bennett, Director of OPE, and Richard M. Rau, Project Monitor, whose collegial assistance we drew on heavily. We caution that this study could not have been completed without the persons and groups cited above, but that none of them is responsible for any errors or other shortcomings of this report.

ABSTRACT

Evaluations are designed to provide useful information including, often, information to aid programmatic decisions. This report discusses decision-analytic approaches to decision making, and some of the limitations and biases resulting from typical approaches to evaluation. We conclude that typical evaluation approaches are likely to seriously bias decisions against programs.

A purpose of the present study was to critique decision-theoretic approaches to evaluation, in terms of practicability and usefulness for decision making. Applications to project-level decisions and to the decision to refund programs administered by the Office of Community Anti-Crime Programs are discussed.

Methodologically, results suggest that decision-theoretic approaches can be successfully and usefully used. Programmatically, results suggest that community anti-crime approaches would have become highly cost-effective had they not been terminated, and that at least some projects achieved a surprisingly high degree of institutionalization.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

ABSTRACT iii

PART I -- INTRODUCTION

1.0 OVERVIEW OF STUDY 1-1

1.1 Methods for Improving Decision-Making Use of Evaluations . . 1-1

1.2 Purpose of the Project: Application of Decision-theoretic Approaches to Evaluation 1-2

1.3 What is the Decision-theoretic Approach? 1-3

2.0 BACKGROUND OF THE COMMUNITY ANTI-CRIME PROGRAM AND ITS EVALUATION. 2-1

2.1 Background of the CAC Program 2-1

2.2 Background of the CAC Program Evaluation 2-2

3.0 WHAT IS DECISION ANALYSIS? 3-1

3.1 Basic Concepts of Decision Analysis 3-1

3.2 Problems with "Traditional" Evaluation Research Methods. . . 3-5

PART II -- APPLICATIONS

4.0 THE GENERAL EVALUATION APPROACH 4-1

4.1 The Planner's Problem 4-1

4.2 The Basic Methodology 4-3

4.3 The Midwood-Kings Highway Development Corporation Evaluation 4-5

TABLE OF CONTENTS (continued)

		Page
5.0	THE MKDC DECISION ANALYSIS	5-1
5.1	Adapting the General Evaluation Approach for Decision Analysis	5-1
5.2	The MKDC Decision	5-5
5.3	Outcome	5-9
6.0	APPLICATION OF MAUT TO A PROGRAM DECISION: THE FUNDING LEVEL OF THE OFFICE OF COMMUNITY ANTI-CRIME PROGRAMS	6-1
6.1	Rationale for the Decision Model	6-2
6.2	Options for Future Funding	6-4
6.3	The MAUT Model for Assessing Benefits	6-4
6.4	Aggregate Measure of Benefits	6-5
6.5	Assessment Procedures	6-5
6.6	The Tradeoff Between Effectiveness and Cost	6-10
6.7	Results of Analysis	6-12
6.8	Sensitivity Analyses	6-17
6.9	Conclusion	6-20
7.0	POST MORTEM EVALUATIONS OF PROGRAMS	7-1
7.1	Institutionalization and Residual Effectiveness	7-2
7.2	The MACC Example	7-3
7.3	The Model for Assessing Residual Effectiveness	7-3
7.4	Data for MACC	7-8
7.5	The MACC "Decision Tree"	7-16
7.6	Implications of Methodology	7-23
8.0	THE DECISION TO TERMINATE THE SEPCPC PROJECT	8-1
8.1	The SEPCPC Example	8-1
8.2	SEPCPC's Decision	8-3

TABLE OF CONTENTS (continued)

		Page
PART III -- METHODOLOGICAL PERSPECTIVES		
9.0	SHORTCOMINGS OF EVALUATION RESEARCH FOR AIDING DECISIONS	9-1
9.1	The View that Evaluations are Empirical, Experimental Exercises	9-1
9.2	A Caveat	9-2
9.3	Deficiencies in Typical Evaluation Approaches	9-2
9.4	Biasing Factors in Programmatic Decision Making	9-5
9.5	The Concept of a Decision Aid	9-6
10.0	COST EFFECTIVENESS AND PROGRAM DECISION RULES	10-1
10.1	The Relevance of Cost-Effectiveness to Program Decisions	10-1
10.2	The Concept of Net Utility	10-2
10.3	Cumulative Net Utility	10-5
10.4	Cumulative Net Utility and Decision Biases	10-7
10.5	Net Utility Models as Decision Aids	10-9
11.0	CONCLUSIONS	11-1
12.0	REFERENCES	12-1

PART I -- INTRODUCTION

1.0 OVERVIEW OF STUDY

Evaluation research, having successfully passed from its infancy to adolescense, has reached a somewhat awkward stage. Many evaluation researchers are asking why evaluations are used primarily in the sense that they contribute to a general knowledge base, but are often not used for programmatic and policy decision making. That is, evaluations themselves often have little direct impact.

There is some controversy about the definition of evaluation research, its purpose, and the kinds of impacts evaluations themselves should have. Many evaluators (e.g., Apsler, 1977) define evaluations essentially as experiments; they tend to stress experimental methodology, typically drawing heavily on concepts stemming from Campbell and Stanley (1963).

Other researchers have tended to define evaluation more generally. Rossi, Wright, and Wright (1978), for example, argue that basic research is to enhance generalized knowledge in a discipline, but that applied research including evaluations is related to a real-world problem. Riecken (1977) suggested that evaluations are undertaken in regard to an action. Bernstein and Freeman (1975) and Rivlin (1971) suggest that evaluations should directly aid decision making, and that cost-benefit considerations may enter into the decision and, thus, should be addressed in the evaluation. And the General Accounting Office (1978) identified several classes of decision makers.

1.1 Methods for Improving Decision-Making Use of Evaluations

Despite rather general agreement that evaluations should affect decisions, there is little consensus about which methodological approaches are most appropriate for this purpose (e.g., General Accounting Office, 1976). It

can be argued, for example, that merely demonstrating that a statistical effect exists is not compelling evidence that a program is worthwhile. For instance, the fact that a decision exists implies that alternatives to the program are being considered. This further implies that the evaluation must be comparative, and (in the view of Rivlin and others) that this comparison must involve cost benefit tradeoffs. Moreover, the existence of a significant statistical effect is not a reliable indicator that an effect is worthwhile. Classical statistical methods are strongly biased against the null hypothesis, and in a sufficiently "strong" design will ensure that even trivial effects are significant (e.g., Bakan, 1967; Savage, 1954).

1.2 Purpose of the Project: Application of Decision-theoretic Approaches to Evaluation

Our colleague, the late Dr. Marcia Guttentag, first suggested the application of decision-theoretic approaches to program evaluation (Guttentag, 1973). She argued explicitly that classical statistical approaches were in certain key ways less appropriate to evaluation research than Bayesian statistical approaches; and she argued that multiattribute utility theory (MAUT) was the appropriate method for capturing cost-benefit tradeoffs and for comparing programmatic alternatives.

The work described herein stems from the ideas in Guttentag's 1973 paper. This project itself resulted from a proposal submitted by her to the Law Enforcement Assistance Administration, which was initiated only just before her untimely death. The project was carried forward by the present authors and by her other colleagues at Research for Social Change, Inc., and at LEAA (now the National Institute of Justice). We acknowledge her inspiration, and that the project without her was not what it might have been.

This is the final report of a project applying what she termed the "MAUT-Bayesian" approach to the evaluation of the Community Anti-Crime Program. A purpose of this project was to explore the use of multiattribute utility theory (MAUT) for assessing program effectiveness across multiple objectives,

and the use of subjective probability and Bayesian statistical concepts to program evaluation. Specifically, we were interested in whether this type of decision-theoretic approach could be successfully applied as an evaluative methodology, and whether it seemed useful for aiding in actual decisions encountered during conduct of the program.

1.3 What is the Decision-theoretic Approach?

A fairly complete description of decision analysis, and the decision-theoretic approach to evaluation research, is contained in Section 3.0 of this report. The discussion in Section 3.0 emphasizes theoretical and conceptual differences with "traditional" approaches. The general approach is also described in Snapper and Seaver (1978), an earlier report prepared under this project.

As described in the 1978 report, the approach briefly is to use multiattribute utility theory (MAUT) to quantify attainment of multiple objectives, and the tradeoffs among them. An aggregate MAUT score is therefore the measure of effectiveness used in decision-theoretic approaches to evaluation.

Recognizing that program effectiveness will vary over time, a priori projections are made; these are compared with actual data. In the intial formulation of this method, probabilistic measures of effectiveness were used; these estimates were revised or updated on the basis of actual data (Snapper and Seaver, 1978). More recent work, described in this report, has relied heavily on subjective projections of future program effectiveness using non-probabilistic measures.[1] Such projections figured heavily in the decision problems we examined, for practical and theoretical reasons we will discuss in the latter sections of this report.

Since the primary purpose of this project was to test the application of decision-theoretic approaches to evaluation, and especially to examine their

[1] Formal Bayesian inference models were not, however, used. A common attitude among Bayesians--as well as others--is that strong data can be used to draw conclusions without benefit of formal inference models. In the examples we were concerned with, data were compelling and formal statistical inference techniques were not used. Our attitude in this regard is similar to that of Edwards, Lindman, and Savage (1963). They refer to the "inter-ocular trauma" test: no statistical inference model is needed because the data "hit you between the eyes."

use in actual decisions, the reader may omit Sections 2.0 and 3.0 and turn to case studies described in Section 4.0 and later. The reader who is interested in a fuller discussion of the Community Anti-Crime Program and the background to this application should read Section 2.0 for details. And the reader who is interested in a fuller discussion of the theoretical motivation for decision-theoretic approaches should read Section 3.0.

2.0 BACKGROUND OF THE COMMUNITY ANTI-CRIME PROGRAM AND ITS EVALUATION

Section 2.0 is divided into two parts. Section 2.1 describes the CAC Program, focusing on the characteristics of the individual local projects. These, it will be seen, vary considerably in terms of the strategies or tactics they use. Section 2.2 describes briefly the evaluation approaches which have been used in conjunction with the CAC Program, and the research organizations that have been involved.

2.1 Background of the CAC Program

The CAC Program, unlike most within LEAA, provides funding directly to community non-profit organizations bypassing state and local governmental agencies. Its purpose is to enable them to conduct anti-crime activities involving community residents.

Major purposes of the CAC Program are to assist citizens and neighborhood community groups in implementing activities aimed at preventing crime, reducing fear of crime, and revitalizing neighborhoods. The Program pursues these purposes by strengthening and expanding existing community organizations, encouraging the establishment of new organizations, integrating anti-crime efforts with community development activities, and improving communication and cooperation among neighborhood residents and criminal justice officials. In order to accomplish these purposes, money is directed to neighborhood "grassroots" anti-crime activities.

The local projects undertake multiple activities which fall into many of the traditional categories of crime prevention activities. Among the more common activities are community organizing, Neighborhood Watch, Operation ID, recreation for youth, and escort services for senior citizens. There are also many innovative activities such as community theatre (including specific productions by neighborhood youth) and provision of childcare training for teenage mothers.

Because a primary purpose of the program is to involve citizens and community organizations in crime prevention, and because activities and objectives vary widely from project-to-project, the CAC Program implied that the community-based project itself was the appropriate unit of analysis. The program stresses the importance of community-based decision making, the need to identify problems in the neighborhood, to determine which anti-crime strategies are likely to be most effective in resolving these problems, to assess project effectiveness, and to enhance attainment of relevant objectives by judicious modifications based on intermediate feedback about results or impacts. Perhaps the major assumption underlying the CAC Program was that "institutionalization of the community anti-crime concept" would occur. This notion of "institutionalization" encompassed a wide range of effects, impacts, or change. It included structural changes in the CJS or other organizations, as well as changes in approach or attitude on the part of persons in those organizations. Examples would include addition of a property engraving service in the police department, or greater cooperation between police and community residents. Physical changes in the environment, such as locks and lights, also embody the CAC concept. Often the most pervasive form of CAC concept institutionalization will be in the residents themselves, and reflected by their change in attitude, awareness of crime, surveillance of public areas, and willingness to participate in community anti-crime activities, among other things. While the modes of change clearly vary, they have in common the notion that the concept, once instilled via the project will survive the project and presumably that changes and benefits therefrom will persist after termination of federal funds.

2.2 Background of the CAC Program Evaluation

Two interrelated approaches to evaluation have been undertaken. Both are designed to track project-level occurrences over the period of project funding. The first is American Institutes for Research's "rationales" approach which is a special type of process or implementation evaluation (AIR, 1979). The approach attempts to build essentially an evidentiary chain between the baseline state of affairs and changes that occur. An advantage of this approach is that, hopefully, it helps attribute change

to the programmatic process. It attempts explicitly to distinguish "program inputs" from "disposing conditions" and exogenous "other events" that impinge on the change process.

The second approach is an impact assessment using decision-theoretic methods and is being undertaken by DSC. There are two different kinds of models used. The first is project-specific. It begins with a listing of a given project's objectives, and develops "multiattribute utility theory" (MAUT) models of effectiveness aggregated across objectives (Snapper and Seaver, 1980a). The project models were designed to track effectiveness over a three-year period--assumed to be a reasonable maximum period of federal support.

The other DSC impact model involved a cost-effectiveness assessment for the overall program and OCAP. It considered projected levels of effectiveness as a function of annual program funding level (Brown, Seaver, and Bromage, 1980). Cost-effectiveness was assessed over a five-year period, the period of reauthorization being contemplated at the time of the analysis. The analysis also assumed implicitly that the program would continue; annual funding levels of $10, $20, and $40 million were considered in the analysis.

Actually, the AIR and DSC efforts grew out of the preceding work initiated by Marcia Guttentag through Research for Social Change, Inc. The original purpose of the RSC grant was to conduct process, decision-theoretic, and intensive evaluations of the CAC Program, under the late Dr. Marcia Guttentag. The process ("rationales") and decision-theoretic approaches were initiated under the RSC grant.

3.0 THE ROLE OF DECISION ANALYSIS IN PROGRAM EVALUATION

Decision analysis has a role to play in evaluations to the extent that they are oriented toward decision making. Traditional evaluations, for analytic and philosophic reasons can be made more responsive to decision making by use of decision-analytic methodologies. In this section we discuss the basic concepts of decision analysis and some of the problems with traditional evaluation methodologies.

3.1 Basic Concepts of Decision Analysis

In general, decision analysis (see Brown, Kahr and Peterson, 1974; Raiffa, 1968) provides for the evaluation of alternatives, focusing on the values (or utilities) associated with outcomes and the probabilities of those outcomes. A complex problem is decomposed or analyzed into clearly defined components, such as alternatives, uncertainties, and values. The uncertainty and value components are quantified using whatever information and expertise are available. Logical implications of the analysis are incorporated into the decision making process.

In the formal model the value and the probability of each of the possible outcomes occurring as the result of selecting a particular alternative are determined. The overall value of the alternative is then obtained by weighting the value of each outcome by its probability and summing across all possible outcomes. The overall values of the alternatives under consideration can then be compared. In practice, decision analysis is often used less formally as an aid to the decision making process. For instance, decision analysis may focus explicitly on only one component of the problem, addressing the less-critical components informally.

Several individuals or groups may become involved, especially if they differ in their expertise regarding different problem components. In addition, decision makers--whether individuals or groups--can use decision analysis to

discipline their informal reasoning, and to facilitate communications among individuals in command structures, committees, or various interest groups. For example, decision analysis can be used to communicate the grounds for a recommendation, to identify sources of disagreement, to focus different expertise on appropriate parts of a problem, and to determine the critically important aspects of a complex decision problem.

Raiffa (1974) has briefly characterized decision analysis as follows:

> The decision analytic approach defines a procedure by which a decision maker can make reasoned choices that will be consistent with his perceptions about the uncertainties involved in a particular problem, together with his fundamental attitudes toward risk taking. Decision analysis pays particular attention to:
>
> - a decision maker's judgments about the uncertainties in his problem;
> - the use of information supplied by expert advisors;
> - the decision maker's feelings about consequences that may result from his choice of action;
> - the decision maker's basic attitudes toward value tradeoffs and risk taking;
> - chronological structuring of the decision problem and flexibility in the timing of decisions; and
> - opportunities to acquire relevant information (possibly through experimentation) before arriving at a final decision, considered in the light of the cost of obtaining such information.

An important and distinctive aspect of decision-analytic approaches pertains to how judgment enters into an analysis. Many non-decision-analytic approaches try to ignore judgmental input altogether, seeking a spurious form of objectivity. Other approaches recognize that the judgments and opinions of the decision maker are valid inputs, but treat them in informal, seat-of-the-pants fashion. Decision analysis is distinctive in part because it involves the careful quantification of both values (using utility theory) and uncertainty (using personal probability theory). It explicitly introduces such subjective quantities into the analysis rather than analyzing only "objective" components. The theoretical basis underlying this use of subjectivity is well-developed, and practical applications are becoming extensive (see e.g., Fischer, Edwards, and Kelly, 1978).

Two theoretical developments form the basis of decision analysis: personal, or subjective, probability theory (e.g. Savage, 1954) and Bayesian statistics (e.g., Phillips, 1974) as the means for quantifying uncertainty; and utility theory, and its extension to multiattribute utility theory (MAUT) (e.g., Keeney and Raiffa, 1976) as the means for quantifying value.

3.1.1 Personal probability. In program evaluation, expert judgment often is the best available information about the likelihood of uncertain events occurring. For example, the attractiveness of a certain program may depend on whether or not it proves more effective than some alternative. Experts familiar with the program area can provide subjective probability assessments which can formally be incorporated into the decision process. Techniques for eliciting such judgments have been developed and tested (e.g., Brown, Kahr, and Peterson, 1974; Seaver, von Winterfeldt, and Edwards, 1978), and procedures are available for combining the judgments of several experts (Seaver, 1976, 1978).

In many instances, some data will be available that provide information on the likelihood of uncertain events occurring. The data, however, may be insufficient to provide an accurate probability assessment. Therefore, it may be desirable to augment the information from the data with expert judgment. Bayesian statistics provide formal procedures for combining expert opinion with actual data. In some instances, these procedures may be useful, particularly with advanced planning. Snapper and Seaver (1978) discuss such formal procedures from a theoretical perspective, in the context of the Community Anti-Crime Program.

It is more likely, however, that informal means of combining expert opinion and data will be appropriate. Experts can be provided with the available relevant data and asked to incorporate it intuitively in their assessments of probability. (Indeed, the approach used in this study was to have people make judgmental estimates, which were replaced by estimates based on the actual data when they became available.)

3.1.2 Multiattribute utility theory. The multiobjective nature of program planning and evaluation suggests the relevance of multiattribute utility

theory as an evaluation methodology. The application of MAUT provides a theoretically justified procedure for assigning value to programmatic alternatives that have multiple effects. MAUT explicitly reflects the relative importance of each objective, and, therefore, the tradeoffs among effects. By doing so, MAUT enables the evaluator to develop a summary measure of program value reflecting many kinds and degrees of "effects."

The use of MAUT begins with the specification of the alternatives that are to be evaluated. Then the objectives that are to be used to evaluate alternatives are delineated. Often complex MAUT models can be structured in a generic, hierarchical form that allows attention to be focused selectively on related objectives, as discussed in Snapper and Seaver (1978). Such a generic model could then be adapted on a case-by-case basis to meet the requirements of particular planning situations.

Each alternative must then be "scored" on each objective. This scoring can be accomplished by different means. Objectives that have underlying quantitative measures (e.g., reduce crime) can be scored using a "single-attribute utility function." Such functions relate the underlying measure to value. Often a simple linear function will suffice as an excellent approximation provided the relationship between the underlying measure and value is monotonic, i.e., value always increases (or decreases) as the measure increases. This type of monotonicity and the appropriateness of linear approximations was assumed throughout this study. To be concrete for illustrative purposes, we assumed that MAUT scores were linear with the degree of reduction in a targeted crime.

Scoring on objectives that do not have underlying quantitative measures is done by direct expert judgment. This entails establishing a scale with well-defined upper and lower endpoints and asking the experts to rate the alternatives on this scale. This approach would generally be used to score the qualitative objectives, such as fear of crime or sense of community.

(The scoring of alternatives on an objective-by-objective basis described in the preceding discussion is formally appropriate only if some assumptions

with respect to the value (not statistical) independence of attributes are satisfied. Keeney and Raiffa (1976) provide the technical details of these assumptions. However, users of MAUT models can generally be assured that the above scoring procedure is satisfactory if the scores (i.e., the relative values) assigned to one objective do not depend on the level of the other objectives. For example, the value of reducing crime by 30% should not depend on how successful a project is with respect to other objectives.)

The next step in the development of a MAUT model is the specification of weights for each objective. The interpretation of weights and the procedure for assessing them depends on the form of the model to be employed to aggregate the single-objective scores. The theoretical basis for a variety of aggregation models has been developed (see Keeney and Raiffa, 1976), but in most cases an additive aggregation rule will be appropriate.

(Weights are assessed by considering the relative importance of moving from worst to the best level on each objective. The assessments are ratio judgments, meaning that if the importance of going from worst to best on one objective is twice that of another objective, the first objective is assigned a weight twice that of the second. This assignment of weights on the basis of "swings" from worst to best is an important feature reflecting the theoretical basis of MAUT that differentiates it from weighting procedures based on some ill-defined concept of "importance." These ratio judgments are then normalized to sum to one to facilitate further calculations.)

The overall value of an outcome of a particular alternative can then be computed by summing the single objective scores weighted by the normalized weights across all objectives. If there is no uncertainty about what the outcome of an alternative will be (or if the uncertainty is not formally incorporated in the evaluation), this overall value represents the relative value of the alternative. If uncertainty about outcomes is explicitly modeled, the value of each outcome must be weighted by its probability and summed across outcomes to provide an overall value measure for the alternative.

3.2 Problems with "Traditional" Evaluation Research Methods

Programmatic decisions clearly involve tradeoffs, value judgments about them, and uncertainty. There is, in fact, uncertainty both about what has happened in a program (and to what it can be attributed), and about what will happen if a programmatic change is made (or if the program continues as is). Insofar as decision analysis addresses these kinds of problems, it is perhaps not surprising that decision-theoretic approaches to program evaluation are no longer regarded as "novel" (though they are still regarded by some as heretical).

The General Accounting Office (1980) has for example, reviewed MAUT models (including the work cited in this report). The GAO has in part made decision-theoretic approaches (and in particular MAUT) part of the "evaluation establishment" by concluding that they are appropriate to the kinds of evaluation problems characterized above. While it is certainly correct that decision-analysis can be used to address components of a traditional evaluation, it is worth emphasizing that there are some fundamental conceptual differences in approach. Indeed, the conceptual differences are such that decision-theoretic approaches could lead to completely different conclusions than traditional approaches. It is useful briefly to discuss some of these conceptual differences, which will be picked up again in the concluding section of this report.

3.2.1 Statistical methods. Guttentag (1973) criticized evaluation research for characterizing evaluations as experiments--when clearly they neither are nor are intended to be. She stated the argument against evaluations-as-experiments forcefully, but methodological problems are hardly limited to the definition of evaluations-as-experiments. Other problems have to do with decision rules: how one relates empirical outcomes to subsequent decisions. Typical classical statistical methods (t-tests, chi-square, F-tests, etc.) are commonly used and they certainly have their proper uses. Yet, as we will discuss below, results of those methods as the output of an evaluation have severe shortcomings as inputs to decision processes.

Edwards, Lindman, and Savage (1963) argued that classical statistical techniques, relying on (sharp) null hypothesis tests, are inappropriate on two counts. First, they argue (p. 235) that "...classical procedures are so ready to discredit null hypotheses that they may well reject one on the basis of evidence which is in its favor..." They argue (p. 235) that Bayesian statistical procedures are more appropriate because they "...can strengthen a null hypothesis, not only weaken it..." and (p. 240) that "...evidence that leads to classical rejection of the null hypothesis will often leave a Bayesian more confident of that same null hypothesis than he was to start with."

The second problem is that decisions--unlike classical tests--are not bifurcated. Actions depend on values (benefits, costs) as well as probabilities. Edwards et al. illustrates this with a simple decision problem with three actions: (1) take a plane, (2) take a plane but buy flight insurance, and (3) don't take the plane at all. They state (p. 214) that: "What action is wise depends on what is at stake. You would not take the plane if you believed it would crash, and would not buy flight insurance if you believed it would not. Seldom must you choose between exactly two acts, one appropriate to the hypothesis and the other to its alternative."

Of course, in program evaluation there are always many more than two, three, or any other small number of actions. Funding level, for example, is a continuous variable but statistical significance alone has little to say about what the "best" funding level is. Decision-theoretic methods make the delineation of decision alternatives explicit, tying the recommended decision to values as well as probabilities.

3.2.2. Values in the decision-theoretic approach. Suppose we knew that a given program has a given "size" of effect, compared to a specified variant of that program. Would we know which to implement? If we knew the respective costs, we might be able to generate a consensus that the extra effect was worth the extra cost, i.e., that the benefit-cost tradeoff was favorable. In practice, there are few examples of such cleanly defined tradeoffs.

There are always alternatives to the current program for which there are no hard effectiveness and cost data. Programmatic decisions invariably open up to consider other funding levels or other altogether different programmatic approaches. Discovery that there are many alternatives implies that the simple tradeoff between size of effect and cost is complicated enormously, and much controversy is likely to arise. Worse, the data for the existing program will usually be incomplete, but that information will have to be used in making tradeoffs against alternatives about which even less is known.

Another complicating factor is the concept of "effect." "Effect," and "size of effect," are fairly common terms in the literature. They do not typically correspond to what is meant precisely by "effectiveness" in program evaluation, however. First, "effectiveness" refers to how fully an objective is attained. The size of an "effect," however, is utterly confounded with sample size. That is, a "large effect" may merely mean that the sample was large--not that the program was particularly effective. Second, "effects" are measured for a given variable, which is likely to be an indirect or imperfect relationship to program objectives (e.g., Francis, 1973). Snapper, O'Connor, and Einhorn (1974) argued that typically there will be multiple, imperfect measures of the same underlying objective. Given that these measures vary in validity, and given that their respective "size of effect" vary widely, how are they to be combined to yield a measure of effectiveness for the objective? More generally, how should data pertaining to multiple objectives be aggregated to yield a measure of program effectiveness?

Multiattribute utility theory (MAUT) provides a mathematically appropriate model for combining data elements into an aggregate measure of program effectiveness. MAUT models make explicit assessments about the relative importance of different objectives, which reflect the relative worth or value. This index of worth or value is then interpreted as the measure of program effectiveness.

Assessments of effectiveness using MAUT are relative not absolute, however. In particular, assessments are relative to the effectiveness associated with alternatives that the decision maker may consider. A common situation is that there are several alternatives, one of which is Do Nothing (i.e., do not implement the program or terminate it if it is underway). In regard to the Do Nothing option, the critical question is whether the difference in effectiveness between doing nothing and the presently implemented option (the Status Quo) is sufficient to justify the cost difference.

The comparisons implied in most evaluations are typically much richer than a comparison between the Status Quo program and Do Nothing. Questions include whether any potential option has an anticipated effectiveness sufficient to justify it over the Status Quo. That is, from the decision-theoretic point of view, the presently implemented option should be compared to several programmatic alternatives, and comparison with Do Nothing is one special case.

From the decision-theoretic viewpoint, even a demonstration that the Status Quo is not very effective compared to Do Nothing is not a valid touchstone for a decision, in itself. Information about the Status Quo option and Do Nothing bears directly only on a decision constrained to a choice among these two alternatives. There are two serious problems in the formulation of an evaluation to address this special two-alternative case of Status Quo vs Do Nothing. First, the decision properly involves a comparison with alternatives not explicitly considered, and a judgment about whether or not any of these potential alternatives would be sufficiently cost-effective to justify adoption in the future. There may be an implicit assumption, of course, that the failure in the programmatic Status Quo rules out an entire class of options: but acceptance of this assumption is primarily a matter of judgment rather than empirical fact.

The second--and probably far more severe--problem with typical evaluation formulations that compare to Do Nothing against the Status Quo programmatic option is that the empirical comparison involves historic data only. While judgments about the worth of a program may clearly be affected by demonstration of its past effectiveness, the decision really depends upon the assessment of future effectiveness! Past costs and effectiveness are "water under the bridge," and decisions about what should be done in the future depend upon one's judgments about what will happen in the future.

Our point is that our purpose--making evaluations maximally useful for decision making--will lead us in this report to emphasize some issues at the expense of others. We will not, for example, focus on historic "null hypothesis comparisons" between the present status of a program and the zero-effect, Do Nothing alternative. We will not, in fact, discuss any classical statistical techniques or formal Bayesian analyses (though our study did in fact use informal analogs of Bayesian analyses). We will, however, emphasize the modeling of objectives and assessing the degree to which they were attained. And we will emphasize the relevance of effective assessments on decisions--and the effect of decisions on future effectiveness.

PART II -- APPLICATIONS

4.0 THE GENERAL EVALUATION APPROACH

As suggested earlier, the decision-theoretic approach to evaluation emphasizes (1) delineation and modeling of objectives using multiattribute utility theory (MAUT), (2) a priori expectations about the future effectiveness of programs, and (3) the comparison of actual results against those prior expectations.

MAUT models were developed for each of eight sites. These sites were intended to "test out" the MAUT approach to see if it proved practicable at the local project level as distinct from, say, a national evaluation. It is the application of the general approach to specific projects that is discussed in this section.

MAUT models were in fact successfully developed for each of the eight projects. This is not to say, however, that each of the eight projects was itself successful. The symptoms of projects that were not successful included inability to specify measures of objectives; specification of patently trivial measures; or inability to supply concrete data or corroborate claims.

The results of the MAUT application to each of the eight projects will not be reported here. As it happened, a project arbitrarily selected among the eight projects developed an interesting decision problem. This problem was addressed using decision-analysis techniques and the results are described in another section of this report. Although this project is definitely not typical, since it has been nominated as an "Exemplary Project," we will nevertheless use it to describe the decision-theoretic approach to developing the "basic" evaluation model for a given project. Doing so will prove economical in introducing the decision analysis described for this project in another section of this report.

4.1 The Planner's Problem

As we (Snapper and Seaver, 1978) characterized programmatic decision making--

generically referred to as the planning process--the initial fundamental decision should specify or prescribe a set of objectives to be achieved by the program. The next decision involves selecting the "best" strategy, presumably based on an analysis to determine which programmatic strategy can be expected to result in maximum achievement given the tradeoffs among objectives. Subsequent decisions are based on evaluative feedback from, say, interim feedback from monitoring the implementation process or the assessment of final, net results. Decisions based on evaluative feedback take into consideration what was initially expected of the program, in regard to achievement of objectives, and the actual results of the program.

In making the "initial fundamental" decision, objectives must be formulated and used to assess tradeoffs among alternative strategies, based upon "prior expectations" about the results or impacts of each strategy. There will often be little hard data upon which to base these prior expectations. Thus, evaluations are most likely to aid those decisions based on "interim feedback" or "final, net results." Difficulties in designing an evaluation useful as a decision aid might be perceived by the introspective planner or evaluator as follows:

- How can objectives and their interrelationships be quantified?
- Since opinions about anticipated results or outcomes and attainment of objectives are fuzzy, how can the complicated tradeoffs among objectives be assessed?
- Given fuzzy opinions, are actual results or outcomes expected or surprising?
- Since actual data are often unreliable, how should prior opinions be revised?
- What modifications could potentially improve the program, i.e, raise future expectations?
- What are the tradeoffs among the current project strategy and any such future alternatives to it?
- What is the "correct" decision, given fuzziness of prior opinions, unreliability of estimates, and resultant ambiguity about tradeoffs?

4.2 The Basic Methodology

The use of evaluative information in decision making depends critically upon "what happened" compared to "what was expected." A decision-theoretic model for explicitly making this comparison was first proposed by Snapper (1974), and is presented in Edwards, Guttentag, and Snapper (1975) with an illustrative application to the National Institute of Education's Career Education Program. This approach helps resolve some of the difficulties referred to above as the "planner's problem." As applied to program evaluation, the decision-theoretic approach consists of (1) constructing a MAUT model including the relevant objectives, their relative importance, and specific measures for the objectives; (2) assessing current status and expectations with respect to project performance; and (3) measuring subsequent actual performance.

The approach used here (Snapper and Seaver, 1980), adapted from Edwards, et al. (1975), includes the following steps.*

Step 1: Determine and Structure Objectives

Project-level objectives are specified by the project director and other people working with the local project. A more general application of this methodology would involve objectives being specified by multiple constituencies with particular interest in the program. Because the overall evaluation model (see Snapper and Seaver, 1978) is already quite complex, and because the local project evaluation must be easily understood by the local project director, we have attempted to keep the number of objectives small (usually four to eight) and have avoided hierarchical models.

Objectives must be non-overlapping; that is, each must be a conceptually distinct factor on which the project would like to achieve some improvement. By carefully defining objectives, we can usually ensure that independence

*Snapper and Seaver (1978) discuss a particular variation of this approach using probabilistic measures and Bayesian procedures for updating measures. The Snapper and Seaver procedure is considerably more difficult to implement, but has some interesting formal properties.

conditions necessary for the additive MAUT model (see Keeney and Raiffa, 1976) are at least approximately satisified.

Step 2: Identify Measures for Each Objective

The project objectives must be operationalized with quantifiable measures that provide an indication of status on the objectives. Lack of resources to collect new data have led us to use only existing data such as that collected by the projects themselves and informed judgment as the basis of our measures. Although this restriction on data collection would usually not exist in a large-scale evaluation of an entire program, it does not seriously affect the local project evaluations. This restriction at the local project level means that the evaluations depend on only the information that projects would normally have available, though possibly augmented by directed informed judgment where there are no existing data.

For each objective, with the assistance of local project personnel, we identify one to three measures that appear to be most relevant, and for which data or informed judgment are available.

Step 3: Assess Importance Weights

In this step, the local project director assigns weights to the project objectives that represent the relative importance of going from worst to best on a specified range of the objectives. These weights are ratio weights so that an objective that is twice as important as another will receive a weight that is twice as large. The assigned weights are normalized to sum to one.

Step 4: Assess Current Status and Expectations

At the beginning of the project, the project's current status on each of the objectives must be determined. This is accomplished by using available data or informed judgment to provide the best assessment on each measure. A scale is developed for each measure where 0 represents the current status at the beginning of the project, and 100 represents the best feasible achievement on the measure, in this case defined as a level of achievement to be reached three years in the future. Where there are multiple measures for a

single objective, the scores for each measure are averaged to provide an overall score for the objective.

In all instances where informed judgments are used as measures, the initial and primary judgment is that of the project director, since this is the person who would use the evaluation for making decisions. To the extent possible, we also obtain confirmation of these judgments from other sources, e.g., government officials, community leaders, etc.

In addition to the assessment of current status, the project director is also asked to use the same measures to assess how the project is expected to perform over its period of existence, usually two to three years. Using the already defined 0 - 100 scales, the project director assesses "prior expectations" on a year-by-year basis for how the measures will change given the existence of the project. Using a similar judgmental approach, we also assess how these measures would have changed without the project. Again these judgments are confirmed by outside experts.

Step 5: Measure Subsequent Actual Performance

After an appropriate period of elapsed time, in this case approximately one year, we assess the project's actual performance on each measure. This performance can then be compared with the prior expectations assessed previously. These comparisons, as evaluative feedback to project directors, may suggest some programmatic changes.

This basic approach was applied to the evaluation of the eight CAC projects. The remainder of this section describes the evaluation of one such project, that of the Midwood-Kings Highway Development Corporation (MKDC) in Brooklyn.

4.3 The Midwood-Kings Highway Development Corporation Evaluation

The approach to the MKDC evaluation involved a visit to the project early in its first year of funding (Year 0). Objectives were elicited from the project staff and tentative measures of objective attainment were identified.

Finalization of objectives and an explicit listing of measures of their attainment was obtained from the MKDC Project Director (Richard Shapiro) on a subsequent visit. Listings of the objectives and of their relative importance (as specified by the Project Director) are included as Exhibit 4-1.

The objectives specified by the project differed in many ways from those which would be specified by others interested in the CAC Program, at least in terms of their weighting. For instance, for the purpose of evaluating the CAC Program, LEAA personnel and others have argued that reduction of crime and fear of crime should receive low weights. Their argument is, essentially, that CAC projects could not be expected to have much of an effect on these phenomena, i.e., the range of potential impacts was small. Since weights are related to the range of the objective considered, these differences do not necessarily reflect differences in how valuable crime reduction is. Rather they show the project has a much more optimistic expectation about what can be achieved.

Other objectives specified by LEAA were not represented, at least explicitly. Some objectives, such as mobilization of resources (resident involvement) and integration with social services were cited as relevant and important both by LEAA and by the MKDC project.

The measures identified for each of these objectives were subject to data collection constraints. Though evaluative approaches that involve special data collection could be designed, an advantage of the modeling approach described here is that it is useful even when evaluation is restricted to existing data. This is particularly important for project-level evaluations where evaluation resources and design skills are often quite limited.

The available measures pertaining to each objective are shown in Exhibit 4-2. A range of data collection modalities is apparent from this list. Data include "hard" statistical series such as crime reports, data collected by special surveys the project has conducted as part of its own management efforts, and (in the case of assessing degree of integration with social

EXHIBIT 4-1

MKDC CAC PROJECT OBJECTIVES

Number	Objective	Importance Weight
1	Reduce Crime	.141
2	Reduce Fear of Crime	.140
3	Increase Police Responsiveness	.119
4	Serve Community Ombudsman Role	.126
5	Increase Resident Involvement	.149
6	Institutionalize Organization	.111
7	Provide Technical Assistance	.104
8	Integrate Other Social Services	.110

EXHIBIT 4-2

AVAILABLE MEASURES FOR MKDC OBJECTIVES

OBJECTIVE	MEASURES/DATA
1. Reduce Crime	Larceny of Motor Vehicles Larceny from Motor Vehicles Burglary
2. Reduce Fear of Crime	Question 18 MKDC Survey Question 24 MKDC Survey Housing Turnover
3. Increse Police Responsiveness	Attitudinal Survey Data MKDC Impact Judgments
4. Serve Community Ombudsman Role	Complaints Processed
5. Increase Resident Involvement	% Blocks Organized Block Club Maintenance % Attendance
6. Institutionalize Organization	% Staff LEAA Supported Cumulative MKDC $
7. Provide Technical Assistance	TA Trips/Month
8. Integrate Other Social Services	Direct Judgment

services) direct judgments about effectiveness supported by descriptive background information.

The evaluation method is illustrated conveniently by considering some representative data and the display of results. Exhibit 4-3 shows the results for the first measure of crime reduction: reducing larceny from motor vehicles in the project area. The half-filled circle shows the level of larceny from motor vehicles in Year 0 (June, 1977 through May, 1978). There were 580 reported instances in that year; with the model we used, that corresponded (by definition) to a MAUT score of zero. The filled circles represent judgmental projections, when actual data are obtained, they are indicated on the display and connected with solid lines. Thus, in Exhibit 4-3, the actual data for Year 1 are displayed. In this case, a substantial reduction in larceny from motor vehicles (and a corresponding increase in the MAUT score) is indicated. The dashed lines show the projections for years 2 and 3. These prior expectations will be replaced by estimates based on the actual data, when those data become available in the future.

The dotted lines and open circles in Exhibit 4-3 indicate judgments about what would have happened, had there been no project in the area. Clearly, from the point of view of experimental rigor, this type of judgment is not a "control" and it is not intended to be interpreted as such. Instead, it provides a basis for the cognizant project and program staff to compare actual results against what, in their opinion, most likely would have happened. This provides a basis for judgments about the magnitude of effects attributable to the project, and, ultimately, for judgments about whether the project or program is worthwhile. In the example shown here, the projections about "no project" results reflect judgments that there would be an annual 6% decline in crime--based on the assumption that trends in this crime over the past few years in the MKDC area and Brooklyn would continue through the time frame being considered.

The 100-point on the scale in Exhibit 4-3--and for each of the other measures--corresponds to the "maximum plausible" value. In this case, the 100-point corresponded to projections about asymptotic impact which would be achieved after roughly three years of the project. In other words, the 100-point on

the MAUT scale represents the best that the project could probably accomplish, and intermediate levels of attainment are scaled relative to that 100-point and the 0-point. Over this range, a linear utility function was used. Thus, for example, in Exhibit 4-3 the reduction to 310 reported larceny cases represents a MAUT score of 76. An advantage of this scoring procedure is that it permits a common interpretation of all scales, as spanning the "maximum plausible range" between initial starting point and realistic prior expectations regarding impacts.

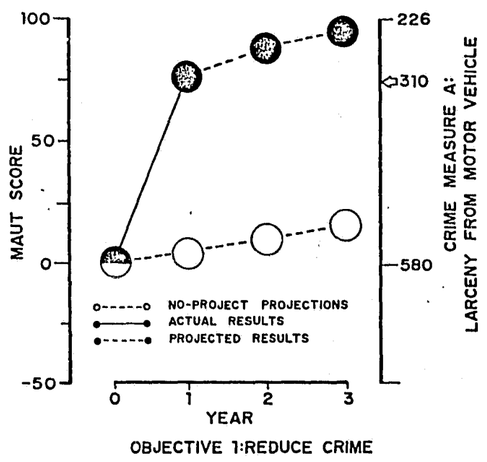
Exhibits 4-4 through 4-7 provide some further illustrative examples of the data from the MKDC project. In each case, there is evidence that the project has in fact made substantial progress over the first year of project operation. Indeed, the data for each of the objectives (and the judgments pertaining to Objective 8) indicate essentially the same degree of objective attainment reflected in the foregoing example.

Exhibit 4-8 displays the aggregate results, collapsed across all measures and objectives. An additive combination rule was used in which measures of objectives were equally weighted to arrive at summary measures for each objective. Summary measures for each objective were then aggregated, using the weights indicated in Exhibit 4-1. The results clearly indicate what each of the separate measures implied individually: the project has been rather successful during the first year; and that, in the opinion of project management, additional increases in effectiveness would occur for each of the following two years for which projections were obtained.

This decision-theoretic evaluation of the MKDC project serves multiple purposes. It provides an assessment of the effectiveness of this particular CAC project. It feeds into the broader evaluation of the CAC Program. And it provides a decision-aiding tool for project management. As a decision-aiding tool, it can be used to project likely results of programmatic changes, and subsequently to assess what actually happened as a result of the change in comparison to prior expectations about the results of the change. In the next section, we describe the application of decision-theoretic models to such interim programmatic decisions.

EXHIBIT 4-3

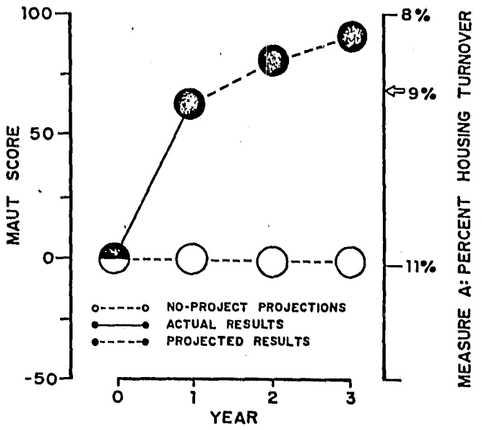
COMPARISON OF CRIME LEVEL IN YEAR ONE AND PROJECTED CRIME LEVEL IN YEARS TWO AND THREE WITH PROJECTED NO-PROJECT CRIME LEVEL.



OBJECTIVE 1:REDUCE CRIME

EXHIBIT 4-4

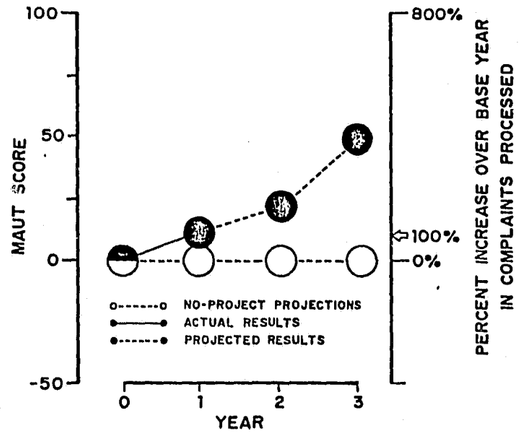
COMPARISON OF FEAR OF CRIME IN YEAR ONE AND PROJECTED FEAR IN YEARS TWO AND THREE WITH PROJECTED NO-PROJECT FEAR OF CRIME.



OBJECTIVE 2: REDUCE FEAR OF CRIME

EXHIBIT 4-5

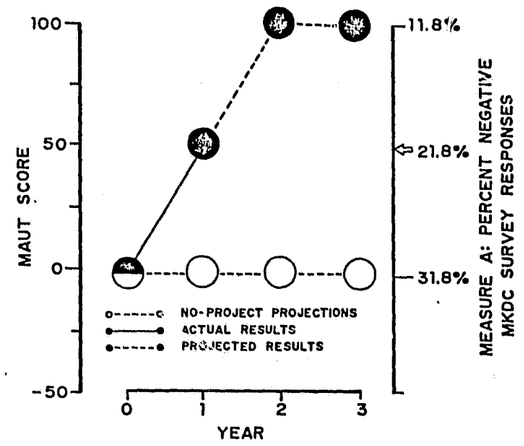
COMPARISON OF MKDC ROLE AS COMMUNITY OMBUDSMAN IN YEAR ONE AND PROJECTED ROLE IN YEARS TWO AND THREE WITH PROJECTED NO-PROJECT ROLE.



OBJECTIVE 4: COMMUNITY OMBUDSMAN ROLE

EXHIBIT 4-6

COMPARISON OF POLICE RESPONSIVENESS IN YEAR ONE AND PROJECTED RESPONSIVENESS IN YEARS TWO AND THREE WITH PROJECTED NO-PROJECT POLICE RESPONSIVENESS.



OBJECTIVE 3: INCREASE POLICE RESPONSIVENESS

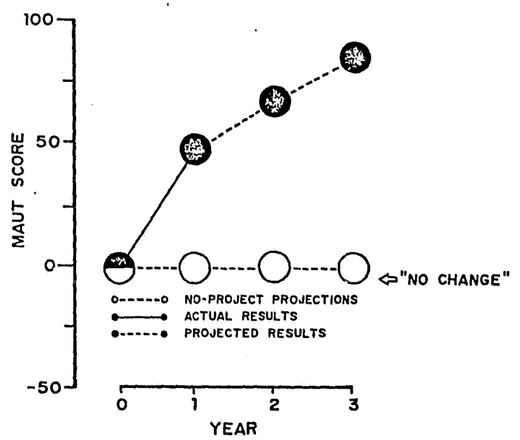
EXHIBIT 4-7

COMPARISON OF RESIDENT INVOLVEMENT IN YEAR ONE AND PROJECTED INVOLVEMENT IN YEARS TWO AND THREE WITH PROJECTED NO-PROJECT INVOLVEMENT.

OBJECTIVE 5: INCREASE RESIDENT INVOLVEMENT

EXHIBIT 4-8

COMPARISON OF TOTAL MAUT SCORE IN YEAR ONE AND PROJECTED SCORE IN YEARS TWO AND THREE WITH PROJECTED NO-PROJECT SCORES.



TOTAL MKDC MAUT SCORE

5.0 THE MKDC DECISION ANALYSIS

Is--as was hypothesized--the decision-theoretic approach to evaluation supportive of decision making? If significant fundamental decisions occur, they (as opposed to incremental fine-tuning) usually occur in projects (or programs) at the time the project is refunded or whenever it is up for re-funding. This application did not, however, cover these critical periods. Serendipitiously, however, a project-level decision presented itself.

This decision involved the Midwood-Kings Highway Development Corporation (MKDC) discussed in Section 4.0. The basic decision was whether or not to expand the project area. The concern on the part of MKDC was whether it would be possible to integrate the anti-crime efforts without diluting resources to the point they became ineffective. Before describing this problem, its analysis, and the ultimate decision more fully, we will briefly discuss the adaptation of the basic evaluation model (see Section 4.0) for decision-analytic uses.

5.1 Adapting the General Evaluation Approach for Decision Analysis

As discussed in Section 4.0, decisions are made at various points in the life of a program (or program). At the outset of the program--when there is little if any experience with the program and its results--decisions are typically based on opinions about what will in fact happen. It is sometimes alleged that decisions based on evaluations, unlike those made at the outset of the program, are "objective" insofar as they are presumably predicated on "hard" data. But are decisions based upon actual data in fact more "objective" than those decisions at the outset of the program, based on fuzzy opinions about results? At one extreme, where decisions involve only relatively minor, incremental changes from the programmatic status quo, the data may have rather direct, "objective" implications for

how the program may be fine-tuned. At the other extreme, where one is considering rather fundamental changes--as opposed to incremental changes as distinguished by Etzioni (1965)--there will be little if any information about large departures from the programmatic status quo. One must again rely upon essentially subjective estimates about the future degree of objective attainment associated with each alternative to the status quo.

In those decisions where there is a fundamental decision to be made, future effectiveness of the program in relation to alternatives must be considered. This means that despite a possibly rich data base, prior expectations about what would happen in the future again determine the decision! Even rich evaluation data bases in such decisions can, in principle, "objectively" address only a part of the problem (i.e., how effective the program has been in the past) which is a starting point for assessment of future effectiveness. We pointedly mentioned this topic in Section 3.0, and will discuss how the general evaluation approach (from Section 4.0) is useful for structuring and analyzing the decision problem.

Often, the need for a fundamental decision--in which major rather than incremental changes from the status quo are considered--is triggered by the difference between actual effectiveness scores and prior expectations. That is, feedback suggests the need for a reconsideration of programmatic strategy. This situation is illustrated in Exhibit 5-1.

The form of display in Exhibit 5-1 is essentially the same as shown for the MKDC data. The difference is that Exhibit 5-1 illustrates the case in which the actual results are markedly lower than the prior expectations for corresponding points in time. Before automatically concluding that such results imply the need to change the program, there are certain judgmental issues to be resolved. Why did the shortfall occur? Were there unrealistic prior expectations, such that the program is in fact performing as well as could be reasonably expected? Was the program simply slow in getting underway, so that if left alone effectiveness scores would begin to approach prior expectations? Are unexpected adverse events responsible for the shortfall, and, if so, would another programmatic alternative prove more robust? Or, is the program simply ill-conceived, and is a fundamental change required?

The decision problem posed by this situation is typical of programmatic decisions made during the life of a program. A representation of the generic problem is shown in Exhibit 5-2. For simplicity, we consider a decision with only three alternatives: continuation of the program as is, and two variants of the program.

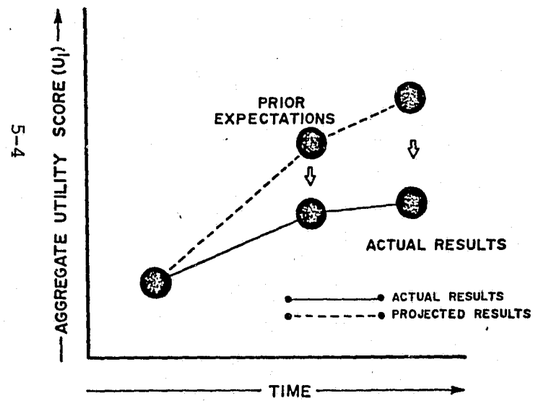
The comparison of the three alternatives is essentially a judgmental matter--despite the fact that data are available for the existing program. In regard to the status quo alternative, the question is: what are the prior expectations for continuance? Will the effectiveness scores increase if the program is allowed to continue (e.g., because adverse circumstances will eventually be overcome), or will scores continue about the same or even decline? Exhibit 5-2 illustrates the "straight-line" case, but increasing (and decreasing) scores are also plausible.

In regard to changes in the programmatic status quo, what are the prior expectations for alternatives? Two alternatives are illustrated which offer to increase the effectiveness score, though alternatives could also fall short of the status quo. Even if alternatives which are identified have higher prior expectations than continuance of the status quo, a change may not be appropriate if, for example, there are costs not represented in the utility scores for the program. For instance, there may be some high initial start-up costs incurred when a fundamental programmatic change is made.

Thus, the decision-theoretic evaluation model provides a general framework within which various project decisions can be addressed. As a monitoring and feedback mechanism, it can assist in the identification of the need for programmatic change. Other events may also trigger decisions about whether to change programs. Changes in funding level, the nature of the problem being addressed, or defined target population are examples of the kinds of events that may trigger such decisions. Regardless of what triggers the decisions, the potential programmatic changes then can be evaluated and compared within the general framework.

EXHIBIT 5-1

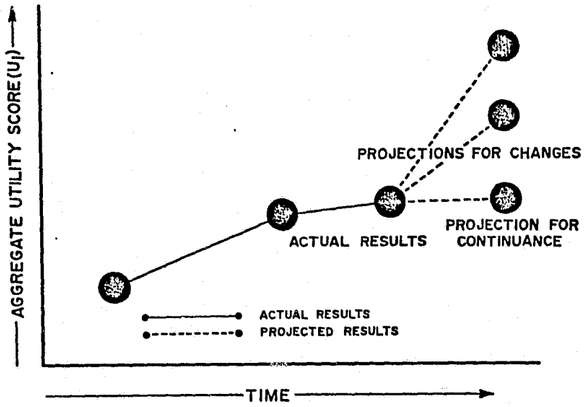
USE OF MODEL: TRACKING RESULTS AND COMPARING PRIOR EXPECTATIONS WITH ACTUAL RESULTS.



WHY IS THERE A SHORTFALL?
- --UNREALISTIC PRIOR EXPECTATIONS ?
- --SLOW PROGRAM STARTUP?
- --UNEXPECTED ADVERSE EVENTS?
- --PROGRAM ILL-CONCEIVED?

EXHIBIT 5-2

USE OF PRIOR EXPECTATIONS TO MAKE PROJECTIONS ABOUT CHANGES IN THE PROGRAM.



SHOULD THE PROGRAM BE CHANGED?
- --WHAT ARE PRIOR EXPECTATIONS FOR CONTINUANCE?
- --WHAT ARE PRIOR EXPECTATIONS FOR CHANGES?
- --IS THE DIFFERENCE IN PRIOR EXPECTATIONS LARGE ENOUGH TO INDICATE THAT A CHANGE SHOULD BE MADE?

5.2 The MKDC Decision

The decision problem facing MKDC arose not as a result of comparing actual results with prior expectations--since the results in Section 4.0 clearly indicate substantial achievement of objectives--but from the occurrence of an outside event that was perceived to have the potential substantially to affect the CAC project.

The city of New York had adopted a policy of "coterminality" in which police and other service delivery areas were to become aligned or "coterminous" with community districts. As a result of coterminality and subsequent political maneuvering, MKDC was placed in a situation where it was considering expanding its project area. The expansion would encompass all of the area served by the Midwood Civic Action Council (MCAC), one of five local civic associations that work closely with MKDC. Exhibit 5-3 illustrates the potential expansion, showing how MCAC would like to move from the 12th to the 14th community district, leaving MKDC with the question of whether or not then to include this new area in their project target area.

As suggested by Exhibit 5-3, the expansion would be roughly 50% both in area and in population served. No additional LEAA funds were available so a primary MKDC concern was the dilution of resources and associated loss of effectiveness. This concern was balanced against the political considerations that made a continued good working relationship with MCAC desirable.

The immediate decision about whether to support integration of the additional MCAC area into the project, depended on expectations about how significantly the project's effectiveness would be impaired. This decision was analyzed during a full-day working session with the MKDC Project Director, Richard Shapiro. In conducting the analysis, we considered two extreme alternatives: Full Integration of MCAC into the project, and No Integration of MCAC (i.e., maintenance of the status quo). These two extreme alternatives are represented in Exhibit 5-4. Of course, there are a number of intermediary strategies

EXHIBIT 5-3

MAP OF MKDC PROJECT AREA AND POSSIBLE ADDITIONAL AREA

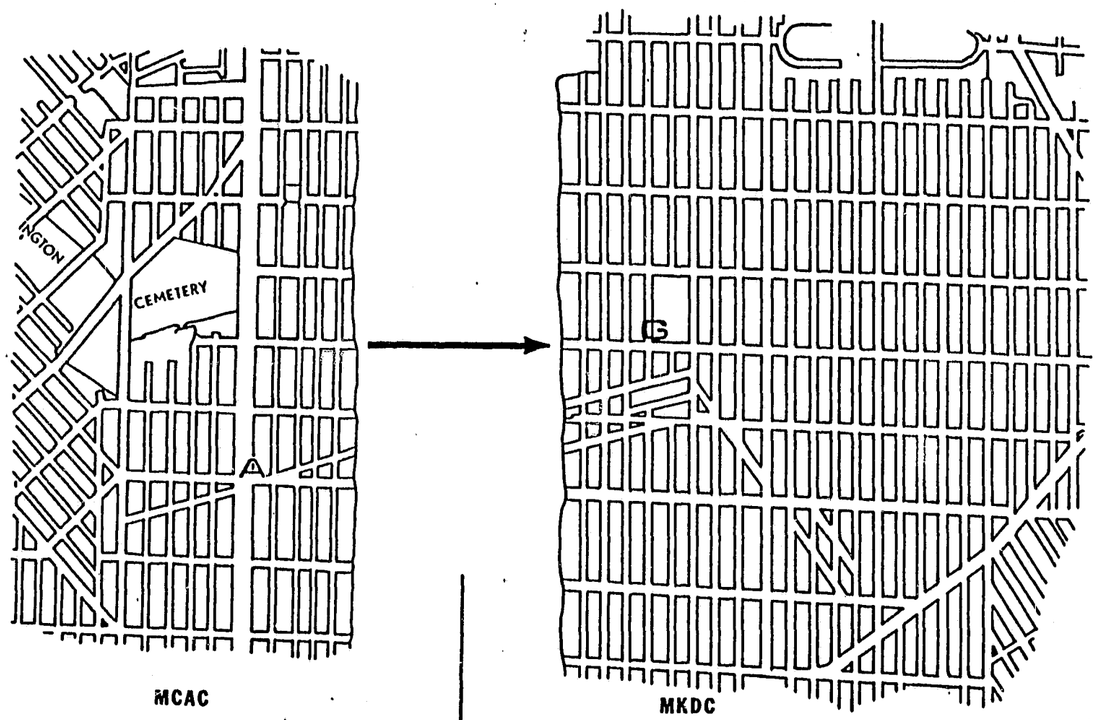
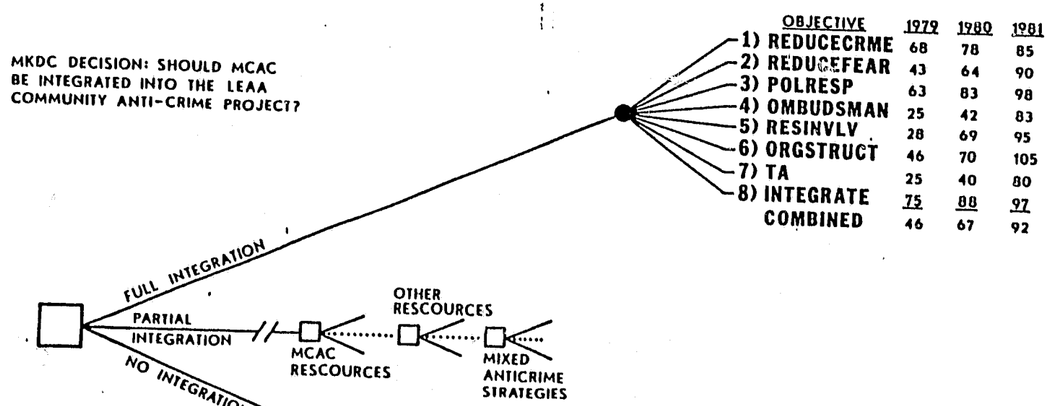


EXHIBIT 5-4

THE MKDC COTERMINALITY DECISION PROBLEM



OBJECTIVE	1979	1980	1981
1) REDUCECRME	68	78	85
2) REDUCEFEAR	43	64	90
3) POLRESP	63	83	98
4) OMBUDSMAN	25	42	83
5) RESINVLV	28	69	95
6) ORGSTRUCT	46	70	105
7) TA	25	40	80
8) INTEGRATE	75	88	97
COMBINED	46	67	92

OBJECTIVE	1979	1900	1981
1) REDUCECRME	68	81	89
2) REDUCEFEAR	43	71	97
3) POLRESP	63	84	100
4) OMBUDSMAN	25	50	100
5) RESINVLV	28	85	100
6) ORGSTRUCT	46	66	100
7) TA	25	50	100
8) INTEGRATE	75	90	100
COMBINED	46	73	98

which could be employed, and which might minimize the degradation in effectiveness. As shown in Exhibit 5-4, Partial Integration strategies could be identified which minimized the depletion of MKDC resources by requiring MCAC to generate resources itself or from other sources, or by adopting a "mixed" strategy that allocated only the less-critical MKDC resources to the MCAC project area.

Only the two extreme alternatives (Full Integration versus No Integration) were considered. To keep the analysis tractable within time constraints, we decided that the fuller analysis of Partial Integration strategies would be conducted provided that a large difference between the two extreme alternatives in fact emerged, causing MKDC to reject outright the full integration of MCAC.

The analysis involved making projections about project effectiveness within the current MKDC project area itself for 1980 and 1981. These projections were the judgments of the MKDC Project Director, and were conditional upon the decision. In fact, the evaluation results shown previously in Section 4.0 as the solid circles in Exhibit 4-8 show the effectiveness projections for 1980 and 1981 (years 2 and 3, in that exhibit) based upon the assumption of No Integration. When the Project Director initially made his projections about effectiveness, he implicitly assumed that the project target area would remain fixed. Also, at the time the decision analysis was undertaken, he saw no reason to modify his projections, provided that MCAC was in fact not integrated into the project. The MAUT scores for each objective (and the combined or aggregate scores) for 1979, 1980, and 1981 are shown at the tip of the No Integration of MCAC branch.

The Project Director's projections about effectiveness provided that the MCAC area is in fact integrated into the project are shown at the tip of the Full Integration branch. The MAUT scores for 1979 are the same as for No Integration of MCAC because at the time of the analysis the effectiveness starting point is of course the same for both alternatives. In 1980 and 1981, however, differences between the two alternatives appear. These are apparent on an objective-by-objective basis, as well as in the combined scores. The projections for 1980 and 1981 were made by the project

director after careful consideration of how resources would be allocated under Full Integration and what that would mean to on-going activities.

A comparison of the combined MAUT scores for 1980 and 1981 shows that there would be an expected six-point drop in effectiveness associated with full integration of MCAC into the project. The magnitude of this drop in effectiveness is shown in Exhibit 5-5. The Project Director viewed this drop in effectiveness as much smaller than he had intuitively expected prior to any analysis. He had been initially quite certain that for the good of the project the target area should not be expanded, and had considered expansion primarily for political reasons. As a result of the analysis, however, he decided that the expected reduction in effectiveness was not large, and was more than offset by other considerations.

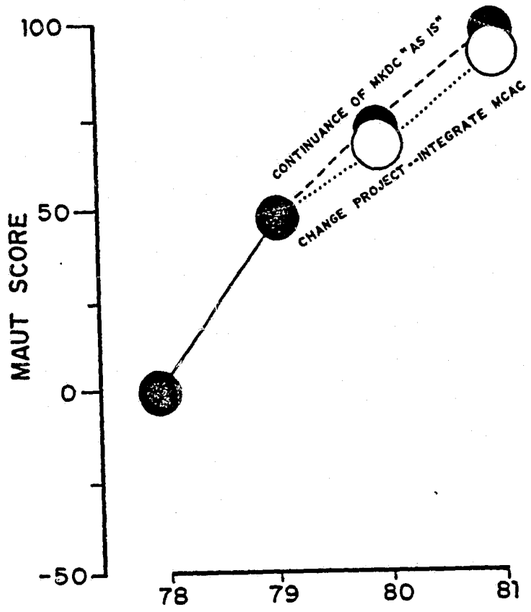
Subsequent political activities in New York City delayed implementation of coterminality. MKDC, however, petitioned LEAA to expand their target area and to integrate MCAC, regardless of the outcome of the coterminality issue.

5.3 Outcome

The decision to integrate MCAC into the project was in fact carried out. MKDC felt that there was indeed some loss in effectiveness but that they did not fall below what they had expected in terms of performance.

EXHIBIT 5-5

COMPARISON OF PROJECTED MAUT SCORES FOR MKDC AREA WITH AND WITHOUT ADDITIONAL MCAC AREA.



6.0 APPLICATION OF MAUT TO A PROGRAM DECISION: THE FUNDING LEVEL OF THE OFFICE OF COMMUNITY ANTI-CRIME PROGRAMS*

As mentioned, this study involved development of MAUT models for selected projects to assess the usefulness of decision-theoretic approaches. The scope of work also included development of a MAUT model to assess overall program effectiveness, as described in Snapper and Seaver (1978). The MAUT model was intended to be used both for monitoring past effectiveness and for making projections about future effectiveness of the program. Because of this two-fold use, the program evaluation methodology paralleled in many ways the applications to individual projects. One major difference, however, was that the American Institutes for Research (AIR) was charged with collecting the actual data regarding past program effectiveness. As with the project-level evaluations, one of the purposes of the program-level evaluation was to assist programmatic decision making.

A decision problem arose in the summer of 1979, when the funding level of the Office of Community Anti-Crime Programs (OCAP) was being considered as part of a broader decision about reauthorization of LEAA for five years, between 1980 and 1984. Congressional deliberations ranged from annual funding levels of $10 million and $40 million, with the possibilities that there might be a compromise (a funding level of, say, $20 million) or that the program might be terminated altogether. Evaluation results from the AIR study were not generally available, and of course the DSC data on eight projects comprised a small sample for purposes of estimating overall program effectiveness. Nevertheless, the decision problem presented itself, could not be delayed, of course, and was analyzed by DSC.

While DSC was assessing OCAP funding level, a broader decision arose about whether and at what level LEAA itself should be refunded, which ultimately superseded the OCAP funding decision. No attempt was made to address this broader question--the reauthorization of LEAA itself--since it was well beyond the scope of the present project. However, the analysis did involve close interaction with Congressional staff, in part

*Portions of this section draw heavily on a report by Rex V. Brown, David A. Seaver, and Robert C. Bromage (1980), all of DSC.

because it was desired that the analysis reflect the Congressional decision, and in part to communicate results to Congress as one of the stakeholders concerned with the Program.

The analysis illustrates a point made in previous sections of this report, namely that decisions depend in large part on expectations about future effectiveness. Empirical data entered into assessment of the current effectivensss and, in part, into estimates of the "dollar worth" of the program. But as with any analysis requiring projections of future effectiveness, much of the analysis was based on subjective judgments of experts.

6.1 Rationale for the Decision Model

6.1.1 Benefits--whose? The question of whether a program is "worth it" is ambiguous from a decision-theoretic viewpoint. "Worth it to whom?" the decision analyst asks. The citizens and residents who receive services? The grantee organizations? The funding agency? Congress? Or the general public? The fact that program benefits can be defined from multiple viewpoints is of course one consequence of the existence of multiple stakeholders.

In part because benefits can be defined in multiple ways, it is necessary to analyze a program decision from a particular viewpoint or perspective. The viewpoint adopted for this analysis was Congress's, since Congress, in fact, was making the critical funding decisions. Congress, of course, is not a homogenous entity and sharply divergent views were known to exist about LEAA and OCAP. Although there were differences in opinion about what ought to be done, the analytic approach attempted to develop a MAUT model that was: (1) consistent with the intent of federal legislation and policies, and (2) could reasonably approximate the "typical" Congressional viewpoint regarding OCAP program effectiveness and benefits accruing therefrom. As described later in this section, this modeling effort was successful.

6.1.2 Benefits--when? If decision makers should focus on the future, not the past, how far ahead should they look? Should benefits be defined in

the short-term, or should one consider only benefits that accrue in the long-term after the program is well-established? Both viewpoints are, in fact, credible, and depend on subtleties in how the purpose of the program is defined. These viewpoints can be clarified by stating the extreme positions.

- Programs-as-services: the purpose of OCAP is to provide services to community residents. Benefits should be assessed in terms of near-immediate effectiveness. Funding should continue only if the program remains effective and benefits out-weigh costs.
- Programs-as-social-experiments: the purpose of OCAP is to test new approaches. While effectiveness should be monitored for research purposes, a net excess of benefits over costs should not be expected in the short-term. If the concept is sound, benefits should be anticipated toward the end of a relatively long funding cycle, at which time Congress will have information to make yet longer-range decisions.

These viewpoints are compatible, and, in fact, one often views a program as having both purposes. Our point is similar to that made by Doherty and Crake (1980) who argue that programs have "research" and "operational" costs and that certain R&D costs should not be counted against the program if it is evaluated as an operational entity for delivering services. An implication is that benefits must be assessed over time, in both the short- and long-term. Also, the decision about whether the program was "worth it" or not depends on some comparison of benefits over time compared to costs over that same period, and on just what the time period is.

Our approach was to use the MAUT model described above to assess expectations about OCAP effectiveness in 1984, at the end of the anticipated five-year authorization period. To assess interim annual 1980-1983 benefits, effectiveness in 1979 was assessed and values for interim years were approximated by linear interpolation between 1979 and 1984 assessments. The 1979 assessments also provide an anchor that aided the judgments required for the 1984 assessments.

The steps in development of the MAUT model, assessment techniques, and results are discussed in the remainder of this section.

6.2 Options for Future Funding

Community anti-crime has been a controversial program area, characterized by continuing suggestions about how adjustments might be made. The fundamental program decision, however, was the OCAP funding level and the present analysis is limited to that decision.

Three funding levels were explicitly considered with termination or zero funding being the implicit fourth option. These were characterized as follows:

- Option A--$40 million: OCAP programs are funded at $40 million per annum for 1980-1984. Approximately 3/4 of the monies would support the CACP; and the balance would go to other urban initiatives. It is assumed that the program would maintain a reasonable balance in terms of communities and type of project. Total population in all project areas between 1980 and 1984 is projected at 40 million persons.

- Option B--$20 million: The CACP is funded at $20 million per annum, with no other significant OCAP program initiatives. This represents a slight increase in funding compared to 1979. The balance among type of communities and projects would remain about the same as for CACP under the $40 million option.

- Option C--$10 million: The funding under Option B is halved. This option implies significant cutbacks in 1980 compared to previous 1979 funding levels.

- Option D--Termination of OCAP: Under this option, funding level would be reduced to zero for 1980.

6.3 The MAUT Model for Assessing Benefits

The analysis assumed that programmatic benefits are a function of effectiveness. Two classes of objectives were identified, against which effectiveness was assessed. The distinction between these two classes was the source from which benefits were derived, as described below.

- Social Effectivenes: defined with respect to altruistic social objectives and benefits accruing to the public. These objectives include reduction in crime and fear of crime; improvement in quality of life; improvement in social processes underlying the crime problem; improvement in community relations with criminal justice systems; institutionalization and improved technical crime prevention know-how (capacity building).

- Political Effectiveness: defined as the political effects accruing from communities or interest groups who received funds, benefit from the program or are supportive of the concept of community-based crime approaches, or who otherwise have a special interest in OCAP programs.

Subobjectives were identified within each of these two categories. These are shown in Exhibit 6-1. For the most part these are self-explanatory. Two subobjectives require special comment: (1.4) Institutionalization and (1.5) Capacity Building. Both refer to benefits that accrue over the five-year period and enhance or leverage effectiveness after that period. Institutionalization refers to generation of resources required to support activities in the future especially after project funding from LEAA runs out. Capacity Building refers primarily to acquired programmatic expertise at federal and local levels to better design, implement, and manage such efforts in the future. Thus, both these objectives reflect the effectiveness of OCAP programs in a given year, although strictly speaking, the benefits therefrom will be realized only in future years.

6.4 Aggregate Measure of Benefits

As mentioned above, it is assumed that benefits are a function of (1.0) Social Effectiveness and (2.0) Political Effectiveness. The weights applied to each of the individual objectives or lowest-level subobjectives are shown in Exhibit 6-2.

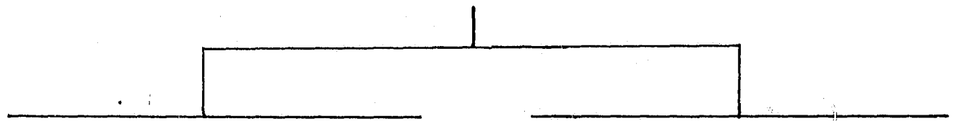
A special procedure was used for assessing effectiveness, and results are shown in Exhibit 6-3. The procedure is briefly described below.

6.5 Assessment Procedures

Experts were used in two distinct ways: to review the formulation of the problem and the analysis, and to provide the judgments which were the basis

EXHIBIT 6-1

TOTAL BENEFIT



1.0 Social Effectiveness

- 1.1 Social Process Impact
  - 1.1.1 Anti-crime Activities
  - 1.1.2 Community Organization
  - 1.1.3 Public/CJS Relations
  - 1.1.4 Reduction of Crime Causes
- 1.2 Crime-related Effects
  - 1.2.1 Reduce Crime Rate
  - 1.2.2 Reduce Fear of Crime
- 1.3 Quality of Life
- 1.4 Institutionalization
  - 1.4.1 Human Resources
  - 1.4.2 Non-OCAP Funding
- 1.5 Capacity Building (Know-how)

2.0 Political Effectiveness

- 2.1 Support of Targeted Communities
- 2.2 Support of Other Interest Groups

EXHIBIT 6-2

OBJECTIVE/SUBOBJECTIVE	IMPORTANCE WEIGHT
1.0 Social Effectiveness	
1.1 Social Process Impact	
1.1.1 Anti-crime Activities	.07
1.1.2 Community Organization	.07
1.1.3 Public/CJS Relations	.04
1.1.4 Reduction of Crime Causes	.05
1.2 Crime-related Effects	
1.2.2 Reduce Crime Rate	.16
1.2.2 Reduce Fear of Crime	.12
1.3 Quality of Life	.09
1.4 Institutionalization	
1.4.1 Human Resources	.16
1.4.2 Non-OCAP Funding	.08
1.5 Capacity Building (Know-how)	.02
2.0 Political Effectiveness	
2.1 Support of Targeted Communities	.07
2.2 Support of Other Interest Groups	.07
	1.00

EXHIBIT 6-3

EFFECTIVENESS SCORES

OPTION AND YEAR		OBJECTIVE/SUBOBJECTIVE												
		1.0										2.0		
		1.1				1.2		1.3	1.4		1.5	2.1	2.2	
		1.1.1 (.07)	1.1.2 (.07)	1.1.3 (.04)	1.1.4 (.05)	1.2.1 (.16)	1.2.2 (.12)	(.09)	1.4.1 (.16)	1.4.2 (.08)	(.02)	(.07)	(.07)	TOTAL UTILITY
1979		10	7	9	2	10	6	4	9	5	15	5	10	7.5
1984	Option A	38	34	24	12	35	38	26	42	31	25	42	50	35.2
	Option B	26	24	18	9	23	23	18	30	23	23	33	45	25.3
	Option C	16	15	13	7	14	13	10	20	16	20	20	40	16.8

Note: Weights from Exhibit 6-2 are given in parentheses.

of the importance weights and effectiveness scores shown in Exhibits 6-2 and 6-3. Experts used in both capacities included NIJ staff, four evaluators working on the overall evaluation, two people providing technical assistance to projects, and three additional researchers in the community crime prevention field. OCAP and Congressional staff also reviewed the analysis to ensure that it represented a "typical" Congressional viewpoint on what should be assessed, and they were also given an opportunity to challenge the inputs (importance weights and effectiveness assessments) to the model. The final inputs to the model were an amalgamation of the judgments of all the experts who participated.

Clearly, the task of assessing effectiveness levels in 1984 is largely a matter of judgment. However, the assessments for 1979 regarding effectiveness were also largely a matter of judgment since "hard" data to be collected by AIR were not yet available. Data from some previous AIR work, the DSC projects, and anecdotal evidence were the primary informational inputs to the experts other than personal experience.

To obtain judgments about effectiveness, a 0-100 scale was defined for each objective/subobjective. The 0-point corresponds by definition to a zero impact level. The 100-point corresponds to the maximum plausible but realistic level of impact attributable to a very effective $40 million program. Where possible, the 100-point was interpreted in terms of specific phenomena. For instance, 100 on the scale for crime reduction was defined as a 30% reduction in project areas between 1984 and 1979 attributable to the program. (That is, it was felt that the maximum plausible impact in project areas would be a 30% crime reduction.) In Exhibit 6-3, the score in 1979 represents a judgment about what percent of maximum plausible effectiveness had in fact been attained at the time of the analysis; the score for 1984 represents projections about effectiveness levels in 1984.

The judgments in Exhibit 6-3 clearly reflect varying perceptions across objectives about what OCAP programs had accomplished in 1979, and what would in fact be accomplished in 1984. To arrive at aggregate MAUT effectiveness scores, experts were queried about importance weights; there was a reasonable degree of consensus and a representative set of weights was used for

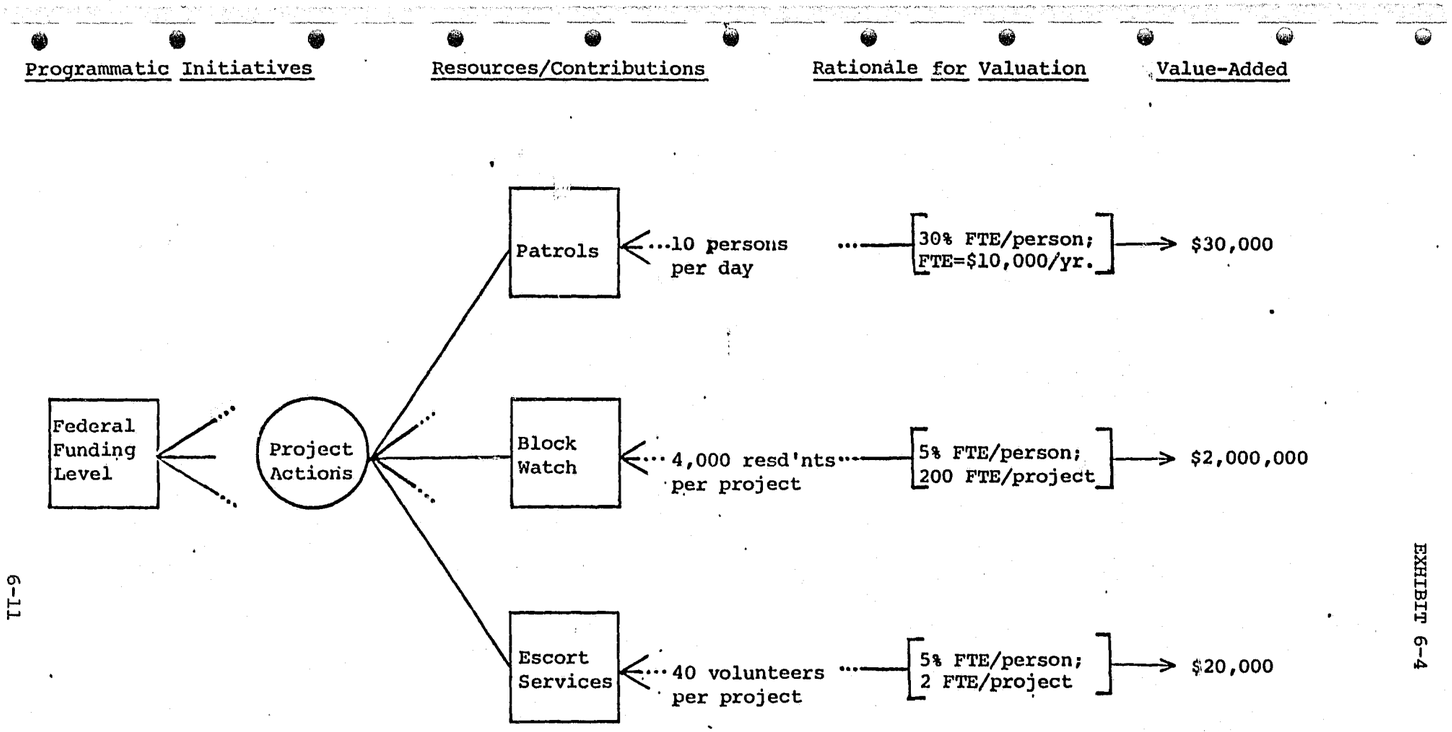
the total utility calculations in Exhibit 6-3. Obvious differences show up between funding options. The question remains, therefore: is the increase in effectiveness worth the additional expenditure?

6.6 The Tradeoff Between Effectiveness and Cost

The technique used for assessing the cost-effectiveness tradeoff was to establish the dollar value of an "ideal" program--one which reached maximum plausible effectiveness on each objective or subobjective.

How much did the experts think should be paid for a "perfect 100"? The range of judgments obtained from the experts indicated little consensus on the issue. Judgments ranged from $15 million to $500 million. It was clear that a factor increasing this spread was strong views about what Congress would in fact be willing to spend on community anti-crime programs. Opinions about Congressional willingness to pay were much more pessimistic than opinions about what should be spent on an ideally effective program, or what the program was "worth." Since the analysis is prescriptive of what should be done (rather than descriptive about what has or will be done) the lower judgments were technically inappropriate. The assessment used for the analysis was $300 million, the median of expert judgments; this figure was substantiated by further analysis of "worth," to double check if it was inappropriately optimistic.

Two techniques for placing a dollar value on program worth are by valuating program outputs or by valuating program inputs. The former would involve looking at, say, crime reduction and placing a dollar figure on it. "Hard" data of that sort were not available, as mentioned. But it was possible, using the second technique, to get a reasonably good assessment on one aspect of program input: value of volunteer time spent on crime prevention activities conducted by projects. The form of the analysis is shown in Exhibit 6-4. We considered only three widely used activities: patrols, block watch, and escort services. For each, we estimated, based on fairly good data, how much volunteer activity there would be for a typical project. Assuming a full-time-equivalent (FTE) was worth $10,000 per year, we calculated the value of these activities to be approximately $323 million.



SCENARIO: MAXIMUM FEASIBLE PROGRAM UTILITY

Federal Funding: $40 million (200 Projects)	Patrols (40 projects)	→ $1.2 million	→ $323 million
	Block Watch (160 projects)	→ $320 million	
	Escorts (70 projects)	→ $1.4 million	

VALUE-ADDED MODEL FOR ASSESSING PROGRAM WORTH

EXHIBIT 6-4

This can be considered a conservative figure since only the three activities were used in the analysis. But this ancillary model does provide some additional support for the $300 million used in the cost-effectiveness tradeoff.

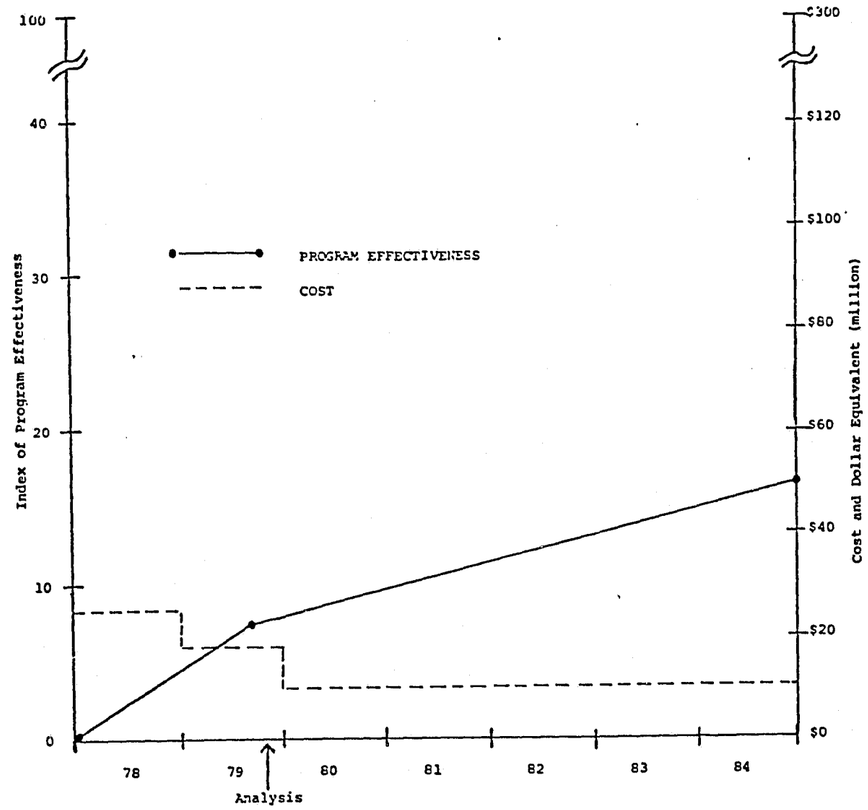
6.7 Results of Analysis

6.7.1 Annual Cost-Effectiveness. The basic results of the analysis are presented in Exhibits 6-5 through 6-7. Exhibit 6-5 shows effectiveness in terms of the percentage of the effectiveness of the ideal program on the left hand scale, and in dollar worth on the right hand scale, and cost for a $10 million program. (Effectiveness scores for 1980-1983 are interpolated between the 1979 and 1984 scores.) Exhibits 6-6 and 6-7 refer to the $20 million and $40 million programs respectively. These exhibits show annual, not cumulative, cost and effectiveness. The calculations of the analysis described above, with the inputs assessed by DSC, were derived using a computer program described in Seaver et al. (1979). (The program performs a variety of needed analyses on the inputs of Exhibit 6-3 and also performs sensitivity analysis on changing weights as discussed later.)

If we consider only effectiveness and cost in 1984, which is the preferred option? One decision rule by which the funding options can be directly compared, is to subtract the cost of the program from the dollar value of the effectiveness. For 1984, this produces net effectiveness minus cost values of $40 million, $55 million, and $65 million for the $10 million, $20 million and $40 million programs respectively. This suggests that, if net surplus of value over cost is the criterion the $40 million funding level would be the preferred option by 1984. Net values on a year-by-year basis are shown in Exhibit 6-8.

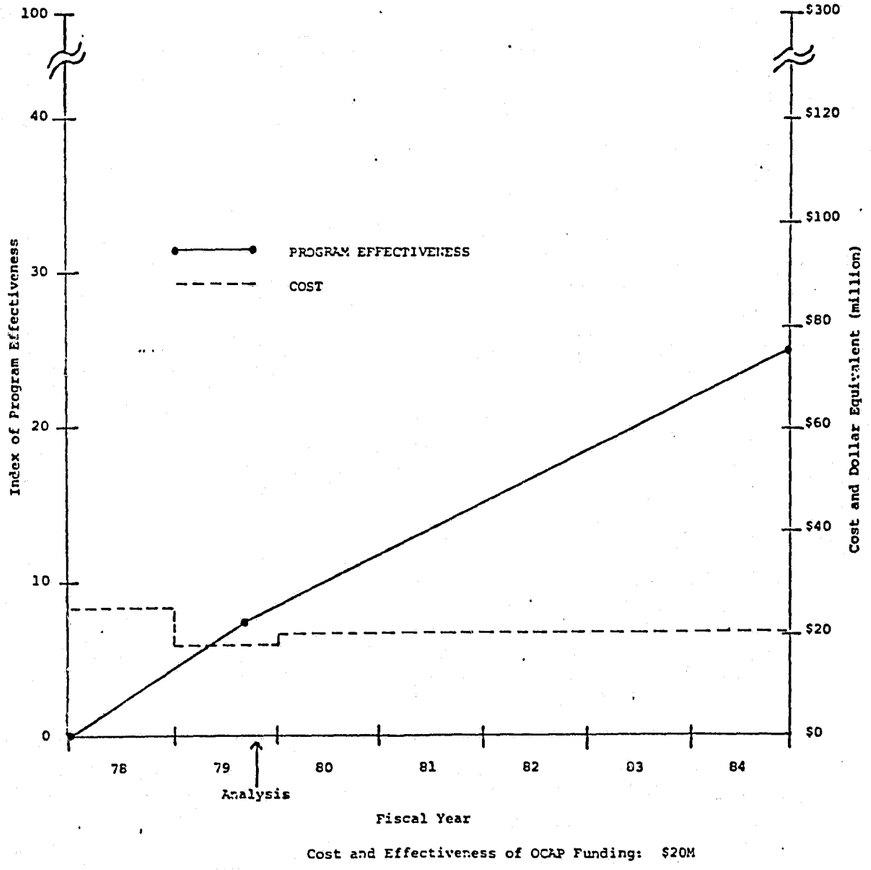
(Under certain conditions, however, a ratio rule rather than the difference rule might be more appropriate for determining the funding level to be adopted. That is, the funding level adopted should be the one that produces the highest effectiveness-cost ratio. The ratio rule becomes more appropriate as the decision becomes one of allocating a fixed amount of resources. If a ratio decision rule were adopted, the $10 million program would be preferred in 1984 with an effectiveness-cost ratio of 5, followed by the $20 million program with a ratio of 3.7, and the $40 million program with a

EXHIBIT 6-5



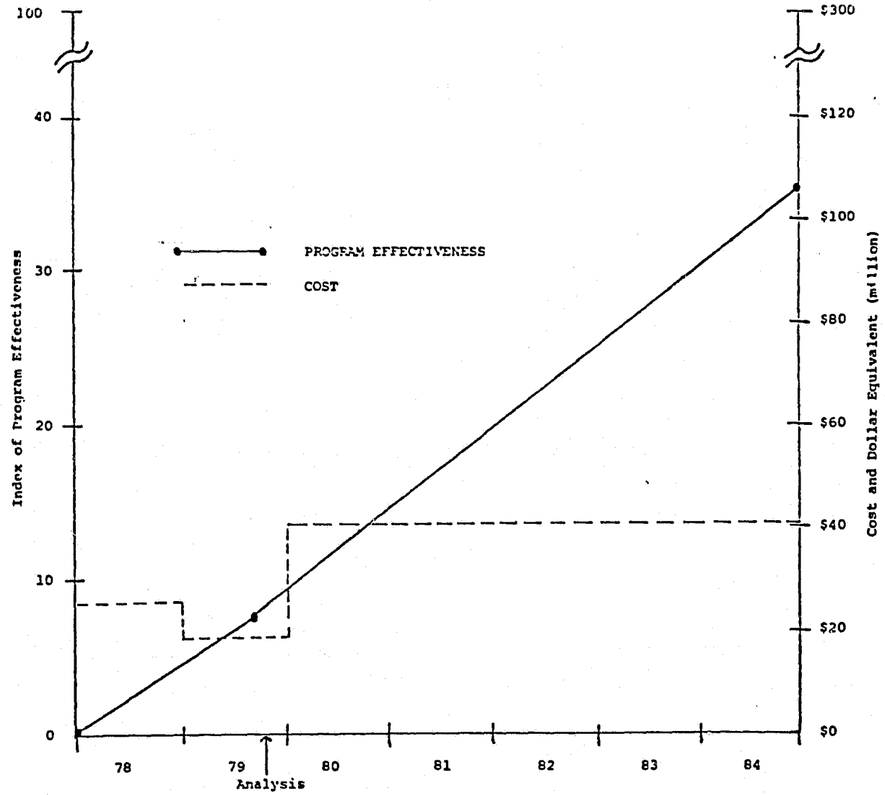
Cost and Effectiveness of OCAP Funding: $10M

EXHIBIT 6-6



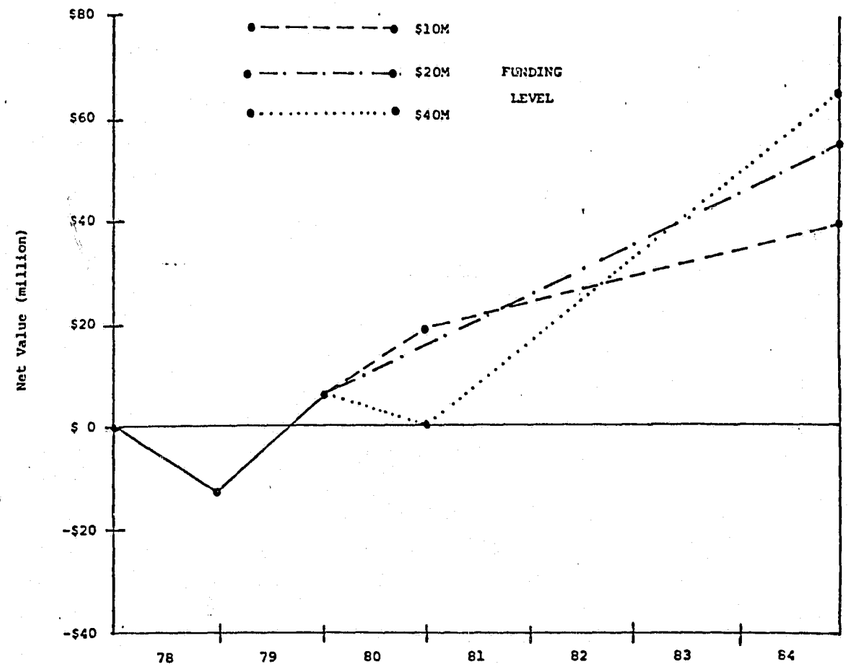
Cost and Effectiveness of OCAP Funding: $20M

EXHIBIT 6-7



Cost and Effectiveness of OCAP Funding: $40M

EXHIBIT 6-8



Annual Net Value (Effectiveness minus Cost)

ratio of 2.6.)

6.7.2 Cumulative cost-effectiveness. The discussion above assumes that a snapshot is taken of the program in 1984. But what if one views the program as a source of services and wants to see a positive "return" early and a rapid cumulation of excess benefits over cost? The decision can also be looked at from this second perspective: that of cumulative cost and effectiveness and how quickly there is a positive return. The effectiveness and cost curves shown in Exhibits 6-5 through 6-8 reflect annual measures and do not show clearly cumulative effects. By adding the areas under the relevant curves, however, these exhibits do suggest that the lower cost program "pays for itself" sooner than do the more expensive options. Exhibit 6-9 shows explicitly the cumulative net value (effectiveness minus cost) of each of the funding level options. At the time of the analysis, the cumulative effectiveness had not yet exceeded the cumulative cost. However, the results indicate that during FY 1980 the cumulative cost-effectiveness would be expected to become positive for both the $10 million and $20 million funding options. The $40 million option would not be expected to have a positive cumulative value until FY 1981.

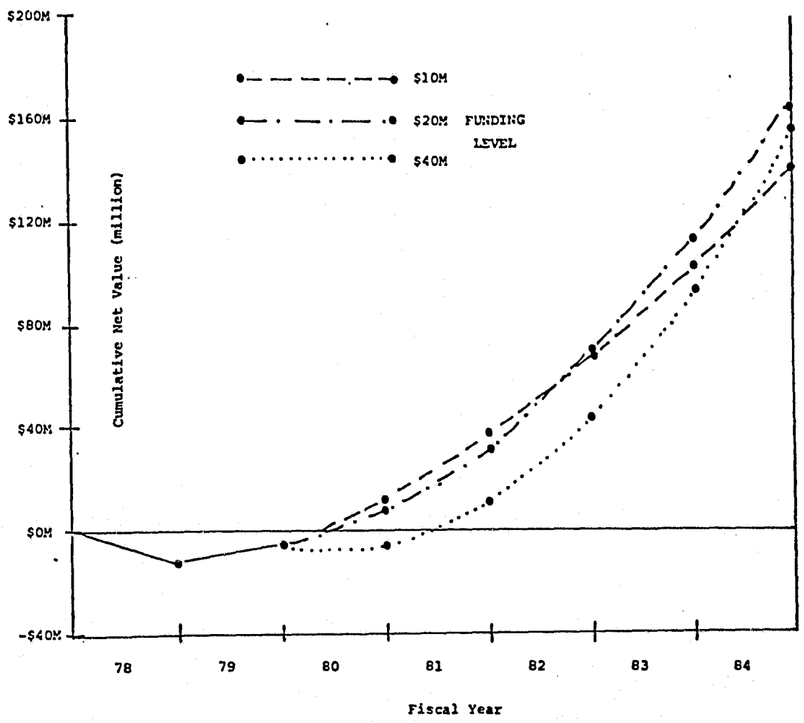
Consideration of cumulative rather than annual cost-effectiveness clearly has ramifications for the funding level decision. In cumulative net value in 1984, the $20 million funding is the preferred option. If a shorter time period were considered, say only through FY 1981, the $10 million funding option would be preferred. If a longer period were considered, the $40 million option would become preferred, since its cost-effectiveness is rising at a faster rate.

This clearly shows the criticality of the time horizon considered in making the funding decision. An emphasis on the next year or two suggests a lower funding level than does a longer term emphasis.

6.8 Sensitivity Analyses

The reader may find it somewhat alarming that the two different viewpoints--and evaluation of cumulative versus 1984-only cost-effectiveness--yield

EXHIBIT 6-9



Cumulative Net Value

different conclusions about which funding level is "best." Actually, we warned in Section 3.0 that the explicitness of the decision rule was characteristic of the decision-theoretic approach, and it is entirely possible and reasonable that the same basic data should lead two or more decision makers to disagree about what should be done. Indeed, in this instance, the analysis suggests that the programs-as-experiments viewpoint (and basing the decision only on expectations in 1984) will lead to higher recommendations about funding level than will the programs-as-services viewpoint (and basing the decision on cumulative net utility). Thus, the decision is sensitive to when one starts "counting" costs and benefits, and when one expects a positive "return." The decision is sensitive to other factors as well, which we will discuss shortly.

However, before discussing factors to which the decision is sensitive, we want to stress one sense in which the analytic conclusion is extremely robust. None of the analyses suggest that zero funding is the right decision: they all suggest that a positive funding level and therefore continuation of the program is called for on the basis of the inputs to the analysis. The analysis in fact suggests that the only case that could be made for termination is by comparing costs and effectiveness for 1978 and/or 1979 and basing the decision wholly on those years, without allowing for potential increases in effectiveness and benefits! The analysis does suggest that cost-benefit differences (and ratios) are small for 1978 and 1979 and so--if one fixates on past performance only--the analysis does not offer compelling argument for continuation. Unless one adopts the position that the program should have proven itself one way or the other by 1978/1979, the analysis suggests strongly that the program should be continued.

The choice of $10, $20, or $40 million funding level is, in a sense, fine-tuning the analysis, and is sensitive to several factors. For instance, a "discount rate" applied to future benefits (or to future net utility), would tend to suggest a lower funding level, since less "value" would be ascribed to the comparatively large utility gains in 1983 and 1984 for example. Truncating the analysis at a time-frame shorter than five years would have the same effect of suggesting a lower funding level.

The analysis is also somewhat sensitive to the assessed worth of the "ideal" program. Recall that this worth was assessed at $300 million, and that that estimate was used as the basis of the annual and cumulative net utility analyses. A computerized sensitivity analysis was used to assess sensitivity with respect to the strongest argument that could be made for the $40 million program: namely, considering only the cost effectiveness in 1984. Considering only cost and effectiveness in 1984, it was established that the $40 million option remains optimal for an assessed worth as low as $200 million, below which the $20 million option becomes optimal. The $10 million option is preferred for an assessed worth lower than about $125 million. Finally, the status quo, or no program, option becomes the best for an assessed worth of the ideal program of less than $60 million per year.

Another analysis of interest is sensitivity to the weight assigned to Political Effectiveness which in the analysis was assigned an importance weight of .14 (compared with .86 for Social Effectiveness) in 1984. Sensitivity analysis shows that changing the weight on this factor has no effect on the preferred option, which remains the $40 million program based on 1984-only data. In the altruistic situation, where Political Effectiveness has no weight, the dollar equivalent values of the three options in 1984 become: $43.7 million for the $10 million option, $68.8 million for the $20 million option, and $100 million for the $40 million option. More generally, the conclusions depended very little on the specific weights assigned; the critical factor is which objectives are considered rather than their exact weights.

6.9 Conclusion

The overall picture that emerges is that OCAP programs were about at "break even" point when the analysis was undertaken. An emphasis only on historic performance, therefore, could lead to a negative decision. The decision-analytic approach, however, explicitly stressed consideration of future costs and benefits. The analysis does not yield an unambiguous, finely-tuned conclusion about the "right" funding level, given the different decision rules one might apply and sensitivities to certain factors in the analysis.

A more detailed and refined analysis would have been appropriate and necessary if we were to attempt unambiguously to specify what the "right" funding level was, from whose viewpoint, and using which particular decision rule. The decision about funding level, as we discussed, is, for example, very sensitive to whether one uses only 1984 cost and effectiveness levels or whether one uses cumulative net cost-effectiveness measures.

A weakness of the analysis with respect to the Congressional decision was that it addressed only OCAP programs, not those of the entire LEAA. The fact that the OCAP decision problem was overtaken by the broader question of LEAA reauthorization was clearly understood and recognized by the Congressional staff personnel who participated in this study. They indicated that the results were compelling, but that it was unfortunate that similar kinds of analyses and arguments had not been undertaken for the entirety of LEAA.

The results of the decision analysis were used, but as part of arguments about "throwing the baby out with the bathwater." Congressional staffers indicated that, despite apparent merits of OCAP programs, OCAP itself was much too small in the scheme of things to warrant special treatment apart from LEAA as a whole. Given that the analysis strongly indicated that at least $10 million should be allocated to OCAP programs, and given that LEAA was terminated whether in fact bathwater or not, the conclusion certainly seems to be that at least the baby was thrown out.

7.0 POST MORTEM EVALUATIONS OF PROGRAMS

An argument in preceding sections of this report has been that decisions do not depend only on past performance. Decisions depend upon anticipated future performance as well.

A corrolary is that evaluation methodologies--if they are to be useful for decision making--should not just take into consideration past performance or snapshots of present performance. Ideally, they should also assess what is expected in the future or how past and present project performance can best be extrapolated. That kind of evaluative information, we believe, would be appropriate and useful in that it would not artifactually lead the decision maker to consider only past performance.

As mentioned in Section 6.0, LEAA was terminated; and in fact LEAA's termination roughly coincided with the end of the grant under which this study was prepared. However, just prior to the end of the present grant we tested a methodology that assessed past and present performance, and permitted extrapolations based on past performance. The application was made to the Metropolitan Atlanta Crime Commission (MACC) CAC project, which was the first project we visited after announcement of LEAA's impending termination.

Briefly, the threatened termination of LEAA (and, within it, the Office of Community Anti-Crime Programs) triggered several decision processes at the local and metropolitan level, as well as at the federal level. Local CAC projects, faced with termination and without prospect of obtaining support through LEAA block grants, had decisions to make about whether the effort had been worthwhile, what initiatives could and should be undertaken, and whether they would be worthwhile. At the federal level, the question was whether the CAC effort (and, indeed, LEAA more generally) had been worthwhile and, as part of this question, whether there would be any persistence of beneficial results, changes, or other impacts.

7.1 Institutionalization and Residual Effectiveness

More generally, very many programs funded by the federal government are intended to become at least partially self-sufficient over time. (For example, the federal "match" or sharing of funding decreases over time for Community Mental Health Projects.) The CAC Program explicitly had this objective, stated as institutionalization of the community anti-crime concept.* "Institutionalization of the community anti-crime concept" encompassed a wide range of effects, impacts, or change. It included structural changes in the CJS or other organizations, as well as changes in approach or attitude on the part of persons in those organizations. (Examples would include addition of a property engraving service in the police department, or greater cooperation between police and community residents. Physical changes in the environment, such as locks and lights, also embody the CAC concept. Often the most pervasive form of CAC concept institutionalization will be in the residents themselves, and reflected by their changes in attitude, awareness of crime, surveillance of public areas, and willingness to participate in community anti-crime activities, among other things.) While the modes of change clearly vary, they have in common the notion that the concept, one instilled via the project, will survive the project and presumably that changes and benefits therefrom will persist after termination of federal funds.

Persistence of effectiveness is clearly a strong determiner of the worth of a program. Decision makers want to know the extent to which a program "works" while at full-funding level, and how much of its effectiveness and benefits will continue to accrue at reduced funding levels. The truly cost-effective nature of programs may not be evident for some period, until after the bulk of funds have been invested and after support has been reduced but a comparatively high level of effectiveness has been maintained.

*Institutionalization of the CAC concept is, for example, discussed as a stakeholder issue in the American Institutes for Research Evaluation Design Plan (AIR, 1979). They discuss CAC concept institutionalization as an issue of concern to stakeholder groups, and, as a concept underlying "effects" and "factors" AIR argued were relevant to the evaluation.

7.2 The MACC Example

As mentioned, the Metropolitan Atlanta Crime Commission (MACC) CAC project was the first project we visited after announcement of the intended termination of LEAA. For MACC and other projects, termination of LEAA meant that other LEAA block grant and discretionary funds would not be available as alternatives to CAC funding. This posed new questions of whether the CAC approach was worthwhile and whether sufficient interest existed to permit continuation. In the case of MACC, the question was which if any of several involved constituent groups would decide to continue support.

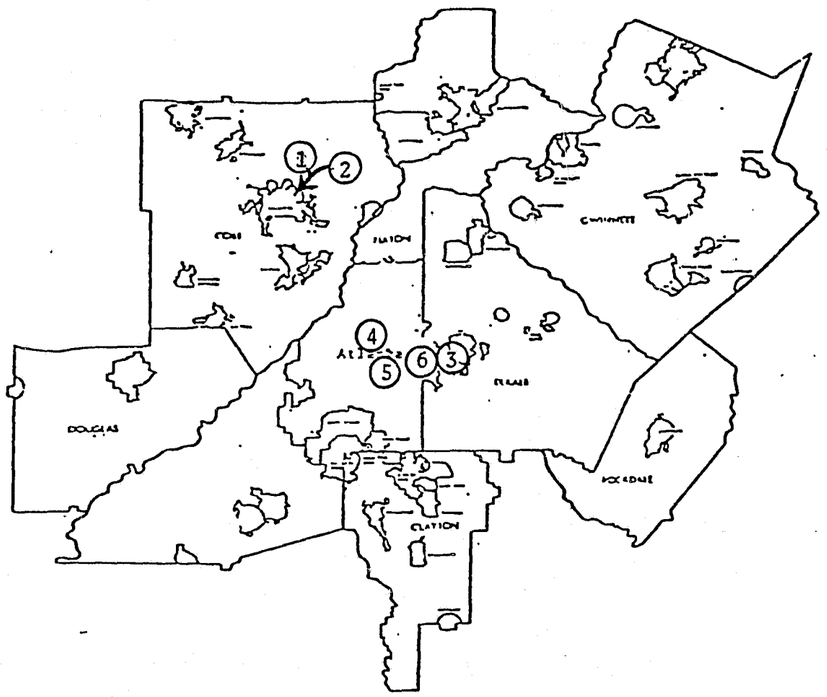
MACC used a "coalition" approach, and embraced separate organizations with distinct activities and objectives. Exhibit 7-1 shows the MACC target area and the constituent groups and their activities. (Not shown is the Lake Claire project, which had become entirely self-supporting at the time of the site visit. Although Lake Claire had become "institutionalized" its data are nevertheless included in the discussion that follows. Lake Claire focused on community organization and community/police relations.) The project had been visited by DSC several months earlier, and had stated its objectives. MACC's objectives are shown in Exhibit 7-2; measures of these objectives are shown in Exhibit 7-3.

7.3 The Model for Assessing Residual Effectiveness

How should the effectiveness of a project during periods of maximum support best be compared to effectiveness at reduced levels of support? And how should projections about effectiveness at reduced levels of support be made? Snapper and Seaver (1980b) discuss the methodology in some detail. Exhibit 7-4 shows the generic "decision tree" model which was developed, and which comprehensively encompasses both full-funding and reduced-funding phases. It provides for an integrated process and effectiveness assessment, in that the model incorporates both process factors ("conditioning events") and effectiveness measures. The left-hand side shows the sequencing during the funding period. After termination of funding, however, there are decisions to be made about whether or not to continue at all and, if so, what strategies to pursue. The critical conditioning event shown in this model is "type and

EXHIBIT 7-1

METROPOLITAN ATLANTA CRIME COMMISSION



(1) South Cobb Jaycees (Victim-witness assistance)
(2) Cobb County YWCA (Battered Women's Shelter)
(3) Candler Park N'hood Assoc. (Block watch, special events)
(4) Northside Shepherd's Center (Security for elderly)
(5) Interfaith, Inc. (Youth counseling, job training)
(6) MACC (Administration)

EXHIBIT 7-2

MACC CAC PROJECT OBJECTIVES

NUMBER	OBJECTIVE	IMPORTANCE WEIGHT
1	Provide Direct Aid to Victims and Witnesses of Crime	.18
2	Provide Material and Emotional Assistance to Battered Women and their Children	.18
3	Prevent and Reduce Crime by Awareness and Cooperation Among Neighbors	.38
4	Reduce Fear and Criminal Victimization Among Elderly	.16
5	Redirect Youth from Criminal Justice System	.10

EXHIBIT 7-3

MEASURES FOR MACC OBJECTIVES

Objective	Measures/Data
1. Provide Direct Aid to Victims and Witnesses of Crime	Percentage of Cases with Direct Personal Contact
	Percent of Dismissed Cases which are Dismissed Because of Witness Non-Appearance
2. Provide Material and Emotional Assistance to Battered Women and Children	Number of Bed-Nights Provided
3. Prevent and Reduce Crime by Awareness and Cooperation Among Neighbors	Number of Events Sponsored (Lake Claire Target Area)
	Crime Reduction--Change in Part 1 Offenses from 1978 (Lake Claire Target Area)
	Crime Reduction--Change in Part 1 Offenses from 1978 (Candler Park Target Area)
	Number of Operating Block Watches
4. Reduce Fear and Criminal Victimization	Number of Mini-Block Watches
	Locks and Lights Purchased and Installed
5. Redirect Youth from Criminal Justice System	Number of School-Referred Clients Participating

"DECISION TREE" MODEL OF INSTITUTIONALIZATION PROCESS AND EFFECTIVENESS ANALYSIS

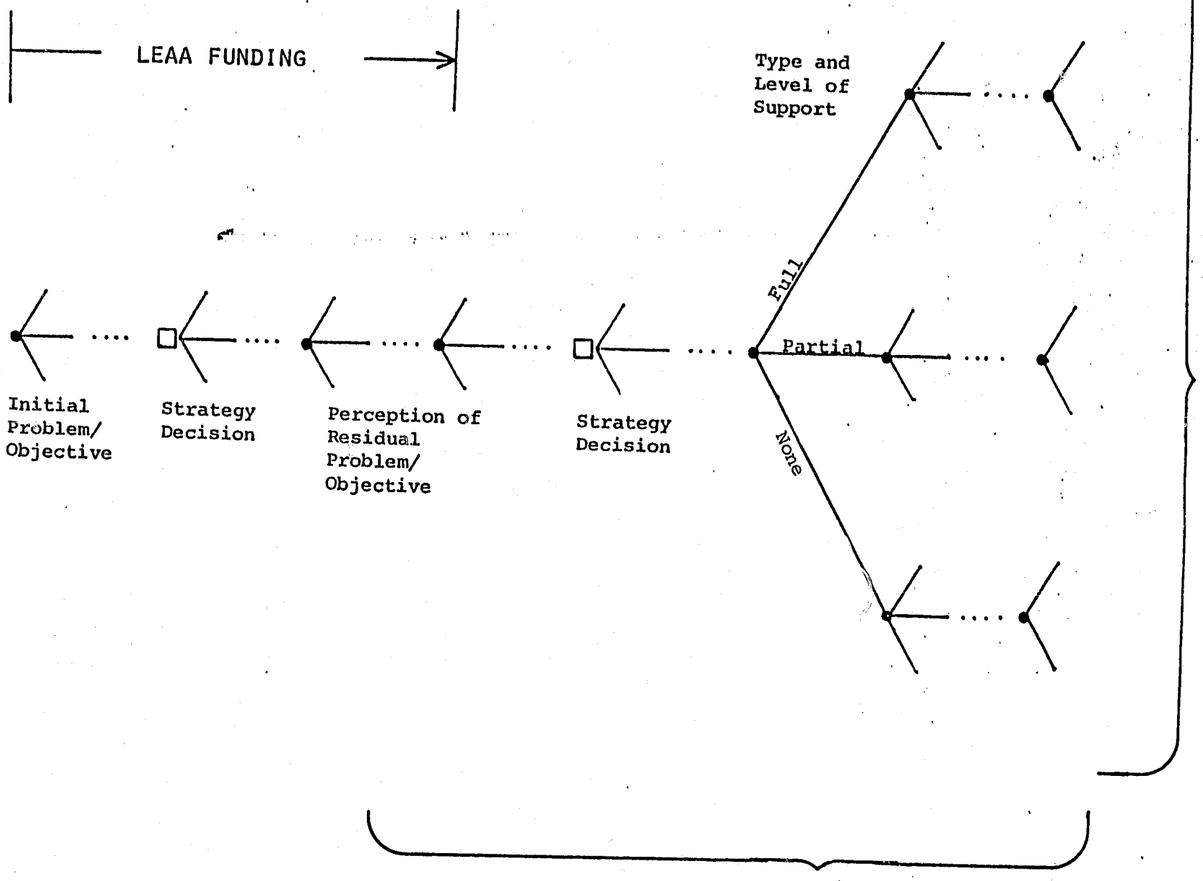


EXHIBIT 7-4

level of support." This could include both volunteer and monetary support, ranging from none to full. Other conditioning events would also be documented and included in the model, insofar as they affected effectiveness.

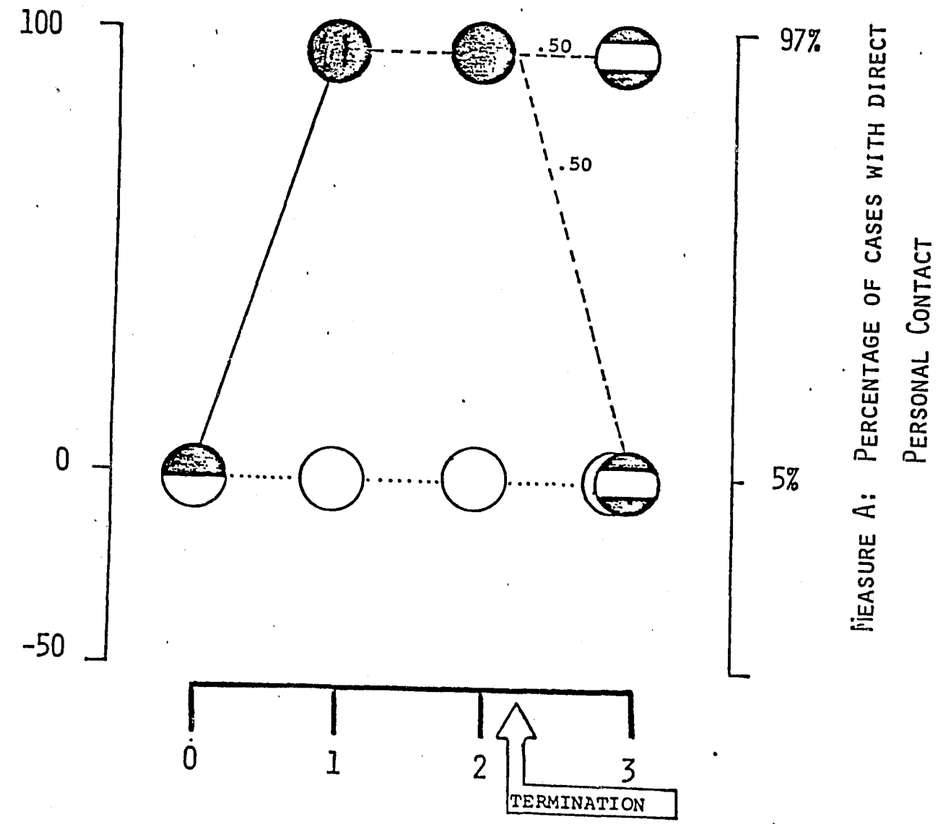
This decision tree model represents ideally how one would assess program effectiveness and provide decision makers with relevant information. Although the present study did not permit application of the full model, it did permit some elicitation from project personnel at MACC about critical conditioning events, especially as they pertain to types and levels of support. The model and its application to MACC also permitted projections about what expected effectiveness would be, contingent upon which conditioning events in fact occur. A simplified version of the model will be considered, based on MACC data.

7.4 Data for MACC

The methodology is perhaps best explained by considering its actual application to MACC in conjunction with presentation of actual data. Interpretation of results is illustrated by Exhibit 7-5. It shows data for the first measure of Objective 1. The right-hand column shows the range over which the measure itself varied, in this case from 5% to 97%. The half-filled circle shows the base rate in Year 0, i.e., at the time the project started. The filled circles represent projections. When actual data are obtained, they are indicated on the display and connected with solid lines. Thus, in Exhibit 7-5, the actual data for Year 1 are displayed. The dashed line shows the projected results for Years 2 and 3. These prior expectations will be replaced by actual data, when they become available.

The dotted lines and open circles in Exhibit 7-5 indicated judgments about what would have happened, had there been no project in the area. Clearly, from the point of view of experimental rigor, this type of judgment is not a "control" and it is not intended to be interpreted as such. Instead, it provides a basis for the cognizant project and program staff to compare actual results against what, in their opinion, most likely would have happened. This provides a basis for judgments about the magnitude of effects attributable to the project, and, ultimately, for judgments about whether the

EXHIBIT 7-5



OBJECTIVE 1: PROVIDE DIRECT AID TO VICTIMS AND WITNESSES OF CRIME

project or program is worthwhile. In the example shown here, the projections about "no project" results reflect judgments that no victim/witness activity would have been initiated without the CAC project, and that personal contact would have remained at about 5%.

The split in Year 3, subsequent to anticipated termination, reflects projections predicated on alternative assumptions about "conditioning events" regarding level of support. The conditioning event, in the case of Exhibit 7-5, is whether or not Cobb County will assume support for the activity in the form of a paid staff position. If so, personal contact was projected to remain at the 97% level achieved by the project; if not, it was projected to fall back to pre-project levels of 5%.

A lesser impact of the conditioning event (i.e., Cobb County staff support) is shown in Exhibit 7-6. The projected drop, under the assumption of no Cobb County support, is less than a return to baseline conditions. Publicity and awareness of the issues, even without continuing support, were judged sufficient to reduce case dismissals about base rates for a significant period of time after termination of all support for the activity itself.

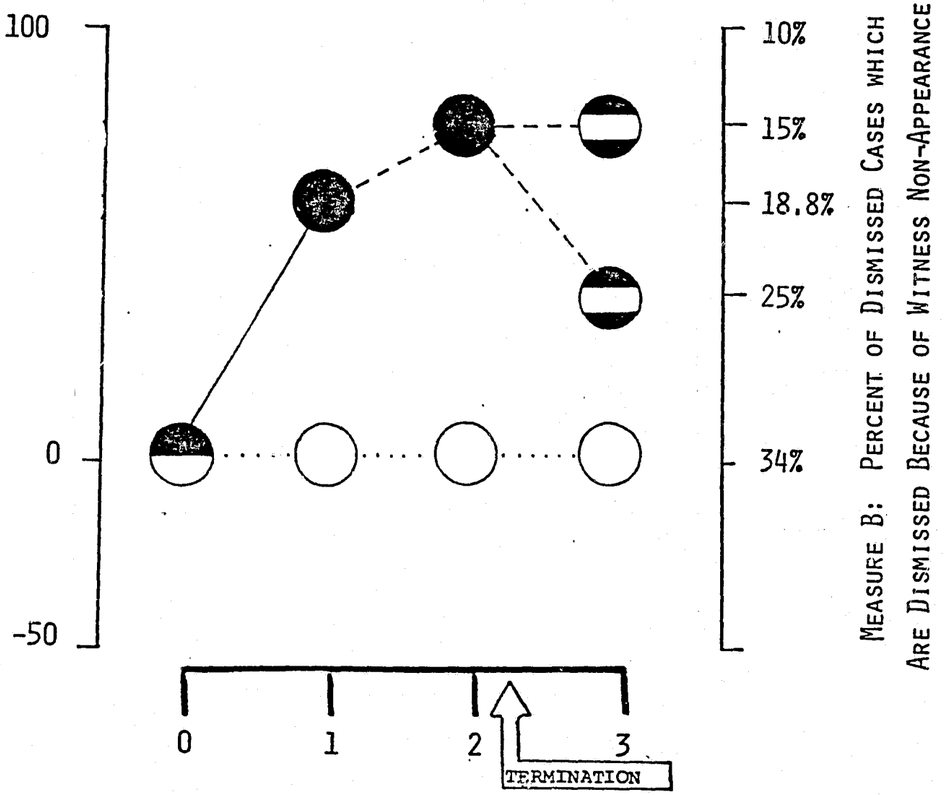
The interpretation of results for remaining figures parallels those for Exhibits 7-5 and 7-6. We will not, however, discuss each in detail and will mention only highlights. Briefly, Exhibit 7-7 shows a case in which an improvement was projected to occur even without the project. Nevertheless, residual impact of the project is shown which would persist even without active additional support. In other words, MACC enhanced a result that would have occurred--an enhancement that is projected to persist after termination.

Exhibits 7-9 and 7-10 are especially interesting because they reflect crime data in two of the project target areas. In Exhibit 7-9 the data show that crime was projected to increase (the open circles) because it was increasing rapidly in surrounding areas. Although crime increased in the project area in Year 1, it increased at a slower rate than it did in the surrounding areas. The data on which Year 1 estimates are based are shown in Snapper

CONTINUED

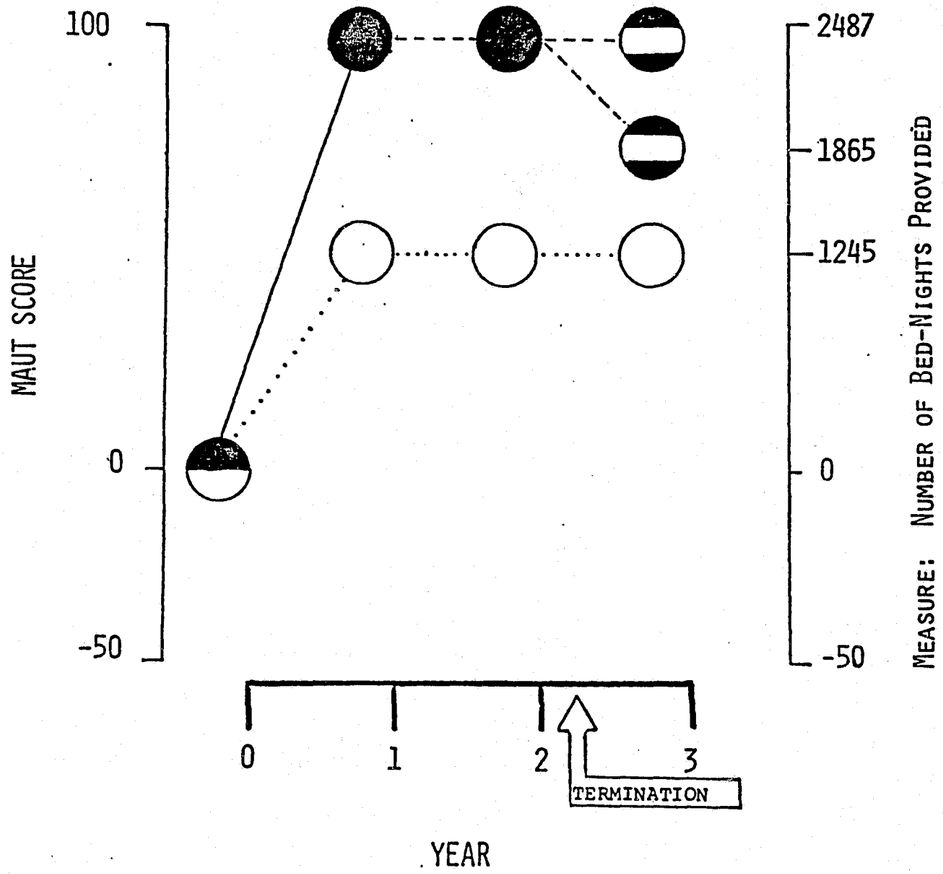
1 OF 2

EXHIBIT 7-6



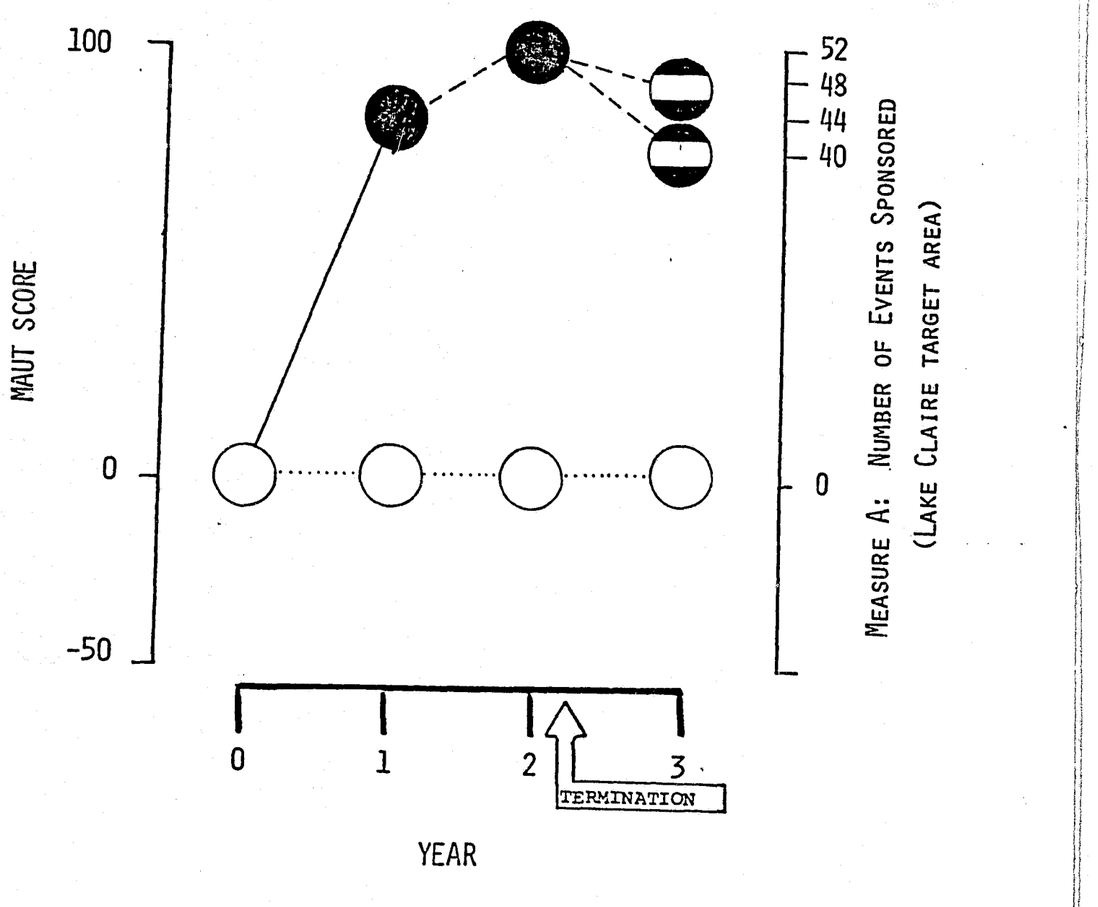
OBJECTIVE 1: PROVIDE DIRECT AID TO VICTIMS AND WITNESSES OF CRIME

EXHIBIT 7-7



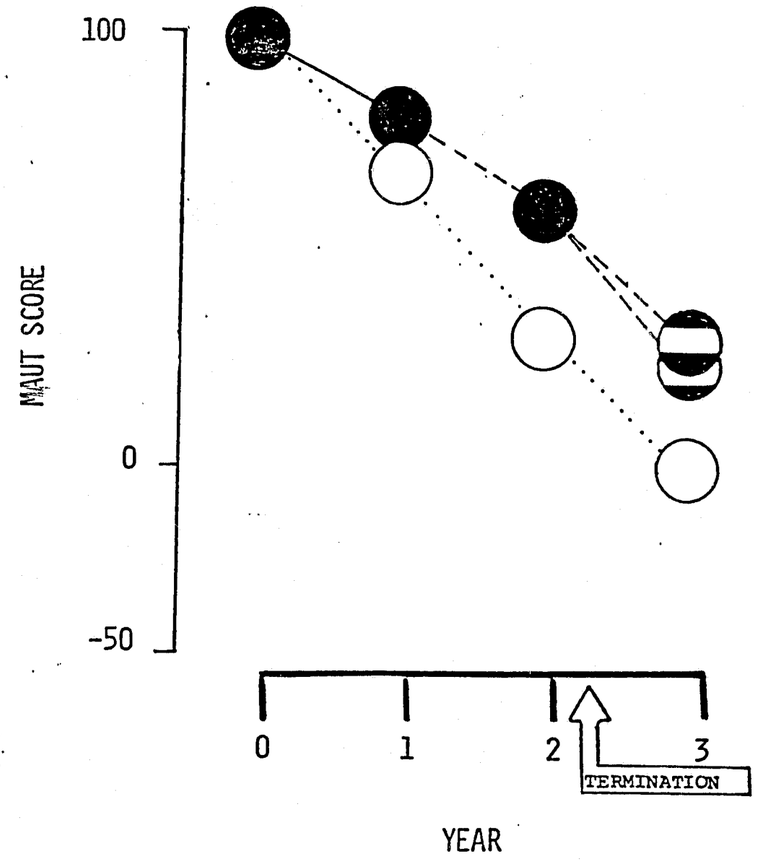
OBJECTIVE 2: PROVIDE MATERIAL AND EMOTIONAL ASSISTANCE TO BATTERED WOMEN AND THEIR CHILDREN

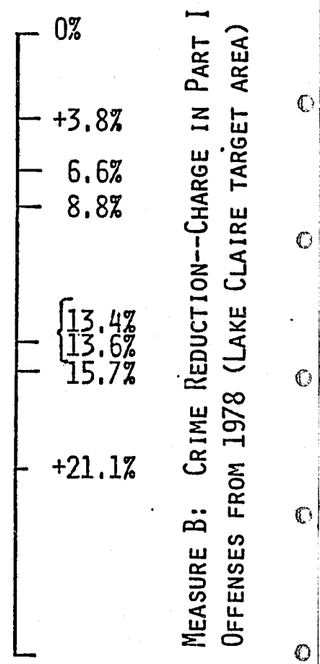
EXHIBIT 7-8



OBJECTIVE 3: PREVENT AND REDUCE CRIME BY AWARENESS AND COOPERATION AMONG NEIGHBORS

EXHIBIT 7-9

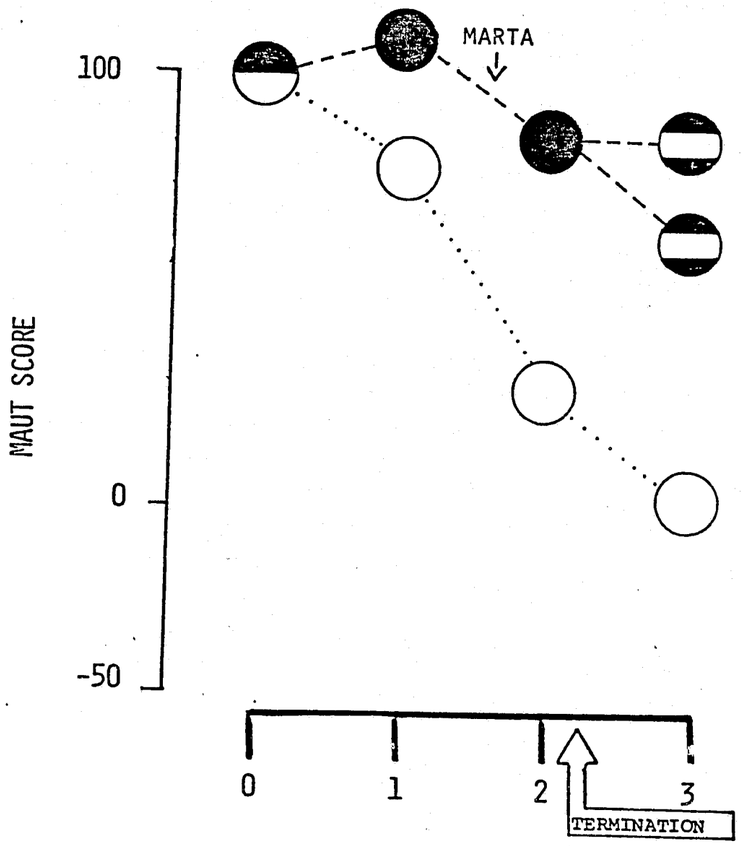


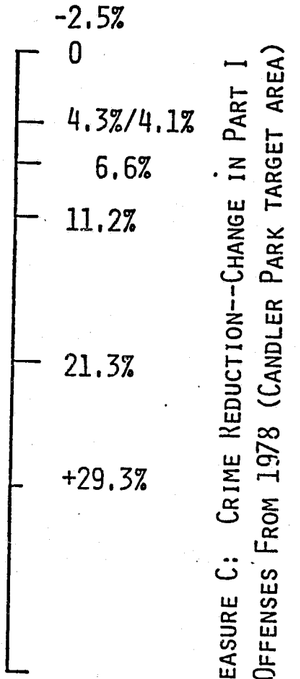


MEASURE B: CRIME REDUCTION--CHARGE IN PART I OFFENSES FROM 1978 (LAKE CLAIRE TARGET AREA)

OBJECTIVE 3: PREVENT AND REDUCE CRIME BY AWARENESS AND COOPERATION AMONG NEIGHBORS

EXHIBIT 7-10





MEASURE C: CRIME REDUCTION--CHANGE IN PART I OFFENSES FROM 1978 (CANDLER PARK TARGET AREA)

OBJECTIVE 3: PREVENT AND REDUCE CRIME BY AWARENESS AND COOPERATION AMONG NEIGHBORS

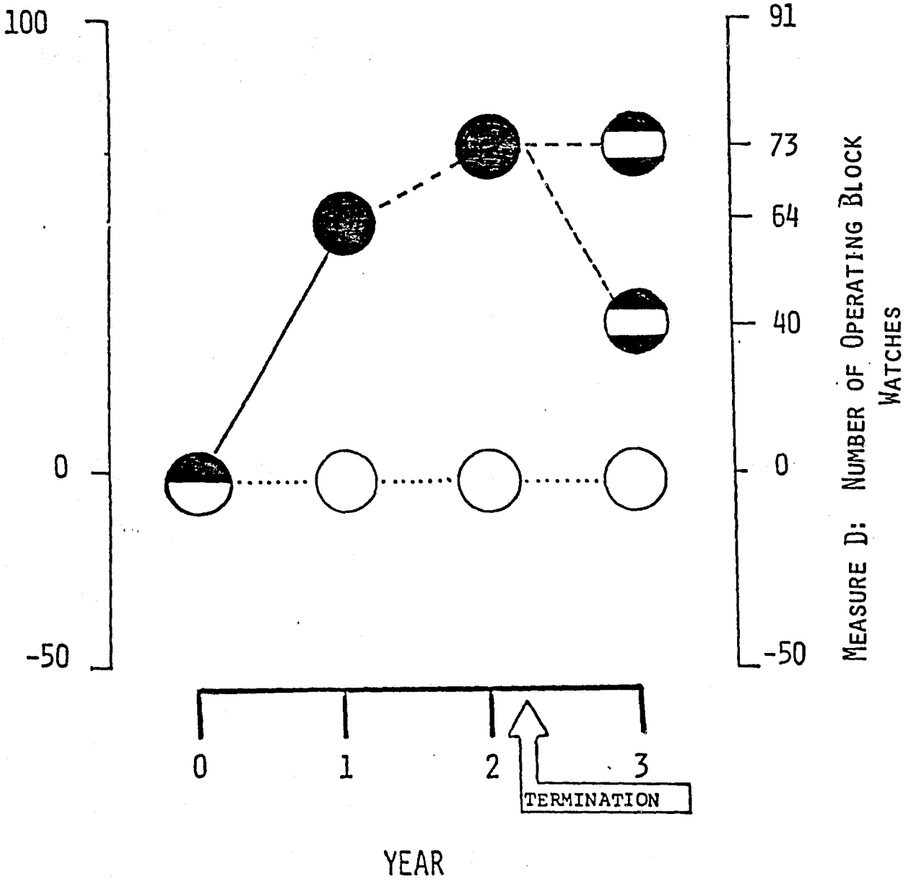
and Seaver (1980b). The projections after Year 1 reflect the opinion that the project area will continue to have slower growth in crime rates, compared to other areas. Exhibit 7-10 shows the case in which crime was actually reduced in Year 1 although it increased in surrounding areas (data are in Snapper and Seaver, 1980b). Projections indicate, however, that the subway station opened by the Metropolitan Atlanta Rapid Transit Authority (MARTA)--which opened after Year 1--would be a significant crime generator. The optimistic projection, following project termination, was that crime would level off if appropriate support and initiatives were undertaken; the pessimistic projection was that crime rates would rise at about the same rate as in surrounding areas.

Exhibits 7-11 through 7-14 show results for other components of the project. While these are comparatively small scale, they were nevertheless regarded as important objectives. Mini-block watches, for example, involved primarily elderly persons in apartments, whose involvement was projected to continue unconditionally, i.e., regardless of whether there was any outside support. The purchase and installation of locks and lights (Exhibit 7-13) was accomplished on essentially a demonstration basis. Once installed, they would of course remain in place. Exhibit 7-14 shows the number of youths participating in diversion activities. The drop from Year 1 reflects a change to a more difficult target population. It was projected that this diversion project would essentially cease if additional support were not available.

7.5 The MACC "Decision Tree"

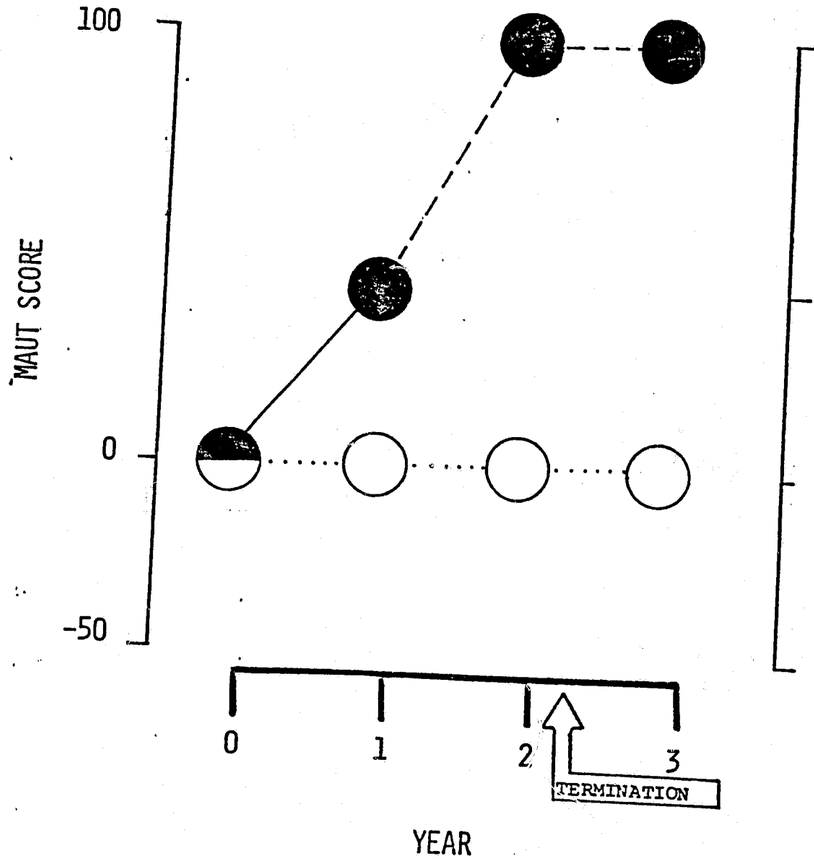
As mentioned earlier, Exhibit 7-4 shows the generic decision tree representation of the proposed method for construct evaluation. The procedure we suggest is to develop such a model prior to termination of a project, based on projections about what will happen. Conditioning events that are identified as strong determiners of results are also included in the model. After termination, the actual occurrence (or nonoccurrence) of conditioning events is determined, along with the actual effectiveness measures and scores. The first half of this approach (but not determination of what actually happened) was possible for MACC during the course of our study, and is shown in Exhibit 7-15. (For convenience of display, we do not

EXHIBIT 7-11



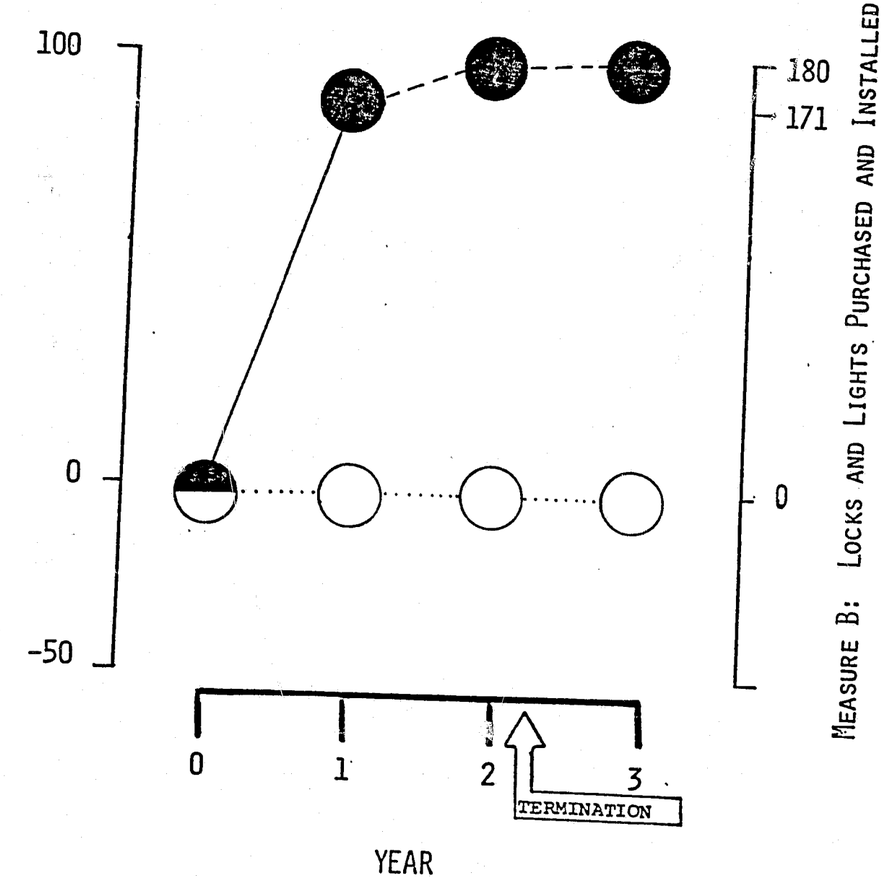
OBJECTIVE 3: PREVENT AND REDUCE CRIME BY AWARENESS AND COOPERATION AMONG NEIGHBORS

EXHIBIT 7-12



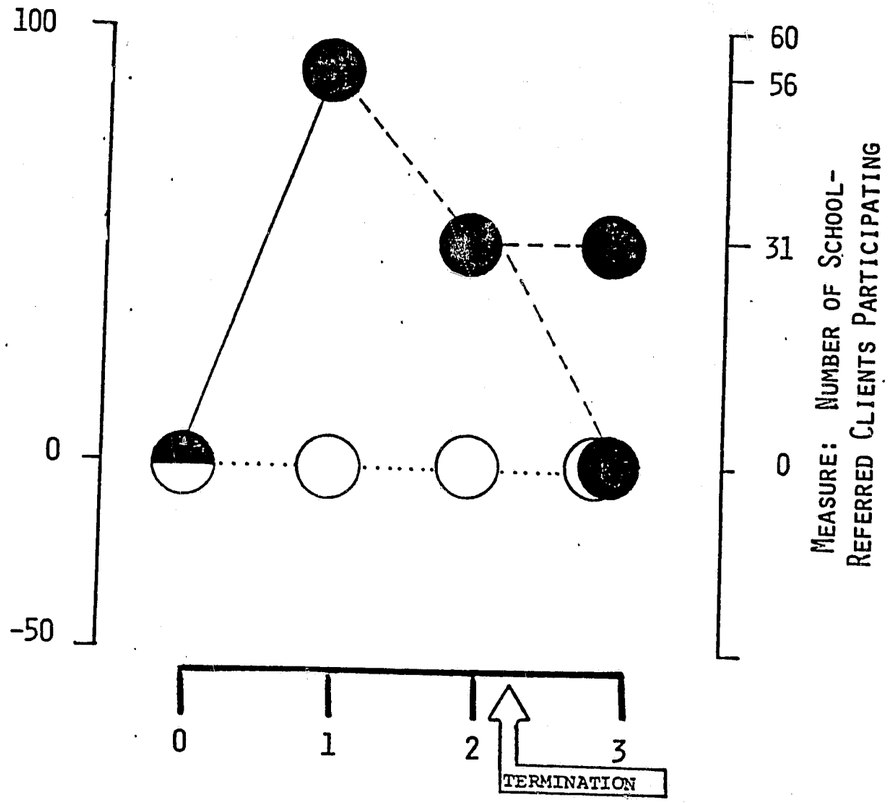
OBJECTIVE 4: REDUCE FEAR AND CRIMINAL VICTIMIZATION

EXHIBIT 7-13



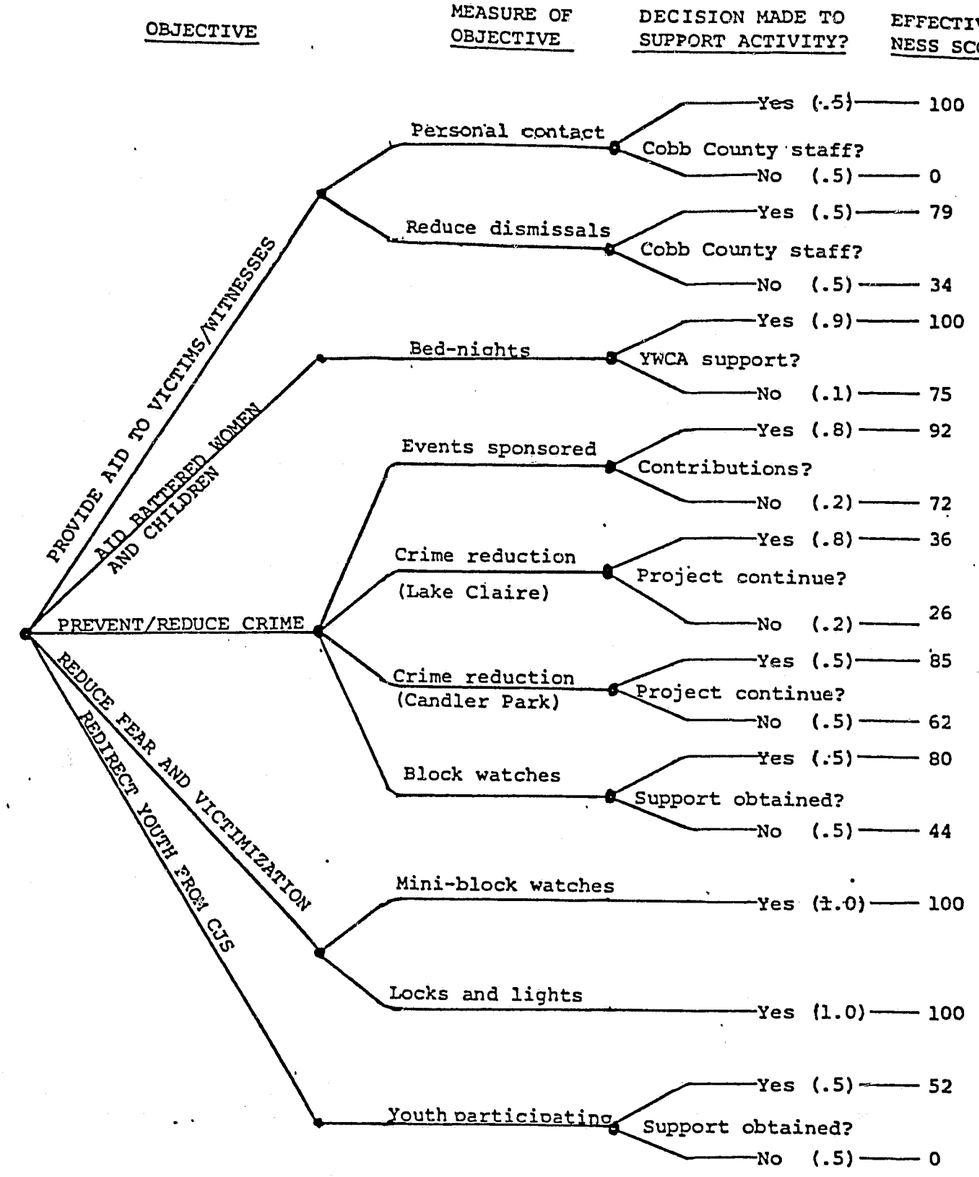
OBJECTIVE 4: REDUCE FEAR AND CRIMINAL VICTIMIZATION

EXHIBIT 7-14



OBJECTIVE 5: REDIRECT YOUTH FROM CRIMINAL JUSTICE SYSTEM

EXHIBIT 7-15



PARTIAL MODEL FOR METROPOLITAN ATLANTA CRIME COMMISSION: CONDITIONING EVENTS AND EFFECTIVENESS

consider the portion of the model pertaining to decisions and activitivies during the funding period. We consider only the decisions or events that determine effectiveness after termination.)

Objectives (shown earlier in Exhibit 7-2) are shown on the left-hand side of the figure. Measures of each objective (shown earlier in Exhibit 7-3) are then shown. Next are shown the decisions (or conditioning events) that were identified as primary determiners of what effectiveness would be one year after termination. (A full post mortem study would of course track actual results for at least a year, and preferably more.)

For convenience, binary decisions or conditioning events were used, and seemed quite adequate for the MACC project. Probabilities were then assigned by MACC project personnel regarding whether there would be a favorable outcome ("yes") or an unfavorable one ("no"). For example, in the case of the victim/witness activity, a probability of .5 was assessed that Cobb County would in fact establish a staff position to continue the activity. The results of that decision--based on project staff projections--are clearly shown on the right-hand side of Exhibit 7-15. We have displayed here the MAUT scores; the corresponding values for the measures can be obtained from Exhibits 8 through 19. For instance, if there is Cobb County staff the MAUT score is projected to be 100 (personal contact: 97% of cases); whereas if there is no Cobb County staff the MAUT score is projected to be 0 (personal contact: 5% of cases).

Although based on judgments, some estimates of degree of institutionalization--as reflected by MAUT scores of post termination effectiveness--can be obtained rather easily. Computationally, we multiply the scores by their probability, summing across the measures for a given objective. This sum for a given measure is then divided by the number of measures, to yield an average score. This average score is then multiplied by the importance weight for that objective (see Exhibit 7-3), and these weighted average scores are then summed to yield the overall expected MAUT score of effectiveness. This score is, in effect, the weighted mean of the actual scores that could result.

To illustrate, the overall expected MAUT score (whose calculation is as described above) is shown in Exhibit 7-16. The pessimistic estimate is based on the assumption that none of the desired decisions or conditioning events occur (i.e., that all the "no's" in Exhibit 7-15 in fact occur). The optimistic score is based on the assumption that all the desired decisions or events occur.

7.6 Implications of Methodology

Clearly, maximum benefits would be obtained if LEAA were to continue support of MACC. One could argue, based on these data, that doing so would in fact be cost-effective. However, the purpose of federal funding was in part to act as "seed money" and for the projects to wean themselves from federal support. If projects maintain a reasonable level of effectiveness (i.e., institutionalization) the federal dollar exerts a high degree of leverage. Such programs may be highly cost-effective in terms of the federal dollar.

The evaluation approach developed for the MACC project helps assess the extent to which projects are institutionalized. (In a later section of this report we will consider separately the implications for cost-effectiveness of the overall national program.) This can serve two obvious purposes. One is to permit early estimates of project (or program) cost-effectiveness in which assessment of future benefits is explicitly modeled and may be at least partially empirically based. (This is in contrast to the OCAP study in Section 6.0, in which a wholly judgmental projection of future effectiveness was made.) The second use of this type of model is in deciding whether a project is institutionalizing itself and when its support should be reduced or eliminated altogether. Uncertainties of course will always exist and the decision about whether to reduce funding will never be risk-free. The method described above, however, will help identify how "risky" termination would be in terms of likely reduction in effectiveness or to identify those activities that should receive continued support should reduction in funds rather than termination be a possibility.

EXHIBIT 7-16

CALCULATION OF OVERALL, WEIGHTED MAUT SCORES, BASED ON VARYING ASSUMPTIONS

Objective	Assumption Underlying MAUT Computation		
	Optimistic	Expected	Pessimistic
1	100	53.3	17
2	100	97.5	75
3	73.3	64.4	51
4	100	100	100
5	52	26	0
Weighted Score	85.1	70.2	51.9

8.0 THE DECISION TO TERMINATE THE SEPCPC PROJECT

In Section 7.0 we indicated that it might be cost-effective to terminate a project--not because it is failing but because it has been successful. The investment of staff resources and money continues to "pay off" over time, thereby greatly enhancing the cost-effectiveness of the initial "investment."

8.1 The SEPCPC Example

The Southeast Polk Crime Prevention Council (SEPCPC) CAC Project was located in a rural area in Polk County, southeast of Des Moines, Iowa. It was a fairly small project in terms of grant size, serving an area of about 14,000 residents. It was also distinctive because it was one of a very few rural projects funded by OCAP. The crime problem, predictably, was different from urban area problems, as were the crime prevention strategies. There was, for example, little typical community organizing because of the large distances between households. Instead, activities included property engraving (to deter theft and enhance recovery and conviction), providing roadside signs with house numbers (to reduce the considerable response time of the sheriff's patrol cars and other emergency vehicles), residential security information dissemination, and a school-based delinquency prevention program.

Objectives of SEPCPC are shown in Exhibit 8-1. Our initial visit to the project in 1978 indicated that it was successfully implementing its activities, and that it was beginning to prove effective in terms of its stated objectives. For example, people were "mobilized" in the sense that they participated in the activities and there was some indication that the crime rate was falling.

The SEPCPC project was not refunded, and in fact the project did not even apply for additional funds and continuation. The project simply terminated itself without fanfare or explanation to OCAP, had its phone disconnected and disappeared.

EXHIBIT 8-1

SOUTHEAST POLK CRIME PREVENTION COUNCIL
CAC PROJECT OBJECTIVES

OBJECTIVE	WEIGHT
1. Mobilize Individuals and Community Groups to Prevent Crime	.24
2. Reduce Crime	.28
3. Reduce Fear of Crime	.14
4. Improve Citizen-Police Cooperation	.23
5. Deter Youth From Crime	.11

8.2 SEPCPC's Decision

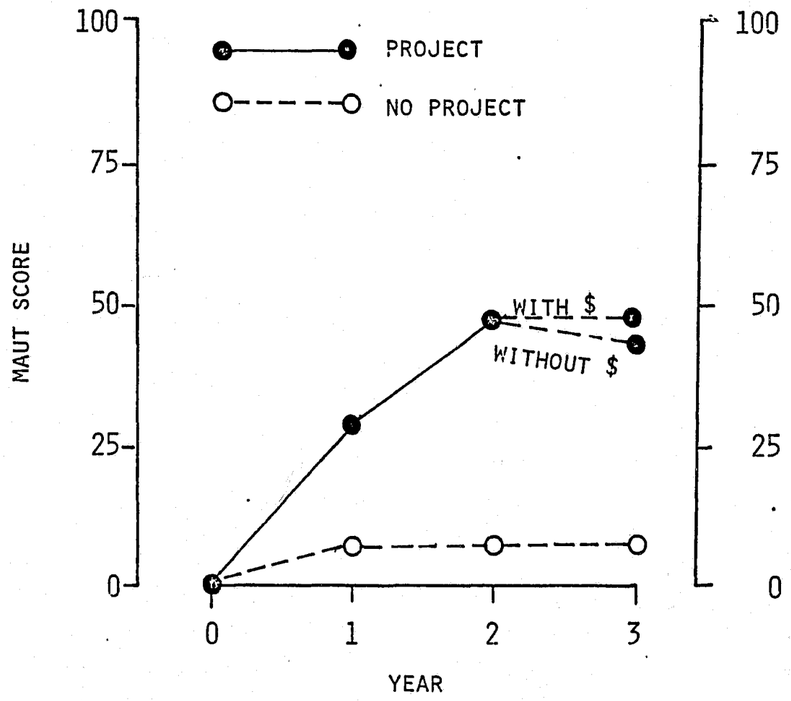
About six months after SEPCPC disbanded, we contacted personally the ex-Project Director and arranged to meet him one evening. He argued, using our terminology, that SEPCPC had terminated itself because its efforts had been effective. He argued that continued OCAP support would have made very little difference. The volunteer fire department had taken over the rural address marking activity. The school system adopted the delinquency prevention effort. And the sheriff's office had hired a full-time crime prevention specialist, and assumed responsibility for public education regarding residential security and for property engraving. SEPCPC had, in effect, entirely institutionalized itself.

Was the decision by SEPCPC to terminate itself a good decision? We analyzed the decision retrospectively, using a slight variant of the modeling approach discussed in Section 7.0. The basic elements of the MAUT model had been identified from previous trips to SEPCPC (though we will not here present data for each objective). All that remained was to have the ex-Project Director make some assessments about where the project was without additional funding, compared to where it would have been had OCAP funding continued. The aggregated results are shown in Exhibit 8-2. Note that the projection with the funds was that the project would have maintained its previously attained level of effectiveness. Without OCAP funds, a slight decrease in project effectiveness was projected, in terms of the MAUT score.

Although we did not formally establish the dollar versus effectiveness trade-off using the type of approach discussed in Section 6.0, the difference in effectiveness for SEPCPC with and without funds seemed small compared to cost. For their part, the project staff clearly concluded that the project's continuation was not worth their time and effort since stated goals had been attained. The slight backsliding in effectiveness that had in fact occurred by the time we visited the ex-Project Director was viewed as minor and not indicative of accelerative erosion of gains, and as the sort of thing that was not unexpected at the time the decision was made to self-terminate. (For example,

EXHIBIT 8-2

SEPCPC TOTAL MAUT SCORE



the sheriff's office placed less emphasis on community organization than SEPCPC would have, had it continued.) In short, the project's decision to terminate appeared to be highly cost-effective, and a case in which the federal dollar had exerted a greal deal of leverage.

That the total MAUT score is near 50 indicates that the project did not accomplish all that it had hoped to as it began. In particular, the rural nature of the target area made mobilization difficult and, in retrospect, the project's expectations were probably somewhat unrealistic in that respect. It had been, however, relatively successful on other objectives, particularly in reducing crime in the target area (compared with surrounding similar rural areas). Thus, the project's decision to terminate itself reflected both recognition that it had institutionalized its achievements and that other achievements (e.g., extensive mobilization) were probably not feasible.

PART III -- METHODOLOGICAL PERSPECTIVES

9.0 SHORTCOMINGS OF EVALUATION RESEARCH FOR AIDING DECISIONS

At the beginning of this report, we stated that a purpose of this project was to test the application of decision-theoretic approaches to evaluation research, especially in regard to decision making. Our claims are fairly modest in regard to how well the decision-theoretic approach worked: some development of the methodology was needed; certain methodological extensions were made; and application was on an opportunistic basis, but on the whole this study demonstrated the practicability and usefulness of the method. And on the whole we felt that we were able to capture the essence of decision problems (though influencing the Congressional decision about OCAP certainly proved an elusive prospect).

We reached much stronger conclusions about the limitations of other evaluation methodologies for purposes of aiding decision making, however. This section will focus on these conclusions, and Section 10.0 will summarize what we believe the appropriate role of decision-theoretic approaches to be.

9.1 The View that Evaluations are Empirical, Experimental Exercises

A powerful assumption about programs is that they are stable, that "treatment effect" is constant, and that a "snapshot" of performance adequately reflects past, present, and future performance. Guttentag (1973) criticized evaluation methods that make such assumptions, arguing that the assumption is invalid for most if not all programs. While many programs may not fit what Guttentag called the "experimental paradigm" because effectiveness is not stable and constant, and because present performance does not adequately represent future performance, the results of the present study suggest to us a gap between evaluations-as-experiments and evaluations-as-decision aids that is in fact procedurally and conceptually greater than Guttentag argued.

The present study explicitly does not question validity of experimental or other approaches to evaluation per se. But, we are explicitly arguing that the results of many such approaches may be inappropriate as inputs to certain kinds of programmatic decision making, and, further, that they may bias decision makers against programs. The extent of this biasing effect will be greatest when the program is in flux and improvement over time is likely, as was certainly the case with OCAP programs.

9.2 A Caveat

A major caveat is in order. There are certain kinds of decisions--usually incremental, fine-tuning decisions--in which past empirical data are extremely salient. These data provide the basis for "adjustments" in the program. Each adjustment is successively evaluated to see if the direction of change is in fact better. The relevance of such data is far less when there is a "fundamental" decision to be made, in which options include radical departures rather than incremental changes from past approaches. When fundamental changes are considered, or when performance may otherwise improve over time, we would argue that past and present performance are in principle inappropriate as inputs to the decision, except in the very limited way in which they may be used to project the effects of the currently used strategy itself. We discussed this distinction rather fully in our evaluation design plan (DSC, 1979), drawing upon Etzioni (1967), and will not repeat our argument fully here.

9.3 Deficiencies in Typical Evaluation Approaches

Guttentag based her arguments largely on what she referred to as the "experimental paradigm." But problems exist in regard to decision-making applications with any method that stresses empirical, historic results at the expense of assessing future effectiveness and benefits. This includes evaluation methods that truncate the assessment of effectiveness at the time federal funding ends, though effectiveness and benefits may persist considerably beyond.

We can illustrate the above with the CAC Program evaluation, as originally conceived. Two interrelated approaches to evaluation were funded. Both were designed to track project-level occurrences over the period of project funding. The first is AIR's "rationales" approach which is a special type of process or implementation evaluation (AIR, 1979). The approach attempts to build essentially an evidentiary chain between the baseline state of affairs and changes that occur. An advantage of this approach is that, hopefully, it helps attribute change to the programmatic process. It attempts explicitly to distinguish "program inputs" from "disposing conditions" and exogenous "other events" that impinge on the change process. The second approach was the MAUT approach for monitoring program effectiveness, discussed in Section 4.0. Although the approach recognized the importance of projecting "prior expectations" about results, to be compared with actual performance, the initial version of the model considered only future effectiveness within the period of funding. It, like the AIR approach, did not explicitly address what would happen after termination of funding. Both AIR and we incorporated some measures or indicators of institutionalization, but stopped far short of arguing that that was perhaps the most critical decision variable and focused on what happened during the project while it was being funded.

This gap in the evaluation approaches used is highlighted by Exhibit 9-1. This exhibit is adapted from Lewis and Greene (1978), and distinguishes between implementation (process), impact, and construct evaluation. Exhibit 9-1 indicates that approaches did not adequately consider construct evaluation which, in the case of the CAC Program, refers to institutionalization of the CAC approach. Institutionalization would be reflected by persistence of activities (a process consideration) or by continued evidence of project effectiveness in terms of its stated objectives.

Construct evaluation is our focal point here. It refers essentially to assessing whether a programmatic concept (i.e., construct) is sound. This may be limited to testing of certain hypotheses about, say, the efficacy of a drug-abuse treatment program. But often it will refer to the program institutionalizing itself or to some other source of long-term "bang-for-the-buck." Whenever institutionalization, long-term effectiveness, or bang-for-the-buck is relevant, it is likely to be appropriate to assess effectiveness

EXHIBIT 9-1

TYPES OF EVALUATIONS, PURPOSES, AND TIMING

Type of Evaluation	When Done?	Purpose	How Done for CACP?
Implementation	Formative stages, ideally during first phase of funding	Management feedback re projects; "best practices"; description	AIR rationales process evaluation
Impact	Last part of project funding, but before termination, when effectiveness is likely to be maximum	Management feedback; some hypothesis testing about what "works"; identify "successful projects"; assessment of how well program "worked"	DSC MAUT project effectiveness models
Construct	If done, usually done as tests of "will it work" hypothesis as part of implementation/impact assessment. Can also be appropriate on follow-up basis if it is assumed there should be institutionalization or long-term "bang-for-the-buck."	Ideally, to modify assumptions and tenents on which program is based and to test whether the programmatic conception (i.e., construct) is sound.	Some description of "how worked"; limited assessment of institutionalization

and costs and benefits on the basis of projections about the future. Thus, while we have considered the evaluation of the CAC Program, our point is likely to be very general in its applicability to evaluation of other programs.

Moreover, even if the conceptualization of the program itself does not call for assessment of future benefits, such assessments are typically relevant to the decision maker and perhaps should be made even if not implied by one's conception of what the program is about. As mentioned earlier, the degree to which effectiveness persists after termination of funding is a strong determiner of the overall worth of a program. Decision makers need to know both the extent to which the program "works" now and what future benefits will be: as will be discussed in Section 10.0, a reasonable level of persistence in effectiveness could easily imply a three-fold or greater increase in a cost-effectiveness sense compared to an assessment based on past and present performance.

9.4 Biasing Factors in Programmatic Decision Making

There is something very seductive about basing decisions on "hard," empirical data. Such data, one is likely to assume, should make decisions "objective" and help correct any biases the decision maker has. And certainly program people hope that hard evaluative data will convince dubious funding sources that their program works and is worthwhile.

There is, however, something very naive about this view, and in fact reliance on "hard," evaluative data can seriously bias the decision. Moreover, the direction of bias will usually be against the program, and in search of what Guttentag called "spurious objectivity" evaluations may do more harm than good in improving programmatic decision making. We will briefly summarize the key points by drawing on the previous sections of this report.

- Decisions are made to influence the future; therefore, possible future benefits and costs are often more critical to a decision than past benefits. Past benefits and costs may actually be irrelevant to a decision.

- Data from evaluations typically bear more on assessing past effectiveness than on forecasting or projecting future effectiveness or benefits. Given the length of time evaluations themselves require, they are likely to present a very dated picture.

- Assessment of the net difference between benefits and costs is likely to be highly dependent on when the program is evaluated. Evaluations after, say, two years are less likely to show positive net benefits over cost than evaluations made later when more benefits have accrued. Large, slow evaluations are likely to show less positive net benefits than smaller more timely evaluations.

- Congressional and other decision makers are keenly aware of current costs and perceptions of net return, and are very likely to "discount" future benefits. Evaluations may encourage this inappropriate fixation.

- If programmatic effectiveness is increasing, heavy reliance by a decision maker on past evaluative data rather than projections may encourage underestimates of net program cost-effectiveness. Consideration of future benefits may increase several fold cost-effectiveness estimates based only on current performance.

- Large programs (like CACP) are likely to take longer than smaller programs to yield a net excess of benefits over cost; small programs are more likely to show positive results than larger ones at an early point in time, therefore. Early evaluations are very likely to weigh against large programs especially if (like CACP) they are new.

- If future benefits are partially discounted, there will tend to be a bias against large programs, since they tend to take longer to yield positive net difference between benefits and costs.

9.5 The Concept of a Decision Aid

The points above suggest both that program decision making can be technically very complex, that costs and benefits must be cumulated and somehow balanced over time, and that evaluations may aggravate the decision maker's problem and actually bias decisions.

A decision aid (e.g., Brown and Ulvila, 1977) is a technique for clarifying all the factors that enter into a decision, which sharpens the decision maker's intuition and, hopefully, leads to better decisions. As we have argued, there is no particular reason why traditional evaluation approaches should serve this purpose well. In fact, we would argue the closer evaluations conform to Guttentag's characterization of the "experimental paradigm,"

the less likely they are to be useful as decision aids! Decision-analytic techniques, in contrast, have been specially developed as decision-aiding techniques and can be used in conjunction with more traditional evaluation approaches.

In the final section of this report, we will discuss some further implications and recommendations about methodological approaches to improving programmatic decision making.

10.0 COST EFFECTIVENESS AND PROGRAM DECISION RULES

In this section we will summarize conclusions we reached regarding evaluations as aids to decision making. The specific type of decision we will focus on is program funding decisions. We will not consider incremental decisions about how to fine-tune a program, since we believe that many approaches to evaluation are useful for that purpose. We will instead focus on fundamental decisions of the sort: should we terminate this program? should we reduce its funding level? should we fund a new program and, if so, what should its funding level be?

The type of funding decision we are concerned with here is typically a resource allocation problem. It involves a decision about how to allocate resources across programs or other options. The OCAP decision, to illustrate, involved an implicit Congressional decision about the size of the "federal pie" (i.e., public vs. private sector allocation), and allocation of resources in the federal budget.

10.1 The Relevance of Cost-Effectiveness to Program Decisions

Usually, when a program decision is being made, there is no explicit representation of each of the other options for using the resources. In part the reason for this is that there is a very large number of alternatives, and decisions are made about each option separately and usually over a considerable period of time. Congressional funding decisions clearly illustrate this point: Congress does not list all its options, then select those it will implement. Rather than one comprehensive resource allocation decision, involving a single decision rule for allocating these resources, each option is considered separately and there is in effect a decision rule for each option considered.

It is in the type of context described in the above paragraph that cost-effectiveness or cost-benefit assessments are most useful. Basically,

cost-effectiveness methods are useful because they provide a decision rule that can, in principle at least, be uniformly applied across many different decisions. That is, a program would be funded only if it were sufficiently cost-effective, and budgetary constraints would result in raising or lowering this hurdle. Viewed this way, cost-effectiveness assessments are surrogate, piece-meal procedures for allocating resources used in lieu of a comprehensive resource allocation decision.

Needless to say, Congressional and other decision makers usually have only the vaguest intuitive comprehension of a particular option's cost-effectiveness. Information about effectiveness may help, but the concept of cost-effectiveness is still likely to prove elusive.

10.2 The Concept of Net Utility

As suggested in earlier sections of this report, net utility is a cost-effectiveness measure defined using MAUT. It reflects the utility associated with program effectiveness, less the disutility associated with costs. In the remainder of this section, we will discuss net utility models as programmatic decision aids. Our discussion will considerably extend the methodological approach discussed in Section 6.0.

We will also discuss how the net utility approach helps avoid the kinds of biases and decision errors discussed in Section 9.0.

The net utility approach requires that there be an explicit relationship or tradeoff defined between effectiveness (benefits) and cost. This tradeoff can be achieved in several ways. Two approaches follow:

- Assess Worth of "Ideal" Program. This is the approach used in the OCAP decision analysis, discussed in Section 6.0. The approach begins by defining "ideal" program effectiveness in terms of some reasonably specific results the program could achieve. (One way to interpret "ideal" performance is as the long-run asymptotic performance of a well-executed and managed program.) Decision makers or other experts are then asked to assess in dollars the worth of the program.

- Assess "Break-even" Point. The second approach is to define "ideal" performance, as above, but then to assess the level of effectiveness at which the program is just worth its cost. This can be done retrospectively or prospectively. One begins by determining whether the program is worth its cost at the "ideal" level of performance. One then considers past and projected performance on a year-by-year basis to determine the year in which program benefits first justify costs.

There are advantages to the second approach. It essentially enables the decision maker or expert to consider a realistic amount of money and to state what a reasonable "return" is that justifies the cost. In the case of programs already funded, the decision maker can readily assess whether, in his or her view, the program has already attained the break-even point and if not when projections indicate it will.

Either procedure for assessing the cost-effectiveness tradeoff--along with a few fairly mild assumptions--will be sufficient to generate the kind of analysis illustrated in Exhibit 10-1 and Exhibit 10-2 (though we will not go into the axiomization and mathematical development in this report.

Exhibit 10-1 shows annual net utility for a hypothetical program. Annual net utility is the utility or value associated with effectiveness in each year, less the disutility associated with program costs. The MAUT scores associated with effectiveness are in Exhibit 10-1-A. (Zero effectiveness has, by definition, a MAUT score of 0; maximum plausible effectiveness has by definition a MAUT score of 100.) The disutility associated with dollar costs is shown in Exhibit 10-1-B. Disutility is constant, since costs each year are assumed to be constant. The fact that the disutility is -20 implies that the benefits of the program in the third year just cancel costs; that is, the program "breaks even" in Year 3. The net difference for each year is shown in Exhibit 10-1-C. (Note "break even", or net utility of zero, in Year 3.)

As suggested in Section 8.0, annual net utility is the appropriate decision variable if, for some reason, the decision should depend on the cost-effectiveness of the program in one particular year. For instance, one might "write off" the first 5 years (in this example) as R&D costs, and elect to continue the program if it has a positive net utility in the sixth year. (This may sound reasonable, but developmental costs are rarely completely

EXHIBIT 10-1

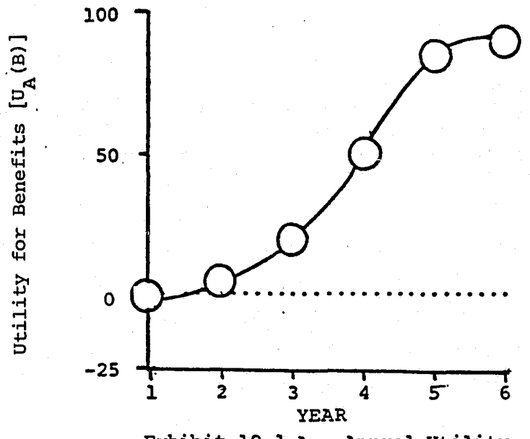


Exhibit 10-1-A. Annual Utility for Benefits

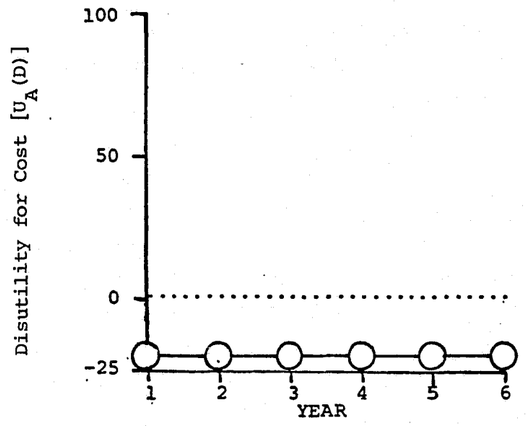


Exhibit 10-1-B. Annual Disutility of Program Cost

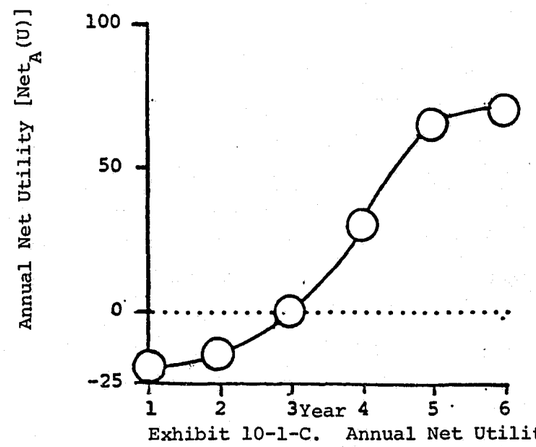


Exhibit 10-1-C. Annual Net Utility [$Net_A(U)=U_A(B)+U_A(D)$]

written off in a decision. For instance, even if in deciding whether to continue the program one focused on net utility in the sixth year, the decision "now" about whether to proceed with this program or some variant would depend on developmental costs over the next five years. That is, one really has a decision "now" about the cost-effectiveness of an R&D initiative over the next five years, and a decision after that period about the cost-effectiveness of continuing that program for purposes of service delivery. In other words, whether one views programs as R&D or as service delivery initiatives, costs and/or benefits over time are relevant to the decision.)

In fact, we suspect that many decisions in actuality depend on perceived net utility at the time the decision is made. For example, it seemed that many people assessed OCAP Programs (see Section 5.0) in terms of perceived performance in 1979, and that a comparison was made between perceived benefits in 1979 and the expenditure of funds. It further seemed to us that these same people had a good intuitive impression of whether, in 1979, the "break even" point had been reached.

A point we have made several times, however, is that it will usually be inappropriate either to base the decision on costs and benefits in a single year or to ignore future benefits and costs. For this reason we are arguing that, in principle, the appropriate cost-effectiveness decision variable will usually be cumulative net utility.

10.3 Cumulative Net Utility

A simple application of the cumulative net utility model is shown in Exhibit 10-2. Exhibit 10-2-A shows cumulative utility associated with benefits, $U_C(B)$. The value for $U_C(B)$ in a given year is the sum of the $U_A(B)$ shown in Exhibit 10-1-A. (The subscripts C and A denote cumulative versus annual values.) Cumulative disutility for costs, $U_C(D)$, is shown in Exhibit 10-2-B.

The cumulative net utility implied by Exhibits 10-2-A and 10-2-B is shown in Exhibit 10-2-C. Cumulative net utility, $Net_C(U)$, falls sharply for the

EXHIBIT 10-2

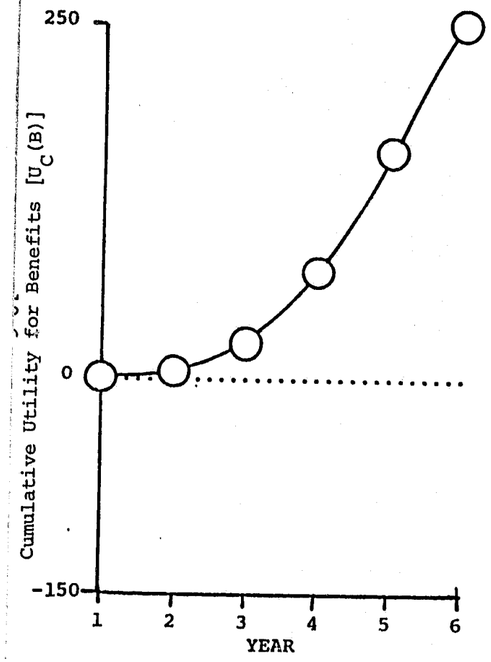


Exhibit 10-2-A. Cumulative Utility for Benefits

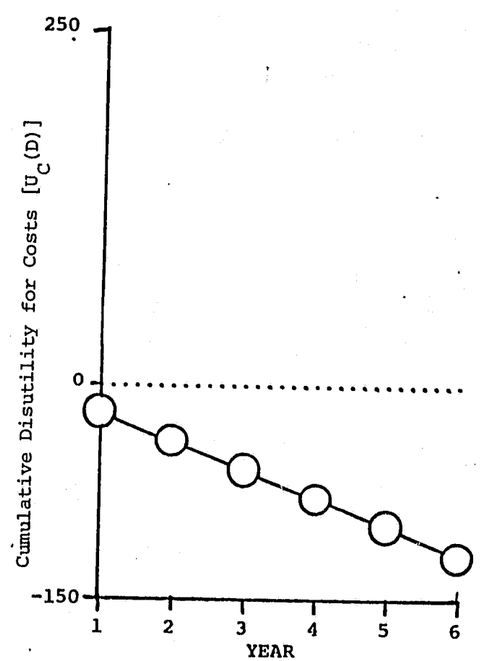


Exhibit 10-2-B. Cumulative Disutilty of Program Cost

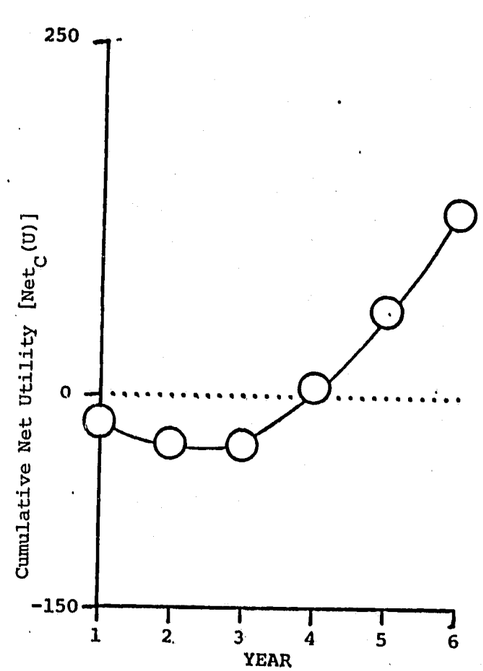


Exhibit 10-2-C. Cumulative Net Utility [$Net_C(U)=U_C(B)+U_C(D)$]

first few years, and does not rise above zero until the fourth year. In fact, Exhibit 10-2-C shows a large jump between Years 3 and 4. This jump, though illustrative here, is perhaps characteristic of many programs, and underscores the prudence either of (1) waiting sufficiently long to collect empirical data on past performance, or (2) considering in the decision projections about future cost-effectiveness.

Exhibit 10-3 highlights the problem of evaluating programs undergoing a start-up phase. Area "A" indicates the period over which cost-effectiveness is below zero, i.e., a "snapshot" would indicate that the program should be terminated. Area "B" is the range over which the program is returning an excess of benefits over cost. Clearly, the worst decision is to fund the program for the first little while only; as long as there is a reasonable prognosis for increasing benefits, the decision should be to stick with the program. The exception to this occurs when equal importance is not ascribed to costs/benefits in each year. Two examples were discussed: the first is the view that programs should produce an early return, and that late benefits should be comparatively discounted; and the second is the view that a program is successful if it ultimately provides an excess return of benefits over costs, and that early costs should be comparatively discounted.

10.4 Cumulative Net Utility and Decision Biases

In Section 9.0 of this report, some biasing factors in programmatic decision making were listed. Below we will discuss some of these potential biases further, along with some of the techniques for helping decision makers avoid them.

Suppose that a decision is made about a program at the indicated point in time in Exhibit 10-4. Further suppose that the decision maker has an intuitive sense that the program has not "paid for itself" and that, as with most programs, there are critics and supporters of the program. Supporters may offfer anecdotal evidence that the program is working, based primarily on personal experience or heresay about what the program is accomplishing. Critics may attack such anecdotal evidence, arguing contrary facts based on their own experience and, perhaps most damagingly, arguing that such data lack scientific

EXHIBIT 10-3

LOSSES FROM PREMATURE DECISIONS TO TERMINATE

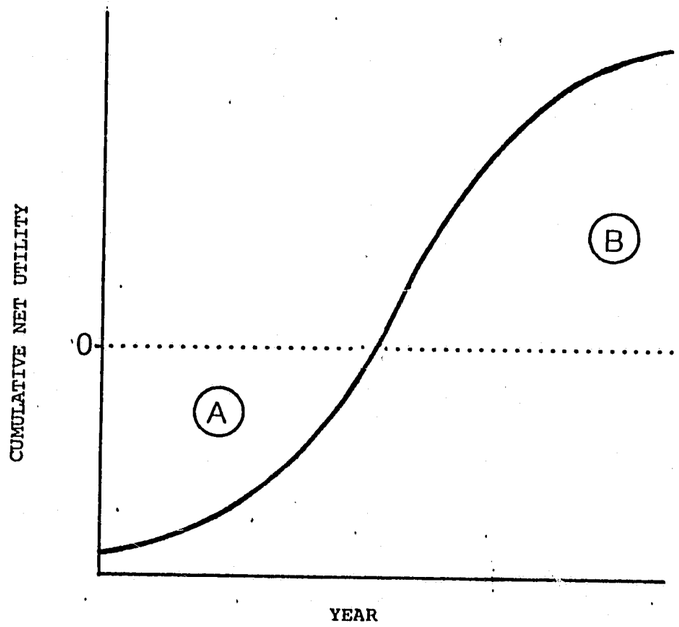
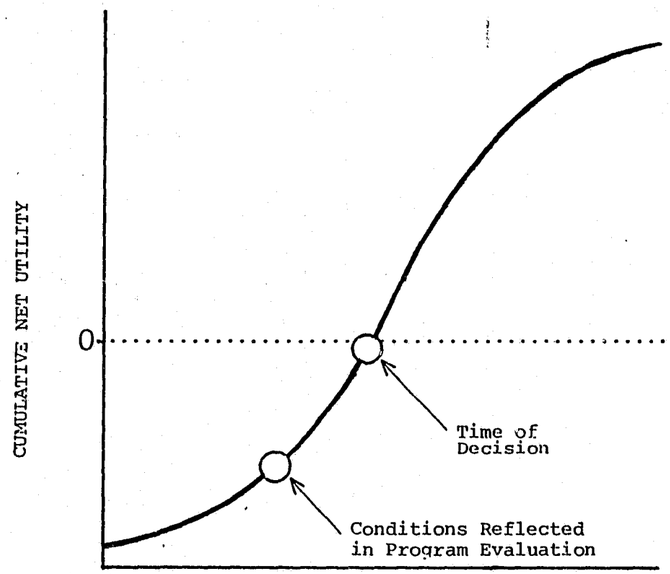


EXHIBIT 10-4

UNDERESTIMATION OF PROGRAM BENEFITS FROM
RIGOROUS (BUT OUTDATED) EVALUATIONS



rigor and objectivity.

Scientific rigor and objectivity are of course virtues quite properly sought. But what effect would a rigorous, empirical evaluation have on the decision? Some evaluations may capture baseline data accurately and may show that change has occurred. More often, the evaluation will provide data for one point in time, and such change data as are available will typically be at least partially equivocal. Moreover, in both cases the data are likely to reflect conditions existing a year or so prior to the decision, even though the evaluation report may be released in a "timely" manner shortly before the decision is to be made.

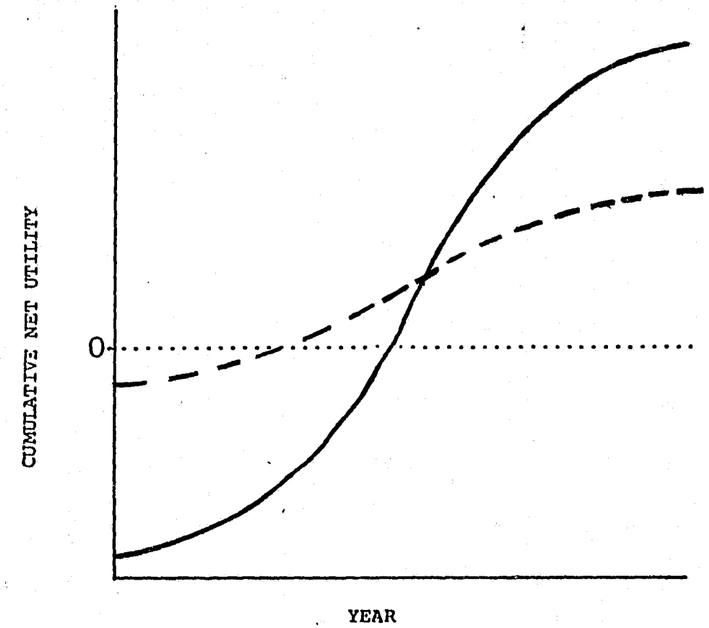
And, regardless of whether a single snapshot or change estimates are provided through the evaluation, both will provide estimates of effectiveness placing the program's worth well below the breakeven point. Even if the decision maker is convinced that change has occurred, assessments of cost-benefit would tend to foster the conclusion that the program was not worth it. The evaluation could therefore tip the decision against the program, on the basis of "hard" data outweighing conviction and anecdote.

In the case of relatively new programs especially, the biasing effects of slow evaluation can be especially troublesome. Delay in presentation of evaluation results, even in the interest of tight design, may introduce more serious biasing effects into the decision than a less-careful evaluation whose results are only a few months old. It could, for example, lead to better decisions if more emphasis were placed on process evaluations--producing timely results--than on impact evaluations whose results are typically more out of date. For new programs, as both the analysis in Section 6.0 and the hypothetical data in Exhibit 10-4 suggest, data one year old are likely to be too old for decision-making purposes. It may in fact be better to do without evaluation data for decision purposes, if those data do not track performance to within a very short period prior to the decision.

Exhibit 10-5 illustrates another biasing factor in program decision making. A comparatively small program, because of its low cost, cannot usually have a large negative net utility (even if it is not very effective). Large

EXHIBIT 10-5

QUICK "RETURN" FROM SMALL OR LESS COSTLY PROGRAMS



programs, in contrast, will typically have large negative net utility. (In the case of the OCAP decision (see Section 6.0), for example, the larger program incurred a period of negative net utility, whereas the smaller programs did not.) Any large programs may have an extended period in which the decision maker may (accurately) perceive that net utility is less than zero. With large programs especially, it is critical to decision making that future cost-effectiveness be considered.

Do evaluations project what the future will be like? Usually not. Evaluations are usually empirical assessments, and the more they are regarded as experiments, the more likely they are to limit themselves to hard, empirical comparisions. Part of the reason for this is probably the fact that there are no sound, generally accepted models for projecting program effectiveness based on empirical data. Regression techniques, for example, are used either to describe relationships in a data set; or, when used in a forecasting model, involve some assumptions that might not be realistic or compelling to most evaluators or decision makers. Projections about future effectiveness (and therefore cost-effectiveness) are highly relevant to the decision process, but are not a part of typical evaluations. Such projections are usually, if not always, judgmental and, indeed, one rarely hears much attention paid to the topic at all. The emphasis is on past performance.

10.5 Net Utility Models As Decision Aids

In this section we will discuss how, in our opinion, cumulative net utility models can be used as decision aids. We will consider first evaluation or experimental design issues. There are many reasons why, in practice, strong designs are not used. However, we know of no instance in which decision-analytic methods would not be enhanced by strong experimental design. Strong, essentially historical fixes on the effectiveness of programs are useful to MAUT analyses. For example, such data would have been useful in establishing the net utility of OCAP programs in Section 6.0. In the following discussion we will assume that the strongest practicable designs are being used, and discuss how decision-analytic aids--and net utility analysis in particular--can be integrated into the evaluation.

10.5.1 Step 1: Develop MAUT model of effectiveness. The MAUT model should reflect all objectives or values about which the decision maker is concerned, and to which the program is relevant. The model should not be limited to data elements that are available, or readily amenable to "hard" measurement.

10.5.2 Step 2: Assess performance longitudinally. To the extent past data are available, the MAUT model should be used to assess changes in performance prior to the time of the decision. When "hard" data are not available, judgments about past effectiveness should be made. Judgments should also be made about the anticipated future effectiveness of the program, including the "ideal" or highest level of effectiveness the program could attain in the long run.

When the assessment is conducted over time, judgmental assessments are replaced by estimates based on the actual data. The judgmental projections are used for determining whether the program is performing as initially anticipated.

Performance assessments can be process evaluations or they can be impact evaluations; both can be accommodated within the MAUT framework. In the interest of timely data, process measures should be obtained.

10.5.3 Step 3: Develop cumulative net utility model. Cost estimates and assessments of the cost-effectiveness tradeoff should be used to develop a cumulative net utility model. Either of the two approaches discussed at the beginning of Section 10 can be used to establish the cost-effectiveness tradeoff.

10.5.4 Step 4: Identify the decision problem and options. Decision makers are not used to being supported by evaluations, and cannot be expected to come to the evalutor for data or to regard evaluations as a useful oracle for practical decision problems. If the evaluation is to be used, the evaluator or someone else must assume the role of the decision analyst. The decision analyst will have to contact the decision maker and identify what the decision problem is specifically. This means delineating substantive program changes, changes in funding level, and so on.

10.5.5 Step 5: Use cumulative net utility analyses to evaluate options. The "evaluation" of options obviously does not refer to building an empirical data base: these options are merely possible causes of action, not actual entities like the program in place. Nevertheless, they must be evaluated.

The method is to project for each option its effectiveness relative to other options and to the program in place (if any), using judgments about future effectiveness. Cost data are then used to assess cumulative net utility for each option.

10.5.6 Step 6: Decide. One implication of the present study is that there is no single decision rule. Reasonable people may be in full agreement about the basic empirical facts, about projected effectiveness, about cost-effectiveness tradeoffs, and still be in disagreement about the "best" option. This kind of ambiguity in the decision process is, however, real; what is erroneous is the view that decisions are driven by hard data, and that making a decision is primarily a matter of agreeing about the facts. The decision process is rich, and it takes a specialized analysis to capture its salient aspects.

Step 6 therefore involves formulation of an explicit decision rule. This decision rule will reflect considerations such as: what is the time horizon over which costs and benefits are assessed? What "discount" is applied to future costs or benefits, compared to present benefits? Which option has the highest cumulative net utility, and how high must cumulative net utility be before that option is implemented? How low must cumulative net utility fall before the program is terminated?

An explicit decision rule must be developed with the decision maker; this decision rule will often reflect policy considerations (e.g., the importance of quick versus deferred benefits).

11.0 CONCLUSIONS

As we mentioned at the beginning of this report, its purpose was to assess the applicability of decision-theoretic approaches to evaluation, with particular emphasis on uses for decision making. We wish to stress once more that evaluations have many different purposes, and we agree with the general point made by the GAO (1976) that different methodologies are better (or worse) for some purposes than others. The generally disappointing degree of use of evaluations for decision making led to the present study, and its exploration of the usefulness of decision-theoretic techniques.

The study suggests that MAUT models can readily be adapted to use in evaluations. MAUT models can be used, we believe, with virtually any type of experimental (or quasi-experimental) design. Though its use is flexible in this sense, weak designs will hamper interpretation of MAUT results (much as interpretation would be hampered with other approaches). In particular, MAUT does not solve the "attribution problem," but it does permit explicit judgment about the strength of an effect when controls are not feasible (see Section 4.0).

The basic MAUT modeling approach was readily adaptable to a variety of decision and cost-benefit analyses. We argued that in many decisions cost-benefit (or cost-effectiveness) is likely to be the critical decision variable. We discussed some examples in which MAUT models were used in that manner.

Insofar as the purpose of this study was testing the applicability of decision-theoretic approaches to evaluation is concerned, the study was a success. Moreover, the analyses were interesting and useful, and a good case could be made for more routine kinds of application.

We formed some less sanguine views, however, about the limitations of evaluations as decision inputs. If a program has stabilized, and one's decisions are essentially incremental or fine-tuning, a number of evaluation approaches may be used and produce quite happy results. But if the program is in "flux" (e.g., a new program whose effectiveness is increasing), or if a fundamental decision is involved (e.g., significant changes in funding level), evaluations in the usual sense have a limited role. We discussed some of the reasons for this in Section 9.0. The limitations stemmed largely from an intrinsic problem with evaluation--namely that they are empirically based, whereas decision making involves projecting the future based on hypothetical actions taken "now."

Fundamental (i.e., non-incremental) decisions and programs in flux, require a different type of "evaluation." This is an evaluation of the alternative programmatic actions; inputs to this type of evaluation are judgments about future cost-effectiveness. This type of decision-theoretic evaluation is called "decision analysis" and should be clearly distinguished from other approaches and definitions of "evaluations."

There are many different program analysis and evaluation approaches that are useful for decision making, at least according to their proponents. The generic problem with these, from our viewpoint, is that they are "generalist" approaches that usually involve peripheral aspects of the decision while not getting to the core of the decision itself. (Indeed, this emerged as one of the limitations of MAUT applied as a monitoring device only (as described in Section 4.0).

Whereas we had initially hoped that MAUT might be a "quick fix" for evaluations that would immediately enhance usefulness for decision making by an order of magnitude, we have concluded that quite specialized techniques are required to augment typical evaluation approaches. Two features of an appropriate specialized approach seem clear: judgmental projections about future costs and/or benefits are required, and all the decision maker's alternatives and tradeoffs among them must be explicitly evaluated. Appropriate technologies for accomplishing the above simply are not well-known in

evaluation research. Indeed, some persons who take the narrow view that evaluation is experimentation (e.g., Apsler, 1977) might argue that the type of analysis we propose has no place in the field of evaluation research!

We believe, however, that decision-theoretic techniques can be used for monitoring programs in a more-or-less conventional evaluation, and that decision analysis is a quite proper aid to help decision makers when a particular decision problem is identified. The net utility model discussed in previous sections emerged as one method for integrating evaluation results and decision-analytic techniques. The major purpose of this study, however, was not to attempt advances in the theoretical state-of-the-art, but rather to demonstrate practicability and usefulness. Our conclusion in this regard is that MAUT can be readily integrated into evaluations, and that decision-analytic techniques can be brought to bear in a timely manner and at a fraction of the total cost and effort of a typical evaluation.

12.0 REFERENCES

American Institutes for Research. National evaluation of the Community Anti-Crime Program (Draft Evaluation Design). Washington, D. C.: American Institutes for Research, January 1979.

Apsler, R. In defense of the experimental paradigm as a tool for evaluation research. Evaluation, 1977, 4, 14-18.

Bakan, D. On method: Toward a reconstruction of psychological investigation. San Francisco: Josey-Bass, 1967.

Bernstein, I. M., and Freeman, H. F. Academic and entrepreneurical research. New York: Russell Sage Foundation, 1975.

Brown, R. V., and Ulvila, J. W. Selecting analytic approaches for decision situations (Revised Edition). Volume I: An overview of the methodology. Volume II: Case studies. Volume III: Appendices (Technical Report TR 77-7-25). McLean, VA: Decisions and Designs, Inc., December 1977.

Brown, R. V., Kahr, A. S., and Peterson, C. R. Decision analysis for the manager. New York: Holt, Rinehart and Winston, 1974.

Brown, R. V., Seaver, D. A., and Bromage, R. C. An analysis of the community anti-crime funding decision (Technical Report 80-2). Falls Church, VA: Decision Science Consortium, Inc., May 1980.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Decision Science Consortium, Inc. National evaluation of the Community Anti-Crime Program (Draft decision theoretic evaluation design). Falls Church, VA: Decision Science Consortium, Inc., March 1979.

Doherty, N., and Crake, G. Social program costs: Adjusting for research and evaluation activities. Evaluation Review, 1980, 4, 537-548.

Edwards, W., Guttentag, M., and Snapper, K. J. A decision theoretic approach to evaluation research. In E. Struening and M. Guttentag (Eds.), Handbook of evaluation research (Vol. 1). Beverly Hills: Sage Publishing, 1975.

Edwards, W., Lindman, H., and Savage, L. J. Bayesian statistical inference for psychological research. Psychological Review, 1963, 70, 193-242.

Etzioni, A. Mixed scanning: A third approach for decision making. Public Administration Review, 1967, 27, 385-392.

Fischer, G. W., Edwards, W., and Kelly, C. W. Decision theoretic aids for inference, evaluation, and decision making: A review of research and experience (Technical Report TR 78-1-30). McLean, VA: Decisions and Designs, Inc., February 1978.

Francis, W. What social indicators don't indicate. Evaluation, 1973, 2, 79-83.

Guttentag, M. Subjectivity and its use in evaluation research. Evaluation, 1973, 1, 60-65.

Keeney, R. L., and Raiffa, H. Decisions with multiple objectives: Preferences and value tradeoffs. New York: Wiley, 1976.

Lewis, R. G., and Greene, J. R. Implementation evaluation: A future direction in project evaluation. Journal of Criminal Justice, 1978, 6, 167-176.

Phillips, L. D. Bayesian statistics for social scientists. New York: Crowell, 1974.

Raiffa, H. Decision analysis: Introductory lectures on choices under uncertainty. Reading, MA: Addison-Wesley, 1968.

Raiffa, H. Analysis for decision making: An audiographic, self-instructional course in decision analysis. New York: Encyclopedia Britannica, 1974.

Riecken, H. Principal components of the evaluation process. In Evaluating Action Programs: Readings in Social Action and Education. Boston: Allyn and Bacon, 1977.

Rivlin, A. Systematic thinking for social action. Washington, D. C.: The Brookings Institution, 1971.

Rossi, P., Wright, J., and Wright, S. R. The theory and practice of applied social research. Evaluation Quarterly, 1978, 2, 173.

Savage, L. J. The foundation of statistics. New York: Wiley, 1954.

Seaver, D. A. Assessment of group preferences and group uncertainty for decision making. (SSRI Research Report 76-4). Los Angeles: University of Southern California, Social Science Research Institute, June 1976.

Seaver, D. A. Assessing probability with multiple individuals: Group interaction versus mathematical aggregation (SSRI Research Report 78-3). Los Angeles: University of Southern California, Social Science Research Institute, December 1978.

Seaver, D. A., Brown, R. V., Campbell, V. N., Ulvila, J. W., and Chinnis, J. O. Applications and evaluation of decision analysis in water resources planning (Technical Report 79-2). Falls Church, VA: Decision Science Consortium, Inc., December 1979.

Snapper, K. J. Subjective planning and objective evaluation (Technical Report 74-2). Washington, D. C.: George Washington University, Social Research Group, January 1974.

Snapper, K. J., O'Connor, M., and Einhorn, H. Social indicators: A new method for indexing quality. In K. J. Snapper (Ed.), Studies in the analysis of complex decisions. Washington, D. C.: Information Resources Press, in press.

Snapper, K. J., and Seaver, D. A. Application of decision analysis to program planning and evaluation (Technical Report 78-1). Washington, D. C.: Decision Science Consortium, Inc., November 1978.

Snapper, K. J., and Seaver, D. A. The use of evaluation models for decision making: Application to the Community Anti-Crime Program (Technical Report 80-1). Falls Church, VA: Decision Science Consortium, Inc., January 1980(a).

Snapper, K. J., and Seaver, D. A. Cost-effectiveness and construct evaluation: A method for programmatic post mortems (Technical Report 80-3). Falls Church, VA: Decision Science Consortium, Inc., 1980(b).

U.S. General Accounting Office. Evaluation and analysis to support decision making (PAD-76-9). Washington, D. C.: U.S. General Accounting Office, September 1976.

U.S. General Accounting Office. Status and issues: Federal program evaluation (PAD-78-83). Washington, D. C.: U.S. General Accounting Office, October 1978.

U.S. General Accounting Office. Evaluating programs with multiple and conflicting objectives: A framework and some methods (PAD-80). Washington, D. C.: U.S. General Accounting Office, February 1980.