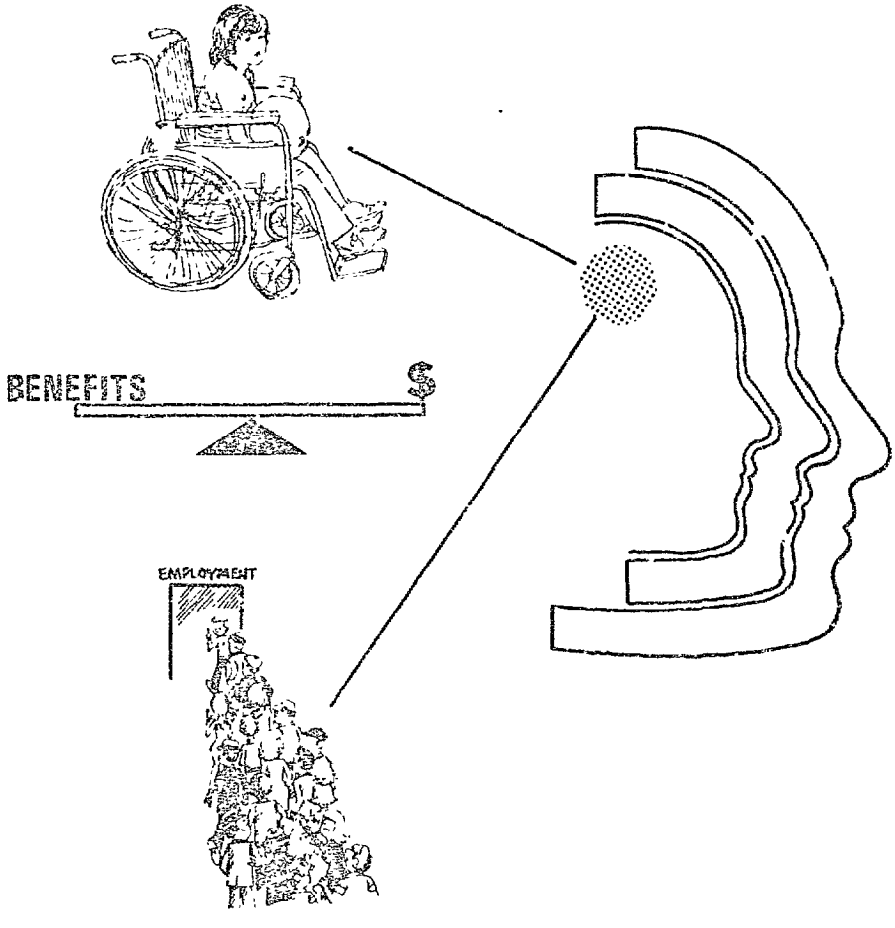
096894 76-0678

EXPOSURE DRAFT

EVALUATION AND ANALYSIS TO SUPPORT DECISIONMAKING



UNITED STATES
GENERAL ACCOUNTING OFFICE

DEC. 9, 1975

OPA-76-9

702507

COMPTROLLER GENERAL OF THE UNITED STATES
WASHINGTON, D.C. 20548

For a number of years, the General Accounting Office has been expanding the scope of its work and the capability of its staff to review the results and effectiveness of Government programs.

The Comptroller General is given very broad responsibility and authority for the evaluation and analysis of Federal programs and activities under the Budget and Accounting Act, 1921, and the Accounting and Auditing Act of 1950. Section 204 of the Legislative Reorganization Act of 1970 supplemented our authority and indicated current congressional interest in analyses of programs. The amendment to section 204 by Title VII of the Congressional Budget and Impoundment Control Act of 1974 further strengthened the congressional emphasis on the evaluation of programs, the statement of legislative objectives and goals, and improved methods of evaluation. Such work is carried out by all parts of GAO in connection with our general review, evaluation, analysis, and audit functions.

The 1974 amendments to the 1970 Act require, among other things, that "the Comptroller General shall develop and recommend to the Congress methods for review and evaluation of government programs carried on under existing law." This draft document is a first step in collecting and disseminating general concepts on these activities and how they are related to other activities in the continuum of decisionmaking about Government programs. The document adds to guidance contained in Standards for Audit of Governmental Organizations, Programs, Activities and Functions, issued in 1972, which includes in the full scope of such audits a review to determine whether desired results are effectively achieved.

This document is intended to be of value to the novice and the experienced practitioner whether engaged in financial audits, program review or in program evaluation or analysis. There are many good text books available for use by the sophisticated evaluator or analyst and therefore it was decided to address this document to those professionals included in various types of reviews in order to place in perspective the need for and value of program evaluation and analysis to support decisionmaking.

We are requesting comments from major Federal agencies and certain committees and Members of Congress on this draft document. Because of the wide general interest in this subject matter we have elected to publish the draft as an "exposure draft" to make it available to other interested parties as well. We would welcome comments and suggestions which any reader may care to offer on this document before it is finalized. If you care to offer comments, please send them to the Director, Office of Program Analysis.

Comptroller General
of the United States

BLANK

Contents

CHAPTER 1

INTRODUCTION

[illegible] years, a multitude of new programs have been [illegible] with the problems of our society, and Government [illegible] expenditures have increased at a very rapid rate. [illegible] working? Are the funds being spent on [illegible] achieving the results desired? Are there better ways [illegible] and solve society's problems? And, [illegible] old programs commensurate with the funds [illegible]

[illegible] and agencies entrusted with public resources and the authority for applying them have a responsibility to answer these questions--to render a full accounting of their activities. Government managers have a responsibility to show not only the purpose for which public resources were used, but also to demonstrate the effect of their use.

The responsibility for rendering such an accounting rests first on the executive agency administering the program. Administration requires that program managers know what [illegible] to them are accomplishing and whether results might be improved. Central management agencies--the Office of Management and Budget or the White House, for example--need to know programs are working either through their own efforts or through the review of the appraisals of managing agencies. Finally, the legislative branch is responsible for monitoring and overseeing Federal programs. [illegible] legislation, appropriations, [illegible] oversight and investigations indicate [illegible] legislative branch to make its own appraisals of programs and to make use of appraisals made by the executive branch.

The ultimate choices about programs--decisions about whether to do or not to do something--will be policy choices. However, political leaders, public administrators, and the public need as much information as possible on the choices that must be made. This need has stimulated the development of an art [illegible] program evaluation and policy analysis. This art [illegible] efficiently developed to permit presentation [illegible] covering "how-to-do-it" in every situation. It is [illegible] solving. While recognizing this, we also recognize that decisions must and will be made by legislators and executives faced with the task of formulating and reformulating programs to deal with the problems of our society.

Thus, we offer this document as a first step in collecting and disseminating lessons learned in GAO and elsewhere about analysis and evaluation. While a somewhat more technical definition is set forth in the body of the document generally speaking, we offer this guidance for the use of anyone who is "evaluating" programs and "analyzing" policy choices in the sense of engaging in a careful appraisal of what happened, why it happened, what choices are available for future actions, and what the implications are of those choices. We intend it for use by all persons who are concerned with this process whatever the academic discipline or professional background from which they approach the problem. While this statement is addressed primarily to practitioners, we hope it will also be a useful reference document to those who, as legislators or managers, for example, are interested in the products of evaluation and analysis.

The concepts and guidance which we offer must be adapted to specific program situations. We recognize, for example, that program objectives are seldom as clearly stated or agreed upon as would be desirable for evaluative purposes; that no program operates in isolation from other social or economic events; and that data and measurement techniques are almost always less adequate than desired. It is in the adaptation of the ideal and the theory to the specific situation that the persons doing the work show their worth. The judgments involved in the identification of objectives, the selection of data and measurement techniques, and the evaluation of external factors transcend in importance any "rules" which can be prescribed in a document of this sort.

The chapters which follow discuss the framework within which these activities are performed and provide both conceptual and practical guidance.

CHAPTER 2

THE PUBLIC DECISIONMAKING PROCESS

There are many reasons for governmental activities. For example, national defense is provided because there is no suitable private alternative. The provision of free public education reflects society's preferences and its notions about its own long term welfare. Health and safety considerations lead to the regulation of private activities to avoid adversely affecting the public's well-being. Also, there are tax incentives such as the investment tax credit which affect the performance and growth of the U.S. economy. The validity and relevance of such governmental activities may be questioned. Analysis, evaluation, and related activities should help decisionmakers in resolving these questions.

The private market place has limitations, particularly because it does not produce certain goods with high social value. It does not always provide adequate information, sufficient competition, efficient designs or qualities of certain goods, desirable distribution of income and wealth, or desirable modifications of consumption patterns. But, resorting to governmental action because of these failings does not automatically insure that the same or other failings will not occur. The effectiveness and efficiency with which the Government performs its functions must also be weighed in deciding whether governmental action should be enhanced, changed, or in some cases, is warranted at all.

Each year as legislative, budgetary, and appropriations decisions are being considered, the practical issue remains: What does the public need and how are priorities established? In a democracy, the political process is relied upon to examine and determine public needs each year and to set priorities as to how such needs are to be met from public funds. Elected officials are responsible for learning and reflecting their constituents' needs and proposing programs or program changes with requisite funding levels to assist in determining priorities for action.

But for the work of elected officials to have meaning, accurate and relevant information must be available and useful debate must take place. The approaches and techniques of analysis and evaluation can be used, not as a replacement for, but in support of the judgment of those persons and groups involved in public decisionmaking.

THE RESOURCE ALLOCATION PROBLEM

Government actions generate benefits and incur costs. These benefits and costs should be broadly defined to include both social and private aspects. The key elements of the problem of choice are:

--Government objectives are achieved by developing, adopting, and implementing policies and by creating and operating programs, all of which consume or transfer resources--tangible and intangible.

--There are many public needs. These needs are large and constantly changing. Demands for resources are much greater than the resources available.

--Decisionmakers must choose among competing objectives and among the alternative programs and policies capable of meeting the chosen objectives at desired and affordable levels of achievement.

Thus, decisionmakers are involved in the process of allocating available resources among competing demands so as to achieve the greatest overall level of net benefits possible. At the same time, full consideration is given to the requirements of justice, equity, and political reality.

ISSUES IN RESOURCE ALLOCATION

Most programs are interdependent and affect more than one of society's goals. This leads to a need for use of evaluation and analysis in two resource allocation issues: (1) choices within a major program area, and (2) choices among major program areas.

For choices within a program area, the following questions are often posed: Is there an appropriate level of a given objective and are there preferred alternatives for reaching that level? Are there obstacles to acceptance and implementation of an otherwise preferred alternative and what would be the costs of overcoming the obstacles? Are there equity considerations connected with the leading alternatives?

For choices among major program areas, similar questions are relevant. Judgments that are made by decisionmakers concerning relative importance of the various objectives will affect the assignment of resources among those objectives.

Resolution of resource allocation issues should result in any of a number of actions: (1) continue, modify, or abandon existing policies, (2) adopt new policies, or (1) continue, modify, expand, reduce, or phase out current programs, and (2) create new programs.

The decisionmaking process and the relationships among its component functions do not necessarily follow a predetermined sequence. In a real sense, the decision process and the functions of evaluation and analysis in particular have no beginning and no end. They are concurrent and continuous processes.

CHAPTER 3

THE EVALUATION AND ANALYSIS CONTINUUM

The perspective of evaluation and analysis covers a wide range of activities and purposes and can provide an important and continuing contribution to the ongoing decisionmaking process. Important contributions are made by reviews known as program audits, budget examinations, management analyses, planning, institutional research, program budgeting, systems analysis, and engineering economics, as well as by program evaluation, policy analysis, cost-benefit analysis, etc.

The specific type of review needed depends, in part, on the particular issues raised or questions asked and the focus of the inquiry on the part of the decisionmaker. If financial information is needed about a Government agency or corporation, an audit of the results of operation and financial condition for a given period may be the appropriate form of review or analysis needed. In another situation, there may be concern about the management of a program, in which case a management review or analysis may be appropriate that is directed more towards the effectiveness of the organization, management, and staffing.

General program reviews that look at program effectiveness and consequences, as well as management effectiveness, are useful to the decisionmaker in determining whether the program is meeting the established objectives and whether there are changes needed to improve the program efficiency and effectiveness. A comprehensive review of an existing program may consider the overall performance of the program, including an evaluation and analysis of performance on any number of criteria. Consideration of alternatives may also be included.

It should be recognized that traditional perceptions of the terms evaluation and analysis tend to overlap. In some cases, the two terms are used interchangeably; in other cases, evaluation is viewed as a source of information needed to perform analysis. The distinctions between evaluation and analysis have tended to focus more upon the phase of development or implementation of a program or activity than upon the processes or techniques used.

According to these perceptions, evaluation attempts to appraise and measure the actual inputs, processes, outcomes, and operational settings of one or more ongoing programs or policies in order to compare these findings with those which were anticipated or assumed; it then seeks to explain the

differences and to suggest alternatives for improvement. Analysis searches for alternative policies and programs for achieving public objectives and attempts to assess and compare their anticipated costs and benefits over time, as well as their other consequences, in order to provide the basis for better future choices.

CONCURRENT PROCESSES IN THE CONTINUUM

A more general view is obtained by starting with the two basic questions which decisionmakers, and their staffs, face: (1) What actually has happened as a result of past or current policies and programs and what have we learned? and (2) What should be done in the future and what are our options? Answering these questions can, in turn, be roughly translated into broad classes of activities: appraising the results of policies and programs and assessing alternative policies and programs. These broad activities include respectively, activities called policy and program evaluation, and policy and program analysis. The broad activities also include the variety of review types listed previously when they performed studies having similar purposes.

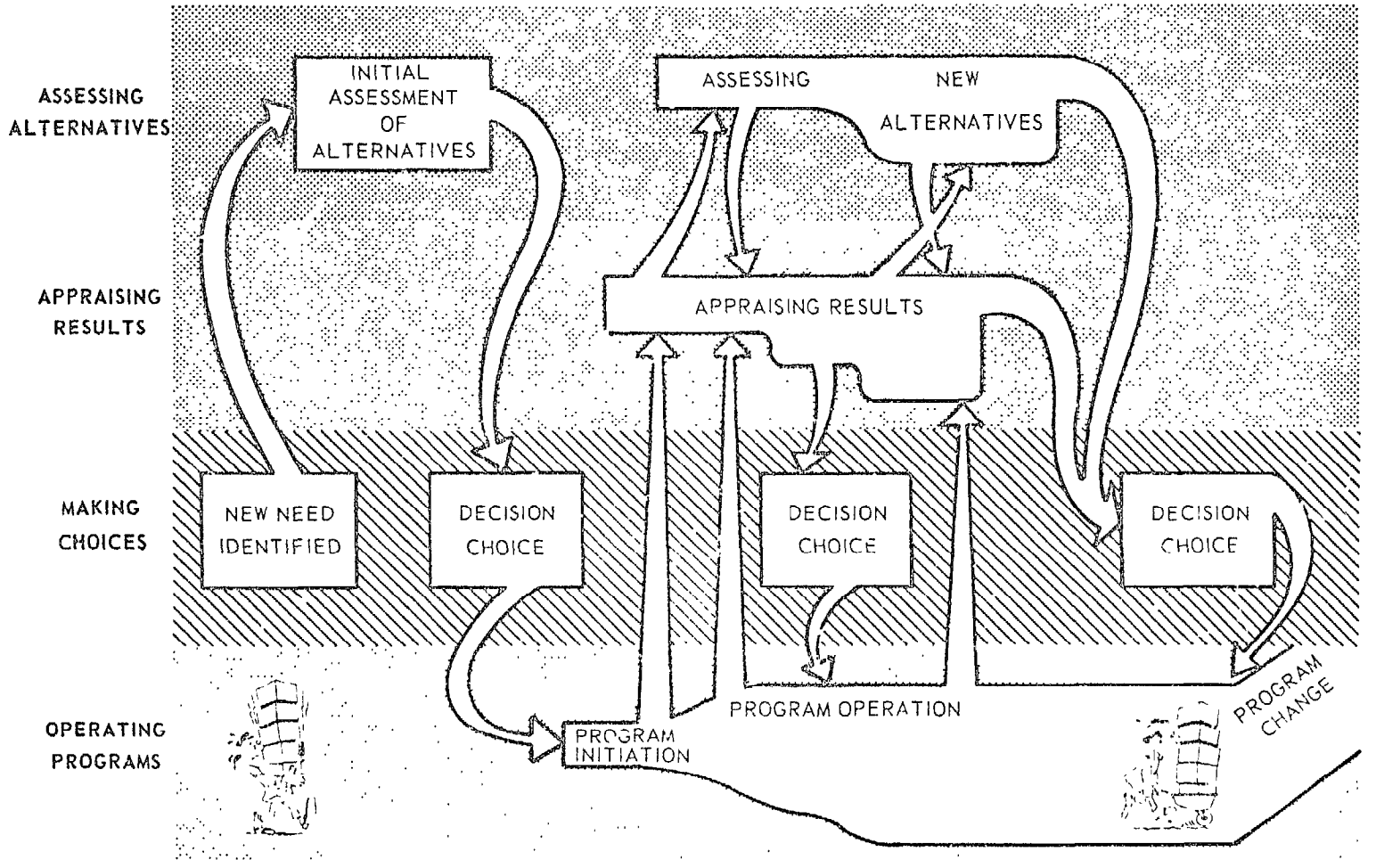
An example of the interaction of these concurrent processes in the continuum is shown in the following illustration.

After the issue or new need is identified, assessment of alternative solutions is undertaken. A systematic attempt should be made to assess the conditions resulting in the issue and to analyze ways to improve the situation. Alternative approaches, developed as part of the assessment, do not normally contain the level of detail that would be needed for actual implementation. For example, after the decision choice is made on a program to be implemented, program initiation would also require a detailed plan that specified some or all of the following: physical resources to be acquired, the acquisition schedule, appropriate processes and technologies to be employed, investment and operating funds needed, and capital facilities to be in place to provide the scale of operations desired.

An important consideration is to build into the implementation plan, specific provisions for gathering information necessary for a comprehensive and valid appraisal of results. During the implementation phase, such information should be gathered, including, where possible, the effects of any changes in the implementation plan.

As required to support decisionmaking, further assessment of alternatives should take place concurrently with appraisal of results during the operation of the program. Appraisal of

EXAMPLE OF THE CONCURRENT PROCESSES IN THE CONTINUUM (FOR A NEW NEED AND NEW PROGRAM)



results without such assessment provides only limited insights into the desirability of new alternatives; assessment of alternatives without sound appraisal of results lacks credibility. A high degree of interaction exists not only within the various phases of an assessment or appraisal, but between these two processes. If these interactions and feedbacks are ignored, the quality of both efforts suffer.

At the conclusion of any major effort in the concurrent processes, the original or modified issue or need should be reviewed again and the question asked: Could a comparable level of results, effectiveness, or benefits have been achieved at lesser cost; or for the same costs could greater effectiveness, or benefits, have been realized? Could these more desirable outcomes have been achieved by reshaping or redirecting the current program or policy, adopting some previously suggested, but rejected, alternative; substituting some wholly new alternative; or creating some combination of the foregoing possibilities?

Real world decisionmaking and many of the activities which support it are complex, uncertain, hurried, and subject to all kinds of constraints; some understandable and some seemingly arbitrary. Evaluators, analysts, and other reviewers can do little about these difficulties, but if they understand the complexity of the situation, they can perform in a way which is of maximum usefulness under the circumstances. These difficulties are compounded by the existence of competing or complimentary objectives (or their related programs) which may also have to be considered.

ROLES PEOPLE PLAY

The distinction between the concurrent activities on the one hand and the people or organizations that perform them on the other hand is an additional source of confusion. As noted earlier, appraising policy and program results and assessing alternatives are related and mutually reinforcing processes. These processes, however, can be, and frequently are, performed within a single organization and often by the same person. This has the significant advantages of efficiency and of keeping the practitioners of the various skills aware of useful interactions.

In the real world situation a variety of people are involved who have different backgrounds and call themselves by various titles. Many of these people move between appraising the results of policies and programs and assessing the alternatives for improved choices in the future. At times, these people may be emphasizing the assessment of alternatives, at other times the appraisal of results, and sometimes they may

be doing both together. People who gain their first experience in one activity move throughout the continuum and interact with other people having other experience when working on a particular study.

This document is aimed primarily at the variety of staff personnel who perform the activities just discussed. The most technically oriented evaluators and analysts should also find it of value, but the emphasis here--as it is in actual practice--is not on advanced quantitative techniques but on essential concepts and basic approaches. In this respect, no group or profession has a monopoly on the talents required of a good evaluator or analyst. The basic prerequisites are (1) an inquiring, skeptical, challenging mind, (2) the ability to think systematically and rigorously, and (3) an openness to new ideas. This mind-set obviously needs to be coupled with an appreciation of the uses, powers, and limitations of such fields as economics, statistics, accounting, operations research, etc. When high levels of skill in these and other areas are judged appropriate, the practicing evaluator, analyst, or other reviewer should not hesitate to call on the needed experts.

The following section focuses on those ideas, concepts, and approaches which are basic to appraisals of policy and program results and to more insightful assessments of alternatives for improvement.

CHAPTER 4

APPRAISING THE RESULTS OF POLICIES AND PROGRAMS AND ASSESSING ALTERNATIVE APPROACHES

The previous chapters discussed problems and issues in public decisionmaking with particular emphasis on resource allocation, and on the continuum of evaluation, analysis, and other review functions which support decisionmaking. This chapter discusses the methods and concepts associated with evaluation and analysis.

The discussion that follows focuses on

--understanding fundamentals in appraising results and assessing alternatives;

--appraising policy and program results; and

--assessing policy and program alternatives.

Obviously, the degree to which the methods can be applied in a particular case depends on the specific problems to be considered.

UNDERSTANDING FUNDAMENTALS

The activities of appraising results and assessment of alternatives share a number of common methods and techniques and also share certain fundamental concepts in which the mode of inquiry is essentially the same. These fundamental concepts include:

--Ascertaining users' needs.

--Defining the nature and scope of the problem.

--Determining valid objectives.

--Specifying comprehensive measures.

Ascertaining users' needs

An initial task in either appraising program results or assessing alternatives is to develop a clear understanding of the decisionmaking needs. These needs can generally be summarized by answering a series of questions.

--What is the problem that is at issue?

--Is there dissatisfaction with effectiveness or consequences of the policy or program?

--What uses are to be made of the information to be collected?

--When is the report needed?

It is helpful to make some distinctions among the various participants in the decisionmaking process. Some may already have an understanding of the nature of the problem they face, what they want to know, and why. Others may only have a general perception of the problem and what needs to be done about it. In the latter case, a more extensive discussion of these fundamentals may be needed to develop the basis for a study that will be useful to these decisionmakers.

It is important to recognize different viewpoints and interests among participants in the decisionmaking process. The official sponsor may be a congressional committee, whereas the real user of the study may be one member of the committee or the committee staff. Other participants in the process include the manager of the program being evaluated or analyzed and those with only a minor or peripheral interest. Another congressional committee, the Office of Management and Budget, or a private organization, such as an association or government contractor, may become increasingly interested as the study progresses.

In developing a clearer idea of any of these participants' needs, attempts should be made to elicit and clarify information on the nature of the problem or issue as it is currently understood, the general context of the problem, and parts that appear to require special emphasis. Specific attention should be given to the order of priority in approaching the various parts of the problem and to particular points of information or insight essential to making the decision or meeting the decisionmaker's needs. The bureaucratic or political context in which the decision will be made needs to be understood. The time available for the study effort, and the critical points at which specific items of information are needed should be ascertained.

Defining the nature and scope of the problem

It is essential that the participants in the decision process, together with those persons responsible for the study, share a common understanding of the nature and scope of the issues at stake.

A full and correct understanding of the nature of the problem will be aided by (1) considering its origin, if known, (2) reviewing legislative hearings, reports, and acts associated with it, (3) inquiring into the history of programs designed to deal with the problem, and (4) examining past analyses, evaluations, audits, and budget examinations of the same or related issues.

The desirable scope of the effort is dependent not only on identifying the questions which it would be useful to answer, but on the availability of methods and data which will provide insights into the solution. There must be a balance between desired scope and such things as precision in the methods. The magnitude of policy implications--such as widely held opinions that there is a need for changing the direction of a program--and the degree to which the conclusions of the study could affect decisions should be ascertained. A shared understanding of the scope of the study and the objectives and measures of the policy or program are the foundation for defining the initial direction of the study effort.

Similarly, understanding is needed of the coverage required in geographical terms (regional, State, local); on areas, populations, individuals or units to be included; and on the scope (how many individuals, approximately how much information from each, etc.). The scope of coverage together with the timing provides for the logistics of the work. For example, indepth work may be undertaken at a small number of locations, less detailed work at a larger number of locations, or some combination of these.

Determining valid objectives

The objectives--the benefits desired to be achieved--inherent in the policy or program at issue frequently are not stated clearly and precisely. The original sponsors of the policy or program may not have had a precise idea of the end results desired. Formal statements of objectives may be intentionally ambiguous if it is easier to obtain a consensus on action. Value judgments underlying the objectives may not be shared by important groups. Consequently, the end results intended may be perceived by some as implying ill effects for them. Furthermore, explicit statements of objectives require a specific assignment of priorities and commitment of resources.

Statements of objectives should:

1. Capture a complete understanding of the intended benefits.

2. Recognize any unintended adverse consequences.

3. Include important qualitative aspects, even though measuring degrees of attainment may be exceedingly difficult.

4. Take account of multiple objectives which may be in conflict.

The importance of taking such a comprehensive view of objectives cannot be overstated. Oversimplified statements (1) will not capture all essential aspects of the effects intended, and (2) may contain implied conflicting consequences for groups other than the intended beneficiaries (e.g., "to eliminate hunger" or "to achieve energy self-sufficiency"). Implied objectives may represent desirable end results. For example, a summer employment program aimed primarily at increasing earnings of young people may be viewed as reducing the prospect of civil disorders. Even desired end results may not all be achievable simultaneously and may be interdependent.

Oversimplified statements may result if activity milestones are contained in them (e.g., "to increase the number of emergency rooms by 20 percent by 1978"). An objective stated in this way overly constrains an assessment of alternatives, the purpose of which is to determine efficient levels of attainment of an ultimate benefit. A better statement might be, "To reduce deaths, additional complications, disability, and suffering of persons with acute injuries or disabling conditions by improving the availability and quality of emergency care."

In appraising results of ongoing programs, if standards or activity milestones have been furnished to managers, they should not be accepted uncritically (e.g., a specified student/teacher ratio). An attempt should be made to find whether deficiencies in attainment of the milestones are caused by unrealistic expectations or by the way the program was implemented or operated.

A shift in objectives can occur over time and care must be taken to assure that statements of objectives currently in use are still accurate. For some, the objective associated with the national 55 miles per hour speed limit has changed from energy conservation to safety.

Determining valid objectives is a complex and frustrating task. A study may have to proceed without fully satisfying these requirements. If this is the case, objectives should be reexamined and clarified as the study progresses.

Specifying comprehensive measures

Valid measures of policy and program consequences are required for both appraising results and assessing alternatives. Objectives and measures of consequences are interdependent. The quality of each depends on the other. Measures should be used which cover all aspects of a given objective. Ideally, measures should

--quantify the extent to which the objective(s) are or would be met--"effectiveness" measures;

--capture qualitative aspects of the consequences--"intangible" measures;

--quantify, to the extent possible, side effects--"externality" measures;

--quantify, to the extent possible, the differences of impact on the beneficiaries and the cost-bearers--"distribution" measures.

When appraising program results, it may be decided for practical reasons to exclude externality and distribution measures, since it is difficult to sort out these effects and directly attribute results to a specific policy or program. For intangible measures, some qualitative indication of relative magnitude should be used (e.g., ratings by clients reflecting their satisfaction with the quality of a service).

Data may not be available on the desired measures, or if available, it is obtainable only at high cost. In these cases, surrogates will have to be used. For example, the scholastic aptitude test is used to measure likely achievement in college. When surrogates are used, their validity should be established.

There is a temptation to define quantifiable measures, especially of effectiveness, too rigidly or narrowly. For example, in evaluating a public employment program, a successful participant might be specified as a person who is employed 1 year after completion of training. If the participant worked one day less than a year, would he be viewed as unsuccessful? Suppose the participant only occasionally held a job, but happened to be working a year after the program. Should this be counted as a success? The range and distribution of outcomes would be appropriate in this case. For example, data on the percentage of persons holding jobs for various lengths of time after training would provide a more meaningful picture of real outcomes.

There is a high degree of interaction among these fundamental concepts. A clear understanding of what is needed for the decisionmaking process, of the nature of the problem, and statements of objectives is necessary in order to assure that a meaningful and feasible set of measures has been specified.

APPRAISING POLICY AND PROGRAM RESULTS

The process of appraising results should begin concurrently with policy or program implementation and continue as needed during the operation.

After the fundamental concepts discussed above are understood they must be further developed through application of other more specific concepts and methods including:

--Making valid comparisons.

--Developing needed information.

--Interpreting program results.

--Checking the completeness of the appraisal.

Making valid comparisons

Comparison is the essence of appraising program results. The measures need to be compared with some goal or base. Various bases for comparison must be developed with the decisionmakers who know whether their focus is on resource input (e.g., mixture of paraprofessionals, nurses, and physicians); operational process (e.g., scheduling of surgeries); outcomes (e.g., disability days averted); the operational setting (e.g., interaction with other outpatient and inpatient facilities); or some mix of these.

The measures that are selected in a specific study need to be compatible with and directly relatable to acceptable bases for comparison, for example, to legislation. This may not be possible if the bases and objectives are not clear, changes are taking place, or for other reasons the bases are arbitrary.

Other sorts of comparisons may be useful. For example, comparison of planned variations of an existing program may help to identify important characteristics. A comparison of similar programs will assist in identifying potential improvement.

Program comparisons are attractive for several reasons. They (1) provide information on effectiveness of alternatives in comparable terms and for the same time period, (2) reduce the need to rely completely on the elusive "control" of experimental methods applied to one project, (3) provide a richer source of knowledge for judgmental interpretation, (4) help generalize the results if widely distributed "representative" projects can be included, and (5) offer an opportunity to identify exceptional performance and to study what is operationally different about those projects.

On the other hand, program comparisons comprehensive enough to yield the above advantages are costly and difficult to manage. For example, although "planned variations" must be carefully documented at the outset, once in operation they will seldom be free of further changes, which also must be documented. It should be noted whether such changes are "positive" (efforts to apply even better methods) or "negative" (resistance to adopting the prescribed methods).

Once the nature of comparisons is established, a series of additional questions directly relating to the problem at hand should be raised. Some of these are posed as hypotheses which the appraisal aims to prove or disprove. When either formulating hypotheses (for appraising results of experimentation, planned program variations, or pilot studies) or questions, care must be taken to assure that they allow an appraisal of whether consequences or impacts are attributable to the program or to some other causes. A decision needs to be made whether only descriptive findings will suffice or whether it will be necessary to demonstrate significance of results or differences in effects.

Choice of a comparison approach depends both upon the questions to be asked and the availability of data. This is not only procedural but involves also questions of access, comparability, restrictions on collection and use of confidential data, etc. These problems may be more severe than many evaluators, auditors, examiners, and others realize. This means that analytical methods which maximize the analytical value of each bit of data are needed.

It is not always possible to use the best theoretical method because of data problems. Some methods may be impractical if data are too highly aggregated, incomplete or missing, or it may require "patch up" efforts after the evaluation is underway.

Some major comparison methods are:

Experimental methods--attempts to measure the results of the program as though everything else is held constant. This is done by measuring the difference, in terms of the measures of success, between those affected by the program and a control group who are not. This is the preferred method for evaluation of social experiments, but it can also be used for any evaluation when the essential requirements of random assignment and control are feasible. This is the approach that was used in the New Jersey Negative Income Tax experiment. In that experiment, several different amounts of monetary incentives were given to different groups of families in the same situations to see what effect the incentives had on work and spending habits. Responses were compared with the habits of families in the same situations which received no monetary incentive from the experiment during the same time.

Experimental designs require that the affected group and the group not affected possess similar characteristics. This is the reason for a strict requirement that the potential participants be randomly assigned so that each one has the same chance of assignment to either group before the program begins. Unless randomization is achieved, there is no assurance that the results are attributable to the program. For example, unless randomly assigned, eligible participants in a social program might enroll because they are more perceptive and desire the benefits more than others who are eligible. This biases any comparison of the response or performance of the two groups because their motivation and other characteristics were not the same.

Non-random comparison group methods--are commonly used when the requirement for strict randomized control cannot be satisfied. Attempts are made to make the comparison group as similar to the experimental group as possible by matching individuals with the same sex, age, racial, or socioeconomic characteristics. The differences in results between the two groups, the experimental and the matched comparison group, are held, as in experimental designs, to be attributable to the results of the program. However, without random assignment there is greater danger that the observed results are attributable to nonprogram influences. Other difficulties with the method include potential bias resulting from self-selection by participants.

Time series--involves a series of measurements at periodic intervals before the program begins and during the program. For example, in evaluating the safety results of Connecticut's crackdown on speeding, it was possible to use time series data collected for several years before and after this new policy change. An abrupt change in such trend data is strong evidence that the action taken caused the observed change in the trend. If measurements can also be obtained in another setting treated as a comparison group, additional insights are possible.

Careful interpretation is needed when using time series data. There may be a time lag between receipt of services and the impact on the services. The question should be asked as to whether there is one or more cyclical phenomena, such as unemployment levels, which cause part of the trend?

The methods discussed above are not exhaustive and there are other ways of making useful comparisons. However, the methods discussed are generally considered to be more reliable than others in determining whether observed results were due to the program or to some other cause.

Developing needed information

Many information systems are not structured to routinely capture data necessary for making valid comparisons. Consequently, a certain amount of ad hoc data collection will be necessary and repeated appraisals of the same programs will be aided by building procedures to capture the desired data. For newly implemented programs, specific provisions for gathering information, necessary to a comprehensive and valid appraisal, can be incorporated into the implementation plans. This requires decisions on:

--Precisely what questions are to be answered.

--Specific items of data required for analytical methods to be employed.

Selection, design, and implementation of data collection instruments may be the least attractive aspect of any appraisal but it is one of the most important. Major sources of data include:

--Interviews.

--Mailed questionnaires.

--Onsite observations.

--Peer group ratings.

--Standardized written tests.

--Project and other program records.

--Federal and State government statistics, such as Census Bureau and Bureau of Labor Statistics.

--Performance tests or other physical evidence.

--Clinical examination.

--Financial, cost accounting, and operational management information.

--Documents such as minutes, progress reports, public releases, etc.

Usually, it is helpful to use several sources, and there are opportunities for creativity in design of collection instruments and in analytical designs which merge data from several sources. It may be very helpful to merge data obtained from personal interviews (condition of home, etc.) with data in program files (achievement scores, etc.). Careful design, in a technical sense, must be coupled with careful consideration of preserving the confidentiality of data about individuals. (See ch. 5.)

Interpreting program results

The key point in interpreting the data is to ascertain the degree to which results, consequences, or impacts are attributable to the program(s) or to other external influences. Frequently the data will reveal only small impacts. Even small effects are important, however, because they may be the only clue available to the potential for larger effects which were either obscured in the data or are achievable only through greater change in the program. Because of the potential for large effects to be obscured by the data, it is important to examine small effects very carefully.

Even if the results obtained are inconclusive, insights into the structure needed in further research and evaluation should be noted. If valid, dependable results are obtained, insights are usually generated concerning a redirection or possible termination of ongoing activities, policies, and programs. These insights should also suggest the need for additional assessments of new and different alternatives.

Those making appraisals have a responsibility to provide systematic information about the results of policies and programs, and the degree of confidence attached to these results. Where a high degree of uncertainty exists, it may preclude firm recommendations concerning policy and program actions. When recommendations are made in these circumstances, the uncertainty must be clearly communicated. Further appraisals can frequently reduce the uncertainties and provide a basis for firm recommendations.

Checking completeness of the appraisal

It is helpful in preparing an interpretive summary of a policy or program appraisal to view the interdependent concepts which have been discussed as a checklist.

Some of the questions which should be contained in such a checklist are:

--Were the reasons for the study found to be valid? Was the cause, scope, and intensity of the original problem or issue redefined as part of the study or as a result of the study? Why did it need attention at this time? Was full consideration given to the expressed needs of all potential users?

--Were the objectives clearly identified? Did they shift over time? Were there implicit objectives?

--Were any special problems, either conceptual or practical, encountered in using input, process, output, efficiency, or effectiveness measures? Were valid standards for comparisons used? Was it necessary to employ surrogate measures and what was the rationale for their choice? What other quantifiable or intangible consequences were measured and how?

--Were data collection instruments sufficient under the circumstances?

--Are findings statistically significant and practically important? Do they answer questions posed at the beginning of the study?

--Were the hypotheses accepted? Were uncertainties resulting from problems with data identified and properly considered? Compared to other studies or evidence, do data and conclusions agree or disagree? If not, why not?

--Were lessons learned identified? Can suggestions be made for immediate improvements?

--To what extent can the performance of this program be generalized to other settings within which the program takes place or may take place? What should and should not be done in the future in other locations or in similar programs? Are these conclusions based on demonstrated causal relationships? Are reasons for program weaknesses indicated?

--Have recommendations been developed for alternatives to be analyzed and compared?

--What is still left to be studied? What new questions were raised that require further research? Which areas of research still need further exploration? What research methods need to be developed or improved in order to make future appraisals more authoritative?

ASSESSING POLICY AND PROGRAM ALTERNATIVES

As in the case of appraising policy and program results, the methods used in assessing policy and program alternatives build on the fundamentals discussed at the beginning of this chapter. In this case also, there are additional concepts and methods which are needed, such as:

--Developing a range of alternatives.

--Screening the preliminary alternatives.

--Estimating the measurable consequences.

--Assessing provisional orderings.

--Determining the impact of constraints.

--Reassessing the orderings of the alternatives.

--Checking the completeness of the assessment.

Developing a range of alternatives

It is essential to search out a wide range of alternatives. The initial search for alternatives should not be constrained. Continuing, modifying, expanding, reducing, or abandoning an existing program, as well as completely new alternatives, should be included. With regard to the existing program, consideration should be given to reexamining the validity of the existing objectives. The process of developing alternatives should include a thorough questioning of the need for any governmental intervention, which may be justified on any of the following grounds:

--Absence of suitable private alternatives or absence of a private marketplace in which the needed service can be distributed.

--The benefits to society resulting from universal use of services or facilities, such as sewage disposal.

--Equal availability of a service, such as public education.

--Distribution of benefits to disadvantaged people, such as health benefits through Medicare and Medicaid.

--Regulation of private activities, such as the certification of effectiveness and purity of drugs.

--Provision of incentives for desired private activities, such as development of energy resources.

Broad classes of approaches which show potential for solving the problem being analyzed should be initially identified. One or more promising alternative approaches from each of the broad classes should be developed. If broad classes are not examined, alternative approaches are usually unnecessarily limited to relatively small incremental changes from existing programs. For example, analysis of an incremental change in eligibility standards for the food stamp program is more narrowly defined than an analysis of overall income security or nutrition policy.

All reasonable, and hopefully well-defined, alternatives suggested by governmental agencies, legislative committees, advocacy or interest groups should be considered. Issue papers, such as described in chapter 5, can be useful at this stage of an assessment.

Screening the preliminary alternatives

A preliminary analysis of the likely consequences associated with the range of alternatives, including the status quo, should now be undertaken. This initial screening is intended to eliminate obviously inferior approaches and to reduce the original list of alternatives to a manageable size. It is helpful to make approximate calculations of cost and consequences, of breakeven points, and of technical feasibility, etc. Alternatives should not initially be ruled out based on implementation difficulties, including organizational or procedural changes.

Modifications and combinations of alternatives usually become apparent and frequently provide the basis for new and superior alternatives. The search for alternatives is a continuing activity and, as the analytical effort proceeds, opportunities to invent or discover other alternatives will arise.

Estimating measurable consequences

Estimates must be made of anticipated measurable consequences as well as all costs and resource inputs under various conditions and levels of available resources. Measurable consequences include effectiveness, externalities, and distribution considerations. In making such estimates, the data on actual costs and effectiveness found in prior appraisals of similar programs should be used together with actual operating data. It may also be necessary to use well developed causal models to make such effectiveness estimates.

Some effort should be made to estimate side effects (externalities) and their resource impacts. An estimate is needed, to the extent possible, of the differences of impact on the beneficiaries and the cost bearers (distribution considerations). Approximations must be used for externalities and distribution estimates and various value judgments are involved in weighing both.

When analyzing costs which should be associated with effectiveness, various cost concepts are needed, and information on these costs is usually available. When analyzing costs which should be associated with externalities and distribution considerations, total as well as incremental costs should be developed. Frequently such costs are incomplete. They should be checked for reasonableness and consistency across the alternatives of interest.

Information at the margin, as contrasted with information on total quantities, is very important in resource allocation decisions. Approximations of incremental costs, however, are more easily obtained than are approximations of the marginal aspects of other measurable program consequences. A reasonable effort should be made to estimate the direction and magnitude of the variations of program consequences over relevant ranges.

Information on measurable consequences obtained from audits, evaluations, or other studies should be used. Historical and trend data may provide information concerning how the various consequences are affected by the scale of activity.

Assessing provisional orderings

Once the total and incremental consequences of the alternatives have been estimated the alternatives should be arrayed in some order. This ordering may be based on one of several available approaches.

One approach is "cost-effectiveness." This approach focuses on resources expected to be consumed and how well the objectives are achieved. Using this framework, a preferred alternative is identified as one which produces the largest achievement for a given level of costs or which minimizes resources expended for attaining a given level of effectiveness.

While the cost-effectiveness approach provides a basis for ordering competing alternatives, it does not clearly allow for comparisons of alternatives associated with multiple, possibly conflicting, objectives and does not treat other consequences of alternatives--externalities and distribution considerations--as an integral part of the analysis.

A second approach to ordering alternatives is "cost-benefit" analysis. Externalities and distribution considerations are incorporated in this approach. Major consequences, or benefits, are measured in dollars, and differences between monetary benefits and costs provides the basis for choice among alternatives. Cost-benefit analysis is more useful than cost-effectiveness analysis in treating differing as well as conflicting objectives. The streams of benefits and costs can be discounted to their equivalent present values, thus accounting for the effects of time. Conceptually, decisionmakers would select programs based on rankings of net present value benefits until the total available resources were exhausted.

Another approach is "cost-value" analysis. This is a technique for obtaining generally acceptable quantitative weights for use in comparing the value of the alternatives. In this approach, the weights assigned to various outcomes are based on value judgments obtained from the decisionmakers.

The cost-value method combines elements of cost-effectiveness and cost-benefit analysis. Externalities and distribution considerations can be incorporated with effectiveness. Because the value judgments of decisionmakers differ, various sets of value judgments should be used and the ordering(s) of alternatives should be tested for their sensitivity to these differences. In such analysis both the array of consequences associated with each alternative and the ordering based on the various value systems should be presented to decisionmakers.

Each approach has both strengths and limitations, but all share certain limitations. One such limitation is uncertainty caused by such things as variations in assumptions and the quality of information on the alternatives. Because uncertainty is always present in anticipating future outcomes, undue reliance should not be placed on small differences in ordering(s) of alternatives. The quantitative analysis which has been discussed should be supplemented with an analysis of nonmeasurable consequences. A serious attempt should be made to indicate the significance of nonmeasurable consequences.

Determining the impact of constraints

Special efforts should be made to assess the impact of actual and potential legal, financial, and political constraints. Programs and policies must operate within the framework of law. Alternatives which may appear theoretically desirable must also operate within the law. Consequently, the alternatives considered for adoption must conform to this framework.

In addition to these sorts of constraints, there are constraints resulting from conflict with other objectives. An example of such constraints is the conflict between environmental, transportation, and energy objectives.

However, constraints are not inflexible. If decisionmakers were clearly aware of the potential opportunities foregone resulting from existing constraints, those constraints might change.

Decisionmakers must consider possible public reaction to alternative policy and program options, strategies that might increase their acceptability, and what administrative or other operational barriers to implementation exist. The problems of implementation and of acceptability may, to some degree, be dealt with in analysis. Usually, assistance can be provided to decisionmakers in identifying the "second or third best" alternatives which may have higher prospects for being accepted or implemented.

It has been argued that if acceptability considerations are avoided, the assessment of alternatives becomes more objective, less parochial, and less tailored to fit preconceived positions. On the other hand, it may be argued that if acceptability considerations are not included, the analysis may prove to be irrelevant.

Reassessing the orderings of the alternatives

Orderings of alternatives are always provisional. They are determined within the context of the factors and values considered to be important during the course of the analysis. The assumptions and values underlying the various orderings of the alternatives must be clearly presented to decision-makers. Furthermore, even when the analyst thinks the study is completed, decisionmakers may raise new issues, ask new questions, request further study, and ask for additional comparisons. As these requests are answered, the orderings of alternatives may shift.

Although attempts should be made to include as many factors as possible, other considerations properly impact the final policy and program choices. Some of these considerations may be completely beyond the analyst's knowledge or ability to estimate, even qualitatively and belong in the province of the decisionmakers' judgments.

Checking completeness of the assessment

Some of the questions which should be considered in preparing an interpretive summary of a policy or program assessment are contained in the following checklist:

--Were the reasons for the study found to be valid? Was the cause, scope, and intensity of the original problem or issue redefined as part of the study or as a result of the study? Why did it need attention at this time? Was full consideration given to the expressed needs of all potential users?

--Were the objectives explicitly stated and validated? Did they change during the course of the assessment? If so, why?

--Were there any potentially interesting alternatives eliminated early in the analysis? If so, why? Under what circumstances might they become attractive?

--Were any special problems, either conceptual or practical, encountered in specifying an adequate set of quantifiable measures? How reasonable were the dollar values attributed to physical measures, if that was done? Were qualitative indicators properly identified and used?

--Do the effectiveness measures accurately reflect the degree of attainment of the objectives? Were they consistently used among all of the alternatives? Is the effectiveness data reliable? Has uncertainty in the data been properly considered?

--Were side-effects and distribution considerations adequately considered? Are there significant differences among the alternatives?

--Were all of the cost implications captured? How reliable are they? What is the range of uncertainty?

--To what assumptions or data is the ranking of the alternatives sensitive? Are there any actions which can make the leading alternatives significantly less affected by the uncertainties?

--Are there any special problems connected with gaining general acceptance of the apparently preferred alternatives? Will implementation of any of these pose particular difficulties?

--Is it likely that additional information about the leading alternatives would change the ranking? How, when, and at what cost could this information be obtained? Can the policy or program decisions be held open while new studies, evaluations, or research efforts are completed? What long-term evaluation or research efforts need to be initiated to meet similar or related problems in the future?

CHAPTER 5

PRACTICAL ASPECTS OF MANAGING AND PERFORMING STUDIES

Previous chapters have discussed the conceptual aspects of appraising the results of policies and programs and of assessing alternative solutions to the issues related to them. In deciding which problems to study and in carrying out actual evaluations or analyses, however, certain practical questions arise which need to be addressed. This chapter discusses some of these practical aspects and offers suggestions for coping with them. The list is not extensive, but is indicative of the very real problems faced in this type of work.

FORMULATING AN AGENDA OF STUDIES

One of the most important responsibilities facing any manager of an evaluation, analysis, audit, or other program review staff is developing an overall work plan for the organization. Planning a program of studies which will be of maximum benefit to decisionmakers should involve two principal tasks:

--Identifying problems or issues which are evolving as major areas of concern.

--Deciding which of the many candidate problems the organization should commit itself to studying.

Identifying emerging problems

A contribution can be made to resource allocation decisions by raising problems and exploring their ramifications in "issue papers." These focus on programs or policies which, there is reason to believe, will become the subject of a full-scale evaluation or analysis.

The ability to recognize emerging problems for issue papers depends on experience and good judgment. It also requires (1) a deep understanding of the program area in question, and (2) changes in the external environment.

An issue paper may follow the format and style appropriate to a full-scale evaluation or analysis but is limited to "state-of-the-art" assessment of what is known about the particular program area. An issue paper could be as short as a few paragraphs or as lengthy as a document which covers all or almost all of the points required in a full evaluation or analysis--the latter would not have the scope or definitiveness of a finished study. The conclusion of an issue paper

should emphasize recommendations on the nature and pace of study efforts. For example, whether the problem should receive high priority attention (and why), whether it should be pursued but on a long-term basis (and why), or whether it should be abandoned (and why).

Deciding which problems to study

Many problems, programs, and policy issues are in need of systematic study. Yet the scarce resources of evaluation and analysis staffs need to be allocated to the most productive projects.

Issue papers can identify policy and program problems worth evaluating or analyzing. However, a complex series of judgments is still needed to select that particular group of problems which, if solved, would maximize the anticipated payoff. While it is relatively easy to list the factors influencing these choices, it is seldom feasible to appraise all of them in a formal quantitative fashion.

A qualitative, yet systematic weighing of the following factors will be helpful:

1. The anticipated payoff of successful evaluation or analysis.

   This payoff can take several forms: an ineffective program can be canceled and costs saved; a mismanaged program can be reshaped with consequent improvements in effectiveness, reductions in costs, or both; better alternatives can be substituted for current programs and policies with gains in effectiveness, reductions in cost, or both.

2. The chance of the successful performance of an evaluation or analysis.

   This judgment depends on a basic understanding of the fundamental causal relationships, the requirements for additional information, the adequacy of current analytical methods, the quality of staff, consultants or contractors, and the time and money available.

3. The chance that a preferred course of action can actually be implemented.

   This judgment depends upon such a thing as newness, simplicity, visibility, coverage, and timeliness of the preferred course of action.

4. The need for resolving the problem or issue.

This need depends on the nature and relative importance of the problem and the time remaining before a meaningful decision has to be made.

5. The cost of the evaluation or analysis.

BEGINNING A STUDY

Certain tasks need to precede the time when the major commitments will be made of staff and other resources. These tasks included preparing a study plan, obtaining necessary agreements, selecting the study team, establishing lines of communication, and selecting appropriate methods.

Preparing a detailed study plan

A substantial effort should be devoted to drawing up a comprehensive and thorough study plan which will serve as a guide for all subsequent work. A study plan that is too broad in scope or loosely stated is almost certain to create false expectations for some interested groups. Clearly, trade-offs have to be made between the time devoted to planning versus doing a study; and within the planning period, between a detailed and a general study plan.

As the study progresses, it is likely to deviate from original expectations. Perhaps, the issue turns out to be different from that originally postulated; the objectives may not have been stated precisely enough; a working assumption may not prove viable; other alternatives to the program emerge; new facts come to light; hoped-for data cannot be obtained; and so on. All of these developments call for some modification of the study plan. Changes should be made, as appropriate, to the study plan.

Essential elements of the study plan would appear to be:

--A clear statement of the problem to be studied and questions to be answered.

--A careful listing of constraints or assumptions.

--A statement of methodological approaches to be used.

--A specification of the resources to be committed (including identification of the key staff members and any contracted tasks required).

--The frequency and format of reports and to whom.

--Procedures for amending the study plan.

--The time frame for the major components of the study and the final deadline.

When a study or a major part of it is to be performed by contract, there should be discussion and understanding by the parties concerning the essential elements of the study plan. This is likely to require lengthy dialogue with the decisionmakers. Persons with official responsibility for the policy or program and for the study should assess feasibility and validity of the study plan. Any unresolved differences between them and the study group should be noted. Large segments of time and effort are worth investing to arrive at a workable understanding.

Selecting the study team

Most analyses or evaluations require contributions from several key persons. For large studies, subteams for particular aspects may be required. As in any group effort, someone must be in charge to (1) provide guidance, (2) manage the work on a day-to-day basis, (3) report to higher authority, and (4) generally be responsible for meeting the terms of the study plan. The coordinator or director should be experienced, with a technically sound but broad background, an instinct for the principal issues, and the leadership abilities that elicit from the team members their best efforts.

Leadership is important for the team effort. A team studying any complex policy or program should be composed of experienced persons from various disciplines.

Regardless of their origin, however, all should be made to feel as coequal members of an exciting intellectual experience and a useful endeavor. It usually turns out that the team coordinator or director will have to be principal editor of the final report--so expository writing skill is also a necessary characteristic.

One way to create such an environment--at the same time avoiding duplication of effort--is to have an initial briefing on the terms of the study plan with all team members. Important aspects such as concepts, assignments, schedules, basic assumptions, need for personal and agency coordination, and reporting requirements should be fully understood and agreed upon in advance. Provision should be made for periodic briefings by each specialist to the team as a whole so that everyone has both a grasp of overall progress and a chance to offer facts or insights on any aspect of the study.

It is often helpful to obtain reviews by competent and widely recognized independent professional analysts [illegible] evaluators and experienced program administrators. This adds a seasoned viewpoint which may improve the technical aspects and may assist the supervisor in assessing the technical adequacy of the work of staff members trained in different disciplines.

Establishing lines of communication

If the study effort is sufficiently large, official points of contact among various interested groups and users of the study should be designated. This should insure that communications of all kinds flow quickly and clearly among the groups having a major interest in the progress of the study. Open communications provide the basis for a more complete assessment or appraisal and a climate in which recommended changes are more likely to be accepted and implemented.

Selecting appropriate methods

Comparison or analytical methods which yield valid and unequivocal results should be used. However, the method must also satisfy the constraints of time, money, and data peculiar to the study. If the constraints imposed are so rigid that the study would be compelled to use methods judged to be analytically inappropriate, the study should be undertaken only after fully informing responsible authorities of the risk that reliable conclusions and recommendations are not likely.

No particular approach or technique is inherently the appropriate one. In practice, there are too many [illegible] to mold the policy or program issue to fit a specific technique. This should be avoided. For a specific study various approaches, each having its own particular logic, should be considered. Usually, a blend of methods and techniques will be required to provide insights into the full consequences of the various alternatives. Reasons for selecting a particular approach or blend of approaches should be clearly stated so others can understand the rationale for the particular choice.

Whatever approaches and methods are selected, they should satisfy the following criteria:

1. Validity--does a high degree of confidence exist that inferences made occur in the real world?

2. Relevancy--are insights obtained useful to decision-makers?

3. Significance--can a reasonable number of nontrivial inferences be drawn?

4. Efficiency--does the value of the insights exceed the costs of using the approach?

5. Timeliness of results--will the analytical information be available within the time available and will it be in time to meet a management or legislative decision point such as renewal of expiring legislation?

Modeling and statistical inference are two related methods which are particularly useful and are frequently used by evaluators and analysts.

A model is an abstraction from or a representation of the key elements in some real world system. If the key elements and their relationships are adequately specified, relevant, and valid, a model can predict the consequences of untried alternatives and variations in data and assumptions.

Statistical inference techniques are widely used to analyze the implication of data obtained from the various collection instruments as well as the results obtained from analytic models. There are conditions and assumptions that must be satisfied before appropriate applications can be made. Mistakes can occur, for example, if prepackaged computer programs are used without understanding the technique itself.

CONDUCTING A STUDY

In the performance of any evaluation or analysis, practical decisions of many types must be made, and practical problems are frequently encountered. Some of the most common ones are discussed here.

Collecting relevant data

In performing studies, there is often a temptation to collect all of the information which might be of use. While every piece of information may have some value in the right place, the question is, Is it relevant and worth what it costs to acquire it? Questions which should be continually applied to any data collection effort are: Exactly what question is this piece of data intended to answer? What analytic model demands it? What calculation cannot be done without it?

Testing the reliability of data

An attempt should be made to estimate whether data are reasonable at the time they are first generated; i.e., how

does this new piece of data square with everything else that is known or can be deduced relating to it? This is especially important when complex calculations are involved. How does the answer compare with rough calculations? The exercise of making rough calculations frequently gives the staff member new insights into the data.

There are numbers of one kind or another which are widely published. In some cases, essential data will have to be "constructed" or "extracted" from secondary sources. Everyone seems to use them unquestioningly. However, a careful analysis has often demonstrated that some data have different interpretations than what is commonly supposed.

Occasionally, an attempt may be made to withhold information. It is not uncommon to hear that data

--is too hard to assemble;

--doesn't exist in the form wanted;

--is only a working paper; or

--is privileged.

When faced with this type of situation, the analyst should (1) consider the value to the study of the information, (2) attempt to obtain a release of the appropriate information if needed, and (3) propose to the study coordinator that a formal request be sent for the needed information.

Frequently, data collected from different sources about the same subject matter will be in apparent conflict. The first practical step in getting the right data, is to reconcile the apparent conflicting interpretations of the data. An appropriate question may be: Are they truly two different sets of values describing exactly the same event or situation? A second step would be to examine how the data were derived. The apparent conflict may be a simple function of the data collection methodology. After these procedures have been employed, it may be appropriate to use an analytical technique to determine the significance of the differences. Additional assurances may be obtained by having data reviewed by experts in the field.

Protecting the confidentiality of information about individuals

It is often necessary in evaluation and analysis to collect data about individuals. It is important to make

certain that such data is not personally identifiable in the study or in unsecure files. If it is necessary to obtain information from the same individuals in subsequent time periods, special controls and procedures should be required to assure that systems of records do not disclose individually identifiable data. For example, an evaluator may use a unique code for an individual's data. This would enable the data source to furnish appropriate information about that particular person without the evaluator knowing the person's name. Encoding is a common step in such control procedures. A very common code is the social security number. Nonetheless, no right, privilege, or benefit may be denied by anyone to an individual who refuses to divulge his or her social security number except where disclosure is required by Federal law or was required by a law or regulation that predates the Privacy Act (Privacy Act of 1974; Public Law 93-579; 5 U.S.C. 552a).

Federal agencies and their contractors are required to comply with all provisions of the Privacy Act of 1974 to protect the confidentiality of individually identifiable data. These provisions include:

--Public disclosure of the fact that an agency maintains a system of records about individuals.

--Strictly enforceable procedures for assuring that individuals have access to their records and the opportunity to correct them.

--Controls on interagency transfer of individual's identifiable data.

--Administrative, technical, and physical safeguards to prevent unauthorized access to such data and personal liability for civil damages as well as criminal penalties for violation. In planning a study, care should be taken to require individual identifiable information to be collected only when no other approach can enable the issue to be validly studied.

Documenting and referencing

Documenting appraisals of results and assessments of alternatives is important. Basic assumptions should be clearly identified and recorded. The documentation should be sufficient so that another individual or team involved in reviewing the policy or program, by reviewing the documentation, could follow the analysis, and as needed, reconstruct parts of it, or use it in another study. The rationale for using indirect or surrogate measures should also be stated

explicitly. Oral interviews should be summarized in writing, dated, and filed. Original documents should be retained. Complete files of relevant raw data and working papers should be kept and filed so that they can be retrieved easily for review. Information which cannot be readily filed should be adequately described and referenced in the files.

The study team should design, use, and save working papers. Well designed, clearly labeled, and fully legible working papers offer an important insurance policy to the study team. The working papers constitute the evidence gathered. A review of the working papers will show whether the study team has been thorough or whether they may have overlooked an important fact or element of a problem and that all similar elements of the analysis or evaluation have been treated consistently. Developing the total costs of each of a series of alternatives is an outstanding example of the need for, and usefulness of, a carefully designed and clearly labeled set of worksheets. Without them, the chances of missing an important cost element, incorrectly calculating an intermediate result, or costing the competing alternatives inconsistently are substantial.

Working papers should be dated and signed so that a clear trail is established as to who did what and when. The best way to tie it all together is to file with workpapers, one copy of the final report, which cross references significant sections to the workpapers.

Adhering to time schedules

Effort should be made to anticipate some of the possible delays and the time schedule should allow some slippage to accommodate unforeseen delays. In a very real sense, most complex tasks are harder than originally anticipated, and therefore take longer than they were estimated to require.

Leading and coordinating the study team

It is essential to maximize the interaction among the study team members. Physical arrangements which inhibit this result should be avoided or modified if at all possible. When gathering the first list of alternatives or hypotheses, brain storming is extremely useful.

The coordinator should take every practicable step to insure easy access to the decisionmakers who expect to use the analysis or evaluation. A continuing (but not necessarily continuous) dialogue should help to make the products useful and well accepted. The coordinator also needs to impress

on the team the importance of maintaining an open, honest, and amicable relationship with the personnel of the program under analysis or evaluation. It is all too easy for program people to frustrate a study if they have been antagonized or hurt.

Using computer-based models

For most large scale, but routine, quantitative manipulations (statistical analysis, linear programming, etc.) good canned programs are available and should be used. When a program or problem has many complex interrelationships, however, and the effects of altering the assumptions or data are not obvious, a specially designed, computer-based model may facilitate the study. In such cases, creative computer programmers are extremely valuable.

The structure and operation of any model should be reasonably apparent to decisionmakers who want to use the study: its output and workings must be readily understandable to them. Usually, this can be accomplished by carefully diagramming the components of the model and explaining how each component operates and interacts with the others. Users of the study will normally accept the computational competence of the model if the logic makes sense to them and they have confidence in the study team.

COMMUNICATING STUDY RESULTS

Many persons doing studies fail to understand that doing a good piece of work is necessary but hardly sufficient for bringing about a favorable change in the real world. At least two major steps beyond successful completion of a study are required: the results must be clearly, concisely, and cogently communicated to all those affected; and a policy or program decision must be made which results in some kind of action.

Specifying the nature of reports

There are three general classes of problems involved in reporting appraisals of results and assessments of alternatives: (1) to whom should reports be made; (2) when should reports be made; and (3) what style and content characterizes good reports. Each new study will suggest its own individual requirements and should be made a matter of record in the agreed work plan adopted before each study is begun. A few general guidelines can, however, be set down.

Unless special considerations dictate otherwise (e.g., security problems), reports should routinely go first to the team supervisor and others as needed to insure that they meet the organizations' professional standards. Even professionally sound studies, however, may result in disagreements with the managers of the programs being studied. In these cases, the study team should reduce the number of areas of disagreement, and where these continue to exist, the issues should be substantial and clearly defined. Although decisionmakers waiting to use the report can be kept informed of key findings, it will in the end serve them best if the foregoing review process is complete before the final report goes to them.

No report, other than the final version, should be distributed beyond those listed earlier without their corcurrence. Unauthorized release of preliminary, draft, interim, or partial reports can be harmful because frequently, erroneous information, even though corrected later, becomes widely diffused and becomes a source of further error and confusion. Publicly available reports should be free of such errors.

In planning the study, sufficient time should be allowed for writing the final draft report, gathering comments, editing, and securing the necessary approvals. The report writers, in turn, have an obligation to complete the report within the scheduled time. Report outlines should be prepared early. They can provide indications of the most critical data gathering and interpretation tasks yet to be completed in order to have a useful and timely report. Decision points come and go relentlessly and a potentially good, decision-affecting report may lose much of its value because it was not available when needed.

Communicating with clarity and conciseness

Writing a good report is an art and the required skills are probably as scarce as that necessary for evaluation and analysis themselves. The solution is to insist that staff members work at learning to write well. One helpful step is to provide staff members with specific guidance, such as Strunk & White's "The Elements of Style" and insist that they study it and use it as part of their regular duties. In addition, someone on the staff can serve as resident editor. All significant alterations should be discussed with the author: not only to insure accuracy, but to assist the author in learning to write clearer, shorter, and more trenchant reports.

Study reports are typically directed at a reader lacking relevant technical training. Therefore, the main body of the report should be written so that it is readily comprehensible

to the nonprofessional reader. However, material included in the report should be sufficient so that a reader can understand the logic trail in support of the conclusion. Jargon should be kept to a minimum, and where it is used, define it carefully. Supporting technical material should be presented in appendixes. Graphs and tables included in the main body of the text must be clearly labeled and fully discussed in the text. Short reports are typically self-contained, while long ones ought to be accompanied by an executive summary of the study's general conclusions and recommendations.

There will, of course, be differences between the format and content of a report on appraising program or policy results and a report on assessing alternatives, or a report containing both. Within each of these, some variation in format and content is inevitable, depending on the nature of the policy or program issue being studied and the methods used. In general, the format and content of reports should cover what was found through each of the concepts and methods discussed in chapter 5, as appropriate.

Following up

Writing a clear, concise, and informative "final" report is not the end of the "communicating" responsibilities. Usually, some decisionmakers will need assistance in (1) interpreting the report, (2) clarifying aspects of it, (3) getting answers to questions raised by it but not answered, and (4) in general, developing a reasoned reaction to it. Briefings, informal question and answer sessions, and various kinds of supplementary written materials may be needed. In some cases, the communicating responsibility may even extend to preparing the supporting technical parts of whatever decision document emerges from the decisionmaking process.

GLOSSARY

Activity	Any project, task, or process required in carrying out a program. A combination of several activities such as research and development, training of personnel, and distribution of information may be elements in a particular program. Activities constituting a program vary with the nature and purpose of the program.
Activity milestone	The accomplishment of a specific amount of activity or work by a given date (e.g., build 5,000 child care centers by 1978).
Alternatives	The different possible courses of action (policies or programs) with respect to achieving a particular objective.
Analysis	For the purpose of this document, analysis is a systematic procedure which searches for alternative policies and programs for achieving public objectives and attempts to assess and compare their anticipated costs and benefits over time, as well as their other consequences, in order to provide the basis for better future choices.
Appraising policy and program results	For the purpose of this document, the term is equivalent to policy and program evaluation.
Assessing policy and program alternatives	For the purpose of this document, the term is equivalent to policy and program analysis.
Bias	An unintended disturbing influence on the outcome of a program or experiment caused by abnormal conditions or an unplanned variation in the treatment of a control or experimental group.
Case study	A detailed, indepth, appraisal, generally of a single policy or program, focusing on outcomes, processes, management, organization, and operational settings.

Constraints	Limitations on the freedom to choose or use particular alternatives due to problems of political acceptability, technical feasibility, cultural values, legal structures, environmental considerations, etc.
Control group	In a field experiment, persons, groups, or entities, who, through a random assignment, do not receive services associated with a program. The control group acts as a basis for comparing the effects or results experienced by the (experimental) group receiving the services.
Cost-benefit analysis	An analytic approach which focuses on resources expected to be consumed and on those consequences which can be translated into monetary terms. A useful approach allowing the determination of discounted present values and a comparison of widely ranging alternatives. (See McKean in the bibliography for a more complete discussion.)
Cost-effectiveness analysis	An analytic approach which focuses on resources expected to be consumed and on the extent to which objective(s) would be met. A useful approach, but limited in situations involving multiple, possibly conflicting, objectives. (See Quade in the bibliography for a more complete discussion.)
Cost-value analysis	An analytic approach which combines elements of cost-effectiveness and cost-benefit analysis. Effectiveness, externality, distribution, as well as other consequences can be treated through the use of value surrogates which reflect the relative importance of various consequences and allow an overall assessment of the attractiveness of various alternatives.
Costs	A measure of that which is given up in order to achieve some objective. Costs can be measured in terms of the resources used, the alternative uses of those resources, the money equivalent of those resources, or in terms of the benefits foregone in the next best uses of those resources. (See Fisher in the bibliography for a reference work giving a fuller treatment of the subject.)

Distribution measures	Measures which reflect immediate and indirect consequences on the beneficiaries and the cost-bearers resulting from the implementation of a policy or program.
Distributional effect	The results of a transfer payment system which takes resources from one group and distributes them to another group (e.g., the progressive income tax system plus the welfare system result in a transfer of wealth from the rich to the poor.)
Effectiveness measures	Measures which accurately reflect the extent to which the desired objective(s) associated with a policy or program are or would be met.
Equity	The notion of "fairness," based on a particular value system. It is used in judging the appropriateness or desirability of distributional effects.
Evaluation	For the purpose of this document, evaluation is a systematic procedure which attempts to appraise and measure the actual inputs, processes, outcomes, and operational settings of one or more ongoing programs or policies in order to compare these findings with those which were anticipated or assumed. It then seeks to explain the discovered differences and to suggest alternatives for improvement.
Experimental group	In a field experiment persons, groups, or entities who, through a random assignment, receive services associated with a program. The experimental group is compared with the control or the nonrandom comparison group to determine which effects are attributable to the program.
Experimental methods	The attempt to appraise the degree to which observed results are attributable to a policy or program or attributable to other factors not associated with the policy or program through the use of an experimental group and a control group.

Externality measures	Measures which sufficiently reflect the side effects associated with a policy or program.
Hypotnesis	A statement of a presumed causal relationship having empirically testable consequences. The information collected is used to prove or disprove the statement.
Incremental effects	See marginal effects.
Intangible measures	Measures which reflect the qualitative aspects of the consequences associated with a policy or program.
Issue paper	A document which conforms in format to a regular evaluation or analysis but which restricts itself to what is currently known about the issue, problem, or program. It is intended to assist in identifying emerging problems and deciding whether they should become the subjects of a full, formal evaluation or analysis.
Managerial efficiency	A situation in which outputs are maximized for a given technology and level of resources through the actions of the manager in deploying those resources and favorably affecting the behavior of the personnel under his control or with whom he interacts.
Marginal effects (incremental effects)	A measure of the effects (usually costs or effectiveness) which are associated with just the next unit of activity or production; the addition to total costs or effectiveness brought on by the next unit of activity or production. Most analysis and managerial decisionmaking is concerned with changes "at margin"--the costs and other effects produced by a small addition to or reduction in activity from some base point. Strictly speaking, marginal refers to the changes brought about by just a single unit change in production or activity. Incremental effects refer to effects caused by small blocks of change in activity or production.

Measurable consequences — The set of program outcomes which are generally measurable in a quantifiable way. Some of these are effectiveness, benefits, costs, externalities, distributional effects, and where possible, intangible effects.

Model — An abstraction from, or a representation of, the key elements and their relationships in some real world system. The usefulness of any model, as used here, lies in its explanatory capability. A useful way of classifying models is: mathematical models (e.g., linear and nonlinear programming, decision analysis, and markov analysis); simulation models (e.g., sensitivity analysis and operational gaming); and qualitative models (e.g., role playing and scenario writing). Each can provide its own special insights depending on the nature of the problem being studied.

Nonrandom comparison group — In a field experiment, when random assignments to the group not receiving services associated with a program are not possible, a pseudocontrol group acting as a benchmark is established which is matched as closely as possible to the (experimental) group receiving the services.

Objective — A good effect or intended result. Sometimes called a goal or purpose. Statements of objectives specify the good effect or end result intended to be achieved and for whom. In the case of ongoing programs, a third clause specifying the general means by which the effect is to be achieved is also given. In an objective statement for an issue or problem, the third, "how-to-do-it" clause is absent, since that is the matter under study. Objectives may be very general (e.g., maintain or improve the health of the people of Appalachia) or quite specific (e.g., reduce the incidence of tuberculosis among miners in West Virginia). Specific or more operationally stated objectives should be demonstrably related to the larger overall objectives with which they are associated. When specific levels of objective achievement or activity by specific dates are

established, they are more properly called management targets or activity milestones and should be reflected in the appropriate program plan.

Operational settings — The environments, physical as well as political and social, within which a program operates. It is of particular importance in evaluation efforts because elements in the operational setting may have strong and unsuspected influences on program outcomes.

Opportunity costs — The benefits which are foregone in any of the alternative uses of a set of resources which are, in fact, devoted to a particular use. They are the good things which are lost or given up when resources are used in a particular way and not in any of their other alternative uses.

Pilot program — A trial implementation of a proposed program on a small-scale basis in order to assess the likely consequences that would be attributable to the program before full-scale implementation, as well as to obtain an understanding of the functioning of the program in a realistic setting.

Planned variation — A controlled change in the operations of a program or an experiment to appraise what program consequences are attributable to the change itself.

Policy — An assertive statement specifying a decision-rule, a guide for action, or a state of affairs deemed desirable (e.g., equal educational opportunities shall be open to all students regardless of socioeconomic status or race.) All policies, to the extent they are operative, generate program activities and thus consume resources.

Present value discounting — A technique for recognizing time preferences for expenditures and benefits (in monetary terms). Discounting factors which are a function of the opportunity costs and times involved are applied to streams of future costs and benefits to reduce them to their equivalent present values. By doing so, the the time value of money is recognized, and alternatives with differing patterns of

costs and benefits may be validly compared. (See Fisher in the bibliography for a reference providing a fuller discussion.)

Program — A unique collection of people, physical resources (facilities, equipment, and supplies), policies, and technologies which by their integrated operation (organized set of activities) produce an output (a good, service, or a capability) that tends to achieve one or more of the responsibilities assigned to an agency (its objectives).

Program comparisons — A technique for appraising the results of programs and determining their generality by operating similar programs in a variety of locales or with a variety of target groups; or by conducting variations on a basic program in the same setting to determine the range of outcomes associated with the variations. (See Weiss in the bibliography for a more complete description of this technique.)

Program inputs — The resources, both tangible and intangible, needed to create and operate a program.

Program outcomes — All of the effects, both intended and unintended, which come about as a result of program operations. Program outcomes include effectiveness, benefits, disbenefits, externalities (so-called spillover effects), and distributional consequences.

Program outputs — The physical goods, services, or capabilities generated as a result of program processes operating on program inputs in accord with a particular set of policies and technologies.

Program processes — The activities and technologies by which a program uses the resources furnished and converts them to outputs.

Surrogate measures — Measures which are used as an indirect substitute for the measures of interest, because knowledge is insufficiently developed, or information is lacking regarding the determinants of the more basic measures.

Resource allocation	The process by which decisionmakers assign scarce resources among various programs to achieve the greatest excess of benefits over costs.
Sensitivity analysis	A group of related techniques which attempts to determine those data and assumptions to which the ordering of the attractiveness of alternatives is most dependent upon, and the degree of that influence. Included are breakeven analysis, the B-C-P technique, parametric analysis, and *a fortiori* argument. (See Hitch & McKean in the bibliography for a more complete discussion of these techniques and their uses in dealing with uncertainty.)
Statistical inference	Methods by which statements can be made with varying degree of assurance concerning the characteristics of a large group from information collected on a small part of the group. A range of techniques are available. (See Wonnacott and Wonnacott in the bibliography for a more complete discussion of these techniques.)
Time series methods	A series of observations on effects or results are made prior, during, and after policy or program implementation. Only if the effects are strictly increasing during the period of policy or program operations are the effects truely attributed to the policy or program.
Uncertainty	A situation characterized by more than one possible outcome and which meets either of two conditions: (1) all of the possible outcomes are known, but there is no information on how likely any of these outcomes are to occur, or (2) not all of the possible outcomes are known.

AN ANNOTATED BIBLIOGRAPHY

The purpose of this appendix is to list some references that should be useful to those persons having limited experience in the conduct of evaluations and analyses. The references are listed under several categories: basic disciplines, quantitative methods, evaluation, and analysis.

BASIC DISCIPLINES

Baumol, William J., Economic Theory and Operations Analysis. 3d ed. Englewood Cliffs, N.J., Prentice-Hall, Inc., 1972.

An excellent treatment of basic economic concepts and quantitative methods as they would apply to issues of resource allocation.

Downs, Anthony, Inside Bureaucracy. Boston, Little, Brown, 1967.

Very useful insights concerning bureaucratic behavior and setting.

Haveman, Robert H. and Julius Margolis, eds. Public Expenditures and Policy Analysis. Chicago, Markham Publishing Co., 1970.

Various aspects of public expenditure economics are discussed. The economic bases of public expenditures are developed in part I. Part II with its emphasis on institutional considerations is of special interest, as is part III which is concerned with analytic problems in policy analysis. The remaining parts survey the Planning, Programing, Budgeting, and System experience and offer suggestions.

Rivlin, Alice M., Systematic Thinking for Social Action. Washington, D.C., The Brookings Institution, 1971.

A provocative series of essays on the issues involved in attempting to solve the problems of society.

QUANTITATIVE METHODS

Hillier, Frederick S. and Gerald J. Lieberman, Introduction to Operations Research. San Francisco, Holden-Day, Inc., 1967.

Although a knowledge of mathematics is required, the text presents a comprehensive survey of the methods, models, and techniques that are used in analyses.

Levin, Richard I. and Charles A. Kirkpatrick, Quantitative Approaches to Management. 2d ed. New York, McGraw-Hill, 1971.

An introduction to quantitative methods and techniques.

Moroney, M. J., Facts From Figures. 3rd ed. Baltimore, Penguin Books, Inc., 1956.

A very readable treatment of the use and misuse of descriptive statistical techniques.

Raiffa, Howard., Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Reading, Mass, Addison-Wesley Publishing Co., 1968.

A clear exposition of the process of determining best choices under uncertainty and considerations affecting group decisions.

Tanur, Judith M., and others, ed. Statistics: A Guide to the Unknown. San Francisco, Holden-Day, Inc., 1972.

Applications of statistics and probability are developed in a case metnod setting.

Wonnacott, Thomas H., and Ronald J. Wonnacott, Introductory Statistics. 2d ed. New York, John Wiley & Sons, Inc., 1972.

A fairly rigorous, but understandable text of statistical inference, including Bayesian methods and nonparametric statistics.

EVALUATION

Hatry, Harry P., Richard E. Winnie, and Donald M. Fisk, Practical Program Evaluation for State and Local Government Officials. Washington, D.C., Urban Institute, 1973.

A very useful primer on evaluation with good examples.

Isaac, Stephen, and William Michael, Handbook in Research and Evaluation. San Diego, Knapp, 1971.

A compendium of useful checklists, do's and don'ts, and summaries of important concepts and techniques for evaluation.

Riecken, Henry W. and Robert F. Boruch, eds. Social Experimentation: A Method for Planning and Evaulating Social Intervention.

New York, Academic Press, 1974.

An invaluable guide--technically, ethically, and administratively--in using experimental designs for evaluations. Excellent annotated bibliography on experiments.

Suchman, Edward A., Evaluative Research: Principles and Practices in Public Service and Social Action Programs. New York, Russell Sage, 1967.

Considered to be a classic text in evaluation. Chapter IV, Categories of Evaluation, gives framework useful in developing evaluative questions for a proposed study.

Weiss, Carol, Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs, N.J., Prentice Hall, 1972.

An excellent introductory text for evaluation.

Weiss, Carol, ed. Evaluating Action Programs: Readings in Social Action and Education. Boston, Allyn and Bacon, 1972.

Collection of articles dealing with basic concepts and issues in evaluation, especially for social programs.

ANALYSIS

Dorfman, Robert, ed. Measuring Benefits of Governmental Investment. Washington, D.C., The Brookings Institution 1965.

A series of contributed papers concerned with the application of cost-benefit analysis. Wide ranging applications are discussed.

English, J. Morley, ed. Cost-Effectiveness. New York, John Wiley & Sons, Inc., 1968.

A series of papers covering the various aspects of cost-effectiveness analysis.

Fisher, Gene H., Cost Considerations in Systems Analysis. New York, American Elsevier Publishing Co., Inc., 1971.

A comprehensive treatment of all relevant cost considerations that should be treated in analysis. Chapters 1, 2, 3, and 6 are especially useful.

Hitch, Charles J. and Roland N. McKean. The Economics of Defense in the Nuclear Age. Cambridge, Harvard University Press, 1960.

Although the application is military, a classic discussion of cost-effectiveness is found in chapters 7, 9, 10, 11, and 12.

Quade, E. S., Analysis for Public Decisions New York, American Elsevier Publishing Co., Inc., 1975.

A highly important and readable book on all analytic aspects involved in formulating and implementing policy decisions.