Cost-Outcome and Cost-Effectiveness Analysis: Emerging Evaluation Techniques

by

JAMES E. SORENSEN

HUGH D. GROVE

As the United States gradually shifts from an industrial to a service economy, nonprofit organizations are providing an increasing proportion of the national income. In 1971, for example, nonprofit organizations such as government (all levels), health, education, nonprofit associations, and research agencies collectively accounted for over 20% of the total national income (Anthony and Herzlinger, 1975, p. 7). In 1973 these same nonprofit organizations had grown from five to eight times their 1950 income levels (Anthony and Herzlinger, 1975, p. 8). Since nonprofit organizations exist to provide a service rather than earn a profit and because services often are more difficult to measure than profits, choices among alternative courses of action and evaluation of accomplishments become problematic. The increase in services (especially in human services) by nonprofit organizations has created new pressures for more effective performance evaluation and may heighten the roles of both the accountant in gathering appropriate data and the auditor in attesting to such data. A manifestation of these pressures is reflected in the recent General Accounting Office (GAO) statement on performance audit standards which requires a review of efficiency and economy in the use of resources and an evaluation as to whether desired results are being effectively achieved (Controller of the United States, 1972).

The GAO standards specify that the performance audit shall include the following:

- The relevance and validity of the criteria used by the audited entity to judge effectiveness in achieving program results
- The appropriateness of the methods followed by the entity to judge effectiveness in achieving program results and reliability
- The accuracy of the results obtained (Controller of the United States, 1972).

Demands for nonprofit performance evaluation are not all from external sources. Internally, nonprofit service program managers are often faced with limited or decreasing resources and require evaluation of programs for planning and controlling program operations. These managers need a wide range of analyses ranging from

- Assessments of population or client needs, demands for services or incidence of a problem within a specified population or geographic area
- Formulation of organizational objectives and development of program structures to deliver services
- Identification of amounts of resources devoted to or budgeted for services rendered and how they are consumed; assessment of patterns of services and development of aids to monitor service programs
- Evaluation of outcome of specific programs or services on clients or beneficiaries and assessment of impact on larger social systems.

The American Accounting Association has sensed the emerging demand for nonprofit service programs. Several special committees have focused on:

- Not-for-Profit Organizations (1971 and 1974 *Accounting Review* Supplement)
- Nonfinancial Measures of Effectiveness (1971 *Accounting Review* Supplement)
- Measures of Effectiveness for Social Programs (1972 *Accounting Review* Supplement).

While the Accountant's role has not traditionally required active involvement in formulating models to evaluate nonfinancial outcomes and impacts, nor has the auditor been expected to attest to such information, the GAO standards, various nonprofit decision needs, and the AAA committee reports signal a new kind of involvement for the accountant and the auditor. A widespread literature focused upon profit-oriented organizations has left the accounting literature with few operational techniques which are responsive to nonprofit service performance evaluations. In addressing this problem, this paper is accordingly divided into three major parts:

- a brief discussion of problems identified with established performance evaluation techniques
- a general introduction to emerging cost-analytic methods and related outcome assessment issues, and
- a specific development and application of cost-outcome and cost-effectiveness techniques in a mental health example to illustrate a methodology for nonprofit performance evaluation

PROBLEMS WITH ESTABLISHED NONPROFIT
PERFORMANCE EVALUATION TECHNIQUES

Major established nonprofit performance evaluation techniques are social indicators, program-planning-budgeting systems (PPBS), and cost-benefit (C/B) analyses. None of these have been widely adopted for performance evaluation purposes in service programs.

Two major problems impede the application of social indicators. First, the method of data compilation may lead the user to question the meaning of the final social indicator statistic. For example, petty thefts over $100 comprise about 40% of the Federal Crime Index. With increasing affluence in society, many more personal property items over $100 are now owned and susceptible to theft. Therefore, if the Federal Crime Index goes up, it may be an indication of increasing affluence rather than just crime increases.

The second major problem with social indicators is that they are often either inadequate proxies (unrepresentative of the social impact being analyzed) or they are too general to aid in specific performance evaluation. Suicide, for example, is a poor indicator of mental health problems in general (Mechanic, 1975). Continuing the crime index example, an increase or decrease in the index by itself cannot evaluate the effectiveness of a single police program such as a program which uses boy scouts and girl scouts with binoculars on rooftops to spot potential crimes. Rather, a detailed index of crime occurrences in the specific areas at the specific surveillance times is needed. Only these types of specific social indicator statistics can help one attempting to assess the outcomes of various nonprofit and service programs.[1]

Two established techniques which attempt to relate costs to outcome assessment are PPBS and C/B analysis. Cost collection and allocation systems are well established and are the strong points of these two technqiues. The crucial weakness which has impeded general use of these methods is outcome measurement. PPBS has fallen into disuse because its required specification of objectives, such as nuclear parity, cannot be readily transformed into operational outcome quantities or statistics, i.e., how operational and meaningful is a maximum kill ratio from nuclear bombings? C/B analysis has not been generally applied because of its fixation with the improbable: transformation of essentially nonmonetary outcomes, such as lives saved or sicknesses cured, into monetary outcomes, such as the present value of all future earnings of the person whose life is saved. Consequently, both the PPBS and C/B analysis endeavors have often resulted in relatively meaningless and not too useful monetary outcome statistics.[2]

---

[1] A taxonomy with related discussion of social indicators is provided by Biderman (1966, pp. 81-153). An extensive strategy for improving social indicators is provided by Gross (1966, pp. 154-271).

[2] Extensive development of PPBS and cost-benefit approaches is provided in Hinricks and Taylor (1969).

COST-OUTCOME AND COST-EFFECTIVENESS METHODS FOR IMPROVING PERFORMANCE EVALUATION APPLICATIONS

Effective program management requires program evaluation. Program evaluation, an important part of any systematic approach to program management (Wholey et al., 1970; Zusman and Wurster, 1975):

- Assesses the effectiveness of an ongoing program in achieving its objectives
- Relies on the principles of research design to distinguish a program's effects from those of other forces working in a situation
- Aims at program improvement through modification of current operation.

Measuring program outcome is a vital linkage in assessing program effectiveness. Hargreaves and Attkisson (1974) suggest three reasons why program outcomes are important in program management.

> One of the most useful occasions for an outcomes study is when it can aid management and clinical staff in making a specific decision about program change. These are generally time-limited special projects. . . . A second reason to examine outcomes is to routinely detect relative strengths and weaknesses in a system of delivering services. Finally, program managers often need to demonstrate their program's overall effectiveness to funders and other groups who have a stake in the . . . (organization). For these latter purposes, routine monitoring and public accountability, some simple outcome assessment can be a useful part of an integrated management information system.

The brief review of well-known problems of existing nonprofit performance evaluation techniques poses a perplexing question which emphasizes the role of outcome measures: Can performance measures which *avoid* the

- generality or incompleteness of most social indicator outcome measures
- nonoperational PPBS outcome measures
- hindrance of monetary outcome measures sought by cost-benefit analysis

be developed while providing meaningful cost and outcome information to be used in program evaluation and program management?

One possible solution might be to relate monetary inputs (or costs) to *non*monetary outcomes for specific programs. Newer forms of effort-accomplishment measurement systems have recognized the problem of monetary output measurements and have shifted to linking monetary inputs to nonmonetary output measurements to analyze the benefits produced by specific programs in a cost analytic perspective.

General Evaluation Strategy: Cost-Analytic Techniques

When examining various ways to perform evaluation in service programs, three general strategies for evaluation emerge (Tripodi, Fellin, and Epstein, 1971):

- Monitoring techniques
- Social research techniques
- Cost-analytic techniques.

Monitoring techniques include accountability or administrative audits as well as time and motion studies. Heavy emphasis is usually given to resource input and process activities. Service delivery statistics currently exist and can be applied. Process measures deal with an examination of the service process and whether the applied process is appropriate. Historically in human service programs, for example, this has meant:

- Audit of records (e.g., case-by-case evaluation)
- Direct observation of staff/program activities
- Examination of client/patient conditions
- Testing of professional staff (with hypothetical cases)
- Comparisons of actual and desired (or normative) profiles or ratios.

Social research techniques include experiments or quasi-experiments, surveys (e.g., of client satisfaction concerning service processes and results), and case studies (Davis et al., 1973; Tripodi et al., 1971; Campbell and Stanley, 1969). The objectives and properties of various performance indicators which may be used as process and outcome measures are reviewed in Table 1.

Cost-analytic techniques include approaches where resource consumption is a common element of the analysis. These approaches range from cost accounting and cost-finding for programs, units of services, and episodes (e.g., spell of illness) to techniques which link resource consumption to nonmonetary outcome or benefit. These techniques are thus responsive to the performance evaluation strategy relating monetary costs to nonmonetary outcomes.

In developing cost-analytic techniques, service accountability measurement must be broken down into manageable measurement problems:

- Identification of specific service attributes (such as outpatient mental health care)
- Identification and measurement of specific service delivery networks such as the relation of input efforts (e.g., mental health care costs) and outcome rewards (e.g., mental health improvements).

Two types of cost-analytic techniques, cost-outcome and cost-effectiveness (Quade, 1967; Goldman, 1967; Levin, 1974; Fishman, 1974; Yates, 1975), are explored in the paper as a combined operational approach to nonprofit performance evaluation:

- Cost outcome defined as the programmatic resources consumed to achieve a change in a relative measure of performance (e.g., health symptoms, social or role performance)
- Cost-effectiveness defined as the comparison of cost-outcomes to identify the most beneficial outcome to cost of programs, modalities, or treatment techniques (e.g., economy versus intensive day care programs).

Table 1

Properties of Service Performance Indicators

Focus of Measure	Conceptual Content	Tells	Examples
Availability	• Amount and type of service provided	• What can be obtained	• List of services • Number of units of services rendered
Awareness	• Knowledge of user population and other agencies (especially referral agencies) of existence, range, and conditions for which services are appropriate	• Who knows what agency services	• % of user group aware of agency service • Patterns of referrals • Sources of individuals referred to other agencies
Accessibility	• Indicates if services can be obtained by appropriate groups	• Ease of reaching service	• Availability of city transportation • Sloping curbs to accommodate wheelchairs • Average travel time • Hours of service available by day of week
Extensiveness	• Compares quantity of services rendered to backdrop of problem	• "How much," but not "How well"	• Students enrolled in public schools • Clients in outpatient programs • Number and rate of catchment area residents' use of service
Appropriateness	• Correct type and amount of service rendered for presenting problem	• Proper use and quality of service	• Length of stay in hospital exceeded criterion • Mismatch between diagnosis and services received
Efficiency	• Compare resource inputs to specified process or output variable	• How much resource was used such as: • • "How many hours of staff time was used per served client • • "How much something cost per unit" • • "How much something cost in total"	• Chargeable professional staff hours • Cost per person riding RTD • Cost per client served • Cost per spell or episode of illness
Effectiveness	• Compares accomplishment to goals (or what was intended) • • qualitative • • comparative	• Characteristics • Duration • Content • Effect • Proportions served • Variance from standards, budgets, or goals • Ranking of options	• Number of patients cured • Number of trainees employed • Number of clients avoiding institutionization • Traffic accidents per thousand vehicle miles • Unduplicated count of clients served to total number of clients in a specified group • Ratio of actual outcome to planned outcome
Outcomes/Benefits/ Impacts	Identifies social or economic benefit	• Monetary effects • Nonmonetary effects	• Changes in earnings • Welfare payments • Decreases in noise levels or air pollution levels • Change in level of social functioning after treatment • Change in arrests for driving while intoxicated
Acceptability	• Assess match of service to client/citizen preferences	• User satisfaction with services and prediction to use services	• Number of complaints • Willingness to refer friends to service

Source: J. Richard Woy, "The New Community Mental Health Center Law (PL 94-63) and Its Implications for Program Evaluation" (mimeographed), Department of Health, Education, and Welfare, National Institute of Mental Health, Region II, New York, 1976.

Using cost-outcome information as building blocks, cost-effectiveness emerges as the last of five sequential steps:

- Identifying the objective to be achieved (or treatment goals to be achieved) for specific target groups (e.g., social functioning of neurotic depressives in a catchment area admitted to a Community Mental Health Center (CMHC)

- Specifying optional (or alternative) treatment programs to be used (e.g., random assignment to individual vs. group therapy)

- Determining the costs of each program, cost per unit of service, and amounts of service rendered (e.g., use of accrual accounting, operating statistics, cost-finding, and rate-setting)

- Assessing the effect or outcome of the program intervention on the target group (e.g., pre-intervention vs. post-intervention assessment)

- Combining cost and outcome information to present cost-outcome and cost-effectiveness analyses.

While development of appropriate costing for nonprofit services is a challenging task,[3] the accountant is more likely to understand the issues of cost accounting than the issues embedded in assessing outcomes. Because nonprofit performance evaluation currently implies *nonmonetary* outcome assessment, the discussion turns to major issues in outcome assessment.

Outcome Issues

In profit organizations decision making is focused on: "What is the impact on profits?" This single, all-encompassing, widely-understood measure organizes the consideration of multiple resource factors and facilitates comparisons among varying organizational units (Anthony and Herzlinger, 1975). No such comprehensive measure exists for nonprofit organizations; to date, there are few good ways of estimating whether additional inputs (or resources) will produce commensurate outputs. The central problem is the *Inadequacy of output measures*, and assessing outcomes is a first step in remedying this deficiency. The following aspects of outcome measurement are briefly analyzed: time patterns, multiple outcomes, effects among different populations, simple vs. complex evaluations, and research design questions.

TIME PATTERNS

Effectiveness in human service organizations should be measured at multiple points in time. A specific drug treatment program may be evaluated at several points in time with each time frame revealing varying levels of effectiveness. Choosing an appropriate time frame is important. Should it be three months? six months? five years? If the time period is too short, changes may not be observable; but if the period is too long important dynamics may be masked, missed, or identified too late for decision maker intervention. Similar

---

[3] For examples, see Sorensen and Phipps (1972) and Sorensen (1976). Although beyond the scope of this paper, there are significant unresolved cost issues in the nonprofit area, such as the major conceptual issue of whether expense or expenditure is the proper cost measure.

to financial reporting, periodic assessment (e.g., quarterly) with cumulative restatements (e.g., year-to-date) may be one workable compromise.

MULTIPLE OUTCOMES

All human service programs have multiple outcomes. A drug abuse program may decrease 'the use of heroin and increase the use of alcohol. So which outcome or aggregation of outcomes should be used? A theoretical response might be "include measures of all outcomes produced with each outcome weighted by decision maker judgments of value." A practical response might be "identify the most important outcomes using simple unweighted measures." While each approach is fraught with conceptual and methodological pitfalls, operational techniques should be stressed in additional research.

EFFECTS AMONG DIFFERENT POPULATIONS

Who receives the benefits of a program? The benefits of a program to develop early reading skills may not be uniform across all socio-economic levels. In this example, the program effects may be expected at two levels: (1) improved levels and lower variance in reading skills in the *target group* of children from lower socio-economic status and (2) improved levels and lower variance in reading skills among *all* children entering primary schools. Identification of goals for various recipients and corresponding outcome assessment is needed.

SIMPLE VS. COMPLEX EVALUATIONS

The distinction between the evaluation of complex programmatic approaches and simple program operations should be drawn clearly. A long-term study examining the cost-effectiveness of short-term hospitalization as an alternative to long-term hospitalization for schizophrenic patients requires an elaborate design using classical evaluation procedures (Glick, Hargreaves, and Goldfield, 1974). However, ascertaining if a given mental health center target group received services that achieved a treatment plan objective and at what cost during the past three months may be accomplished with less complicated approaches.

RESEARCH DESIGN ISSUES

Simple designs should not be shoddy, however. Controlled evaluations of accomplishment are required whether the evaluations are complex or simple. While this is not a treatise on research design, a few research guidelines seem worth repeating.

- The results of the program should be observable (Gruenberg, 1966; MacMahon, Pugh, and Hutchinson, 1961). This requirement has directed the front line evaluator in mental health and vocational rehabilitation, for example, to examine variables such as social functioning and problem manifestation.

- In any comparison of populations, treatment and comparison samples must be created by random or systematic allocation of individuals to groups (MacMahon et al., 1961; Fishman, 1974). While other procedures can be used (such as matched samples and analysis of covariance), complexity and possible contaminants arise quickly. Randomization avoids self-selection or biased selection, increases objectivity, and offsets variables beyond the control of the experimenter (Gilbert, Light, and Mosteller, 1975). Conducting such studies is far more preferable than just "fooling around with people" in service delivery systems that gather unreliable and misleading data on effectiveness.

- Analysis of improvements of a specific target group must be supported by comparison with similar groups which (may) have received different interventions. In this way, changes which are simply concurrent with treatment during a time span are controlled or randomized for the treatment and comparison group (Campbell and Stanley, 1963).

- Evaluation instruments must be assessed for reliability, especially for inter-rater agreement and for validity (Nunnally and Durham, 1975). If the inter-rater agreement as to the actual service delivery is poor, for example, reliability and the resulting information are bound to suffer. Validity should be examined at the level of local service delivery to assure that measures are reflecting what was intended to be measured.

- Observed differences (Gilbert et al., 1975) are often small. New programs usually create only modest effects and large "slam-bang" effects will be few.

In brief, while highly sophisticated research designs may still be required in program evaluation, these traditional, usually expensive and complex, approaches are not likely to be the ones used by programmatic practitioners and administrators. In the following mental health example, the cost-effectiveness research design is advocated as a pragmatic, operational solution.

OUTCOME ISSUES SUMMARY

Usable outcome evaluations should honor the constraints of simplicity, economy, and utility:

> Evaluations . . . are designed to fit within the operating budget of a service facility . . . used by persons with varying degrees of experience with research methods . . . (and) are to reinforce effective service and to signal the need for changes in delivery techniques or policies (Davis et al., 1973).

As the conceptual approach and role of outcome assessment becomes clearer, improvements in the quality of the output side of the cost-outcome approach can be expected to meet vital external and internal service accountability demands.

COST-OUTCOME AND COST-EFFECTIVENESS PERFORMANCE EVALUATION: MENTAL HEALTH EXAMPLE

The last major section of the paper develops and applies specific cost-outcome and cost-effectiveness evaluation techniques to a mental health example. An activity flow chart for human service performance evaluation (Table 2) suggests desired service activities to be delivered to target groups in three time frames. The following discussion will emphasize the emerging nonprofit performance evaluation activities in Table 2:

Table 2

Activity Flow Chart of Human Service Performance Evaluation:
Mental Health Example

Time Frame	Activity
One	1. Specification of objectives and corresponding programs for target groups which are identified by random assignment or matched comparisons
	2. Assessment of level of functioning at intake or beginning of time period (Table 3)
Two	3. Delivery of program services
	4. Update Management Information System files for service rendered and resources consumed
Three	5. Assessment of level of functioning after intervention (Table 3)
	6. Comparison with Time One level to determine change
	7. Compute costs for units of service
	8. Calculate costs per episode/time frame by using cost per unit of service received
	9. Plot cost-outcome on matrix (Figure 1)
	10. Plot cost-outcome program results on cost-effectiveness matrix (Table 4)
	11. Statistically assess relative cost-effectiveness of competing programs for decision purposes (Tables 5-8)

Source: Adapted from Fishman (1974) and Carter and Newman (1975).

- outcome assessment (activities 2 and 5)
- cost-outcome calculation (activity 9)
- cost-effectiveness calculation (activity 10) and corresponding assessment (activity 11)

Outcome Assessment

In this mental health care output scaling example, the output measures are the mental health professionals who determine the degree of client (patient) impairment or level of functioning at successive points in time. In one such ordinal (ranking) scale procedure, the patients are rated on four major criteria, personal self-care, social functioning, vocational and/or educational functioning, and evidence of emotional stability and/or stress tolerance. These four criteria are used to rate the patients on a nine-point ordinal scale with respect to their level of functioning.[4] Table 3 presents the scale.

The final determination of these ratings is based upon consensus among the professional evaluators. The reliability and validity of such ratings are

---

[4]There are numerous mental health scales. Weissman (1975) has a useful review of the global scales currently available; Waskow and Parloff (1975) have a comprehensive analysis of psychotherapy change measures. This specific nine-point scale has been used successfully in actual CMHC operations (Carter and Newman, 1975).

TABLE 3

Nine-Point Scale for Rating by Level of Functioning

Definitions of the Nine-Scale Levels of Functioning

With regard to the balance of the four criteria (personal self-care, social, vocational/ educational, and emotional symptoms/stress tolerance), the person's ability to function autonomously in the community is at "Level X", where "X" can assume one of the following nine (9) levels.

Level I: Dysfunctional in all four areas and is almost totally dependent upon others to provide a supportive protective environment.

Level II: Not working; ordinary social unit cannot or will not tolerate the person; can perform minimal self-care functions but cannot assume most responsibilities or tolerate social encounters beyond restrictive settings (e.g., in group, play, or occupational therapy).

Level III: Not working; probably living in ordinary social unit but not without considerable strain on the person and/or on others in the household. Symtoms are such that movement in the community should be restricted or supervised.

Level IV: Probably not working, although may be capable of working in a very protective setting; able to live in ordinary social unit and contributes to the daily routine of the household; can assume responsibility for all personal self-care matters; stressful social encounters ought to be avoided or carefully supervised.

Levels V through VIII describe persons who are usually functioning satisfactorily in the community, but for whom problems in one or more of the criteria areas force some degree of dependency on a form of therapeutic intervention.

Level V: Emotional stability and stress tolerance is sufficiently low that successful functioning in the social and/or vocational/educational realms is marginal. The person is barely able to hold on to either job or social unit, or both, without direct therapeutic intervention and a diminution of conflicts in either or both realms.

Level VI: The person's vocational and/or social areas of functioning are stabilized, but only because of direct therapeutic intervention. Symptom presence and severity is probably sufficient to be both noticeable and somewhat disconcerting to the client and/or to those around the client in daily contact.

Level VII: The person is functioning and coping well socially and vocationally (educa- tionally); however, symptom reoccurrences are sufficiently frequent to maintain a reliance on some sort of regular therapeutic intervention.

Level VIII: Functioning well in all areas with little evidence of distress present. However, a history of symptom reoccurrence suggests periodic correspondence with the center; e.g., a client may receive a medication check from a family physician who then contacts the center monthly, or the client returns for bi-monthly social activities.

Level IX: The person is functioning well in all areas and no contact with the MH/MR services is recommended.

Source: Carter and Newman (1975).

checked with special emphasis on the consensus concept. A number of professionals are asked to rate a number of cases and analysis of variance is performed on the results. The amount of variance caused by the different cases is isolated from the variance caused by the different raters. This case variance is determined as a percentage of total variance. This percentage is the reliability coefficient for the professional raters where a low percentage indicates low rater reliability and a high percentage indicates high rater reliability.[5] Validity is investigated through reference to external behavioral criteria from other rating procedures, i.e., a concurrent validity approach.[6] These types of reliability and validity verifications may be applied to help determine output aspects in the cost-effectiveness approach.

When such an estimate is made of a client's level of functioning at each clinical encounter, summaries can describe the changes which may have taken place in a specific group of clients over a given period of time. Adequate cost-finding procedures will enable accumulation of the service costs for the same group of clients during the period of time associated with the changes in level of functioning.

Cost-Outcome Matrix

The matrix of costs and outcomes in Figure 1 was drawn from an adult mental health program using a nine-point level of functioning scale. Clients, grouped by level of functioning, may be traced over time utilizing the matrix.[7]

The matrix mechanics are as follows:

- All clients are assessed at the beginning of the period and placed in the appropriate diagonal cell (e.g., a pre-rating of 3 would place a client in the 3,3 diagonal matrix cell)

- All clients are reassessed at the end of the period and placed in the appropriate horizontal cell (e.g., using the previous example, a post-rating of 4 would place the client in the 3,4 matrix cell, indicating improvement, a post-rating of 1 places the client in the 3,1 cell, indicating regression, a post-rating of 3 leaves the client in the 3,3 cell, indicating no change)

- in-summary:

  improvement is indicated by horizontal movement to right of diagonal

  regression is indicated by horizontal movement to left of diagonal

  no change is indicated by remaining in the diagonal

  vertical movement is not allowed.

Cost data are accumulated by:

- initially computing costs by units of services provided to individuals
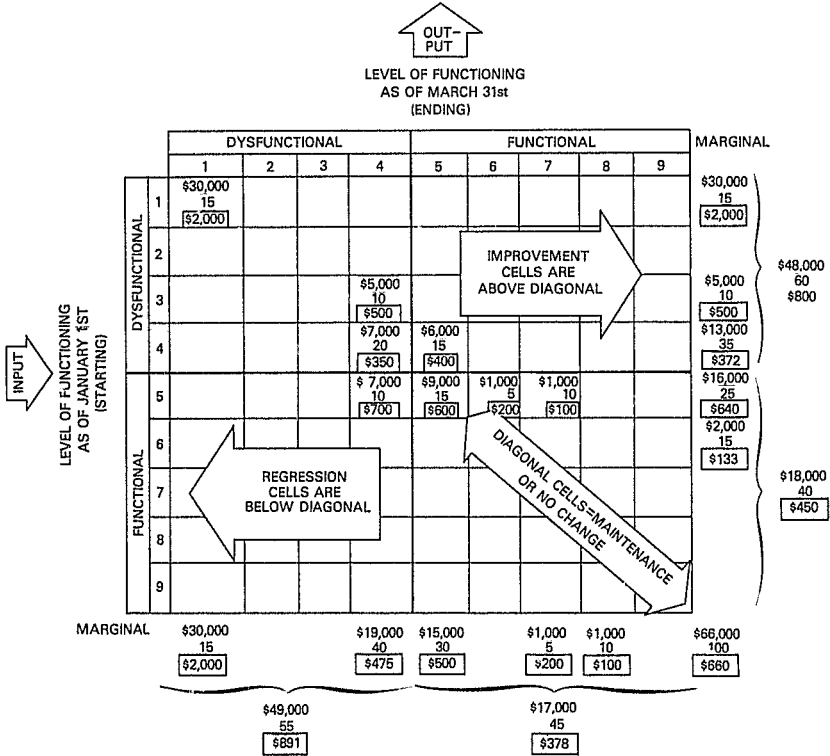
---

[5] Refer to Carter and Newman (1975, pp. 231-240) for a detailed seventeen-step procedure for computing this reliability coefficient based upon Cronbach et al. (1972); also see Nunnally and Durham (1975).

[6] Refer to the nine-step validity procedure in Carter and Newman (1975, pp. 250-253).

[7] This example uses the nine-point level of functioning scale developed earlier.

FIGURE 1

A Sample Cost-Outcome Matrix for Target Group of Mental Health Adults, Ages 45-64
January 1, 19xx to March 31, 19xx

OUT–
PUT

LEVEL OF FUNCTIONING
AS OF MARCH 31st
(ENDING)

INPUT

LEVEL OF FUNCTIONING
AS OF JANUARY 1ST
(STARTING)

		DYSFUNCTIONAL				FUNCTIONAL					MARGINAL
		1	2	3	4	5	6	7	8	9	
DYSFUNCTIONAL	1	$30,000 15 [$2,000]									$30,000 15 [$2,000]
	2						IMPROVEMENT CELLS ARE ABOVE DIAGONAL				$48,000 60 $800
	3				$5,000 10 [$500]						$5,000 10 [$500]
	4				$7,000 20 [$350]	$6,000 15 [$400]					$13,000 35 [$372]
FUNCTIONAL	5				$7,000 10 [$700]	$9,000 15 [$600]	$1,000 5 [$200]	$1,000 10 [$100]			$16,000 25 [$640]
	6										$2,000 15 [$133]
	7	REGRESSION CELLS ARE BELOW DIAGONAL					DIAGONAL CELLS=MAINTENANCE OR NO CHANGE				$18,000 40 [$450]
	8										
	9										
MARGINAL		$30,000 15 [$2,000]			$19,000 40 [$475]	$15,000 30 [$500]		$1,000 5 [$200]	$1,000 10 [$100]		$66,000 100 [$660]

$49,000
55
[$891]

$17,000
45
[$378]

Source: Adapted from Carter and Newman (1975).

- aggregating these unit costs by various programs for the target group at each functional level
- summing these program costs at each functional level.

Observe the kinds of questions that can be answered using this approach:

- How many individuals are in the target group (age 45-64) and what was the average cost of service during the quarter? (N=100; average cost = $660)
- How many individuals were dysfunctional at the end of the time period? (N=55) How many were functional? (N=45) Did these two groups consume the same amount of resources? (No, the dysfunctional consumed almost three times as much as the functional: $891 vs. $378 on the average and $49,000 vs. $17,000 in total)
- How many individuals were dysfunctional at the beginning of the quarter? (N=60) How many were functional? (N=40) How much in total was spent on each group? ($48,000 and $18,000 respectively) On the average? ($800 and $450 respectively)
- How many individuals were simply maintained over the quarter? (N = 15 + 20 + 15 = 50) At what average cost? ($30,000 + $7,000 + $9,000/50 = $920)
- How many improved? (N = 10 + 15 + 5 + 10 = 40) At what average cost? ($5,000 + $6,000 + $1,000 + $1,000/40 = $375)

To determine if these types of changes in the cost-outcome analysis are significant, chi-square analysis of frequency distributions can be applied. For comparability with other cost-outcome matrices, contingency coefficients can also be computed (Sorensen and Newman, 1976).

Using the foregoing illustration, the cost-outcome approach can be generalized to provide useful insights on questions such as (Carter and Newman, 1975; Burwell, Reiber, and Newman, 1975):

- Did varying target groups receive the services planned for specified levels of functioning?
- How much improvement did members of a target group achieve? At what cost?
- Where are the high and low cost services being used?
- Did the services delivered have the expected impact?
- What are the average values and are certain values out-of-line?

In this example, by combining target groups and programs, the clinician's assessment of client movement from one functional level to another, and costs into a cost-outcome matrix, a basic decision-making system has been formed for effective client and resource management and accountability.

Cost-Effectiveness Matrix

Cost-effectiveness builds on cost-outcome measurement procedures and strengthens the service measurement and evaluation procedure by correcting the design for a deficiency in the cost-outcome approach. The deficiency is that "there is not a way of scientifically documenting whether such change is actually caused by the treatment or is simply concurrent with it" (Fishman, 1975).

The correction in the research design is simple: use a control group. Potential intervening variables, such as history or selection bias, can be

randomized or controlled by using two groups of similar patients which are exposed to different types of treatments, such as token-economy or intensive therapeutic day care. Ideally, the treatment variable is isolated as the only difference between the two groups:

> At the level of cost-effectiveness analysis, the causal relationship between intervention and change over time can be investigated. For example, if two groups of comparable patients each receive different interventions and then are assessed with standard procedures at follow-up, any differences between the two groups at follow-up can be ascribed to the difference between the treatments (Fishman, 1975).

This research design is similar to one of Campbell and Stanley's (1963) "true experimental designs," the pretest-posttest control group design, modified to reflect two different sets of treatments:

Group 1: $R \, O \, X_1 \, O$

Group 2: $R \, O \, X_2 \, O$

Where:

> R = randomization of clients or patients into two groups or matched samples or covariance analysis (Glick, Hargreaves, and Goldfield, 1974)

> O = process of observation or measurement, for example, the functioning level rating at the beginning and the end of the period

> X_1 = exposure of first group to a specific type of therapeutic treatment, such as a special token-economy day care program

> X_2 = exposure of first group to a specific type of therapeutic treatment, such as an intensive day care program

The research design is pragmatic in acknowledging real world demands of human service clients (i.e., immediate service needs) which preclude utilizing the true control group where no treatment is given during the research period.

Table 4 summarizes the logical relationships and choice points for comparing treatments A and B. The decision criteria, based upon this measurement information, is to maximize output at minimum cost, and seven choices are self-explanatory. Because the binary choice in cells 1 and 9 is not obvious, the analysis could proceed along statistical lines by testing for the significance of outcome and cost differences and choosing the treatment outcome or cost with the highest level of significance using a nonorthogonal comparison test (such as Scheffe or Tukey).

The subsequent example statistically assesses the results of one cost-effectiveness matrix example, including the binary choice problem in cells one and nine. The nontechnical reader may desire to skip this section.

Cost-Effectiveness Assessment

Cost-effectiveness leads to ranking of optional choices, e.g., A is more cost-effective than B for decision purposes. A and B may be competing programs within a human service organization (HSO) or they may be the same program, but are in two different HSOs. The ranking process may be changed

TABLE 4
Cost-Effectiveness Matrix

		Cost of A Relative to B		
		A is Less Costly	A is As Costly	A is More Costly
Effectiveness of A Relative to B	A is Less Effective	*11* ?	*12* Choose B	*13* Choose B
	A is As Effective	*21* Choose A	*22* No Difference	*23* Choose B
	A is More Effective	*31* Choose A	*32* Choose A	*33* ?

TABLE 5
T-Tests Between Treatment Programs

Comparison	Program A Mean (N=30)	S.D.	Program B Mean (N=30)	S.D.	Level of Confidence t (Mean)	F (S.D.)
Social Functioning:						
Time 1	4.90	1.062	6.10	1.155	.99	67
Time 2	6.00	.91	7.40	0.814	.99	.72
Changes in Social Functioning:						
Time 1-Time 2	+1.1	0.84	+1.30	0.794	.64	.63
Costs:						
Time 1-Time 2	$500	50.86	$550	25.43	.99	.99

Source: Adapted from Fishman (1974).

depending on the type of measure or statistical analysis employed as well as the level of experimental control. To illustrate, assume the following data in Table 5 using social-functioning measures of outcome.

Using the cost-effectiveness matrix in Table 4 and the *changes*, or perhaps more desirably the percent of change, in social functioning in Table 5, Program A does not appear to be significantly different from Program B on outcome ($p \geq .64$). On the cost-effectiveness matrix, this outcome condition is identified in Row 2. Program A also appears to be significantly less expensive than Program B ($p \geq .99$). On the cost-effectiveness matrix, this cost condition is identified in Column 1. The intersection of Row 2 and Column 1 is Cell 21, which indicates that Program A is chosen for cost considerations.

If the measure is shifted to social functioning at *Time 2* (a cross-sectional approach without any time series considerations or where clients have been randomly assigned to treatment programs), the statistical comparison suggests Program B is superior on outcome ($p \geq .99$). Row 1 of the cost-effectiveness matrix is chosen. Since the costs did not change, Column 1 is chosen again. Row 1 and Column 1 intersect at Cell 11, which does not indicate a clear decision. Program A has lower outcomes but also is less costly. Since both cost and outcome are significantly different, the decision is still open.

A troublesome aspect is the significant difference on the *entry* levels of functioning between the two programs. If the differences between Program A and Program B at Time 1 and Program A and Program B at Time 2 are materially unequal, an *interaction* between the levels of functioning in the various programs and the time of measurement may be expected. Analysis of variance (with repeated measures) reveals significant effects on programs, social functioning, but *no* interaction between program and time of measurement. (Only program and social functioning F-ratios were significant in Table 6 Program A may appear to have lower initial levels of functioning than Program B, but this condition was not revealed in the analysis of variance interactions.)

A further statistical test, analysis of covariance of Time 2 social functioning, can investigate the apparent initial differences between the two treatment programs by adjusting for the initial condition at Time 1. The covariance analysis suggests significant differences between programs outcomes ($p \geq .95$), with Program B achieving higher outcomes even after adjusting for differing initial conditions (Tables 7 and 8). Program B still seems to be superior on outcomes and the choice between A and B is yet unresolved. However, if there were no significant differences in outcomes of Programs A and B from this covariance test, then Program A, which costs less, would have been chosen as the more effective method. In this situation, the choice would revert to Cell 21 in the cost-effectiveness matrix, which then indicates the selection of Program A. These illustrations demonstrate how different concepts of "outcomes" can lead to different decisions: in one case to choose A, another shows indecision, and a third case indicates the choice of B.

A different conceptualization of cost-effectiveness may change the ranking, however, assuming that the following measure is used (Hanson et al., 1974):

TABLE 6

Analysis of Variance With Repeated Measures on Social Functioning
Time 1 Versus Time 2

Source of Variation	SS	df	MS	F
	SOCIAL FUNCTIONING			
Between Subjects:				
Program	1.690	1	1.690	30.92*
Subjects within groups	3.170	58	0.055	
Within Subjects:				
Social Functioning	1.440	1	1.440	128.49*
Program x Functioning	0.010	1	0.010	0.89
Functioning x Subjects within groups	0.650	58	0.011	

* Level of Significance < .01.

TABLE 7

Analysis of Covariance of Time 2 Social Functioning*
(where initial condition = covariate)

Source	df	YY	SS (about)	df	MS
Treatment (Between)	1	29.40			
Error (Within)	18	43.20	23.18	57	0.40
Total	19	72.60	29.92	58	
Differences (for testing adjusted means)			6.74	1	6.74
$F(1, 57) = 16.554$					
$p < .05$					

* To use multiple measure requires analysis of covariance using repeated measures.

TABLE 8

Table of Adjusted Means and Standard Error
(covariate = initial social functioning)

Time	Treatment	Mean	Adjusted	Adjusted SE
Time 2	A	6.00	6.31	.125
	B	7.40	7.08	.125

$$\text{Cost-Effectiveness} \; = \; \frac{\substack{\text{Average Cost} \\ \text{per Episode}}}{\substack{\text{Average Change in} \\ \text{Global Impairment} \\ \text{per Episode}}} \; = \; \substack{\text{Cost per Unit of} \\ \text{Change in Global} \\ \text{Impairment}}$$

When applied to the illustrative data, the results are:

$$\text{Program A} = \frac{\$500}{1.10} = 454.54 \approx 455$$

$$\text{Program B} = \frac{\$550}{1.30} = 423.07 \approx 423$$

Observe that this analysis ranks Program B as more cost-effective (viz., lower cost per unit of change) although no test of statistical significance was applied.

By comparisons of two or more cost-outcome statements (Step 11 in Table 2), the effectiveness of costs for achievement of programs, modalities, or techniques can be assessed. The cost-effectiveness model emerges as the summary technique which incorporates the fundamental service account-ability information requirements.

SUMMARY

In summary, the cost-effectiveness approach uses the cost-outcome measure-ment procedures and severely restricts the possibility that the outcome was caused by nonservice factors occurring concurrently in the treatment period. This advancement in service measurement is achieved by identifying specific types of services or treatment costs with specific results or outcomes and statistically investigating the results of such programs. Decision-making infor-mation about service measurement and evaluation are improved accordingly. But the variety, complexity, and difficulties of interpretation of cost-effectiveness measures should not be underestimated. The approach offered here is neither easy nor obvious when extended in realistic programmatic decision making and accountability environments.

REFERENCES

Anthony, Robert N. and Herzlinger, Regina, *Management Control in Nonprofit Organi-zations* (Richard D. Irwin, Inc., 1975).

Biderman, Albert D., "Social Indicators and Goals," in Raymond A. Bauer, ed., *Social Indicators* (The M.I.T. Press, 1966), pp. 68-153.

Burwell, Barbara, Sue Reiber and Frederick Newman, "The Client-Oriented Cost/Outcome System of the Fayette County Mental Health/Mental Retardation," State-wide Conference on Cost Analysis Systems in MH/MR Programs (Hershey, Pa., June 25, 1975).

Campbell, Donald T. and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Rand McNally, 1963).

Carter, Dale and Frederick Newman, *A Client-Oriented System of Mental Health Service*

Delivery and Program Management: A Workbook and Guide, Department of Public Welfare, Commonwealth of Pennsylvania (Philadelphia, Pa., 1975).

Committee on Accounting for Not-For-Profit Organizations, *The Accounting Review,* Supplement to Vol. 46 (1971), pp. 81-164.

Committee on Measures of Effectiveness for Social Programs, *The Accounting Review,* Supplement to Vol. 47 (1972), pp. 337-398.

Committee on Nonfinancial Measures of Effectiveness, *The Accounting Review,* Supplement to Vol. 46 (1971), pp. 165-212.

Committee on Not-For-Profit Organizations, 1972-73, *The Accounting Review,* Supplement to Vol. 49 (1974), pp. 225-249.

Controller General of the United States, United States General Accounting Office, *Standards for Audit of Governmental Organizations, Programs Activities, and Functions* (1972).

Cronbach, L.J., G.C. Gleser, N. Nanda, and H. Rajartnam, *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles* (John Wiley and Sons, 1972).

Davis, Howard R., Carol Weiss, Karen Louis, and Janet Weiss, *Planning for Creative Change in Mental Health Services: Use of Program Evaluation,* U.S. Department of Health, Education, and Welfare, National Institute of Mental Health, Publication No. (HSM) 71-9057 (Rockville, Maryland, 1973).

Fishman, Daniel B., "Development of a Generic Cost-Effectiveness Methodology for Evaluating and Patient Services of a Community Mental Health Center," National Conference on Evaluation in Alcohol, Drug Abuse, and Mental Health Programs (Washington, D.C., April 3, 1974).

_____, "Suggested Guidelines for Utilizing 2% Evaluation Funds in the Emergent Community Mental Health Center Legislation: A Cost-Effectiveness Approach," National Council of Community Mental Health Centers (Washington, D.C., February, 1975).

Gilbert, John P., Richard J. Light, and Frederick Mosteller, "Assessing Social Innovations: An Empirical Base for Policy," in R. Zeckhauser et al., eds., *Benefit-Cost and Policy Analysis, 1974* (Aldine Publishing Company, 1975), pp. 3-65.

Glick, Ira D., William A. Hargreaves, and Michael D. Goldfield, "Short vs. Long Hospitalization," *Archives of General Psychiatry,* Vol. 30 (March 30, 1974), pp. 363-369.

Goldman, Thomas A., ed., *Cost-Effectiveness Analysis* (Frederick A. Praeger Publishers, 1967).

Gross, Bertram, "The State of the Nation: Social Systems Accounting," in Raymond A. Bauer, ed., *Social Indicators* (The M.I.T. Press, 1966), pp. 154-271.

Gruenburg, Ernest M., ed., "Evaluating the Effectiveness of Mental Health Services," *Milbank Memorial Fund Quarterly,* Vol. 44, No. 1 (1966), Part 2.

Hanson, Marshall R., Henry H. Horn, Jr., Katherine Ruland, Edward D. Shafer, and Melvin H. Voyles, *Final Report: Five County Cost-Effectiveness Study,* Department of Health, State of California (October, 1974).

Hargreaves, William A. and C. Clifford Attkisson, "Outcome Studies in Mental Health Program Evaluation," (University of California, 1973).

Hinricks, Harley H., and Graeme M. Taylor, eds., *Program Budgeting and Benefit-Cost Analysis* (Goodyear Publishing Co., 1969).

Levin, Henry M., "Cost-Effectiveness Analysis in Evaluation Research," Evaluation Consortium (Stanford University, 1974), pp. 8-10.

MacMahon, B., T.F. Pugh, and G.B. Hutchinson, "Principles in the Evaluation of Community Mental Health Programs," *American Journal of Public Health,* Vol. 51, No. 7 (1961), p. 963.

Mechanic, David, "Evaluation in Alcohol, Drug Abuse, and Mental Health Programs: Problems and Prospects," in J. Zusman and C. Wurster, eds., *Program Evaluation* (D. C. Heath and Co., 1975), pp. 3-25.

Nunnally, J.C. and Robert Durham, "Validity, Reliability and Special Problems in Evaluation Research," in E.L. Struening and M. Guttentag, eds., *Handbook of Evaluation Research*, Vol. 1 (Sage Publications, Inc., 1975), pp. 289-352.

Quade, Edward S., "Introduction and Overview," in T.A. Goldman, ed., *Cost-Effectiveness Analysis* (Praeger Publishers, 1967), pp. 1-16.

Sorensen, James E., "Uniform Cost Accounting in Long-Term Care," in J.H. Murnaghan and K.L. White, eds., *Long-Term Care Data*, Supplement to *Medical Care*, Vol. 13 (1976).

——————, and Frederick L. Newman, "Using Cost-Outcome Analysis to Develop Cost-Effectiveness for Program Evaluation," (Unpublished paper, University of Denver, 1976).

——————, and David W. Phipps, *Cost-Finding and Rate-Setting for Community Mental Health Centers*, U.S. Department of Health, Education and Welfare, National Institute of Mental Health, Publication No. (HSM) 72-9138 (Rockville, Maryland, 1972).

Tripodi, Tony, Phillip Fellin, and Irwin Epstein, *Social Program Evaluation* (F.E. Peacock, 1971).

Waskow, Irene E., and Morris B. Parloff, eds., *Psychotherapy Change Measures*, U.S. Department of Health, Education, and Welfare, Publication No. (ADM) 74-120), National Institute of Mental Health (Rockville, Maryland, 1975).

Weissman, Myrna M., "The Assessment of Social Adjustment," *Archives of General Psychiatry*, Vol. 32 (March, 1975), pp. 357-365.

Wholey, Joseph S., John W. Scanlon, Hugh G. Duffy, James S. Fukumoto, and Leona M. Vogt, *Federal Evaluation Policy: Analysis of Effects of Public Programs* (Washington, D.C.: The Urban Institute, 1970).

Woy, J. Richard, "The New Community Mental Health Center Law (PL94-63) and Its Implications for Program Evaluation" (mimeographed), Department of Health, Education, and Welfare, National Institute of Mental Health, Region II, New York, 1976.

Yates, Brian, "Towards Cost-Effectiveness Analysis in Applied Psychology: Compiling Cost-Effectiveness Indices for Behavioral Treatment," (Western Psychological Association, April, 1975).

Zusman, Jack, and Cecil R. Wurster, *Program Evaluation* (D. C. Heath and Co., 1975).

END