The Canadian Journal of Program Evaluation Vol. 14 No. 2 Pages 145–160
ISSN 0834-1516

A MULTI-DIMENSIONAL PROGRAM EVALUATION MODEL: CONSIDERATIONS OF COST-EFFECTIVENESS, EQUITY, QUALITY, AND SUSTAINABILITY

William A. Reinke
John Hopkins School of Hygiene and Public Health
Baltimore, Maryland

Abstract: Program evaluation has become increasingly multi-dimensional to include considerations of cost-effectiveness, equity, quality, satisfaction, and sustainability. The various aspects are interrelated but not necessarily mutually compatible. For example, services rendered cost-effectively to an easy-to-reach urban population are not likely to be distributed equitably. The consequent trade-offs cannot be assessed objectively unless measures of equity, quality, and sustainability are made as explicit as indicators of cost and coverage. This article presents an algorithm for integrating the evaluative considerations and offers suggestions for refining the system of measurement and analysis.

Résumé: L'évaluation des programmes est devenue de plus en plus multidimensionnelle, étant donné qu'elle doit comporter l'analyse des coûts et rendements aussi bien que des considerations d'equité, de qualité, de satisfaction, et de durabilité. Ces divers aspects, bien qu'étroitement liés, ne sont pas necessairement compatibles. Ainsi, des services procurés de façon économique pour une population urbaine facilement accessible ne se prêteront probablement pas à une distribution equitable. L'évaluation des compromis necessaires demande l'identification explicite des exigences d'equité, de qualité, et de durabilité en tant qu'indicateurs du coût et de la distribution des services. Cet article donne un algorithme pour l'intégration des considérations évaluatives et offre des suggestions pour affiner le système de mesure et d'analyse.

Program evaluation is becoming increasingly multi-dimensional (Edwards & Newman, 1986; Kaplan & Norton, 1992). There was a time when effectiveness concerns dominated. Then, economic concerns over the efficient use of scarce resources led to cost-effectiveness analyses (CEA) (Reynolds & Gaspari, 1985). In

recognition of the special importance of intangible benefits in the health sector, economic assessment was broadened to form cost-utility analysis (CUA) (Torrance, 1986). The notion of health for all then focused attention on the distribution of benefits, as well as overall levels of achievement; thus equity became a prime concern (Montoya-Aguilar & Marin-Lira, 1986; Reinke, 1995; Ubel, DeKay, Baron, & Asch, 1996). More recently we have come to appreciate the labour-intensive nature of health care and its effect on recurrent costs (Prescott & DeFerranti, 1985). One-time externally supported development costs to improve service coverage can lead to a substantial ongoing recurrent cost commitment to maintain the capability that has been developed. Thus, issues of program sustainability and affordability have come to the forefront of attention (Bossert, 1990). This interest has been heightened by growing concern over issues of quality, which can be costly (Berwick, Godfrey, & Roessner, 1990; Reinke, 1995; Walton, 1986). The question is raised whether improved quality, achieved at a cost, carries with it a level of client satisfaction that leads to willingness to pay for the added cost. Thus, issues of cost recovery enter the equation (Gilson, 1997).

These multiple desires are not necessarily mutually compatible. For example, the most cost-effective strategy in a given situation could focus attention on a readily accessible urban population, but this approach would be unlikely to satisfy even minimum criteria of equity. Sound and reliable appraisals of the trade-off between cost-effectiveness and equity demand explicit measures of each evaluative indicator being compared. While cost-effectiveness measures have been well-defined, the meanings of equity, quality, and sustainability are still somewhat vague and need to be refined. The dual purposes of this article are to present a systematic, unified approach for integrating the multiple dimensions of evaluation and then to propose operational measures for assessing the separate components and interrelated trade-offs.

INTEGRATION OF EVALUATIVE MEASURES

To facilitate integration of the various aspects of program performance, the following six-step analytical framework is recommended.

1. Define improvements in health outcomes sought and the activities and services that should be initiated, maintained, extended, and/or expanded in order to achieve the outcomes. Projects are not launched in a vacuum; they are initiated under circumstances that are more or less satisfactory in

some respects and unacceptable in others. Decisions are to be made, therefore, regarding new activities to be undertaken, current endeavours that are to be maintained or strengthened, and present efforts to be abandoned.

2. Determine, or at least form a judgment about, incremental levels of resources to be made available for the coming project period. Desired objectives may seem to be unachievable because of resource limitations. On the other hand, if it can be shown what inputs are required to achieve desired results, the necessary resources might be made available. Thus, the planner must move back and forth between objectives specification and resources determination until a realistic relationship between the two is established. Initial "ball-park" budget estimates can help to focus attention on reasonable magnitudes of effort to be contemplated for project fine-tuning.

3. Determine the cost of prospective new services and the marginal cost of service improvements, along with expected utilization levels for the new and improved services package. Total cost depends upon volume of use, of course, and utilization is affected by cost. Despite this dilemma, rough appraisal of market conditions in terms of crude demand curve estimates is usually feasible. To illustrate, demand is low in many clinics because of uncertain availability of drugs. Because staff are present but underemployed, quality improvements in the form of dependable supplies of drugs could be made at the small incremental cost of the drugs. Recognizing drug shortage as the principal bottleneck and noting patient willingness to pay for drugs, substantial increases in demand could be anticipated at a small increase in total cost and an actual reduction in unit costs.

4. Make estimates of how the anticipated costs might be borne. This requires establishment of a clear, quantitative link between quality and cost, association of quality with patient satisfaction, and implications regarding willingness and ability to pay. On the basis of cost-recovery estimates thus derived, the consequent budget needs of government can be determined and "affordability" established. To the extent that budgetary requirements are excessive, the planner must return to Steps 1 and 2 and modify projections arising from them accordingly.

5. Within the framework of realistic budget limitations, identify the most cost-effective mix of activities and services. Having established limits on what is affordable, i.e., having fixed the amount of input anticipated, the mix of outputs that is expected to maximize outcome is determined.

6. Ascertain trade-offs between cost-effectiveness and equity considerations in order to derive an acceptable balance between the two. Explicit measures of equity must be introduced to determine the extent of incompatibility, if any, between cost-effectiveness and equity aims. While the optimal trade-off cannot be mathematically prescribed, quantification of both dimensions permits sound subjective judgments to be made.

The stepwise approach we have suggested is deceptively simple. In practice, one cannot move directly from Step 1 through to Step 6 in linear fashion. Planning requires dynamic movement back and forth among the various steps until the results at each are mutually compatible. Furthermore, the present state of the art is such that quantification and integration of the entire evaluative process is not yet entirely straightforward. Considerable judgment remains to be exercised in bringing together the various dimensions of evaluation that we have cited. Nevertheless, development of clearer, more explicit measures of cost-effectiveness, equity, quality, satisfaction affordability, and sustainability outlined in the following sections should help to unite the various components into a single, comprehensive strategy for evaluation.

REVIEW OF CEA AND CUA

Cost-effectiveness measures are quite well defined, so that here we need only review the main features and call attention to one possible ambiguity that is of importance for the present purposes. Considering that evaluations relate inputs (IP) to outputs (OP) and outcomes (OC), the ambiguity arises over whether to focus on the cost per unit of output or per unit of outcome. To illustrate, are we concerned about the cost per child immunized against measles or the cost per measles death averted? The answers will differ, of course, according to program emphasis, but in any case a general clarification of terms would be helpful. Let us define OP/IP as efficiency and introduce the notion of impact as OC/OP. Then cost-effectiveness is defined as OC/IP, and we have

Cost-effectiveness = Efficiency x Impact
OC/IP = OP/IP x OC/OP

Distinguishing the two components of cost-effectiveness can be useful in making equity comparisons. It might be expected that an underserved population would be hard to reach (physically or culturally) and therefore command a high cost per unit of service. Thus, it is inefficient to serve such a population (OP/IP low). On the other hand, service provided to members of this especially needy group is likely to have considerable impact (OC/OP high). Whether the product of efficiency times impact for the underserved is fully cost-effective depends upon whether inefficiency is outweighed by impact. If not, a trade-off between cost-effectiveness and equity is required.

CUA extends CEA thinking to recognize the quality of life as more than its economic value (Torrance, 1986). Program benefits are typically measured as quality adjusted life years (QALY) gained. The utility associated with a specified state of well-being is judged by standard gamble and other techniques of decision analysis that have been developed and successfully applied to circumstances of health. The QALY concept has been modified to produce disability adjusted life years (DALY) in connection with national comparisons of the burden of disease (World Bank, 1993). An important feature of DALY is that a year of life (of whatever quality) during childhood or among the elderly is considered less valuable than a year of life at, say, age 25. The rationale is that, because persons in their 20s and 30s usually have relatively high family and social responsibilities, disability at that age is especially burdensome to society.

Both CEA and CUA methods include the discounting feature which assigns less value to the future than the present. This feature, combined with the DALY assumption that child years are of relatively little current value, leads to the assignment of less priority to efforts to reduce child mortality, since the future years of productive life lost are discounted, and near-term years are not worth much in the first place. Lives saved at older ages are given relatively little value for somewhat different reasons. These attitudes toward the young and the elderly seem contrary to societal views and raise the question of whether evaluative measures should reflect societal values, or whether they should be more “objectively” derived and then society educated to accept the professional judgments.

EQUITY CONSIDERATIONS

Reference to inequity suggests that there is a maldistribution of resources between a relatively advantaged (ADV) segment of the population and a disadvantaged (DIS) group. The precise nature of the maldistribution is seldom made explicit, however (Montoya-Aguilar & Marin-Lira, 1986; Reinke, 1994; Ubel et al., 1996). Is the problem an inequity in the distribution of the resources (budgets) themselves, in the benefits derived from the resources (outcomes), or based on some other criteria? Suppose that the infant mortality rate in urban areas is 60, while the rural rate is 80. Would a strategy of resources allocation that produces a reduction of 20% in both areas be considered equitable, even though the two rates remain unequal at 48 and 64? The dilemma is compounded, as indicated earlier, when the cost of producing a specified effect in rural areas exceeds that of producing a similar benefit in urban areas.

To make explicit the dilemma and its possible resolution, consider the following hypothetical scenario (Table 1). Members of a target group of 12,000 advantaged children can be immunized against measles at a cost of $2 per capita. If left unvaccinated, all would contract measles and 3% would die. The corresponding immunization cost for 8,000 disadvantaged children is $4 each, and the mortality rate among the unvaccinated is 4%. How should a budget of $20,000 be allocated?

The most cost-effective approach would be to vaccinate 10,000 in the ADV group (assuming constant marginal cost for the sake of the illustration) and none of the DIS children. Three hundred lives would be saved as a result, but this is not likely to be acceptable for equity reasons. If the decision were made instead to equalize the budget per capita in the two groups, the ADV group would get $12,000 to vaccinate 6,000 children (50%) and avert 180 deaths. The DIS group would also get $1 per capita but because of higher costs only 2,000 children (25%) could be covered, and 80 deaths would be averted. This arrangement likewise seems inequitable and, in addition, the 260 deaths averted represent a reduction of 40 deaths, or 13% below the most cost-effective situation. The strategy would probably be unacceptable on any grounds.

Calculations based on definition of equity in outcome, rather than input, terms show that $10,000 should be assigned to each group. As a result of vaccinating 5,000 ADV children, 150 of their lives would be saved; in the DIS group, 2,500 children would be immu-

Table 1
Considerations of Equity

Basic Information on Possible Strategies

	ADV	*DIS*
Target Population (000)	12	8
Cost per Immunization ($)	2	4
Death Rate per Thousand	30	40

Strategy		(IP) Budget ($000)	(OP) Immun. (000)	(OC) Deaths Averted	%C-E
	ADV	20	10.0	300	
Cost-Effective	DIS	0	0	0	
	TOTAL	20	10.0	300	100
	ADV	12		6.0	180
Equal IP/Capita	DIS	8	2.0	80	
	TOTAL	20	8.0	260	87
	ADV	10	5.0	150	
Equal OC/Capita	DIS	10	2.5	100	
	TOTAL	20	7.5	250	83
Equal Residual	ADV	6	3.0	90	
Death Rate	DIS	14	3.5	140	
	TOTAL	20	6.5	230	77

Comparative Indicators

Strategy		IP/Capita ($)	% Covered	Deaths Avert/000	Residual Rate/000
	ADV	1.67	83	25.0	5.0
Cost-Effective	DIS	0	0	0	40.0
	TOTAL	1.00	50	15.0	19.0
	ADV	**1.00**	50	15.0	15.0
Equal IP/Capita	DIS	**1.00**	25	10.0	30.0
	TOTAL	**1.00**	40	13.0	21.0
	ADV	0.83	42	**12.5**	17.5
Equal OC/Capita	DIS	1.25	31	**12.5**	27.5
	TOTAL	1.00	37	**12.5**	21.5
Equal Residual	ADV	0.50	25	7.5	**22.5**
Death Rate	DIS	1.75	44	17.5	**22.5**
	TOTAL	1.00	32	11.5	**22.5**

nized, and 100 lives would be saved. The total of 250 lives saved is 17% below optimum in cost-effectiveness terms. Moreover, the residual of 27.5 deaths per 1,000 DIS children (220 deaths among 5,500 unvaccinated) is still substantially higher than the ADV rate of 17.5 per thousand.

In principle, the most truly equitable strategy would presumably assign first priority to elimination of excess mortality among the DIS. Then, having equalized death rates in the two groups, any remaining budget would be devoted to an equal further reduction in each group. Calculations in this direction indicate that ADV should receive $6,000 and DIS should be allocated $14,000. The consequent 3,000 ADV immunizations would save 90 lives, and the 3,500 DIS immunizations would save an additional 140 lives. The total of 230 is 23% below the most cost-effective level. Moreover, the assignment of 70% of the total budget to a segment of the population that represents only 40% of the total and probably has little influence may not be acceptable for political reasons.

While the illustrative case is artificial and oversimplified, it points out the importance of establishing a reasonable, explicit definition of equity and then determining how it impacts on cost-effectiveness considerations and what trade-offs must be contemplated.

QUALITY AND SATISFACTION

Considerations of service quality are receiving increasing independent attention, but realistically they should be incorporated into cost-effectiveness analysis, since quality improvement presumably enhances effectiveness and changes the cost structure. As with equity, we need to be clear what we mean by quality improvement in a given situation (Berwick et al., 1990; Reinke, 1995; Walton, 1986). In particular, two dimensions of the discussion must be recognized. First, are we interested in a higher level of performance, or do we seek greater dependability at existing levels? An MRI may produce more definitive diagnoses than x-rays and lead to more effective treatment but at higher cost. In contrast, a new screening test with greater specificity than the current one may not affect the outcome of treatment but is more dependable in that it reduces the number of false positive cases and the cost of their unnecessary treatment.

The second dimension of concern is the distinction between technical quality and client satisfaction. A certain orally administered

medication of professionally proven effectiveness may not be appreciated by patients who have come to expect relief only from injections. The definition of quality should specify either a provider or client perspective, or ideally combine the two.

Interest in industrial quality assurance is driven in large measure by possible cost savings from reduction in the number of product defects and consequent customer dissatisfaction and cost of rework. Health services quality improvement, on the other hand, more often involves added costs for elevated levels of performance that are not readily apparent to patients who therefore are reluctant to share the burden of higher costs; they may be more interested in the length of the wait for service. Thus, industrial experience with quality improvement should not be transferred uncritically to the health sector without systematic analysis of its relevance.

To help conceptualize the introduction of quality dimensions into health services evaluation, we refer to figures 1–3. Figure 1 presents a typical provider-oriented input-process-output-outcome evaluation model with quality improvement explicitly added to the linked relationship. Available resources are organized to provide a service capability that is partially or fully utilized to produce an impact on health. Quality improvement requires a different mix and amount of resources that are, or are not, provided to alter service capacity, utilization, and impact.

Figure 2 offers the beneficiary's perspective of the system. Service capabilities offered by the provider (Figure 1) are acted upon according to willingness and ability to pay and generate overt demand for care that yields a degree of satisfaction. From the client perspective quality improvement efforts potentially modify the level of satisfaction derived and consequent willingness to pay for services. Willingness and ability to pay are distinguished because willingness is directly related to level of satisfaction envisioned, whereas economic and other factors intervene to affect ability to pay and, therefore, effective demand.

Improved technical quality is expected to enhance client satisfaction, but the two are by no means the same. The client is likely to be interested, for example, in the way services are delivered (location, timing, etc.), as well as in the health benefits derived from them. Hence quality is depicted separately from satisfaction in Figure 3, which is a composite of figures 1–2. In bringing the two prior figures together in the composite, certain modifications are appropriate. Available resources

Figure 1. Provider-Oriented Evaluation Model

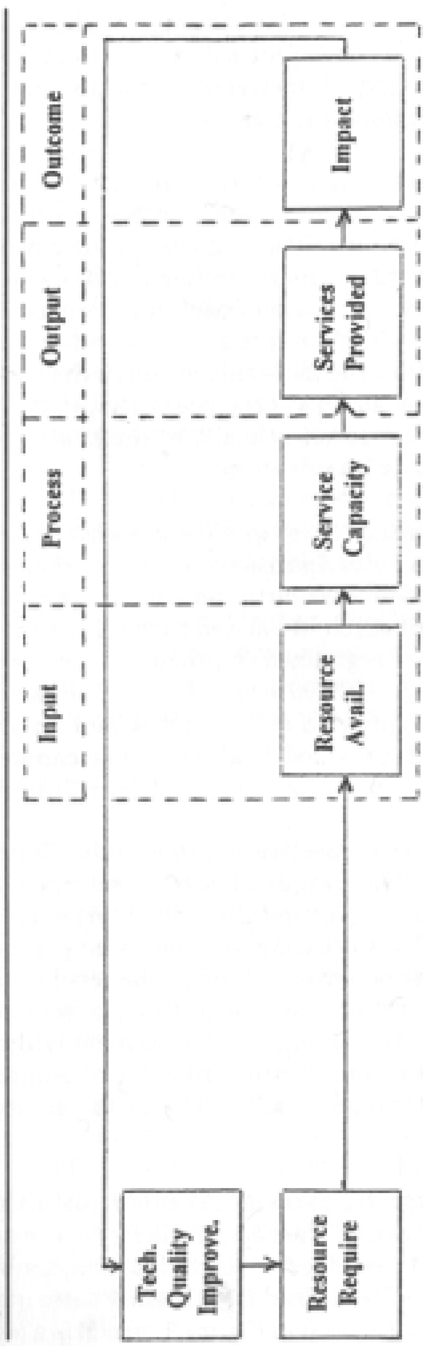
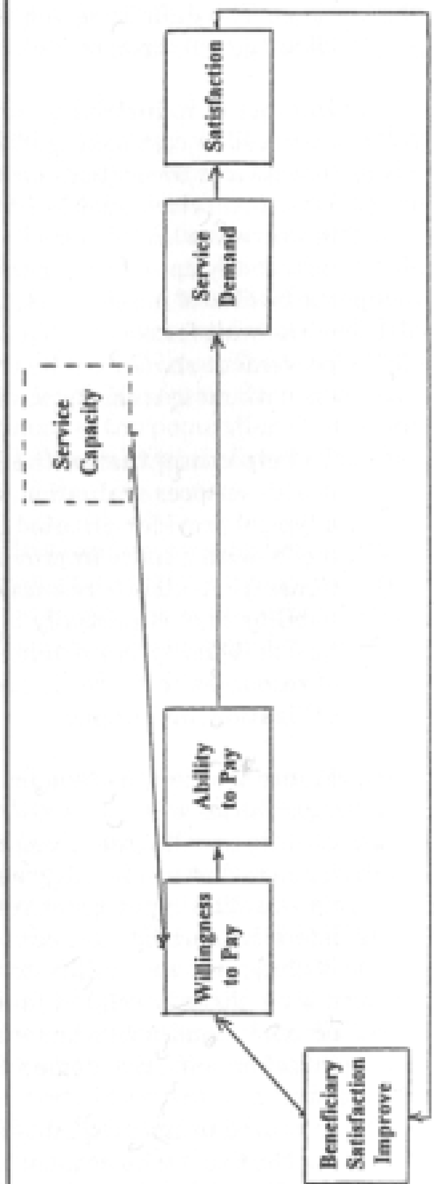


Figure 2. Beneficiary Perspective on Evaluation



are broken down into two categories to reflect cost-recovery considerations. Some costs are borne by public agencies; others are in the form of user fees based on willingness and ability to pay. It may also be useful to distinguish between services sought (demanded) and those actually offered and provided. The latter produce a health impact and degree of patient satisfaction. Though these terms have somewhat different meanings, as we have suggested, they are not mutually exclusive; it is to be hoped that providers have concern for patient satisfaction, and patients seek health benefits. Therefore, the two concepts are shown as overlapping in Figure 3.

We also note in the figure that evaluative indicators of input, output, and outcome, as well as quality and satisfaction, must be formulated in a manner that permits comparative appraisal of cost-effectiveness, equity, and continued sustainability.

Important relationships indicated in Figure 3 between quality/satisfaction, utilization, cost, and cost recovery are made more direct in Figure 4. To begin with (baseline), we suppose that BC(0) services are utilized at unit cost AB(0) to generate total costs that are of a magnitude depicted by the area of rectangle ABCD(0). Fees of BE(0) are charged, meaning that a portion BCFE(0) of total costs is recovered from users; the remainder, AEFD(0), is borne by the provider system, i.e., government.

A highly desirable form of quality improvement results in more dependable services that are provided at reduced unit cost AB(1) and are so popular with the public that utilization increases to BC(1), despite the fact that user fees are not reduced to reflect the lower costs, and may even be increased to BE(1). The new total cost rectangle ABCD(1) is smaller than ABCD(0), shown in dotted lines for comparison. Moreover, the cost-recovery portion BCFE(1) has increased, so that the costs borne by government, AEFD(1), are substantially lower.

Quality improvement does not always produce such uniformly favourable results, of course. We consider two among the several possibilities. In the first case, quality improvement adds to unit costs, but the improvements are so well received that users are willing and able to pay more than enough to cover the added costs and still increase demand. Although the increase in demand and cost per unit of service raise the level of total costs ABCD(2), the cost-recovery portion BCFE(2) is so large that the government portion AEFD(2) is lower than at baseline.

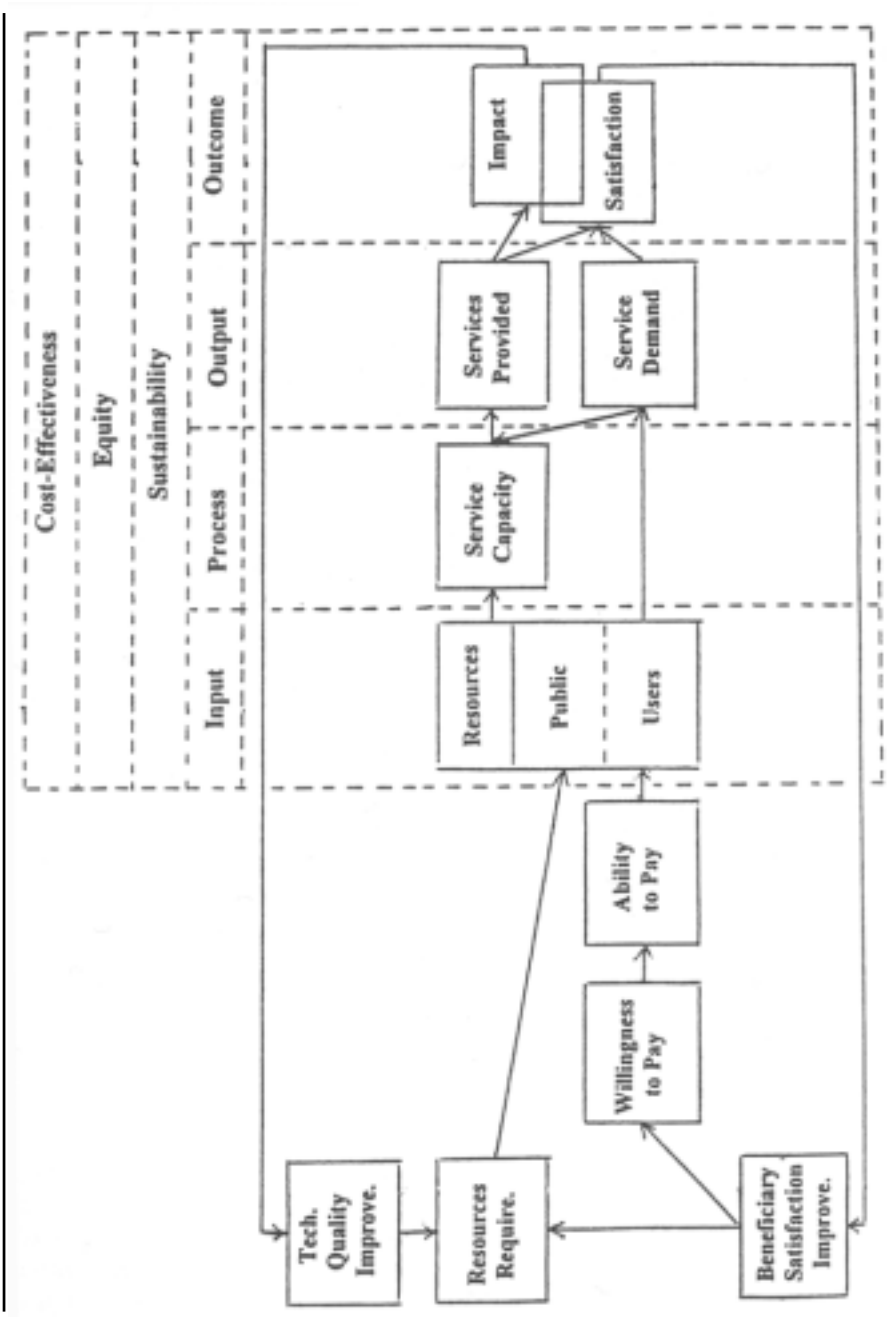
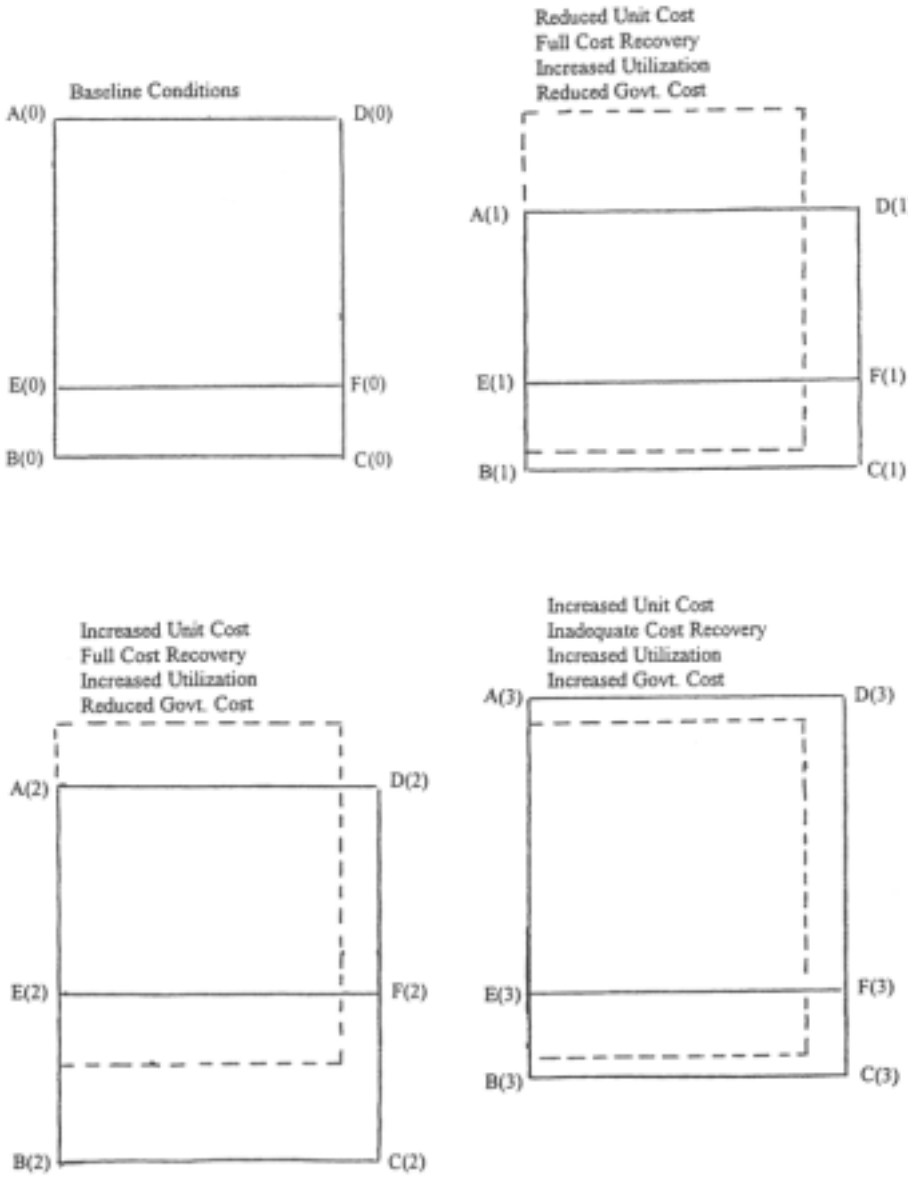


Figure 3. Composite Evaluation Model

Figure 4. Quality-Utilization-Cost Relationships



Although we have shown an increase in unit cost due to quality improvement, this need not be the case. When utilization increases, fixed costs are allocated over more services, so that the amount charged to each unit is reduced. This reduction could conceivably more than offset the cost of improved quality. If this is the case, the budgetary demands on government will be lower than shown in the diagram.

In the final scenario, costs increase due to quality improvement, and although utilization increases in response to the improvement, consumers are unwilling or unable to absorb the added costs. Thus, all of the increase in unit costs must be borne by government, and the net cost per unit to government, AE(3), is borne over a larger number of units, BC(3). Because of increased utilization, the amount of funds recovered from user fees, BCFE(3), is somewhat larger than BCFE(0), but total costs ABCD(3) are much higher, and the portion borne by government is therefore increased to AEFD(3).

SUSTAINABILITY

Suppose that a reasonable balance has been established between cost-effectiveness and equity in the provision of quality health care that is sufficiently satisfying to users that they share enough of the cost to make the program affordable to the government. Can this level of success be indefinitely maintained? Effectiveness, quality, and affordability undoubtedly contribute to sustainability, but the matter deserves closer scrutiny than we have given it thus far. Once again we encounter the difficulty of clear definition. Which aspects of the program *should* be continued?

Even programs that are considered to be generally effective usually have some dead wood in them that needs to be pruned. One primary care program, for example, trained medical assistants at the health-centre level, village-level volunteers, and health communicators, several of whom were to function within each village. The first two categories were used effectively, but the role of the communicator in relation to the volunteer was unclear. The latter workers were therefore felt to be superfluous and not to be retained. A good monitoring system was needed to identify and distinguish useful program components from unnecessary features.

Other aspects of sustainability likewise require definition. Is it enough simply to continue essential inputs (e.g., training programs)? Must the inputs continue to be productive (services utilization main-

tained)? Should the program piloted in one district be adopted by others? Should the program be self-supported by the community, or is it acceptable for government support to replace donor funding? What if support from one donor is replaced by aid from another donor? In short, future desires for continuity should be made explicit so that the evaluation of program status can produce unambiguous results.

Factors additional to effectiveness, quality, and affordability are likely to contribute to sustainability and should be monitored (Bossert, 1990; Prescott & DeFerranti, 1985). For example, there is evidence that a project's institutional base is important. If the project is established within the existing government administrative structure and commands the active participation of government officials, the effort is more likely to be maintained after project termination than if it is set up as a separate entity, even though the separate entity would have the advantage of closer scrutiny and control during the project period. Likewise, active community support and participation from the outset can be expected to contribute to long-term sustainability.

In short, concrete desires for sustainability should be defined in advance insofar as possible, factors expected to contribute to the fulfillment of those desires should be identified, and appropriate measurements made in the course of continuous monitoring.

CONCLUSION

We have noted that meanings attached to program evaluation continue to expand, so that aspects of cost-effectiveness, equity, quality, satisfaction, affordability, and sustainability are now included. We have cited the need for greater clarity in the definition of each of these terms and contributing factors, so that comparative analyses can be carried out objectively. The importance of such comparative analyses has been illustrated with respect to possible trade-offs between cost-effectivenss and equity. Comparisons involving patient satisfaction and program sustainability are more problematic because of difficulties in measurement, but they are no less needed.

REFERENCES

Berwick, D., Godfrey, A.B., & Roessner, J. (1990). *Curing health care: New strategies for quality improvement*. San Francisco: Jossey-Bass.

Bossert, T.J. (1990). Can they get along without us? Sustainability of donor-supported health projects in Central America and Africa. *Social Science and Medicine, 30*(9), 1015–1023.

Edwards, W., & Newman, J.R. (1986). Multiattribute evaluation. In H.R. Arkes & K.R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 13–37). Cambridge: Cambridge University Press.

Gilson, L. (1997). The lessons of user fee experience in Africa. *Health Policy and Planning, 12*(4), 273–285.

Kaplan, R.S., & Norton, D.P. (1992, January-February). Putting the balanced scorecard to work. *Harvard Business Review*, 71–79.

Montoya-Aguilar, C., & Marin-Lira, M.A. (1986). International equity in coverage of primary health care: Examples from developing countries. *World Health Statistics Quarterly, 39*(4), 336–344.

Prescott, N., & DeFerranti, D. (1985). The analysis and assessment of health programs. *Social Science and Medicine, 20*(12), 1235–1240.

Reinke, W.A. (1994, April). Program evaluation: Considerations of effectiveness, efficiency and equity. *Journal of Family and Community Medicine, 1*, 61–71.

Reinke, W.A. (1995). Quality management in managed care. *Health Care Management: State of the Art Reviews, 2*(1), 79–88.

Reynolds, J., & Gaspari, K.C. (1985). *Cost-effectiveness analysis*. Chevy Chase, MD: Center for Human Services.

Torrance, G.M. (1986). Measurement of health state utilities for economic appraisal: A review. *Journal of Health Economics, 5*, 1–30.

Ubel, P.A., DeKay, M.L., Baron, J., & Asch, D.A. (1996, May 2). Cost-effectiveness analysis in a setting of budget constraints: Is it equitable? *New England Journal of Medicine, 334*(18), 1174–1177.

Walton, M. (1986). *The Deming management model*. New York: GP Putnam.

World Bank. (1993). *World development report, 1993: Investing in health*. New York: Oxford University Press.