Cost-benefit analysis for health

Zweifel, P; Telser, H
1. Introduction and motivation

Much of the health care services are transacted within the public sector, are financed by health insurance, or provided by members of professions where barriers to access are high. In these situations, observed prices are not informative in the same way as on open markets, where willingness to pay by consumers is at least equal to the market price and market prices reflect the value of the resources used in production. Cost-benefit analysis (CBA) is an imperfect substitute for market valuations. It holds the promise of indicating to policy makers whether a health program is worth its cost.

However, CBA requires a measure of valuation by consumers, which is hard to come by. For this reason, alternatives to CBA have been developed in health economics, notably cost effectiveness analysis (CEA) and cost utility analysis (CUA). These alternatives are discussed in section 2 and found deficient in important aspects. Section 3 is devoted to a particular challenge that health behavior seems to pose to the application of CBA, the often-heard quote being that consumers "do not give a trifle for health when healthy but are willing to spend their fortune when sick". It is shown that this does not necessarily reflect instability of preferences (which would render the application of CBA indeed difficult). Issues surrounding the measurement of true cost are discussed in section 4, while section 5 returns to the benefits side. The human capital approach is criticized for not being compatible with standard microeconomics. The alternatives are Contingent Valuation and Discrete-Choice Experiments (DCE). The latter is argued to be superior on several grounds. A rare empirical application of CBA is presented in section 6 and conclusions are offered in section 7.
2. Alternatives to CBA as applied to health

There are different approaches to evaluating health programs. They usually are distinguished by the units in which positive and negative effects of an intervention are measured. The simplest alternative is Cost-Effectiveness Analysis (CEA). Here, the effects are measured by natural units on a one-dimensional scale. This can either be a clinical parameter such as the lowering of blood pressure in mm Hg or the lengthening of life in years. CEA allows evaluating and ranking two or more mutually exclusive interventions, with the intervention having the lowest cost per unit of positive effect to be preferred. However, as soon as there is more than one health effect, CEA cannot be used. Furthermore, even if an intervention is cost-effective, it fails to answer the question of whether the program “is worth the money”, i.e. should be implemented at all.

Cost-Utility Analysis (CUA) goes one step further by taking the multi-dimensionality of health into account. CUA reflects both lengthening of life and change in health status. It uses a cardinal utility function defined in terms of health (but no other goods), which maps these two dimensions of health into a scalar index, which then can be compared to the cost of the program. The best-known and most frequently used index is quality-adjusted life years (QALYs).\(^1\)

To derive QALYs, all conceivable health states are evaluated on a scale from 0 (death) to 1 (perfect health). The other values are defined in such a way that for any number \(x\) between 0 and 1, a representative individual is indifferent between the following alternatives, “survive one year in a health state with a utility index of \(x\)” and “survive the fraction \(x\) of a year in a state of perfect health”. In this way, all health effects of an intervention are made comparable, permitting them to be aggregated into a single number, which can be interpreted as the ‘gain in QALYs’.

The QALY concept is easy to apply. However, it has no sound decision-theoretic foundations since it is based on several restrictive assumptions. Preferences for health states and trade-offs against other objectives must be stable over the whole life cycle, there must be risk neutrality with respect to length of life, and preferences w.r.t. health states with durations of zero have to be equivalent (see Zweifel et al., 2008, ch. 2.3). There are several empirical studies suggesting that the requirements of the QALY concept are in general violated (see Dolan, 2000 for a survey).

If QALYs are used as a utility index in a CUA framework, the decision rule is to pick the intervention, which maximizes the number of QALYs for a given budget. This rule is based on two fundamental value judgments,

\(^1\) Alternative concepts are disability-adjusted life years (DALYs), see World Bank (1993), and for a detailed exposition Murray (1994); or healthy years equivalents (HYEs), see Mehrez and Gafni (1989).
(1) The welfare of the affected person enters the collective decision rule exclusively through QALYs gained;

(2) It is irrelevant who experiences the increase in QALYs.

This makes CUA not compatible with the usual welfarist position adopted in economics, viz. that collective decisions should be based on overall utility of the affected persons and not only health-related utility (see Boadway and Bruce, 1984). The use of QALYs and CUA in general may be justified if the health budget is fixed. However, there is empirical evidence suggesting that the distribution of QALYs matters, in the guise of a moderate or even strong preference in favor of the young rather than the old [see e.g. Johannesson and Johansson (1997a, b) for a specific case; or Schwappach (2002) for an overview]. In principle, it is possible to account for the distribution of QALYs in a generalized CUA framework. However, this requires considerably more information than a basic CUA.

Compared to CEA, CUA has the advantage of being applicable both to medical interventions of different types as well as non-medical interventions because it makes effects measured on different (e.g. clinical) dimensions comparable by mapping them into a single utility index. It can be used to construct so-called ‘league tables’ of medical interventions, which have repeatedly shown that similar increases in QALYs can be achieved at very different costs. This type of information is useful to political decision makers who are responsible for allocating a fixed budget within the health care sector.

However, CUA has its limitations, too. Besides weaknesses in its decision-theoretic foundations and its neglect of the distribution of QALYs among the members of society, CUA is conditional on whose utility function goes into the determination of QALYs. Furthermore, like CEA, CUA provides only a rank ordering of mutually exclusive interventions but fails to provide an answer as to the cutoff value of cost per QALY above which realization of a program cannot be justified. This is important as soon as the total amount of money to be allocated to health is not predetermined.

In Cost-Benefit Analysis (CBA), both costs and benefits are measured in money, making it suitable to evaluate interventions occurring inside or outside the health care sector and having any health and non-health consequences. If the money value of benefits exceeds that of cost, implementing the program is worthwhile. The application of this rule can be justified using the welfare economic criterion of ‘potential Pareto-improvement’ [also called ‘Kaldor-Hicks’ criterion, see, e.g., Zweifel et al. (2008), ch. 2.4].
3. Why CBA is particularly challenging in the case of health

The crucial feature of health is that it does not constitute a tangible stock that can be held, controlled, and traded [see, however, the stock formulation by Grossman (1972)]. This means individuals cannot sacrifice ‘health’ in exchange for something else – apart from the hesitation of most to say that they are consciously opting for a level of health that is below the maximum attainable. However, individuals do trade a (small) reduction in the probability of being healthy in the future against other objectives (that will be called ‘consumption’ for simplicity). Let \((1 - \pi)\) denote this probability during a short period (a week, say). Accordingly, the probability of being in bad health is \(\pi\). Let the individual use expected utility \(EU\) as the decision criterion, with utility derived from consumption depending on whether one is healthy \(h\) or sick \(s\). Therefore, expected utility is given by

\[
EU = \pi \cdot u_s[C_s] + (1-\pi)u_h[C_h], \quad \text{with} \quad u_h[C] > u_s[C].
\] (1)

Therefore, at a given level of consumption, utility in the healthy state exceeds utility in the sick state – a very natural assumption. Now an indifference curve in \([(1 - \pi), C]\)-space can be constructed by varying \(\pi, C_s, \) and \(C_h\) while holding \(EU\) constant, resulting in the condition,

\[
dEU = 0 = d\pi \cdot u_s[C_s] + \pi \frac{\partial u_s}{\partial C_s} dC_s + (1-\pi) \frac{\partial u_h}{\partial C_h} dC_h + (1 - d\pi) u_h[C_h].
\] (2)

This can be solved for the slope of the indifference curve (setting \(dC_s = dC_h := dC\) and writing \(u'\) for marginal utilities),

\[
\frac{dC}{d(1-\pi)} = -\frac{u_h[C_h] - u_s[C_s]}{u'_h[C_h] + (1-\pi) u'_s[C_s]}.
\] (3)

This slope (the marginal rate of substitution MRS) is negative since \(u_h[C_h] > u_s[C_s]\), as stated above. Moreover, utility increases with consumption in both states \((u'_h, u'_s > 0)\). Without further justification, assume MRS to be decreasing as usual, resulting in convex indifference curves (see Zweifel et al. [2008], ch. 2.4 for details). Now consider a sequence of \(T\) periods, in each of which there is a probability \((1 - \pi)\) of being in the healthy state. It follows from the binary distribution that the expected number of periods in the healthy state (before a change to the sick state occurs) is \(ET_h = 1/\pi\). Since this is a monotonic transformation of \((1 - \pi)\), indifference curves can be drawn in a \((C, ET_h)\)-space rather than in \((C, 1 - \pi)\)-space without loss of generality (see figure 1). In the following, the argument will be couched in terms of consumption \(C\) and expected number of healthy periods \(ET_h\).
Up to this point, the existence of a MRS between consumption and ‘health’ (more precisely, the expected number of healthy periods over a future comprising $T$ periods, with $ET_h \leq T$) has been established. But this MRS is nothing but the individual's marginal willingness-to-pay (MWTP) for more time in good health because it indicates how much in terms of consumption (or income) he or she is prepared to give up for more time in good health. According to the model presented here, MWTP is predicted

1. to depend on the initial endowment in terms of consumption and expected number of healthy periods [originally chances for good health $(1 - \pi)$];

2. to decrease when $ET_h$ approaches its maximum value of $T$;

3. to increase but at a decreasing rate when a larger gain in $ET_h$ is being considered.

It can also be shown that (4) MWTP increases with initial consumption (or equivalently, wealth; see Zweifel et al. [2008], ch. 2.4 for details).

Now, members of the medical profession and many laymen would doubt the existence of a stable preference field as displayed in figure 1. Indeed, human behavior often is decried as, "Not giving a trifle for health as long as one is healthy, but willing to spend one's entire fortune once one is sick". This seems to be evidence in favor of unstable preferences, with MWTP strongly state-dependent. If true, the argument would severely limit the usefulness of evidence with regard to WTP for policy purposes. However, it can be shown that the state-dependence of observed MWTP may follow from a state-dependence of the transformation curve between $C$ and $ET_u$.

In figure 1 consider the transformation curve with index $h$, holding for a current period in good health. From the origin, the curve starts with a positive slope because when $(1 - \pi)$ increases, the expected number of healthy periods $ET_h$ increases as well, providing for healthy time that serves either to generate labor income (which can be used to finance consumption goods) or leisure time (which can be used to produce consumption services). Thus, for small values, healthy time has the character of an investment good, permitting to increase consumption services as well. The transformation curve reaches its maximum where the additional resources spent on increasing the chances of being in good health (and hence, expected healthy time) result in an equivalent gain in resources in terms of time available for work and investment in health (see Zweifel et al. [2008], ch. 3.3 for details). As is evident from figure 1, at the optimum (symbolized by $Q^*$), health has turned into a consumption good, implying that more time spent in good health entails a sacrifice in terms of consumption.
Figure 1 contains a second transformation curve relating to sickness \(s\) during the current period. In bad health, one’s productivity both in the labor market and in household production is reduced for at least part of the planning period \(T\). Accordingly, this transformation curve runs lower and reaches its maximum value \(ET_h\) (and also \(C\)) sooner than its counterpart for good current health. In order to demonstrate that this difference rather than an instability in preference may be the source of observed state-dependence of MWTP, two assumptions are made:

(a) preferences are homothetic;

(b) sickness in the current period affects the individual as a producer of future healthy time more strongly than as a producer of future consumption.

*Figure 1: State-dependence in the production of healthy time*

Given these assumptions, the two optima \(Q^*\) (good current health) and \(Q^{**}\) (bad current health) can be compared as follows. When moving towards the origin on the ray \(0Q^*\), the slope of the indifference curve remains the same according to assumption (a). However, the slope of the transformation curve must become steeper on average for assumption (b) to be satisfied. This means that the optimum \(Q^{**}\) cannot possibly lie on the ray through the origin; indeed, it must lie above that ray, where the slope of the indifference curve is greater (in absolute value). However, this also implies that at \(Q^{**}\), the marginal rates of transformation and substitution must be larger than at \(Q^*\). In other words, the revealed MWTP for additional healthy time is greater if the current period is one of bad health than if it is one of good.
health. The observed MWTP thus turns out to be state-dependent not because of any instability of preferences but because of the dependence of productive capabilities on current health status, which seems a very natural assumption.

**Conclusion 1:** The observed instability of revealed marginal willingness-to-pay for health (low when healthy, high when ill) need not be caused by an instability of underlying preferences but may well be caused by the dependence of individuals’ productive capabilities on the current state of health.

Thus, the argument that individuals' preferences with regard to health and health care are too fickle to provide a basis for policy making need not be accepted. This is not to say that measuring true MWTP for health is without problems, quite to the contrary. While in principle, it is possible to infer individuals' MWTP from the equality between the marginal rate of substitution and the marginal rate of transformation (and hence actual behavior) at the optimum points $Q^*$ and $Q^{**}$, there are at least two reasons for important deviations. First, the rate of transformation is biased because of insurance coverage. With a coinsurance rate of 20 percent (say), the sacrifice of one Dollar’s worth of consumption in effect buys five times as much medical care as without insurance coverage. Thus, the transformation curve looks five times flatter than in figure 1 to an insured individual who considers using medical care. The second reason is that as soon as individuals rely on medical care, they are not likely to actually reach the optimum $Q^{**}$ (bad current health). This is due to the fact that they act under the influence of physicians, who pursue their own interests. To the extent that physicians are not perfect agents of their patients, they will misrepresent the trade-off between $C$ and $E_{Th}$ and/or influence their patients’ MRS, both of which cause $Q^{**}$ to be missed.

### 4. Issues with measuring true cost

In CBA, the costs of health care programs fall in three categories. First, there are direct costs arising from the use of resources within the health care sector, in other sectors and by patients and their families. Second, there are indirect costs due to productivity changes. Third, the so-called intangible costs as from suffering or pain. It is both important and difficult to avoid double counting, taking into account only costs that are caused by the program and not incorporated in the benefit measures. If for instance an intervention reduces the pain of patients, intangible cost of the disease is reduced. This would be accounted for separately in a CUA. However, in a CBA derived from experiments (see next
section), these cost reductions are already incorporated in the WTP values. There are additional issues to bear in mind when measuring cost of health programs as part of a CBA.²

**Vantage point of the analysis**

A natural question to ask is, “Whose costs, whose benefits?” This may be society as a whole, a third-party payer, a physician, a hospital, or a patient. Clearly, costs and benefits differ between these vantage points. While, e.g., travel to obtain care is clearly a cost component from the patient’s and society’s point of view, they do not have to be taken into account from a hospital’s point of view. Since CBA may serve different kinds of decision-makers, this is not a weakness but reflects the versatility of the tool.

**Direct costs**

Direct costs represent the value of all resources consumed by the intervention analyzed, including the neutralization of its side effects and other current or future consequences. There are two different types of direct costs.

Direct health care costs include physician services, hospital services, and drugs, used. Direct non-health care costs accrue outside the health sector, such as care provided by family members and transportation costs to the hospital. In quantifying direct health care costs, it may be problematic to use readily available hospital or physician charges, since the health care market is not a competitive one, causing the charges to be a poor reflection of true (opportunity) cost (Finkler, 1982).

Different studies explored this issue in the context of hospital fees in the United States. While there were large differences in the magnitude of the cost estimates obtained from the various estimation methods, the ranking by cost was not attended [see e.g. Cohen et al (1993); Taira et al., (2003)]. Therefore, failure to use true resource cost may not be too much of a problem in CEA and CUA (which are within health), this is not true of CBA (which pits health against other objectives).

**Indirect costs**

If the vantage point of society at large is adopted, indirect costs (i.e. productivity losses) need to be accounted for. They include lost work hours due to absenteeism or early retirement, impaired productivity at work, lost or impaired leisure time, and premature mortality. Intangible costs of pain, suffering, and grief are real yet very difficult to measure. Depending on the scenario description (see below), they are incorporated in the benefits (WTP values) of a CBA and therefore should be neglected in the costs of a program.

² For a detailed overview of cost estimation in health-related CBA see Johannesson (1996), ch. 7; or Drummond et al. (2005), ch. 4.
**Time horizon**

Choice of time horizon can importantly affect the costs of a health care intervention, which may accrue irregularly over time. For instance, Henderson et al. (1998) examined costs of percutaneous transluminal coronary angioplasty (PTCA) and coronary artery bypass grafting (CABG) as a function of time. While after one year, the costs of CABG were substantially higher than those of PTCA, this difference was not statistically different anymore after five years (see Figure 2). The time horizon should be chosen in a way that does not bias the analysis in favor for or against the intervention considered. It should be long enough to capture important deferred effects (including unintended ones), amounting to a disease episode, a patient’s remaining life expectancy, or even several cohorts of patients or individuals as in the case of HIV transmitted to children. Quantitative modeling approaches may be needed to estimate costs and outcomes that are beyond those of officially available data. Of course, the higher the discount rate used, the less important are future costs.

**Figure 1:** Cumulative costs of percutaneous transluminal coronary angioplasty (PTCA) and coronary artery bypass grafting (CABG) over time (confidence intervals indicated by the bars)

![Cumulative cost graph](image)

**Source:** Henderson et al. 1998

**Average vs. marginal cost**

Economists stress the importance of using marginal cost in evaluation because most decisions are not of the “on/off” type but of the “more/less” type. In that situation, economic analysis may reveal that,
beyond a certain level of spending, the additional benefits are no longer worth the additional (marginal) cost. For example, as shown in Table 1, the average cost per desired outcome of an iterative screening test may appear to be quite acceptable (e.g. $2,451 per case of colorectal cancer detected at six tests per person), whereas the marginal cost of adding the sixth test would be an astronomical $47mn.

Table 1 Cancer screening and detection costs with sequential guaiac tests (source: Neuhauser and Lewicki, 1975)

<table>
<thead>
<tr>
<th>No. of tests</th>
<th>No. of cancers detected (in 000s)</th>
<th>Additional cancers detected (in 000s)</th>
<th>Total cost ($) of diagnosis</th>
<th>Additional ($) cost of diagnosis</th>
<th>Average cost ($) per cancer detected</th>
<th>Marginal cost ($) per cancer detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65,947</td>
<td>65,947</td>
<td>77,511</td>
<td>77,511</td>
<td>1,175</td>
<td>1,175</td>
</tr>
<tr>
<td>2</td>
<td>71,442</td>
<td>5,496</td>
<td>107,690</td>
<td>30,179</td>
<td>1,507</td>
<td>5,492</td>
</tr>
<tr>
<td>3</td>
<td>71,900</td>
<td>0,458</td>
<td>130,199</td>
<td>22,509</td>
<td>1,810</td>
<td>49,150</td>
</tr>
<tr>
<td>4</td>
<td>71,939</td>
<td>0,038</td>
<td>148,116</td>
<td>17,917</td>
<td>2,059</td>
<td>469,534</td>
</tr>
<tr>
<td>5</td>
<td>71,942</td>
<td>0,003</td>
<td>163,141</td>
<td>15,024</td>
<td>2,268</td>
<td>4,724,695</td>
</tr>
<tr>
<td>6</td>
<td>71,942</td>
<td>0,000</td>
<td>176,331</td>
<td>13,190</td>
<td>2,451</td>
<td>47,107,214</td>
</tr>
</tbody>
</table>

**Discounting**

Costs that occur in the future have less present value than costs that materialize today (since the money can be put to a bank e.g. to earn interest). Discounting therefore reflects the opportunity costs of capital, i.e. the returns that could have been achieved elsewhere. Discounting makes flows of costs and benefits having different time profiles comparable in terms of their present value. The Capital Asset Pricing Model (CAPM) predicts that a project is competitive in the capital market if its expected rate of return \( E_r_i \) satisfies the condition,

\[
E_r_i = r_f + \beta_i (E_{r_M} - r_f)
\]  

(4)

Thus, the benchmark is the so-called risk-free interest \( r_f \), usually equated to the rate of return on government bonds. Since returns of shares on the capital market exceed the risk-free rate on expectation \( (E_{r_M} > r_f) \), the (share of) firm \( i \) must achieve a higher rate of return than \( r_f \) if \( \beta_i > 0 \). Now \( \beta_i \) is the slope parameter in a regression of the firm’s rates of return \( r_i \) on the market values \( r_M \). Usual estimates hover around \( \hat{\beta}_i = 1 \). However, as soon as \( \hat{\beta}_i < 1 \), the (share of) firm \( i \) has a risk diversification effect. Now to the extent that the health intervention is sponsored by the government, \( \beta_i = 0 \) is the appropriate value, resulting in \( E_{r_i} = r_f \). Therefore, the rate of discount becomes equal to the interest rate on government bonds. However, the value of \( r_f \) should refer to bonds having about the same time horizon of the health care intervention. A discount rate of 3 or 5 percent is typically used (Gold et al. 1996).
Conclusion 2: In Cost-Benefit Analysis, measuring the true resource cost poses several problems. Of particular importance are the choice of time horizon, the distinction between average and marginal cost, and the use of a rate of discount that is in accordance with capital market theory.

5. Alternatives for measuring benefits

To assign a money value to an improvement in length or quality of life, two entirely different concepts have been developed, the human-capital approach and the willingness-to-pay approach.

*Human-Capital Approach*

The human-capital approach [see e.g. Mushkin (1978)] places money weights on healthy time using market wage rates. Therefore, a life is valued in terms of the present value of the individual’s contribution to the Gross Domestic Product, which equals his or her labor income. There are several problems with this approach. First, wage rates should reflect marginal productivities of workers. However, there are often imperfections in labor markets, which cause wages to deviate from productivities due to discrimination by age or gender. Second, from a societal perspective, healthy time gained that is not sold for a wage has to be valued, too. This raises the question on how to place shadow prices on non-marketed resources. Finally and as a matter of principle, the human capital approach is not compatible with microeconomic theory (see e.g. Mishan, 1971). After all, it is the rest of society who (through the labor market and employers) values an individual, resulting in a zero value of life for pensioners and others who are unable to work, whereas in microeconomics, it is the individual who (subjectively) values consumer goods, jobs, risks, and life. Because of these economic and ethical flaws, we do not deem the human capital approach an appropriate method for placing money weights on health benefits. The main focus will therefore be on the willingness-to-pay approach.

*Willingness-to-pay Approach*

The willingness-to-pay (WTP) approach is based on the assumption that a person’s utility depends both on disposable income and on length and quality of life. In contrast to CUA, it does not impose restrictions on the utility function. Length and quality of life are elements of the vector $\theta_i$, while $y_i$ denotes disposable income (or consumption). Thus, the utility of person $i$ is given by

---

3 Two particular ways to attach a shadow price to healthy time are to use opportunity cost (see Klarman et al., 1968) or to use replacement cost (see Weisbrod, 1964).
Suppose that an intervention causes $\theta$ to increase from $\theta_i^1$ to $\theta_i^2$. Then the willingness to pay $Z_i$ for this intervention is defined by the maximum amount of money individual $i$ would pay in order to obtain the intervention. This means that utility is the same in the two states,

$$U_i[\theta_i^1, y_i] = U_i[\theta_i^2, y_i - Z_i],$$

where $[a]$ means “evaluated at a”.

There are various ways to measure WTP. The main distinction is between revealed and stated preference methods. While revealed preference methods try to derive WTP values from actual behavior, stated preference methods elicit them by asking people more or less directly. In general economists prefer to infer WTP from actual behavior. The market price provides a lower bound for WTP because consumers buy a product only if their WTP exceeds or equals the price paid. However, markets for medical care are both influenced by health insurance and heavily regulated, causing observed price to be too low and estimated WTP too high as a rule.

This leaves the possibility of measuring WTP by asking directly. On the one hand, Contingent Valuation confronts participants with one hypothetical scenario, reflecting the health care intervention to be evaluated. The only attribute that varies is price. By way of contrast, Discrete-Choice Experiments (DCE) present participants with a series of yes/no choices between the status quo and an alternative that differs with regard to several attributes, not only price.

Revealed as well as stated preference methods have their specific pros and cons, which will be discussed in the following.

**Revealed Preference**

Inferring WTP indirectly from market data has the important advantage of being based on actual consumer choices involving chances for health versus money, rather than hypothetical scenarios and choices. Risk preferences can be inferred from actions designed to avoid risks. A well-known example is the choice (or avoidance) of a job that is known for its risks to life and health (examples are truck drivers, miners, or electricians). The basic idea is to estimate the compensation required for accepting a
higher risk of death from the difference in wage rates for occupations with and without an increased threat to life. However, everyday decisions like putting on the safety belt also reveal risk preferences.

Yet, using revealed preference for measuring WTP has its problems as well. Wage-risk trade-offs are very context and job specific, and disentangling the many factors that confound the relationship between wage and health risk is difficult. One hardly finds two occupations that are identical except for their risk to life and health. Differences in wage rates also reflect differences in educational requirements, mental and physical demands experience, and many other characteristics of both workers and occupations.

Furthermore, there is a fundamental concern that risk-money trade-offs may not reflect the kind of rational choice revealing preferences that the WTP concept is based on. Notably, there are imperfections in labor markets (collectively negotiated wages) and limited rationality (biased estimates of occupational risk; for an overview, see Viscusi [1992]). As fatal accidents are relatively rare, it is questionable whether those affected know the frequencies, let alone use them as a basis for their subjective probability estimates. Surveys among drivers indicate that most underestimate their personal risk relative to the frequency of accidents observed in the driver population. Another question is whether observed behavior actually can be interpreted as the outcome of expected utility maximization as required by theory. Empirical evidence (presented already by Eisner and Strotz, 1961; see also Starmer, 2000) indicates that when dealing with relatively small risks, individuals systematically violate this rule. However, the known alternatives fare no better (Hey and Orme, 1994).

Finally it can be questioned whether persons in risky occupations are representative of the total population. The fact that they chose such an occupation and no other presumably reflects a basic preference for it, implying that the wage differential could constitute the upper limit of their compensation asked for bearing the increased risk. Members of that particular profession thus may have a special preference for risky situations (perhaps for the thrill that comes with them) that is not shared by the rest of the population. Conversely however, the observed wage differential marks the lower limit for the rest of the population because it would take a higher one to draw additional workers into their particular occupation. In conclusion, observed data may not be informative enough to permit estimation of WTP values.
Contingent Valuation

The traditional alternative to the method of revealed preference is Contingent Valuation (CV)\(^4\). The CV method measures the value of a non-market good by directly asking individuals how much the good (with attributes usually fixed) is worth to them. Typically the survey begins by giving some background information to respondents on the good in question. Next, they are told the way the money would be collected to finance the good. Finally, respondents are asked about their WTP and possibly upper and lower bounds. Since the mid-1970s, CV has also been applied to health care, resulting in rapid growth of the number of CV studies published.\(^5\)

Nevertheless, the CV method in health economics is still reluctantly used, for two main reasons. First, there are conceptual issues. The results of CV are criticized for lacking validity, since they depend heavily on the choice of the elicitation and estimation method. Indeed, both continuous and discrete CV formats are used, which vary in their familiarity to respondents and in their potential for biasing WTP values (Donaldson et al., 1998; Johannesson, 1996; Mitchell and Carson, 1989, Nocera et al., 2002).

The two main approaches to elicit WTP values by using a continuous CV format are open-ended questions and the payment card format. In the open-ended variant, individuals are directly asked to indicate their maximum WTP. As this may be too demanding (see e.g. Johannesson et al., 1991), the payment card variant confronts respondents with an ordered sequence of bids, with respondents indicating the maximum acceptable bid. In the discrete CV format, closed-ended techniques are prevalent. Here, respondents are only asked whether or not they would pay a single price out of a range. By varying the price in different subsamples, the proportion of respondents who are willing to pay the price can be calculated, and by multiplying this proportion by the number of respondents, a demand curve for the good can be estimated. Closed-ended techniques thus attempt to create a situation familiar to respondents by asking just yes-or-no questions.

The question of whether open-ended or closed-ended techniques should be preferred in CV studies is still an unsettled issue. More importantly, however, the CV approach suffers from its susceptibility to bias, in particular of the following types.\(^6\)

---

\(^4\) For an overview of more than one hundred CV studies, see Cummings et al. (1986) and Mitchell and Carson (1989).


\(^6\) For a comprehensive overview of possible sources of bias in CV studies, see Mitchell and Carson (1989).
• **Bias caused by reference values and the ordering of questions**: CV is prone to so called ‘anchoring’ effects. Respondents indicate their WTP relative to some reference value rather than their own value. For example, in the payment card format, stated WTP frequently depends on the starting point (*starting point bias*). When several projects are presented at a time, the answer to the first question can influence all the following ones (*question order bias*, see Boyle et al., 1996).

• **Sensitivity to wording of questions**: The results of CV studies are very sensitive to the wording of the questions. For instance, the allocation of property rights (*ownership bias*) or the payment vehicle used can affect WTP values. At an even more basic level, there is the risk that respondents perceive the presented good or program quite differently from what investigators intended.

• **Attitude towards the object of investigation**: When applying the closed-ended technique, there is the danger of ‘yea-saying’. To express their general agreement with the object of investigation, respondents accept prices in excess of their true WTP (see Blamey et al., 1999).

Considering in addition the general problems associated with interview studies, it seems doubtful that reliable WTP measurement can be obtained from CV studies. Nevertheless, this method has yielded theoretically plausible results when applied to health care (see e.g. Klose, 1999). Specifically, individuals with high incomes have been found to have a higher WTP than others. Moreover, measured WTP increases with the quantity of health services offered by a program.

However, external validity of stated WTP derived from CV studies in health care has not been established because WTP values differ substantially with regard to the elicitation technique and estimation method used. WTP values from discrete CV methods usually are much higher than those from PC formats (see e.g. Nocera et al. [2003], ch. 8.2) noted differences in a magnitude of up to 500 percent]. This is a problem since WTP values determine the absolute magnitude of welfare estimates and hence the outcome of a CBA [Boyle et al. (1996); Ready et al. (1996); Welsh and Poe (1998)].
**Discrete-Choice Experiments**

Discrete-Choice experiments (DCE) are a variant of conjoint analysis, which was developed in psychology in the late 1960s (see Luce and Tukey, 1964). This method attempts to explain and predict consumers’ behavior on the basis of their preferences for the attributes of a good. It is based on the New Demand Theory (see Lancaster, 1966), which defines preferences in terms of attributes rather than quantities of goods. Respondents in the survey then simply have to choose between programs differing in various attributes. In contradistinction to the CV approach, the status quo and the hypothetical alternative differ with regard to several or all attributes rather than price only.

Since the beginning of the 1980s, DCEs have been applied in transport economics and more recently, to environmental economics (see e.g. Hensher, 1997; Bennet and Blamey, 2001). In the mid-1990s, the method was implemented in health economics as well (Ryan, 1995; Ryan and Hughes, 1997). In the meantime, quite a few DCE studies have been conducted [see Ryan and Gerard (2003) for a first overview].

The behavioral assumption is that rational individuals will always choose the alternative with the higher level of (expected) utility. According to random utility theory (Manski, 1977; McFadden, 2001), the decision-making process within a DCE can thus be seen as a comparison of induced utilities, $V_g$,

$$V_g = v(a_j, p_j, y_i, s_i, \varepsilon_g)$$

(7)

Here, $v(\cdot)$ represents the indirect utility function of individual $i$ for a good $j$ described by a vector of attributes $a_j$ and a price denoted by $p_j$. The income of individual $i$ is $y_i$, the socioeconomic characteristics are denoted by $s_i$, and the error term, by $\varepsilon_g$. Given an additive error term, the individual will choose alternative $j$ over alternative $l$ if:

$$w(a_j, p_j, y_i, s_i) + \varepsilon_g \geq w(a_l, p_l, y_i, s_i) + \varepsilon_l.$$  

(8)

Here, $w(\cdot)$ is the deterministic component of the utility that can be estimated, while the error terms reflect unobservable factors that vary between individuals and alternatives. The utility function $w(\cdot)$ can be inferred from observed choices by assuming that the probability $P_{ij}$ of choosing alternative $j$ over $l$, given the vector of attributes. The probability of this occurring is assumed to equal the difference in utilities, and therefore

$$P_{ij} = \text{Prob} [\varepsilon_l - \varepsilon_j \leq w(a_j, p_j, y_i, s_i) - w(a_l, p_l, y_i, s_i)].$$

(9)
The assumption thus amounts to the probability of the noise \((\varepsilon_u - \varepsilon_m)\) being dominated by the systematic difference of utilities \([w(\cdot) - w(\cdot)]\). The marginal rate of substitution (MRS) between two attributes \(k\) and \(m\) is given by the ratio of their marginal utilities,

\[
MRS := \frac{\partial v_j / \partial a_k}{\partial v_j / \partial a_m}.
\]  

(10)

The MRS shows the subjective importance of attribute \(k\) relative to attribute \(m\). If the \(m\)-th attribute is price (and hence the negative of disposable income net of price), MRS indicates the marginal WTP for attribute \(k\). Therefore, relating back to equation (6), one has

\[
U_i[\theta^i, y_i - Z_i] = U_i[\theta^i, y_i] - \left(\theta^i - \theta^j\right) \cdot \frac{\partial v_j / \partial \theta^i}{\partial v_j / \partial y_i} = U_i[\theta^i, y_i] - \left(\theta^i - \theta^j\right) \cdot MRS.
\]

(11)

The model defined by eqs. (7) to (10) is usually estimated by logit and probit techniques, depending on the assumption being made on the distribution of the error terms. For a more detailed explanation of discrete choice models and their application, see Louviere et al. (2000) or Train (2003).

A DCE therefore amounts to tracing out an indifference curve in attribute space, with the status quo defining the reference point. A preferred combination of attributes must lie above the indifference curve (or surface, respectively in the case of more than two attributes), a rejected one, below. Through repeated choices, the indifference locus can be interpolated.

An important advantage of DCE over CV is due to the fact that respondents tend to evaluate all attributes of a program rather than focusing on price only. This makes DCE less susceptible to strategic behavior and other biases. Moreover, being based on an estimated utility function in attribute space, the results of a DCE can be used to determine WTP for any program that has the same set of attributes. However, being hypothetical, DCE are subject to the criticism that they may fail to elicit reliable and valid WTP values. Outside health, Louviere and Woodsworth (1983) have presented evidence suggesting that DCE can be used to explain modal choice in transportation, while Ghosh (1986) has found that they contribute to explaining the choice of a shopping mall. To this day, few studies have investigated the validity and reliability of DCE in the health care field. However, first results obtained by Bryan et al. (2000), Ryan et al. (1998); Telser and Zweifel (2002); Zweifel et al. (2006), and Telser and Zweifel (2007) indicate that DCE may be a valid and reliable approach to WTP measurement in the case of health as well.
Conclusion 3: The main alternatives are contingent valuation and Discrete-Choice Experiments when it comes to use stated rather than revealed preference for estimating benefits. The latter has important advantages, notably less susceptibility to a member of biases and greater similarity to actual decision-making situations.

6. Empirical example

Even if there are many papers in health care incorporating the term “cost-benefit analysis” in their title, only few CBAs were conducted in this area until now. Indeed, the label CBA is used in a very imprecise way in the health care literature. Most of the published so-called CBAs are not full economic evaluations but rather costing studies where cost savings due to a health care intervention are treated as benefits. However, these cost savings constitute a lower bound of a health program’s benefits at best. This misinterpretation leads to only programs that exhibit cost savings to be recommended whereas a true CBA may suggest even a program with higher costs to be worthwhile if it generates positive net benefits. According to Zarnke et al. (1997), only about 30 percent of all studies labeled CBAs valued health outcomes in money units. Since most of these used a human capital approach to value health outcomes, only about 10 percent of alleged CBA studies thus would meet the standards of cost-benefit analysis.

Moreover, even if outcomes of a program are valued in money units, they usually are not compared to cost. Most of the WTP studies are feasibility or pilot studies focusing on methodological issues but not on comparing cost to benefits (Drummond et al., 2004). Therefore, in the following we will present one particular CBA as an example of a full economic evaluation.

Nocera et al. (2002 and 2003) conducted a CBA of three programs against Alzheimer’s Disease (AD) for Switzerland. First, the program ‘care’ tries to ease some of the strain on informal caregivers, which provide the bulk of long-term care to people suffering from AD. To elicit WTP for ‘care’, a scenario was built in which informal caregivers receive training in caring for demented patients. Additionally, they had the possibility to engage a professional nurse for a few weeks per year for free. Respondents were asked whether they would be willing to pay higher income taxes for such care program to be implemented.

Second, the program ‘diagnosis’ focused on an early detection of AD. While no current therapy can reverse the progressive cognitive decline caused by AD, several pharmacological and psychosocial treatments exist which may delay the proceeding of the illness (Small et al., 1997; Mayeux and Sano, 1999). For these treatments to be effective, an early diagnosis of AD is important (Callahan et al., 1995). However, diagnosing AD is a relatively difficult task since there is still no validated test available.
Currently, a first diagnostic tool are screening interviews, which inquire into personal details, family contact, and health state. Additionally, short quantified screenings of cognitive function such as the Mini-Mental State Examination (see Folstein et al., 1975) is used. If there are signs of dementia, a more comprehensive neuropsychological assessment conducted by specialists is needed (Small et al., 1997).

To elicit WTP for ‘diagnosis’, Nocera et al. (2003, ch.2) designed a scenario consisting of a routine dementia screening test, which is currently not reimbursed by Swiss sickness funds. Therefore, respondents were asked whether they would be willing to pay a higher health insurance premium for such a program to be included in the list of benefits.

Third, the program ‘research’ focused on research into curing AD. Intensive research all over the world has led to an increasing understanding of the primary factors causing AD [see e.g., Vassar et al. (1999)]. But despite these advances, there exists no causal therapy for AD to date. The scenario asked respondents for their WTP to intensify research on AD at Swiss universities financed with tax money.

To elicit WTP values for these AD programs, different elicitation and estimation techniques were applied within a CV study. One aim was to analyze methodological problems in applying the CV method to measure people’s values for the outcome of health programs. As expected, there were big differences in WTP values according to the elicitation technique used (see Table 2, column 3).

After eliciting WTP values, Nocera et al. (2003, ch. 6) additionally investigated whether the three programs against AD should be implemented from a societal point of view. Therefore, the costs of the three programs were estimated. In the case of the program ‘care’, the cost per AD patient (consisting of a two-day course for informal caregivers and the hiring of a professional nurse for four weeks) sum up to about CHF 9,300 per year (1 CHF equals some $0.84 at 2007 exchange rates). In Switzerland, 32,000 AD patients are cared for by informal caregivers at patients’ homes. Therefore, if all of these patients were to take advantage of the care program, maximum possible costs of CHF 298 mn. would arise. These costs were then compared to estimated benefits, calculated by multiplying elicited WTP values with the number of the Swiss population of 18 years and older (see Table 2).

### Table 2 CBA of the programs against Alzheimer’s Disease in Switzerland

<table>
<thead>
<tr>
<th>Program</th>
<th>Maximum Cost (millions CHF p.a.)</th>
<th>Total WTP (millions CHF p.a.)</th>
<th>Net Benefit (millions CHF p.a.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Care</td>
<td>298</td>
<td>319–1568</td>
<td>21–1270</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>175</td>
<td>17–919</td>
<td>−158–774</td>
</tr>
<tr>
<td>Research</td>
<td>561</td>
<td>561–1056</td>
<td>0–495</td>
</tr>
</tbody>
</table>

Source: Nocera et al. (2003, ch. 6).
While the choice of the elicitation technique affects results in a substantial way, net benefits for the program ‘care’ are always positive. Therefore, from a societal point of view, the implementation of this program can be recommended since it generates more benefit than cost. The same is true for the program ‘research’, while the program ‘diagnosis’ can be recommended under some circumstances only. In a pessimistic scenario, it is possible that cost are higher than benefits resulting in a net benefit of CHF –158 mn. per year.

Conclusion 4: In an empirical application of Cost-Benefit Analysis, one finds programs ‘care’ and ‘research’ (but not ‘diagnosis’) to be worth the money for dealing with Alzheimer’s disease.

7. Conclusion and outlook

Microeconomic theory predicts that consumers optimize the ratio of marginal utility to price or marginal cost. Cost-Benefit Analysis (CBA) is nothing but the attempt of implementing this rule for goods and services that are not marketable, such as public health programs. Since (marginal) utility is difficult to measure, health economists have developed Cost Effectiveness (CEA) and Cost Utility analysis (CUA) to skirt this difficulty. However, these alternatives to CBA have the disadvantage of failing to inform policy makers whether a public program is worth its money, possibly even at the price of an expansion of the budget devoted to health.

An important recent innovation in CBA is the experimental estimation of utility in the guise of marginal willingness to pay, which makes benefits commensurable with cost. This contribution argues that the conventional Contingent Valuation (CV) approach frequently is too prone to biases of several types to be worthwhile. Rather, it proposes Discrete Choice Experiments (DCE, a variant of what is known as conjoint analysis in the marketing literature) as the tool of choice. In a DCE, all relevant attributes of the program are varied simultaneously, among them its price (usually in terms of a changed contribution to health insurance or to the public budget). Since an actual CBA using DCE evidence does not seem to have been performed by the time of writing (2007), the contribution presents a study based on CV in spite of the reservations stated above.

But even with all the recent progress in measuring the benefits side of CBA, there are at least two issues that need to be addressed in the future. The first concerns the modeling of the decision-making situation facing individuals in the experiment. Invariably, CBA (but also CEA and CUA) scenarios have been depicted as one-period, one-shot problems. However, individuals may be aware that their decisions
today affect the range of choices open to them (and possibly even their preferences!) later on. If true, this would call for a dynamic modeling approach that spells out the restrictions defining an optimal control path over time. The second issue is distributional. Implicitly, CBA gives equal distributional weight to all individuals affected by a program. The problem of course is to weigh benefits (and possibly costs) according to some social welfare function. Fortunately, DCE are able to provide estimates of these weights as well. After all, they can also be used to infer the willingness to pay of citizens to let some other members of society benefit from a public program (and to burden them with its cost).

In sum, refining CBA for health especially using experimental evidence is an exciting field of future research. Apart from its scientific challenges, such an endeavor has great potential value in guiding policy makers in their decisions regarding the allocation and distribution of resources devoted to health – always in competition with other objectives they and their constituencies may pursue.

References

Klarman, H. E., J. O. S. Francis, and G. Rosenthal (1968), Cost-Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease, Medical Care 6, 48–54.
Manski, C. F (1977), The structure of random utility models, Theory and Decision 8, 229–254.
Mehrez, A. and A. Gafni (1989), Quality-Adjusted Life-Years, Utility Theory, and Healthy-Years Equivalents, Medical Decision Making 9(2); 142–149.


