CHAPTER 11

EVALUATING HEALTH CARE PROGRAMS: EFFICIENCY, EFFECTIVENESS, AND COST

PUBLIC POLICY AND PRIVATE INTERESTS

The discussion thus far has emphasized that one cannot describe, analyse, or hope to understand and predict the behaviour of either users or producers of health care in isolation from the other. But we have also seen the public sector, the state, intervening in a number of ways to influence the behaviour of both. Either directly or through delegated authority, the state regulates access to health care markets by buyers and sellers. It prohibits, regulates, or mandates the sale/use of particular commodities or services, determines who shall and shall not be permitted to offer or provide them, and regulates or empowers self-regulation of the conduct of providers. It provides information to assist decision-making by private transactors through public information campaigns, certification of provider qualifications, product labelling, and providing technical information to providers.[1] It taxes or subsidizes particular forms of consumption or investment in particular types of capacity -- grants for construction, training, or research. Some forms of services it provides directly, superseding the market -- public health programs everywhere, Canadian hospital and medical insurance, the Saskatchewan Dental Service, the United States Veteran's Administration hospitals, the British National Health Service. No other sphere of economic activity, with the possible exception of defense, is so extensively subject to public intervention, in so many different ways.

In the previous sections, however, we have encountered these interventions as they impinged on the behaviour of users or providers -- our focus was on the analysis of these actors. Now we change our perspective, walking around to view the elephant from a different angle, as it were. But we pursue the same approach, regarding government as a behaving entity within society, pursuing various objectives by means of particular strategies and under certain constraints.

In the process, we shall take a rather traditional view of the objectives of government, treating it as an instrument for the furthering of collective, public purposes. Implicitly we assume the existence of some sort of general public interest, with all the problems of identification and definition which that implies. We thus regard the proper role of government as the designing of public policy to serve the public interest.

An alternative viewpoint, which has enjoyed a recent vogue among economists, views government as a transactor in the marketplace cheek by jowl with the others, differing only in that it sells coercively enforced rules of the economic game in return for votes or the wherewithal to buy votes. From this perspective, the public interest is an empty concept; there exist only various private interests which are furthered through private or public institutions. And indeed many of the actions of governments *are* better explained as responses to the interests of powerful private groups, not to broader public interests. Critics of the public interest approach thus refer to the alternative as a "positive" as opposed to normative, analysis of government, returning to Machiavelli's claim to describe what is done rather than what ought to be done (Stigler 1971; Posner 1974).

The study of professional regulation conveniently demonstrates the contrast of approaches. From the public interest point of view, regulation is intended to protect interests which for various reasons would be unduly vulnerable amid the free play of market forces (Trebilcock *et al.* 1979). It responds, or should respond, to specific breakdowns in the conditions under which markets lead to desirable outcomes in terms of resource allocation and/or output distribution. The "positive" theory, however, notes that in fact professional regulation is almost always sought by the profession itself, not by the supposed public beneficiaries. Following Shaw, "all professions are conspiracies against the public," it views regulation as a form of market power, economically advantageous to the regulated, "purchased" by them from government in return for various forms of political support.

The process of public economic regulation, as it actually goes on in and out of health care, provides a number of examples which strongly support the "positivist" view. Agricultural marketing boards, for example, admit no other plausible explanation. But even the most enthusiastic positivists concede that its predictive power is unsatisfactory -- the conspiracy approach explains why all want self-regulation, but not why only some get it. Nor can it explain regulation which the regulated find onerous and energetically lobby against -- drug safety testing, for example. And the theoretical models of government as a quasi-economic transactor -- buying votes and selling rules or enforcement -- suppress significant parts of the institutional reality of political behaviour. If there is a political "marketplace," it is obviously oligopolistic, and economists' efforts to formulate plausible theories of oligopoly behaviour have been notoriously unsatisfactory. Information plays a very different role in this "market" -- to be seen to be purchasing or selling regulation may be very costly to the seller, at least. Overt political bribery can cost votes, and in some forms is illegal. Politics is not just economics carried on in a slightly different language. There *is* a difference between a government and a Mafia enforcement agency, which is what the legitimation of sovereignty is all about.

Moreover, the logical limit of the positivist approach is, in fact, quietism. If all supposedly public policy is simply a collection of private transactions, then there are neither criteria for analysis and criticism nor means for modification. What is, is. The tendency for positivist analyses of government to be associated, in practice, with advocacy of *less* government and more reliance on private market forces is wholly illogical. If one is unwilling to critique extant distributions of wealth or influence or power, then presumably whatever set of markets arises to support economic or political transactions is the best available: Positivism meets Pangloss. As for changing or reducing the role of the state, that will occur, on the positivist theory itself, *only* if it serves the interests of the dominant private interest groups. And the advocates of such policy, in the positivist view, are simply the "hired guns," or public relations arm, of those groups. The development of United States political and academic culture in the 1970s and early 1980s, and the role of positivist economic analysis of public policy in this process, serve to illustrate the argument (Evans 1982*b*).

The public interest approach, however, grounds the analysis of public policy firmly in the normative. Policy *ought* to be devoted to general public ends. In many cases it is not; these are cause for criticism and, perhaps, correction. Perfection is not anticipated, in this life at least, but by assumption a process of analysis, education, and exhortation can lead to improvement. Nor is the policy-maker, the philosopher-king, the only target of this process. "Bad" policies such as professional regulations which benefit only the profession and harm the general public, or similarly, agricultural marketing policies, probably *are* the result of balances of political interests, not ill-informed or stupid policy-makers (Hartle and Trebilcock 1983). In a democratic state, the informational component of public policy analysis extends to the whole society. The

analyst of policy from a public interest perspective may better serve, not as advisor to the Prince, but as preacher or teacher to the multitude. But what are the alternatives?

PROGRAM EVALUATION: A GENERAL REPRESENTATION

Accordingly, our perspective on government assigns it the role of choosing among policies and programs on the basis of their consistency with the public interest -- cost-benefit analysis writ large.[2] Each possible act has both positive and negative consequences across the society as a whole. The task of policy formulation involves the identification and measurement of all of these benefits and costs, and their aggregation in such a way as to determine whether the policy is on balance good or bad. Designing mixes or "portfolios" of policies, whose effects interrelate with each other, adds an additional level of complexity. In this chapter, however, we will look at the government's problem in terms of a single project or small scale program -- an immunization campaign or the funding of a new diagnostic facility -- before extending the general approach to such larger issues as public insurance systems, or strategies for research funding or health promotion. In some cases individual projects may form the building blocks for more general programs, in others, universal public insurance, for example, they do not.

The analytic representation of the public sector program evaluation problem is very simple. One has only to test whether:

$$\sum_{i,k,t}\left[\frac{P_{ikt}B_{ikt}}{(1+R)^t}\right] - \sum_{j,k,t}\left[\frac{V_{jkt}C_{jkt}}{(1+R)^t}\right] \begin{matrix} > \\ - \\ < \end{matrix} 0 \qquad (11\text{-}1)$$

If the expression above associated with a particular project or program is positive, then it should be carried out; if not, not.

As is typical of economic decision rules, however, the formal correctness of an expression is not much help, except perhaps as an organizing framework for future work. The really interesting issues, and the difficulties, arise when we try to give content to the various symbols above, to define and measure them, and to deal with the complications when this proves impossible.

To begin, however, we must define the symbols themselves. *B* and *C* are the measurable impacts or results of a particular program. The benefits *B* might be measured in terms of lives saved, days of ill-health averted, or disabilities reduced by a health program, or they might be extra tons of grain produced by an irrigation project. Costs, *C*, on the other hand are the things used up by this project -- human time, energy, and skills, physical resources, capital equipment services -- which are therefore unavailable for other purposes.[3] The pattern of benefits and costs, *B* and *C*, generated by a project is also a description of the project itself in technical terms, whether the technology be engineering, agricultural, or clinical. If resources of *C* are applied in the manner described by the project, desired outcomes *B* will result.

But neither resources used nor benefits generated are uniform, homogeneous entities. Thus they are indexed, in three different dimensions. The subscripts *i* and *j* run across the different forms of benefit and cost. A project with several different types of benefit would have them measured, in some appropriate unit, as B_1, B_2, B_3 ..., B_i, and the corresponding different types of costs or resources used up are C_1, C_2, C_3 ..., C_j.

The benefits and costs also, however, impinge on different people or groups, and the evaluation process may have to be sensitive to these distributional questions. Hence both *B* and *C* are indexed over the people in the relevant community; B_{ik} represents the *k*th person's receipt of an amount *B* of the *i*th type of benefit.

Finally, projects have a time dimension. Most commonly, though not necessarily, costs tend to be higher at the beginning of a project, and benefits come later. In any case, it matters at what time period costs and benefits are incurred or received, and thus C_{jkt} represents *C* units of a cost of type *j*, incurred in period *t*, by person or group or agency *k*.

The definition of the *B* and *C* patterns associated with a project embodies a critical step, which is all too often passed over unconsciously. The state of the world in the absence of the project has simultaneously been defined as a situation in which particular consequences *do not* occur. In general, it is desirable to specify this alternative explicitly, since otherwise the project being evaluated may be compared with an irrelevant or implausible baseline. And one must specify *all* the project consequences, intended or unintended. Particularly common in health care are claims that particular projects, seatbelt use for example, or home care of potential hospital patients, or new diagnostic interventions, will pay for themselves by dramatically lowering costs in some other area of care. Implicitly this means that suppliers of these competitive forms of care will be put out of work and will leave the industry. If that is not, in fact, anticipated, then full specification of the *B* and *C* for the project must include some description of what is likely to happen to redeployed substitute resources, and the costs or benefits of their new activities. If, as is frequently the case, new projects merely add on to existing activities for which they are nominally substitutes (cimetidine *plus* antacid therapy, CT scanning *plus* other imaging techniques, long-stay or day surgery beds *plus* a maintained level of inpatient activity), then there will be no cost offsets. And the benefits of the new activity must be measured relative to what was previously occurring, not a zero base line.[4]

In aggregating these benefits and costs across the whole society, one must employ various weighting factors to add up the apples and oranges involved. These weights are in the form of prices, converting units of measurement to units of account, and are represented by the P_{ikt} and V_{jkt}. In a society using dollars as currency, it is natural to think of P_{ikt} as the dollar value per unit assigned to benefits of type *i*, accruing to individual *k*, in time period *t*, but any accounting unit will do as well. In some cases benefits and costs will have well-defined prices determined in private markets, which may be appropriate for use in equation 11-1. In other cases, particularly in health care applications, there may exist no market for the benefit (*e.g.*, human life), or the market price (volunteer services, professional services if monopolized, subsidized commodities) may not represent the true resource cost or opportunity cost of the commodities involved. The analysis may then require calculation of shadow prices, synthetic values of P_{ikt} and V_{jkt} which attempt to represent their true rates of exchange with other goods or services.

The indexing of *P* and *V* across *k* enables the evaluator also to build in different values for benefits or costs affecting different people. Thus, if some particular groups -- low-income people, the elderly, children -- are considered particularly deserving, augmenting the P_i and V_j for those subsets of the society has the effect of favouring programs which benefit them. On the other hand, if benefits and costs are to be counted equally "to whomsoever accruing," then P_{ikt} and V_{jkt} will be constant for all *k*.

DISCOUNTING, INFLATION, AND CONSISTENCY

The time discount rate, *R*, plays a special role in the aggregation process and interacts with the time-pattern of the *P* and *V* in a way which can lead to serious errors of analysis. As noted, the consequences of many projects arrive over time. But one cannot aggregate present and future consequences, even if they are measured in the same units, because as capital markets remind us, time is productive. A dollar now is worth more than a dollar in the future, and the amount of this advantage is reflected in the interest rate, *R*.[5] If, for example, the going rate of interest on loans for a length of time spanning the main effects of the project in question were currently 10 percent, then $1,000 worth of benefits accruing five years from now could be "bought" with a present investment of:

$$\frac{\$1{,}000}{(1.10)^5} \approx \$620.92$$

That sum, invested now at 10 percent (compounded annually) would be worth $1,000 in five years; hence the present value (*PV*) of $1,000 five years hence (whether benefit or cost) is $620.92. Present values *are* additive, so we can evaluate a project yielding $1,000 worth of benefit in five years and requiring a current outlay of $600 and say that (at 10 percent interest) it is worth doing. If the current outlay required were $700, it would not be. Furthermore, if interest rates on five-year loans were 20 percent, the *PV* of $1,000 five years hence would be

$$\frac{\$1{,}000}{(1{,}20)^5} \approx \$401.88$$

so the project would no longer be worthwhile at $600 outlay.

Failure to discount future benefits, or cost offsets, can be particularly deceptive in the evaluation of programs with long-term effects, such as prevention of handicap. Rubella immunization, for example, or screening for Down Syndrome (*plus* abortion of affected foetuses) may reduce future support costs over many years. But to add up undiscounted cost offsets as if a dollar twenty or thirty years hence were equivalent to a dollar now, when in fact such a future dollar could be purchased now (in capital markets) for a much lower sum, would grossly overstate the cost offset component of program effects.

It is obviously an abstraction to refer to "the" market rate of interest, there being a wide array of such rates; and there has been extensive debate over the appropriate choice for public investment projects. Moreover, while financial flows, or flows of benefits or costs which (like tons of grain or man-hours) have well-established money prices, must be discounted at some rate, the case for discounting non-traded or non-tradable things such as human lives or illness states is less clear. One cannot trade present for future lives, either directly or through some intervening financial market, as one can trade, for example, wheat. The person whose life is lost now is *not* compensated by the saving of one, or even 1.5, lives some years hence. We will return to this point below.

What *is* clear, however, is that whatever discount rate is chosen, it must embody some treatment of inflation, and thus must interact with the time-pattern of P_t and V_t. Market interest rates are nominal rates; they include an estimate of the rate of inflation over the relevant time period. The faster money falls over time in purchasing power, the higher must be the nominal rate of interest to compensate for this. If R = 20 percent, but prices are rising at 12 percent per

year, the true or real interest rate is only 7.14 percent.[6] Real resources given up now can be traded for only 7.14 percent more resources next year, even if for 20 percent more money.

Thus the value chosen for R, and the values assumed for P_t and V_t, must be consistent. If one chooses a 12 percent discount rate, which would have been reasonable in mid-1984 based on then current market interest rates, one must keep in mind that this embodies an expected inflation rate of about 6 percent. Accordingly, the values of P_{it+1} and V_{jt+1} should be about 6 percent above P_{it} and V_{jt}, for all t. If inflation is implicitly in the denominator of equation 11-1, as it is when one uses a market interest rate, then it *must* be built explicitly into the numerator. Of course one may believe, for various reasons, that particular P_i or V_j will rise faster or slower than the general price level. But to fail to escalate them at all, and then to use a market discount rate, amounts to assuming that the prices of B and C will remain steady in nominal dollars while everything else rises in price at the expected inflation rate built into the market interest rate. In inflationary times, this creates a serious (and erroneous) bias against projects with long payoff periods and high initial costs.

One can, of course, ignore inflation in the numerator of equation 11-1 and then discount by a real rate in the denominator, *i.e.*, a market rate reduced by expected inflation. One might, for example, compare current (annualized) inflation rates with long-term government bond rates to estimate (on the assumption that current inflation rates are expected to persist) the current real rate of return. Such a procedure emphasizes the inherently uncertain nature of returns on all non-indexed investments. One cannot define real returns without assuming (guessing at) future inflation rates.

Such rates on, for example, long-term bonds, have historically tended to fall in the 2-4 percent range, well below the 8-12 percent used in official cost-benefit guidelines. Short-term fluctuations may move well outside this range -- real rates were negative for part of the early 1970s and in the early 1980s reached 8-10 percent -- but these were temporary aberrations. Few professional investors have consistently earned 4 percent real over long periods of time, let alone 10 percent. Anyone recommending 10 percent as a discount rate for public projects, without inflating the numerator, should be invited to try running a portfolio![7]

COSTS AND BENEFITS TO WHOM? WHERE YOU STAND DEPENDS ON WHERE YOU SIT

This way of setting up the program or project evaluation problem is sufficiently general as to be merely a particular representation of the rules of rational behaviour. It is a framework for identifying all the consequences, good and bad, of an activity, reducing them to a common metric (dollars) by weighting factors (prices) which reflect their differences in time of occurrence as well as intrinsic characteristics, and then adding them up to see whether or not the good outweigh the bad. The evaluation of public projects is in this respect no different from that of private projects. The individual firm or consumer in economic theory is represented as going through a precisely similar calculation.

The private firm, for example, evaluates a production decision or a long-term investment project in terms of the stream of saleable outputs, goods and services, which will flow from it, the B_i, and the costs in terms of productive resources which will be used up in the process, the C_j. It then estimates the prices at which the B_i can be sold, P_i, and the prices V_j which it will have to pay for the C_j, and adjusts for the time-pattern of receipts and expenditures. Equation 11-1 represents the firm's net revenue stream, or its profits from the activity. Assuming profit-

maximizing behaviour, the firm chooses to carry out only those projects for which 11-1 is positive.[8]

Similarly the rational consumer, in allocating her budget over consumption alternatives, present and future, or more generally in choosing patterns of resource supply (allocation of productive time and skills, as well as other owned productive assets) and commodity consumption, is modelled as carrying out a utility-maximizing process in which the B_i are commodities consumed, the C_j are resources supplied to the market, and the P and V are composed of market prices and personal marginal utilities or disutilities.

The principal difference from the evaluation of public projects, however, is that in the private calculations $k = 1$. The private firm or consumer calculates costs and benefits to him/her/it alone, disregarding "external effects" or consequences of the project, positive or negative, which fall on others. If it is indeed the case that B_{ik} and C_{jk} are very small or zero for all k other than the decision-maker, then clearly private and social decision processes would give the same result. But if there are many k for which B and C are significant, then obviously calculations made over only a subset, or one, of the relevant k can be erroneous in either direction. In such cases, as discussed in chapter 3 above, private markets lead to faulty resource allocation decisions and public programs develop to respond to this market failure. Thus public projects frequently involve calculations of effects over a large number of people or institutions.

Accordingly in any cost-benefit analysis it is critical that one be clear about the perspective chosen -- *whose* benefits and costs? Health care programs, for example, will affect patients, providers of care, public budgets, and taxpayers. Particular individuals may play several of these different roles. And a program or project may, frequently will, serve to redistribute costs and benefits among members of society as well as to generate them. A successful project to shorten hospital stay, for example, might serve to lower hospital costs and thus benefit governments who pay such costs -- and through them, taxpayers. But if it generated significant costs in the home for patients, it might in fact raise costs overall. Similarly, a government which negotiated a reduced fee schedule for physicians in return for permitting extra-billing of patients would have lowered costs in the public sector and presumably benefited taxpayers. Patients, of course, lose; providers gain (insofar as fees plus extra-billing exceed the fees they would otherwise have received) at least on average, and at the society-wide level the whole transaction may wash out as a pure income redistribution, depending on whether or not actual patterns of care utilization are affected. Full evaluation of any project thus requires the identification of all costs and benefits, wherever they impinge or accrue.

But it is often useful to identify also the partial sums across k, the subgroups of winners and losers. It is frequently the case that one man's cost is another man's benefit, and much conflict over program evaluation has its roots, not in disagreement over society-wide effects, but over partial distributions of gains and losses. A program which involves increased demand for the labour and skills of particular types of workers will be viewed favourably by the owners of those skills, even though from a society-wide perspective the using up of those scarce resources of time and skills is a cost. Cost to society is income to the resource owner.[9]

We have seen this phenomenon above in several contexts. The overhead cost of a universal public health insurance system is known to be much less than that of multiple-source private insurance, but higher costs are the revenues of private insurers. A public dental service for children using dental nurses is, from a society-wide perspective, much less costly than public or private insurance for private dentists' services. But the costs of the latter program are dentists' incomes.

Hence programs which seem obviously efficient from a society-wide perspective are frequently blocked because they impose losses on a powerful subgroup whose interests cannot or

will not be compensated, while other programs whose costs viewed globally seem clearly beyond their benefits are adopted because of their partial payoffs. The cost-benefit analyst may treat all winners and losers equally; the political system does not.[10]

PRICELESS DOES NOT MEAN VALUELESS: SHADOW-PRICING LIFE AND LIMB

In addition to their impact on a range of different people, whose interests pro or con are often difficult or impossible to represent in a market framework, public programs frequently involve benefits, and sometimes costs, which are very difficult if not impossible to represent in a common metric of dollars, or anything else.

The framework of equation 11-1 assumes that a P_i or V_j can be identified for each B_i or C_j, such that costs and benefits can be summed and compared. This process is straightforward if B_i and C_j are traded in well-functioning markets where their relative prices are established. This will be true, for example, of the costs of new construction associated with a public project, which can be contracted out to the private sector or organized in-house, but in either case at well-defined dollar costs. Similarly the benefits of a land irrigation project may be identified in terms of increases in yield of particular crops, which will in turn be sold at clearly identified prices.

Somewhat more complex are the cases in which prices or costs of inputs and outputs are available, but are systematically distorted from true resource costs or market values. A public program might receive administrative or building services from some other branch of government, at no budgetary cost. Yet building space and administration use up real resources, so the evaluation process would have to use a "shadow price" reflecting the true cost of such inputs. An irrigation project might increase output of a commodity for which there was a price-support program; a government agency might be accumulating surplus stocks to hold market prices up. In this case, the apparent market price overstates the value to society of more (surplus) output. A hospital might receive donated equipment or space, or use volunteer labour, which was "free" in its budget, but such inputs are clearly costly, resource-using, from a society-wide perspective, and should be shadow-priced.[11] In general the particulars of how a program is financed, from whose resources, should not affect the society-wide evaluation of its costs and benefits, though clearly the interpersonal pattern of gains and losses will be affected.

The problem of shadow-pricing, however, shades into that of evaluating costs and benefits which have no obvious financial metric at all, conceptually or in practice -- the "life-and-limb" problem. Shadow-pricing assumes that a conceptually appropriate price for each of the consequences of a program exists, though it happens not to be reflected in existing market prices or budgetary entries and must therefore be calculated directly. For consequences involving life and death, pain and suffering, physical, psychological, and social functioning, it is not obvious that any shadow price exists.

Yet it is clear that programs with mortal and morbid consequences must be evaluated. The human life may be priceless, but it is not infinitely valuable -- we implicitly assign it a finite value in any number of settings. Individually we drive cars (or worse, ride bicycles), ski, climb mountains, smoke, thus trading life expectancy for other satisfactions. Collectively we establish policies on factory, highway, or product safety, knowing that risk-reduction is costly and can only be pursued up to a point. On the other hand, lives are not valueless either, nor is the avoidance of pain, suffering, or grief. To leave them out of an evaluation process because their values are ill-defined is bad economics as well as bad policy.

The life-and-limb issue has been addressed in two distinct ways. The first is to try by a variety of techniques to calculate shadow prices for life and limb. Since all actual program decisions which involve life and limb implicitly place values on them, it is argued that better, or at least more rational, decisions could be made if the values used were explicit and consistent. It is not difficult to show, for example, that at present public program decisions are quite inconsistent. In some fields (airline safety) resources are used to save lives or prevent deaths at a much higher cost per life saved (at the margin) than in others (highway safety). If all saved lives (or averted deaths) are equally valuable regardless of the nature of the threat, then we could be better off, so goes the argument, if we redeployed our lifesaving resources.

The alternative approach is to abandon efforts to evaluate life and limb explicitly, as productive of more confusion and misinformation than enlightenment. Cost-benefit analysis, as represented in equation 11-1, in which all components are weighted in dollar terms, is replaced by cost-effectiveness analysis in which some at least of the consequences of a project are measured only in terms of their natural units. The transition from cost-benefit to cost-effectiveness analysis in the United States was most clearly demonstrated outside the health care field when program evaluation techniques were extended from water-resource projects to weapons systems. Payoffs to irrigation activities -- increased crop yields, flood control, recreational benefits -- can be calculated using market or shadow prices. But the dollar value of megatonnage delivered to downtown Moscow is difficult to measure in any meaningful sense.[12] Cost-effectiveness analysis takes the objective as given -- extending lives or terminating them, and measures the relative costliness per unit of desired consequence from alternative projects for achieving the same or comparable consequences. Such alternative projects might be different approaches to the same problem -- submarine versus land-based missiles, school immunization campaigns versus subsidies to private physicians -- or they might be different scales of operation of the same program.

VALUING LIVELIHOODS: ESKIMO ECONOMICS

Attempts to identify explicit shadow prices, money values, for life, disability, pain, and suffering began from the human capital view of the person as a set of productive capacities whose value was represented by earning power. A life cut short before its normal time is a loss of economic product, equal to what that person would otherwise have earned (on the assumption that markets ensure that a person's earnings equal the economic value of her production) over the remaining expected life, discounted back to its present value at point of death.[13] Lives saved by a project are then valued at the present value of their possessor's subsequent expected earnings. Disability costs can be measured the same way, by the present value of economic output lost if the disabled person must take a lower-paying job, or none at all.

This "value of livelihood" approach has the advantage of yielding concrete results, apparently firmly grounded in economic reality.[14] But it is now largely discredited among students of program evaluation because of its fundamental inadequacies in both theory and application.[15]

The decision rule implicit in the value of livelihood measure can be summarized as "Eskimo economics," embodying the priorities (alleged to be) characteristic of pre-contact Inuit culture. When the game fails, and starvation threatens the community, the elderly population must be sacrificed. If conditions remain bad, the children go next, followed by women, and then men in

reverse order of hunting prowess. The best adult male hunter is the last to go, as he has the best chance of finding food for whatever is left of the band -- in the end, him.

These are precisely the priorities of the human capital approach to project evaluation. The elderly, living on pensions and other assets, have no earned income and are worthless. (They may be credited with some non-market production of rather ill-defined value in a more sophisticated analysis.) Children are worth something, but their future product is discounted back to the present and thus, for very young children particularly, becomes very small.[16] Earnings of women, in most cultures, are below those of men, so the value of saving a woman's life is correspondingly lower. Unemployed time is worthless, so value of life must be adjusted for expectation of employment. In the United States, blacks earn less on average than whites, so black lives are worth less. Specific examples of these powerful value judgements, masquerading as objective quantitative analysis, can be found in the cost-benefit literature.

But these priorities do not in general govern resource allocation in our society. The elderly and/or disabled receive a far higher share of health care (per capita) than do the general population, largely because they are sicker. The classic dictum of the lifeboats is: "Women and children first." If we, as a whole society, lived on the margin of subsistence, we might follow the same rules as the Eskimo did. We do not.[17]

The difficulty with the human capital approach is that, when capital assets have zero or negative present value, they are not repaired but scrapped. Indeed some versions of the livelihood approach take it to its logical conclusion, and measure value of life by the individual's net production, earnings less consumption, adjusted for taxes and use of public services, and for gifts and bequests. Does society, *excluding* the individual in question, profit (in dollar terms) from that individual's existence? Obviously, this figure is negative for almost all retired persons, indicating that a low-cost euthanasia program, not just for the terminally ill, but for all the retired, would be the highest valued public health program one could mount.

At this point one can insert an embarrassed footnote that of course other considerations apply, which they certainly do, but the absurdity of the conclusion is *not* simply an unfortunate result in a particular application. It arises from the fundamental assumptions of the technique, which are themselves absurd. For in fact, human capital is *not* indifferent to being scrapped; lives are of value to the living person, as they are to his or her relatives, friends, associates, and in general (in decreasing amount) to a wide ring of those who are aware of the life. In the same way disability, pain, and suffering, or rather their absence, are valuable and valued independent of their economic consequences. Loss of the left hand may have devastating consequences for the earning power of a pianist or surgeon; little or none for that of a (right-handed) academic or hospital administrator. Yet it does not follow that the latter would be indifferent to such loss, or what is the same thing, would place no value on saving the hand. And the individual sufferer in question is one of the *k* in equation 11-1. Her preferences matter, even if no one else cares (which in general they will, only less).

In theoretical terms, equating value of life with livelihood is erroneous because, as we recall from chapter 1, the economic value of anything is defined in terms of what some individual or group is willing to give up in return for it. In a competitive market, under various stringent conditions on demand and supply, this value will tend to be equated with cost of production, *i.e.*, what must be given up to produce the commodity. But there is no competitive or any other production of lives, independent of the people concerned. My life is unique, no other lives compensate me for the loss of it.

In a slave society, the market price of a slave would tend to equal the present value of net earnings, production less consumption, and in long-run equilibrium this would in turn equal the cost of producing slaves of a given age and skill, in accordance with the value of livelihood

model. But the slaves would not be counted among the *k* of equation 11-1, and their asset values would tell us nothing about the proper value of *masters'* lives.

YOUR MONEY OR YOUR LIFE? WILLINGNESS-TO-PAY AS A SOURCE OF SHADOW PRICES

Efforts to value life and limb explicitly have therefore shifted to attempts to measure willingness-to-pay for life-saving consistent with the basic concept of value in economic analysis. Such approaches encounter two problems immediately: (1) whose willingness-to-pay? and (2) for what?

The first is usually answered in terms of individual choice, the potential sufferer. There have been some suggestions that value of life, or pain and suffering, might be inferred from decisions by legislatures concerning public projects with mortal or morbid consequences (positive or negative) or damage awards by courts, but it is generally recognized that this is a circular procedure. Such agencies look to analysts for guidance as to the values *they* ought to use. The inconsistency of their decisions is a principal justification for efforts to find an explicit monetary value by quantitative analysis.

But if the individual potential sufferer is to be the source of information, it is clear that the problem must be posed in statistical terms. Faced with the certainty of life or death, the individual's willingness-to-pay is bounded only by her wealth, plus whatever she can beg, borrow, or steal. The proponents of this approach have focussed on individual valuations of *probabilities* of death. One can explore how much an individual would pay to avoid a 1 percent chance of death over the next year, for example, and then sum over a hundred such people to determine what value the group would place on saving one (expected) life.

The willingness-to-pay approach opens up a number of interesting research possibilities for determining individuals' values. Life insurance purchases were suggested, and quickly rejected as reflecting the individual's estimate of her financial obligations, not value of life. Wages paid in hazardous occupations, explicitly as hazard pay or implicitly as above-average wages, can be compared with the degree of hazard to calculate the implied value of life. But this assumes full information and perfectly competitive labour markets, which students of labour economics find implausible; further if people differ in their valuations one would expect to find those with lowest life values (or least information!) in hazardous occupations. Thus generalization is suspect. Other measurements have been based on seat-belt use, balancing estimated values of time against risk. Still others try questionnaires.

Apart from the intricate problems of measurement, however, it is clear that this approach raises some deeper problems. First, individual valuations are not linear. If I will pay $1,000 to avoid a 1 percent chance of death in the next year, but not $1,001, it does not follow that 10 percent is worth only $10,000, still less that I would prefer certain death to paying $100,100! Thus whatever results emerge for a particular risk level cannot be generalized to other levels.

This fact has been given a rather peculiar twist in discussions of prevention versus cure. It is sometimes alleged that individuals will forego relatively cheap preventive activities, yet spend very large amounts to save or prolong lives *in extremis*, thus valuing lives or years of life very differently in the two settings. This apparent inconsistency, however, can be "explained" in terms of standard consumer theory. When one has many (expected) life years ahead the marginal value of life years in terms of "other things" is less than when one has very few. Hence it is "rational" to spend little on prevention and much on cure.

This argument has a number of problems, including the facts that information about the efficacy of prevention is notoriously poor even among specialists in the field, that expenditure on terminal care in every country is covered by public or private insurance, and that death is a discrete event. Groups have life expectancies, individuals have lives, and when the current life year is given up, the rest go with it.

But on its own terms, the argument illustrates a problem of micro-rationality leading to macro-irrationality. Suppose a group of people, a society, make their allocations between (efficacious) prevention and cure on the basis of individual marginal rates of substitution. By spending less on prevention, and more on cure, they may as a group have both shorter life expectancies *and* higher expenditure on life-prolonging care. (Any parallel with United States health care is accidental, though the micro-rationality argument does come from the United States.)

If an investment decision presents itself strictly in terms of statistical lives, the individual willingness-to-pay approach may be useful. But health care problems rarely present themselves in that form. There is usually a stage in the illness process in which sufferers are known, flesh and blood individuals, and the project decision will take the form of whether or not to care for *this* person in *these* specific circumstances. At the theoretical level it can be shown that *ex ante* optimal decisions are not necessarily optimal *ex post*, and conversely (Evans 1983); in practice, it is obvious that people who have made a particular decision as to the value of statistical lives will wish to recontract when the time comes.

Nor will those around them be keen to enforce the contract. Apart from the inherent unpleasantness (in most cases) of letting someone else die, still more of expediting the process, there is the difficulty that lives are of value to other than the individual concerned. And the individual willingness-to-pay criterion, like market processes generally, provides no way of taking those preferences into account. Each of us is, to a greater or lesser extent, a public good (though Donne said it rather better).

As a further complication, individual valuations of lives will of course be very sensitive to individual income or wealth levels. If public program evaluations are based on individual preferences, they will favour lifesaving for the wealthy. One could, of course, average values across society, but this undercuts the theoretical basis for the individual approach.

Finally, recent psychological research on decision-making under uncertainty has cast serious doubt on the empirical validity of the expected utility model. Decisions in uncertain situations seem to be quite sensitive to the setting or framework in which the decision is taken; changes in this setting which are formally irrelevant to the payoffs available can lead to reversal of choices (Schoemaker 1982).

COST-EFFECTIVENESS ANALYSIS: LET THE OUTCOME STAND FOR ITSELF

For all these reasons, and perhaps others as well, most people evaluating health care programs have abandoned the attempt to establish explicit monetary values for life and limb, pain and suffering. It remains true that any program decision, go or not go, implicitly places a value or at least a one-side bound on the values of such consequences. But cost-effectiveness evaluations do not attempt to build such values into the analysis, rather they confront the decision-maker with a menu of costs and consequences, relying on the political process to yield the "right" choice.

The cost-effectiveness approach to evaluation is most straightforward to apply when a choice is to be made between two alternatives whose consequences are equivalent. Given that, for unspecified or unevaluated reasons, a certain level of megatonnage is to be delivered, which system gets it there at least cost? Two forms of health care, inpatient and daycare surgery, for example, may be therapeutically equivalent for a particular set of conditions which, it has been determined, are to be treated. This form of evaluation is really only a cost analysis and comparison, measuring the second term of equation 11-1; the decision to do something has already been taken.

In this comparison, different programs may be essentially similar in process, or radically different in concept and activity. Dietary versus surgical approaches to heart disease, for example, or prevention of rubella syndrome by either mass inoculation or "rubella parties" to ensure that female children have the disease before child-bearing age, or prenatal (or even pre-conception) dietary supplementation versus neonatal intensive care, each represent totally different activities. But the intended results of each are the same, and they can therefore be compared in cost and reliability of achieving that result.

Measurement of project effects is necessary in the more general case, however, in which either no action is a possible option, or projects have different scales of operation. ("Do nothing" is simply operation at zero scale.) In these cases, project consequences in physical terms must be compared directly against costs: X lives saved for Y dollars, $X + A$ lives saved for $Y + B$ dollars. The outcomes are in commensurate physical terms, but the judgement as to whether X lives are "worth" Y dollars, or if they are, whether A more lives would be worth an additional B dollars to scale up the project, is ultimately, in this context, a political choice. In general one expects that, following Figure 1-3, projects of particular types will show diminishing returns. The dollar cost per life saved will rise as the program expands; $Y/X < B/A$. But if there is always some possible project with a potential positive effect on health status, then as in Figure 1-3b, the person or body with decision-making responsibility will have to set a cut-off value for lives.[18]

Still more difficult is the case of multiple types of non-monetary consequences -- projects which influence the full spectrum of lives lost, temporary or permanent disability, and pain, fear, and suffering. Consistent with the basic concept of cost-effectiveness analysis, the analyst can then only present the alternative menus of differential miseries associated with different activities and let the politician or administrator choose. But obviously the possibility of informational overload grows with the complexity or ramifications of projects.

A number of attempts have therefore been made to find ways of combining the diversity of dimensions of project outcome into a single index, an explicit "health status" measure, which could serve as an ordinal or even cardinal measure of the total payoff to a particular program. More ambitiously, such an index could even measure the progress of a population and its health status through time, or permit inter-regional comparisons. But while, as discussed in chapter 1, such a concept seems to be implicit in all health policy, its explicit representation is fraught with conceptual and measurement difficulties (Culyer 1978; Berg 1973).

A LIFE AFFIRMED, A DEATH CONFRONTED: LIFE YEARS AS OUTCOMES

A promising solution to this problem, within the context of cost-effectiveness analysis, begins from the observation that while we speak loosely of life-saving projects (probabilistically or with certainty) the brute fact of existence is that we can do no such thing. The most solidly established regularity in epidemiology is that one out of one dies (known also as the Law of

Competing Risks). Death can be postponed, but life cannot (at least by non-spiritual means) be saved.[19]

If death is postponed, however, then *years* of life are saved, and this represents a natural physical unit in which to measure outcomes. It is consistent with our general sense that saving the life of a child is in some way more significant than terminal care for the elderly. This sense is sometimes expressed in equity terms -- the child has not had a chance to live -- but also reflects the obvious quantitative comparison of life expectancies.

The significance of different sources of mortality is dramatically rearranged by a refocussing from lives to life years; cancer becomes much less significant because it usually strikes late in life, while accidents which carry off the young represent a much higher cost in lost life years. Even if cancer were to vanish completely, life expectancies in North America would rise by only 2-3 years (and presumably mortality from heart disease would leap upwards).

Measuring the physical consequences of health programs in life years may lead in some cases to results which parallel those of the livelihood approach, since the present value of one's expected earning stream, like one's life expectancy, falls with age. Both will tend to give less weight to programs benefiting the elderly. But there are several very important differences.

First, and most obviously, the value of life saved does not fall to zero at retirement. The livelihood approach values the life of a sixty-four-year-old man who happens to earn $50,000 a year at $50,000; next year when he retires, he is worthless.[20] But the life-year approach merely subtracts slightly less than one from the stock of expected remaining life years represented by that person, when he ages by one year.

At the other end of the scale, a child's life saved now represents a large number of added life years. The present value of livelihood for a child, however, is substantially reduced by the discounting process -- earnings will not accrue until the child grows up and enters employment. Thus the life-year approach increases the weights at the high and low ends of the age spectrum -- precisely where health spending (per capita) is in fact greatest.

Finally, the life-year approach does not weight by earning capacity. One person's life year is as good as another's. The fact that women or members of some ethnic groups have in North America lower average earnings than white males (which may reflect either lower economic product or discrimination -- or both) is not built into the process of project evaluation so as to assign lower values to projects benefiting them. As a practical example, the estimated relative payoffs to screening for breast cancer, hypertension, and tuberculosis will differ significantly according to whether one values outcomes by life years or by livelihood.

But the focus on life years raises an awkward point concerning cost offsets which can more easily be neglected in the "life-saving" context -- the living use more health care than the dead. The person whose life is saved today lives to grow old, or at least older, with increased probability of using health care later on. The anti-smoking campaign which saves lives and costs of treatment for lung cancer is generating future costs of heart disease, as well as treatment for senility. A quick and fatal heart attack at sixty-two is, in the long view, cheap, not only in saving future pension costs, but also in "saving" future health care interventions which the sufferer might otherwise have needed over the next twenty years.

This is *not* an argument for euthanasia. Life is, we strongly believe, better than death,[21] and is worth paying for. But consistency in project evaluation requires that we record and cost *all* consequences. And if a program saves life years now, those years have cost implications in the future which must also be recorded.

How they should be treated in decision-making is, however, another issue. Stason and Weinstein (1977), for example, take a "hard-nosed" view that all subsequent health care use by those kept alive through a particular program should be included in the costs of that program.

And certainly anyone responsible for health care planning or delivery should take account of the additions to later utilization which will result from a successful "lifesaving" program. But the proper treatment of health costs in subsequent years, when evaluating that program, is ambiguous.

It is clear that, at one extreme, a successful program addressing a particular condition which requires interventions for that condition in each subsequent year of life should have the costs of those interventions included in program costs. On the other hand, a "one-shot" successful program which extends its beneficiaries' lives will still require that they eat over the years of subsequent life. Such eating uses up resources which might otherwise be "saved," yet we do not include it as a cost of the lifesaving program. (Unless we are taking a purely instrumental view of the person; Stason and Weinstein do not.) At what point do particular forms of resource use during the added years of life cross over the line and become program costs? If all health care costs are included, then the evaluation of a particular program becomes sensitive to patterns of cost and effectiveness over the rest of the health care system; it is no longer being evaluated solely on its own merits. Accordingly, most analyses have tended to exclude from costs subsequent health care use during additional life years gained from a particular program, which is not directly related to that program.

QALYS: NOT ALL LIVES ARE WORTH LIVING

Life years are, moreover, themselves seriously incomplete as a measure of consequences, as they fail to take account of the impact of health care programs on morbid but non-mortal conditions. Further, lifesaving may have very different morbid *sequelae*; a life year as a quadriplegic or on a renal dialysis machine is not quite the same as a year of complete health, defined either in some absolute sense or as normal for one's age and situation.

The concept of the quality-adjusted life year (QALY) has been developed in response to this inadequacy. Weighting factors (between zero and one) are attached to years of life in particular states. If a program keeps one alive but bedridden and/or in permanent pain, those years of life are "marked down" to represent a smaller number of QALYs before being compared with the outcome of some other program which restores its beneficiaries to full health. Programs dealing with non-mortal problems can also have their outcomes measured in QALYs'; improved states of physical function or shortened periods of pain or disability represent increased QALY values for the same expected life years. QALYs thus represent the value or utility of the health status outcome of a program to its beneficiaries; evaluation using such measures is known as cost-utility analysis.

To illustrate, suppose two hypothetical programs are each expected to save ten lives a year. But one benefits children (average age five), the other elderly people (average age seventy-five) and life expectancy is otherwise eighty-five. Furthermore, the program for the elderly leaves them alive but permanently bedridden, and this state is judged only 60 percent as desirable as full health. Comparing lives saved, the two programs are equal, but in life years the first saves 800 and the second only 100. In QALYs, the comparison is 800 to 60.

The politically responsible decision-maker must still determine how much a QALY is worth, and is under no constraint, logical or otherwise, to apply common dollar values to QALYs in different settings. The calculation does, however, assemble information on the physical consequences of programs in a way which seems more consistent with overall social preferences and behaviour than do the livelihood or individual preference approaches.

The measurement of outcomes in physical terms does, however, raise questions about the appropriateness of discounting over time. These arise in two ways. The outcomes of projects or programs arrive over time, so that a particular payoff which represents a life or life-year saving may occur at some time in the future. In addition, an outcome occurring now will have, if its effects are measured in life years, consequences accruing over the expected lives of the present beneficiaries. Is keeping one person alive for the next ten years "worth" as much as keeping ten people alive over the next year, or should the future QALYs be discounted?

If there is a consensus among analysts, it appears to be in favour of discounting, on two lines of argument. First, it is obvious that economic resources, costs, must be discounted because markets exist which generate well-defined prices for future resources in terms of present, and conversely. If one can also trade resources for life years in both the present and the future, through varying the scale and scope of "life-saving" programs, it follows that an indirect market exists in which present and future life years can also be traded, at the same relative price. So one should discount future life years. Second, individuals clearly display time preference -- in concrete (and informed) situations they indicate a willingness to trade expected life years in the future for years now at more than one to one. Cancer victims, if allowed to choose between surgery and radiation therapy, place more weight on short-run surgical risk and less on five-year survival rates than they would if indifferent to time (McNeil *et al.* 1978).

Against discounting, one might note that while particular investments in health care, now or in the future, may in fact represent the trading of lives for dollars, or conversely, that is a far cry from a well-developed futures market in life years. And the life years accrue to different people; a discounting of future life years relative to the present represents a weighting of different people's life outcomes. If the world were fully known, continuous and twice differentiable, all such discrimination could be compensated. In practice it will not be. Moreover it is difficult to separate time preference from diminishing marginal rates of substitution in the interpretation of individual behaviour, again because, for better or worse, one cannot choose to live either five life years next year or one each over the next five years. Life years may be a measure of outcome, but they are *not* commodities, and differ from such in crucial respects. (If you miss this year's life year, you've missed them all -- *pace* Hotblack Desiato.) If there is time preference in individual cases, it is not clear that it should necessarily be reflected in social decision-making (though in the cancer case, since the alternative techniques affect the same individuals, the case for reliance on individual preferences is much more compelling). And in any case, the absence of the necessary direct markets in physical outcomes implies that (real) rates of interest set in financial markets have no particular privileged position as discount factors, if such are to be used.

There is a more fundamental problem, however, with the QALY approach to program evaluation. Where do the weights come from? While it seems beyond dispute that health is better than illness, and some forms of illness or disability are more serious than others, the use of these insights in project evaluation requires their quantification. How does one determine the relative valuations which people attach to life years spent in different circumstances?

One can construct hypothetical scenarios, explain them to subjects, and ask them which situations they would prefer. There are, of course, a number of difficulties with such techniques, including identifying "representative" subjects -- patients with the condition to be studied, providers of associated care, M/WITS? -- and ensuring that the respondent has sufficient information to respond. Alternative techniques include the time trade-off, in which subjects are asked to indicate how many years of life with a particular condition (from the present) they would regard as equivalent to a standard number of years of normal health (followed in each case by painless death). Standard gamble techniques confront subjects with a choice between a particular condition (for a specified period) and a gamble between present (painless) death and

normal healthy future life. The odds in the life/death gamble are adjusted until the subject is indifferent between the choices; the higher the acceptable risk, the worse the condition.[22]

Such techniques are still in the experimental stage. So far they have shown some promise, but are far from representing a finely calibrated tool of measurement. Certain interesting regularities seem to emerge, however. Sufferers in a state of illness seem to regard it as less bad than non-sufferers, suggesting either that anticipation is worse than reality, or that people adapt. To which preferences should cost-effectiveness analysis respond? Moreover, some states appear *worse* than death (severe graft versus host disease following bone marrow transplant, for example). Clearly one should mark down the effectiveness of a program, one of whose possible outcomes is that "beneficiaries" would rather be dead, and an arbitrary cut-off of weights at zero would over-estimate its benefits.[23]

PRIOR CONSIDERATIONS: IF THE PROJECT DOESN'T WORK, EFFICIENCY IS NO HELP

Cost-effectiveness and cost-utility analyses share with cost-benefit analyses the characteristic that they take as given the sets of B_{ikt} and C_{jkt} associated with a project under study. The economic dimension of project evaluation, like economic analysis generally, assumes that the technical, engineering, or clinical work has all been done, and that full information on the technical transformation of inputs into outputs is available before the economist arrives on the scene. If any concession to incomplete information is made, it is in the form of an assumption of probabilistic outcomes, in which the range of possible outcomes and their associated probabilities is well known, and various sophisticated techniques for decision-making under uncertainty can be applied.

But as the evaluator of real-life projects rapidly learns, this idealized situation is very far from the norm. The specification of a project, in terms of what is to be done and what are the expected outcomes in physical terms, frequently proceeds in parallel with the economic evaluation, though it is logically prior. This process can be represented by the triad of concepts, efficacy, effectiveness, and efficiency, which, while most readily applied to drug trials or other well-specified clinical interventions, can be used without too much stretching in the more general world of project evaluation.

Efficacy refers to the impact of an intervention under ideal or laboratory conditions. It is related to the economist's *ceteris paribus* assumption. If all confounding influences are held out of play, and if the drug is administered correctly or the project managed competently, will the desired result be achieved? In our terms, is it possible that a particular set of resources C_{jkt}, manipulated according to present best practice, could yield amounts B_{ikt} of a specified set of benefits? If not, then no further analysis, economic or otherwise, is needed. Projects which do not work are not worth doing at any (positive) price. (Recall the stool guaiac example).

But the point is far from trivial, for two reasons. First, though logically there is no point in the economic evaluation of a project for which efficacy has not been established, in practice the available information is usually rather fuzzy. The range of positive (and negative) effects of a program, and their quantitative significance, will all be open to a greater or lesser degree of doubt, and the economic analyst will frequently have to evaluate (or seek help in evaluating) the quality of the technical data as well. Evaluation may have to be conducted for several possible outcome scenarios, not just for "the" B_{ikt} "known" to be the project result.

Second, attempts to explore the bases for efficacy estimates frequently reveal alternative or additional objectives for a project or proposal. Not infrequently, a project is advocated by a particular group, either because they expect to benefit from it, or because they have views as to the interests of the wider society which they feel those responsible for funding public projects may not share. The project is then justified in terms of alleged benefits, efficacy, which if taken at face value by the economic analyst might well satisfy equation 11-1 or its cost-effectiveness equivalent. But a certain amount of probing of the bases for efficacy estimates can often reveal inconsistencies or unjustified extrapolations or interpretations of existing technical data. While this may not be within the sphere of economic analysis narrowly defined, there is no point in carrying out elaborate analyses on unsound or non-existent efficacy foundations. A little checking helps, and the economist or any other analyst who goes into program evaluation without some understanding of the technical aspects of the project is at risk of looking rather foolish.

Effectiveness extends the concept of efficacy from the laboratory to the field. Efficacy is a necessary, but not a sufficient condition for effectiveness; a program or intervention which will work under ideal conditions or with ideal management may simply be unworkable under field conditions. While efficacy is in principle a purely technical issue, effectiveness has an economic dimension in that one may make more or fewer resources available for program implementation. Two different immunization programs might be built on equally efficacious interventions -- both using the same vaccine, for example, which has been demonstrated under controlled conditions to confer a specific immunity -- but in terms of achieving a particular percentage of population immunity, or drop in the illness rate, one program might be effective and the other not. The difference might be simply one of resource input, effort level, or it might be a difference in organization and approach. The effectiveness of clinical programs, in particular, depends on the extent of patient compliance with therapeutic recommendations, as well as on the success with which the target population can be located and assembled. It is not uncommon in health care to find that better organized programs can be both more effective and less costly. As in the case of efficacy, however, an explicit focus on effectiveness forces one to address the question -- effectiveness for what? -- and can smoke out a number of collateral or alternative objectives which may be smuggled in along with the ostensible benefits of a public program.

The efficiency component of the triad should be self-explanatory; it has been the subject of most of this chapter. Once efficacy and effectiveness have been established, the *B* and *C* patterns specified, then various forms of economic analysis can be applied to determine if the project is worth doing. Such cost-benefit, cost-effectiveness, or cost-utility analyses, however, form only a part of the overall program evaluation problem, and particularly in the health care field one can argue that the problems of efficacy and effectiveness evaluation are at least as challenging. If we know whether the program does what it was supposed to do, the economic analysis is often fairly straightforward. But the clinical and epidemiological problems of establishing what works, under what circumstances, create the real difficulties.

When the appropriate cost and outcome parameters of a project are, as is usually the case, shrouded in uncertainty, the recommended response is sensitivity analysis. The doubtful parameter or parameters are assigned a range of values plausible to the investigator or her expert advisors. Benefits of type *i*, for example, accruing to people *k* in time *t* may be doubtful, but firmly believed to be no lower than B^-_{ikt}, and no higher than B^+_{ikt}. Is the project present value positive at $B_{ikt} = B^-_{ikt}$? Or negative at B^+_{ikt}? Or the discount rate may be varied from 2 percent (or zero) to 8 percent -- does the project pay off (or fail to) for all values?[24]

If moving from "best guesses" to "plausible ranges" for program parameters fails to modify an initial go or no go finding, then obviously one can be more confident about the robustness of

the decision with respect to uncertainty. On the other hand, if the evaluation turns out to be very sensitive to the assumed value of a key parameter or parameters -- project payoff or input requirement, shadow price, or discount rate -- about which there is substantial uncertainty, then obviously one's findings are correspondingly doubtful. The response may be to invest more effort and resources in establishing the value of the crucial parameter(s), but if this is not possible, then the appropriate decision rule is, "When in doubt, don't." If the project is to go ahead, it must be on the basis of some other considerations.

Such "other considerations" -- the need to be seen to do something, continuation of past practices, the desire to generate earnings or to display professional prowess, perhaps above all the need to *act* in unsatisfactory situations -- seem to underlie a remarkably large part of health policy as well as health care services. While formal project evaluation has made enormous strides in the past decade in both methodology and range of application, there remains a great gulf between what is learned from such analysis and what is done.

Not least, of course, this is due to the fact that such analyses are threatening to clinicians, administrators, bureaucrats, and politicians alike. They call into question existing patterns of activity and raise embarrassing questions, not only from outsiders, but in one's own mind. Accordingly, while there are some clear cases -- hypertension screening, for example -- in which formal evaluations have given clear-cut answers that are translated into clinical practice, more generally the results of project evaluation, at least in health care, are ignored or resisted.

And a large proportion of health care interventions are innocent of any formal or scientific evaluation at all, not just of efficiency, but even of efficacy. It is remarkable, and may be without parallel in human activity, that so much effort and resources are devoted through public or private channels to health care whose effectiveness has not been conclusively demonstrated, at least for the purposes claimed, either in the setting of application or at all. Odd.

NOTES

[1] At present the informational role of the state, like that of providers, is difficult to separate from that of marketing. Warnings on cigarette packages are intended to modify behaviour, not to inform, just as cancer or heart disease awareness campaigns or school dental education programs are intended to encourage increased use of practitioner services. No one has ever launched a campaign saying "*Don't* see your doctor or dentist under the following circumstances."

But in principle, at least, the genuinely informational role of the state could be significantly expanded in support of either providers or individual consumers. Certification of members of occupational groups, for example, can serve in place of licensure by providing consumer/patients with authoritative information about provider qualifications while permitting them to choose whether or not to deal with the uncertified. Licensure prohibits such dealing. "Technological assessment," such as the federal reports on cervical screening and periodic health exam, can improve information available to practitioners on the effectiveness of interventions. More broadly, if future health policy involves the development of alternative systems of care competitive with fee-for-service or hospital inpatient care -- Health Services Organizations, Surgicentres, Self Care, Home Care, or Care-by-Parent programs -- the role of the state in disseminating, or at least facilitating, the flow of information may have to be greatly expanded.

[2] What follows is not intended to be a comprehensive survey of the now vast field of cost-benefit and cost-effectiveness analysis. Nor is it a "how-to" guide. For more comprehensive treatments of the principles involved see Drummond (1980) and Warner and Luce (1982). Drummond (1981) also provides a catalogue and brief description of evaluation studies in health care.

The treatment in this chapter owes a great deal to discussions with Greg Stoddart, David Sackett, and other members of the Department of Clinical Epidemiology and Biostatistics, and to work by George Torrance, all at McMaster University. None of them should be in any way implicated in remaining inadequacies or distortions.

[3] The similarity of this discussion to the general description of economic analysis in chapter 1 should be apparent.

[4] Cost offsets can lead to a technical problem, though they need not. If a new project leads to cost-savings elsewhere in a system -- real ones, not just redeployment of resources -- then it obviously makes no difference whether such savings enter the first term of equation 11-1 as a positive benefit, or the second as a negative cost. But in some applications a benefit/cost *ratio* is defined, the first term divided by the second, and this ratio is sensitive to how cost offsets are classified. A program with (present value of) benefits *B*, direct costs *C*, and cost offsets *K* has a present value (*PV*) of $B - C + K$ but a benefit-cost ratio of either

$$\frac{B + K}{C} \quad \text{or} \quad \frac{B}{C - K}$$

Shifting *K* between numerator and denominator cannot move the ratio from above to below unity, or conversely, so the absolute (un) desirability of a project is unchanged, but if the *relative* values of projects, their priority ratings, are based on the ratio, then the analyst can fiddle the rankings. If the budget for all projects is constrained such that not all positive *PV* projects are possible, the better procedure is to choose a project "portfolio" of maximum *PV* subject to the budget constraint, not to rely on benefit/cost ratios to set priorities.

[5] Strictly speaking, *R* is itself time dependent, and future effects should be discounted using the time-pattern of *R* between now and their arrival -- preferably in continuous time. But in fact no one forecasts interest rates well enough to justify specifying a function $R(t)$, and benefits and costs tend to be reported by budget year. With increasingly sophisticated cash management the time profile of *R* will become more important, though not, alas, any easier to predict!

[6] The inflation-adjusted or real rate of interest is calculated from the ratio of the nominal to the expected inflation rate, thus

$$\left[\frac{1.20}{1.12} - 1\right] = 7.14 \text{ percent.}$$

The subtraction approximation of 20 - 12 = 8 percent becomes more inaccurate as inflation rates rise.

[7] There is, as always, more to the story than this. Private-sector returns on capital (after-tax, risk-adjusted) run well above 4 percent. But governments do not borrow at such rates.

[8] Complications arise as projects can be carried out at different scales, different projects interact with each other, and the *P* and *V* may be functions of the *B* and *C*, often negative and positive respectively.

[9] If input markets were perfectly competitive, then resource owners could all sell their resources of labour, skills, or capital to any one of a large number of potential buyers at "the" going market price. They would be undetectably affected by the presence or absence of any one program. This discussion therefore rests on an assumption of imperfect competition in the relevant input markets, monopoly power as well as quasi-rents to investment in industry-specific skill acquisition. A perfect competition assumption is wholly out of place, and a policy analyst who relied upon it would get some nasty surprises!

[10] It is possible, as noted above, to vary the *P* and *V* in equation 11-1 explicitly across *k* so as to weight differentially the interests of particular people and groups. Projects would be favoured which paid off to those groups. While logical, this approach has not been widely used. Its weakness may be that while formally and overtly we would all wish, and wish our political system, to assign high weights to the disadvantaged, poor, elderly, in practice program choices frequently follow the interests of the wealthy, highly organized, and politically powerful. The most potent combinations are programs which appear to, and perhaps do, benefit the disadvantaged, while also providing payoffs to politically powerful interest groups. There are also, however, social costs which result from making redistributional processes, in whichever direction, too overt in operation.

[11] Volunteer labour presents a special conceptual problem, which illustrates the concept of opportunity cost. The argument for shadow-pricing such labour in any particular program is that, if not used in this program, it could have been used elsewhere. Just like wage labour, volunteer labour has a cost in terms of other productive opportunities foregone and its V_j value should reflect this. If volunteer labour substitutes for a category of wage labour, the market wage for that category would be an appropriate shadow price. On the other hand, it is conceivable that the volunteer is willing to serve only in one capacity. A parent on a care-by-parent ward is there to look after his/her child, and no one else's. The volunteer is, of course, foregoing other earning or leisure opportunities, but presumably the satisfaction of looking after one's own child outweighs these costs (or the child would be placed in a regular ward). Thus one should not shadow-price the one-activity volunteer -- the opportunity cost is zero -- but should shadow-price *e.g.*, the members of a religious nursing order.

[12] Whether its value is positive or negative is also rather sensitive to whether or not the evaluator is a Muscovite.

[13] It is only the person's earned income, from labour and skills, which is counted. Income from other assets such as bonds or buildings is unaffected by the death; it is simply transferred to someone else.

[14] It also requires a number of intricate and sophisticated calculations, largely impenetrable to the user of evaluations, which is an advantage for analysts.

[15] Yet particular program evaluations relying on this measure continue to turn up. Erroneous analytic techniques, like inefficacious surgical procedures, take time to work through a group of practitioners!

[16] Thus we find calculations of the "costs" of thalidomide which assign a negative value to phocomelia equal to the present value of the stream of (lost) market earnings of the disabled newborn -- which of course does not start for about twenty years! if in addition one overstates the discount rate as described above, the costs of disabling conditions in the disabled newborn become almost trivial. It is best to put such calculations in an appendix.

[17] Where resource constraints are absolute, *e.g.*, in the case of a life-saving technique in fixed supply and excess demand, allocation rules may be used which are based on age, family situation, etc. But these are related only indirectly to the present value of the expected earnings stream, and as discussed below have other and better bases.

[18] A classic example of the scale problem, and the importance of marginal analysis, is given by Neuhauser and Lewicki (1976). A very inexpensive test for occult blood in the stool will detect a pre-cancerous condition, but it has a false negative rate of about 5 percent. To reduce the chance of missing a "curable" case, one can perform the test several times. The clinically recommended technique of performing the test six times reduces the probability of failure to find a true positive case to $(.05)^6$, and still represents a relatively low cost of testing per person tested. But the probability of finding a case on the sixth test, which was not already found during the first five, is $.95\ (.05)^5$, so the cost per *new* case found, on the sixth test, would be $1 \div [.95\ (.05)^6]$ or 3,040,000 times the cost per test. The marginal cost per case found, on the sixth test, could easily run into the tens of millions of dollars. Maybe six times is too many.

The stool guaiac test studied by Neuhauser and Lewicki has since provided a second lesson. Their analysis rests on the assumption that a true positive finding, at whatever cost, leads to an effective intervention -- "cure." But as of early 1984, research currently underway suggests that intervention is *not* effective, and contrary to previous belief has no effect on life expectancy (David Sackett, personal communication). If so, then the test has no value, the optimal number is zero, and cost data are irrelevant. In this case, better data on therapeutic efficacy have dissolved what appeared to be a "grim trade-off" between lives and dollars, providing an illustration of the general point made previously. Too much fascination with the philosophical and political issues in such trade-offs distracts attention from the possibilities for unambiguous improvements in efficiency.

[19] Sadly, as the military examples show, deaths or megadeaths *are* well-defined consequences.

[20] The present value of income net of consumption for a sixty-four-year old is already large and negative; his consumption during retirement is counted as a net loss to the rest of society.

[21] Despite absence of evidence.

[22] The evaluation of states of ill-health will generally vary from person to person, though if the approach is to be convincing they should be roughly consistent. Weights used in a project evaluation will have to be some average of those of the people questioned. The same aggregation problem, of course. arises with efforts to monetize the value of life and limb; measurement is taken over a population which may not accurately represent the population at risk.

[23] It is not inconsistent with these findings that people in such circumstances continue to live. The act of terminating life, one's own or others', is "costly" in utility terms, so that one might perfectly consistently wish to be dead yet refrain from suicide.

[24] Sensitivity analysis on a parameter-by-parameter basis is of course a crude version of the more formal analysis with explicit probability distributions. As outlined, it assumes rectangular and independent distributions on each parameter, and fails to consider the implications of joint probability distributions even in the independent case. What if several uncertain parameters *all* take extreme values? A formal analysis of probability contours in a multi-dimensional parameter space would be much more elegant; it is also very uncommon. The reason is only partly that analysts (to say nothing of their readers or clients) are not familiar with the necessary "statistical tap-dancing"; there is also a strong (and probably correct) feeling that our relatively crude knowledge about the values of the components of equation 11-1 does not justify the extra effort.