Introduction to health economics for physicians

Martin I Meltzer

Since the 1960s, expenditure on health care in developed countries has risen faster than the general rate of inflation, thus making economic assessment of interventions an integral part of decision making in health services. This paper is the first in a series whose goal is to provide some basic principles of health economics that will allow practising physicians to understand better the economic relations between their practice of medicine, the health-care sector, and the national economy. Some of the most important principles described in this paper include opportunity costs, identifying the appropriate perspective, correctly categorising costs, and discounting costs and non-monetary benefits (eg, lives saved) over time. Economic analyses of medical interventions must also take into consideration the difference between efficacy and effectiveness. Efficacy is the maximum possible benefit, often achieved with carefully controlled trials, and effectiveness is the actual decrease in disease achieved when the intervention is applied over a large, non-homogeneous population. This introduction ends with three methods of assessing the costs and benefits of an intervention—namely, cost-benefit, cost-effectiveness, and cost-utility analyses.

Economic assessments have become an integral part of policy decisions on the use of health-care technologies. This has happened because, since the early 1960s, expenditures on health care in developed countries have often risen faster than the general rate of inflation. Consequently, health-care expenditure has been consuming an ever-larger share of the total economy (figure 1), and politicians have to balance expenditure on health with spending on other sectors, such as education. What is the optimum level of expenditures on health care, and how would society define what is optimum?

Figure 1 does not explain the links between policymakers and individual physicians, and how such links are used to determine the appropriate level of expenditure on health care. These links are often unclear. For example, in the USA the Federal Government’s public health agency, the Centers for Disease Control and Prevention (CDC), recently recommended that all children under the age of 5 who live in states where the incidence of hepatitis A is well above the national average should be routinely immunised against hepatitis A virus. Children under the age of 5 years, however, most commonly have asymptomatic disease. Paediatricians and parents may therefore question the economic wisdom of such a recommendation when it comes to vaccinating their patients or children against hepatitis A. The answer is that both the patients and society could benefit economically because vaccinating young children might protect them at a later age, when their risk of having symptomatic hepatitis A is likely to increase. Society might further benefit economically because vaccinating young children might greatly reduce the transmission to, and thus frequency among, older siblings, parents, and caregivers.

Similarly, physicians and their patients might not readily understand why a health insurance plan, or a government programme, such as the USA’s Medicare (which pays for many health-care costs for those over 65) or the British National Health Service, will not pay for a test even if the patient is at low, but not zero, risk of having contracted the disease. Physician and patient might argue that the use of the diagnostic test could help remove a doubt, however small, and thus eliminate a source of anxiety (surely an important outcome). An insurance plan or government health scheme, however, might respond that they do not have unlimited resources and they must focus on items with “high priority”. An obvious question then is: How does a “higher priority” get defined? Later in this Lancet series, cases studies will allow clinicians to understand more about the economic relations between their practice and the health-care sector and national economy—that is, to learn how economics is used to set priorities in health-care budgets. But to begin with, here are some principles and definitions in economics.

**Financial versus economic analyses**

There are at least two methods that can be used to assess the economic effect of a health-care intervention, the financial and the economic. For example, a physician assessing the financial viability of adopting a new diagnostic test might use as part of the analysis his or her...
usual charge for an office visit eg, US$30. However, an economist looking at the same test will want to assess it in terms of opportunity costs, or what alternative investments could be made with the same health-care resources. The economist might note that large insurance companies and government-run health plans often reimburse physicians at a much lower rate than those physicians would usually charge (eg, $15 for the $30 visit). The lower value ($15 per visit) would be termed the opportunity cost.

The difference arises, in this example, because insurers and government health plans can negotiate lower prices by offering volume. The fact that physicians (and other health-care providers such as hospitals) accept the lower reimbursement levels tells an economist that such reimbursements are indicative of the "true" costs. Of course, the health-care providers will argue that they have no choice but to accept less money. Such arguments will probably continue for as long as there is health care. But the key notion is that the marketplace price for an office visit is not always an accurate measure of the value of the resources used to administer such a visit. There is a wealth of economics literature on how to estimate opportunity costs when economists suspect that the market price is wrong but all physicians need to know is the basic fact that the prices that they might use to run their businesses (financial costs) might differ notably from an economist’s valuation of the resources used. When used in the proper context, the economist’s valuation can be correct.

**Perspective**

The benefits and the costs of using an intervention to prevent or treat a disease depend upon whose perspective is. Differences due to perspective taken are one of the main reasons why there might be disagreements between patients, physicians, health-care payers, and policy makers with respect to the value of using a particular intervention. The differences on which costs are included and excluded with different perspectives are shown in panel 1. For example, for an insured patient who is considering being vaccinated against influenza, the cost of the intervention would be the time off work to go and get vaccinated, any travel costs to and from the site of vaccination, and any cash payment needed to make up the shortfall in what the insurer will pay for (co-payments). The patient would also bear some of the costs in terms of vaccine-related side-effects, as well as costs associated with disease should the vaccine fail to fully protect (no vaccine is 100% effective). The benefits to the patient include avoiding time lost from work because of illness, cost of travel to and from the doctor and the office in the event of illness, and any co-payments for the physician visit and medications. Without insurance, the cost of physician visits and medication is borne entirely by the patient. Also, these costs to the patient, unless he or she is self-employed, will to some extent depend on the employer’s sickness-absence policy.

From the perspective of the payer, the health-insurance company or state health insurance plan, the costs of vaccinating that patient include the vaccine itself, the administration (ie, a physician’s or nurse’s time), treatment for any vaccine-related side-effects, and also the costs of treating a patient when the vaccine fails and the patient gets influenza. The benefit to the payer is the avoidance of having to pay any treatment costs should the patient become ill from influenza (providing the vaccine works) and, in some circumstances, there might be the benefit of the protection that vaccinating one person offers to others. The payer does not, of course, bear the cost of the time lost by the patient from work, household chores, and social activities. A policy maker, on the other hand, might consider the societal

### Panel 1: Inclusion and exclusion of costs, dependent on perspective for economic analysis

<table>
<thead>
<tr>
<th>Examples of costs</th>
<th>Include (+) or not (-) dependent on perspective (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient(b)</td>
</tr>
<tr>
<td><strong>Direct medical</strong></td>
<td></td>
</tr>
<tr>
<td>Physician time</td>
<td>Yes</td>
</tr>
<tr>
<td>Other medical personnel time (eg, nurse, technician)</td>
<td>No</td>
</tr>
<tr>
<td>Drugs</td>
<td>Yes</td>
</tr>
<tr>
<td>Medical devices (eg, syringes, ultrasound)</td>
<td>No</td>
</tr>
<tr>
<td>Laboratory tests</td>
<td>No</td>
</tr>
<tr>
<td><strong>Direct non-medical</strong></td>
<td></td>
</tr>
<tr>
<td>Administration(f)</td>
<td>No</td>
</tr>
<tr>
<td>Physical facility (eg, clinic, office)</td>
<td>No</td>
</tr>
<tr>
<td>Utilities (eg, telephone, electricity)</td>
<td>No</td>
</tr>
<tr>
<td>Patient’s travel costs</td>
<td>Yes</td>
</tr>
<tr>
<td>Temporary hired care-giver(g)</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Indirect</strong></td>
<td></td>
</tr>
<tr>
<td>Time off from work to visit physician</td>
<td>Yes</td>
</tr>
<tr>
<td>Time off work while ill and recuperating</td>
<td>Yes</td>
</tr>
<tr>
<td>Hire temporary household help while ill(h)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

(a) Inclusion of cost item will depend upon chosen perspective; four perspectives (societal is the sum) do not cover all possible perspectives.  
(b) Assumes patient is covered by health-care insurance; physician time and drug costs will involve co-payments.  
(c) Perspective assumed to be that of a physician employed by health-care provider such as hospital.  
(d) Third-party payer who reimburses physician for services rendered that are covered by an insurance scheme (private or public).  
(e) Sum of all perspectives.  
(f) Physician’s practice; hospital, etc.  
(g) Hired to look after family members while adult visits physician.  

perspective, adding up all costs and benefits irrespective of who pays and who benefits. This societal perspective is the most comprehensive one; all others are subsets of the societal perspective.

**Categorisation of costs**

In economic analyses, costs are typically categorised as “direct medical”, “direct non-medical”, and “indirect costs of lost productivity”. Examples are given in panel 1. In financial or accounting analyses, costs are classified differently, as “variable” or “fixed”. Variable costs, such as the physician’s time and drugs administered, vary dependent on the numbers of cases treated, whereas fixed costs do not vary in the short-to-medium term and are unlikely to change with any fluctuations in the number of cases (eg, the cost of a building). Some health economists do use accountancy terms such as “fixed costs” but this is not a serious problem so long as the costs that are included are those appropriate for the perspective chosen (panel 1) and reflect the opportunity costs. However, just one terminology, appropriate for the intended audience, should be used throughout a single analysis.

**Intangible costs**

In economic analyses, there is another category of costs, labelled “intangible”. These include entities such as pain and suffering, and fear, and widely accepted values for them are often difficult to find. A methodology called “willingness-to-pay” can be used to obtain these values, and such estimates are often obtained directly through surveys, although indirect methods are available. Willingness-to-pay estimates for the valuation of health technologies do have their problems, and some authors simply list the readily identifiable intangible costs and benefits that might be associated with the effect of a disease and an intervention. These non-dollar costs and benefits might become crucial in any public debate over the adoption of an intervention designed to prevent, treat, or control a disease.

---

**Discounting costs over time**

A key economic notion is that society places a premium on benefits gained in the present rather than at some time in the future. For example, most people would prefer to invest $1000 today, rather than wait for a year. A 1 year delay would mean losing the returns on investment for that year, and this idea is true even with 0% inflation. To reflect this preference for goods and services that are delivered now, both resources spent and benefits gained in the future are discounted when being compared with resources spent, and benefits gained, in the present. Discounting thus allows for the direct comparison of costs and benefits during different periods. The formula for discounting is in panel 2.

**Discounting non-monetary costs and benefits**

All future non-monetary costs associated with an intervention, such as future deaths delayed, should be discounted. Society also has a time preference for such non-monetary costs, and will usually value the life of somebody living now above the value of a birth at some future time.

**Efficacy versus effectiveness**

For an intervention to eliminate a disease or to cure every patient with a medical condition is rare. The maximum possible reduction in a disease due to the use of an intervention is termed the “efficacy” of the intervention. Efficacy is often measured with randomised controlled trials (RCTs). Such trials achieve the maximum possible reduction because patients are often selected on the basis of who will comply with the protocol, and because trial participants are often free of other diseases or conditions that might interfere with the intervention being studied. Furthermore, clinicians participating in RCTs are usually carefully selected for their interest and expertise, and they work with carefully trained staff who have time for follow-up and record keeping. The benefit that accrues from an intervention that is applied in day-to-day practice to a population larger than that taking part in the RCTs is

---

**Panel 2: Formulae for discounting and for calculating net present value**

To calculate the present value of a “stream” of costs or benefits that extend into the future, the following formula is applied:

\[
PV = \sum_{t=0}^{N} \frac{\$t}{(1+r)^t}
\]

Where

- \(PV\) = Present value
- \(\$t\) = dollar value of cost or benefit in year \(t\)
- \(r\) = discount rate, expressed as in decimals (eg, 3% = 0.03)
- \(t\) = time period, ranging from 0 to \(N\)
- \(N\) = maximum time period being examined

Suppose that a proposed infectious disease control programme will save $15 000 in direct medical costs every year for 5 years (first year = year 0). The PV of this “stream” of savings is:

\[
\frac{15000}{1+0.03^0} + \frac{15000}{1+0.03^1} + \frac{15000}{1+0.03^2} + \frac{15000}{1+0.03^3} + \frac{15000}{1+0.03^4} = ($70 756.58)
\]

The formula for calculating NPV is:

\[
NPV = \sum_{t=0}^{N} \frac{(\text{benefits-costs})}{(1+r)^t}
\]

where

- \(t\) = year, from 0, . . . , \(N\)
- \(N\) = number of years being assessed
- \(r\) = discount rate.
called “effectiveness”. The difference between efficacy and effectiveness can be large, and obtaining realistic measures of effectiveness is a challenge.

**Recognition of both benefits and harms**

Although we often focus on the benefits of an intervention, the associated harms must also be recognised. The difficulty is that many harms, and the costs associated with them, might take some time to become apparent or be so rare that they do not show up in the initial trials and are thus excluded from the initial assessment of an intervention. For example, Guillain-Barré syndrome has been associated with viral, bacterial, and other infections, as well as vaccinations. Although this syndrome is associated with a wide variety of outcomes, including partial paralysis from which a victim often gradually recovers, the risk associated with contraction of Guillain-Barré syndrome from vaccination is about one in a million. Thus, although the risk of this adverse reaction to certain vaccinations is small, the cost of managing the harmful outcome is so high that the side-effect should be considered for inclusion in an assessment.

**Assessment of economic costs and benefits**

The three main methods used to assess the economics of an intervention designed to control and prevent a disease are: cost-benefit analysis (CBA), cost-effectiveness analysis (CEA), and cost-utility analysis (CUA) (panel 3).

**Cost-benefit analysis**

For many applied economists, CBA is the “gold standard” by which the other methods are judged. In its simplest form, a CBA lists all the costs and benefits that might arise as a result of an intervention up to a prespecified time. These costs and benefits are discounted (see earlier) to the year zero. If the total discounted benefits are greater than the total discounted costs, the intervention is said to have a positive net present value (NPV) (panel 2). CBA is most useful under three circumstances. First, when a choice has to be made between two or more interventions, then the logical action is to give top priority to the intervention with the highest positive NPV. Second, a CBA can indicate the economic effect of a single intervention. Third, CBA is useful because it can include an array of important benefits or costs not directly associated with a health outcome, such as time off from work taken by family members to care for sick relatives. In a CBA all costs and benefits must be expressed in monetary terms, including the value of human lives lost or saved as a result of the intervention. Quantifying all the benefits and costs is not easy.

**Cost-effectiveness analysis**

A CEA expresses the net direct and indirect costs and cost savings in terms of a predefined unit of health outcome (eg, lives saved or cases of illness avoided). The total net costs, sometimes called incremental costs, of an intervention are calculated and then divided by the number of health outcomes averted to yield the total net cost per unit of health outcome (eg, net $ cost or savings per death averted). Many of the data required for an economic CEA, with a societal perspective, are the same as for a CBA, the most important exception being that in a CEA no value needs to be put on a life. Nonetheless, explicitly or implicitly, the value of a human life is part of the health outcome used.

A serious limitation of CEA is that there is no numerical valuation of the health outcome. For example, CEA can provide an estimate of the net cost of averting a case of poliomyelitis but it cannot help a physician, a patient, or a society to value each averted case, even in a seemingly similar outcome. How might a community value the prevention of life-threatening influenza in a 75-year-old versus the avoidance of poliomyelitis in a child? CEA is best used when comparing two or more strategies or interventions that have the same health outcome in the same population—eg, is vaccination more cost-effective than chemoprophylaxis in prevention of a case of influenza in people aged 65 or older?

**Cost-utility analysis**

CUA is a special form of CEA, in which the health outcomes in the denominator are valued in terms of utility or quality. A CUA, for example, might attempt to differentiate between the quality associated with an averted case of poliomyelitis and one of influenza. These non-monetary units of valuation include the quality-adjusted life year (QALY; panel 4) and the disability-adjusted life year (DALY). The result of a CUA is usually expressed as the total net cost per unit of utility or measure of quality (eg, net $ cost or savings per QALY gained).

As in a CEA, the value of life itself is implicit in a CUA since the value of life is part of the QALY denominator. One unresolved issue, however, is how to deal with time costs, such as time lost from work because of illness. Thus, when using CUA, the analyst should be explicit about whether morbidity costs such as lost productivity are included in the calculation.

There are other, more fundamental, difficulties with CUA. The techniques used to measure quality-of-life lost because of a disease (panel 4) often focus on long-term disabilities. Are QALYs, therefore, an appropriate tool to measure the value of interventions for infectious diseases such as influenza and dengue that cause short-duration illness in large numbers of people? A related
Panel 4: **Utility and QALYs**

QALYs measure the “usefulness” or utility of a particular health state and the length of life lived under that state. Is the value or quality of living a year with both legs paralysed due to poliomyelitis equal to, say, 0·65 of a year without the polio-induced paralysis? There are three basic methods for obtaining values of the utility of a defined health state: expert opinion, values used in previous studies, and surveys. Surveys can be direct or indirect.

**Direct**
Survey techniques include the “standard gamble”, “time trade-off”, and a “rating scale”. Typically, in the gamble method, an individual is asked to choose between a gamble and a certainty. For example, the gamble might be that there is a probability (p) of dying because of a medical intervention to alleviate polio-induced paralysis and, for the same intervention, a probability (1-p) of healthy life for 29 years; the certainty is to live, say, 30 years with polio-induced paralysis of the legs. During interview, p is changed until the respondent is indifferent between the gamble and the certainty. The p associated with that point of indifference (eg, 0·65) is then equivalent to the QALY of a person living with polio-induced paralysis.

**Indirect**
Here, questionnaires are used that split a health state into subgroups or domains such as opportunity (eg, social, cultural); health perceptions (self-satisfaction with health state); and physical function (eg, mobility, self-care). In every domain, a respondent rates the effect of the disease from a set of descriptions. For example, for mobility, a respondent might state that he or she can walk around the house and neighbourhood without help, but with some limitations. This response is assigned a preference weight, such as 0·9, on a scale of 0 to 1. The preference-weighted responses from all the other domains are then used to construct a single index.

difficulty arises when attempting to use CUA to compare very different diseases and health states. Is it feasible to compare, for example, the loss of utility due to diabetes with the loss of utility due to influenza? This problem becomes evident when “league tables” are constructed. Such tables typically rank interventions according to their cost-utilities, irrespective of the disease. League tables might be used, with all their faults, as a basis for health policy. However, many of these criticisms could be levelled at CBA too, especially when the willingness-to-pay methodology for valuation of intangibles is used.

**Mathematical models and economic analyses**
In an ideal world, economists would prefer to examine health-care technologies by application of a rigorous statistical procedure such as the Student t test or a regression equation, to “real world” data. Unfortunately, in the real world, even after well-done RCTs, the data needed to answer specific economic questions are lacking. Thus, many published articles, reporting economic analyses of health interventions use a particular type of mathematical model to simulate the conditions under which a technology might be used. “Decision tree”, “Markov model”, and “Monte Carlo model” are examples of the tools that can be used in any of the three types of economic analysis. The model chosen usually reflects the question to be answered, the type of data available, the intended audience, and the personal preference of the research group. Every mathematical method has its own advantages and disadvantages but the model chosen should accommodate the economic principles appropriate to the type of economic analysis being done.

**Sensitivity analyses**
Since one of the main reasons that mathematical models are used is because the researchers are trying to bridge gaps in data, it is always appropriate to ask—What if some of the assumptions used in the model were changed? That is, how sensitive are the results to changes in the underlying assumptions? One aim of sensitivity analysis is to find out which variables in the model “drive” the results. Some variables carry greater weight than others; for example, a sensitivity analysis might show that the time lost from work, and the value of that time, are the two most important variables used to determine the economics of routine immunisation against influenza. If the true values of time off work are unknown, a sensitivity analysis would suggest the need to fund research designed to find out how many days off work can be attributed to this illness.

This paper uses material that I have published elsewhere (listed in the suggested reading list), but with additions and revisions made specifically for this article.

**Background and other reading**

**Textbooks**


**Book chapters**


**Journal articles on methodological issues**

Drummond MF, Richardson WS, O’Brien BJ, Levine M, Helyland D. Users’ guides to the medical literature XIII: how to use an article on economic analysis of clinical practice (A)—are the results of the study valid? JAMA 1997; 277: 1552–57.

O’Brien BJ, Helyland D, Richardson WS, Levine M, Drummond MF. Users’ guides to the medical literature XIII: how to use an article on economic analysis of clinical practice (B)—what are the results and will they help me in caring for my patients? JAMA 1997; 277: 1802–06.

Stewart KJ. The challenge of cost-effective decision making. Fam Practice Manag 1996; July/August: 16–17.


Journal articles containing examples of economic evaluation


