1. Introduction to health care evaluation

1.1 INTRODUCTION

The aim of this book is to explain how the theory and practice of cost–benefit analysis apply in the context of making health care expenditure decisions. We start by explaining why evaluation is necessary, defining an evaluation and identifying the main ingredients. Then we show how the ingredients are put together to form four alternative types of economic evaluation. Cost–benefit analysis will be demonstrated to be the primary evaluation technique, so the basic concepts behind this approach will be outlined. We proceed to supply applications (case studies) which illustrate some of the main evaluation principles and provide an overview of the health evaluation field. We close with sections that include problems, a summary and a look forward that provides a guide to the book. All of the themes introduced in this first chapter will be developed in greater depth in later chapters of the book.

1.1.1 Why One Needs to Evaluate Programs

To many non-economists working in the health care field, the idea of taking the trouble formally to evaluate treatment programs seems a total waste of time and effort. Surely, they would ask, health care is absolutely necessary, so we know in advance that we must spend on what is necessary?

However, an economist would counter that it is doubtful that every item of health care expenditure was really indispensable. Moreover, even if it were true that every health expenditure were necessary, it is a fact of life that there are not enough resources to meet every health care demand. It is probable that, in the present state of knowledge, all the world's resources would not eliminate cancer. Even if this cancer goal were feasible, what about resources for other deadly diseases, such as acquired immunodeficiency syndrome (AIDS)? Also, what about resources for non-health necessities, such as food? Very soon one reaches the conclusion that some necessities are more necessary than others and so we must choose among the necessary items to decide which to finance from our limited resources.

Economic evaluation as a subject has been developed to provide a framework for helping to make the choices facing us in the health care field. Only by making an economic evaluation can one be sure that all the other ways of making people better off have been considered and resources allocated to the areas that give the highest net satisfaction.

If anyone requires convincing that one ever needs to carry out an economic evaluation in the health care field, consider this case. The American Cancer Society endorsed the protocol of having six sequential stool tests for detecting cancer of the bowel. Five sequential tests were previously the standard practice. ‘Hey’, you can imagine doctors...
saying, ‘one can never be too careful when it comes to matters of health’. However, it was found that for every extra case detected by the sixth test, the cost was $47 million! One may not know exactly what is the value of detecting the one case of colon cancer, but one can suspect that it is not as high as $47 million, given that for the same amount society can instead feed over 12 000 people for a year (at $10 a day). Only by subjecting health care expenditures to an economic evaluation can we uncover the basis for making worthwhile decisions.

1.1.2 Definition of an Economic Evaluation

An economic evaluation tries to assess the social desirability of a program relative to some other alternative. If no other alternative is being considered, then the program is being described, but not evaluated. A good example of an analysis in health care that does not include alternatives explicitly is that contained in estimating the burden of the ‘cost of a disease’. Consider the $9.8 billion estimate of the cost of prostate cancer in the US in 2006 by Roehrborn and Black (2011). This estimate seems to be implying that, if prostate cancer were eliminated, then $9.8 billion would be saved. But, although such a figure is interesting in an abstract sense, involved with making a ‘thought experiment’ whereby one would like to know what would happen if the world were different than it actually was, the figure has little relevance for public policy purposes. There is no set of programs that exist today that can actually eradicate completely prostate cancer. So no current choices are clarified by knowing the total cost of a disease.

The existence of alternatives is at the heart of an economic evaluation because making choices is central to economics. In fact, economics can be called the ‘science of choice’. When there are no alternatives to consider, there is no need to make a choice, and there is no need to make an evaluation. It is important to understand from the outset the need to make choices, as much confusion over the nature of economic evaluations can then be avoided.

In the first place, it explains why economists carry out evaluations. Economists try to value a life in monetary terms not because they want to ‘play God’, but because one needs to choose how much of the world’s resources should be devoted to saving lives. To claim that a life is infinitely valuable is not helpful (no matter how true it is in a spiritual sense). Devoting all the world’s resources to save one life does not leave any funds for the rest of the 7 billion on the planet who would also desire to survive. The existence of finite resources forces us to make choices and undertake evaluations based on finite valuations.

Secondly, the need to make choices explains why an approach to evaluation can be adopted as ‘best’ even though one may have major reservations with the approach. Some approach must be adopted or else choices cannot be made. A flawed approach may be better as a guide to choice than one that is incomplete because difficult measurement issues have been avoided. Refusing to make difficult estimation decisions does not make the problem of having to make difficult choices disappear.

Prior to an evaluation, one must always check the effectiveness of the programs. That is, one has to establish whether a treatment does actually have an effect on the complaint. There is no point in carrying out an ineffective program efficiently! Nonetheless, having a controlled clinical trial, where one tries to control for all major influences (such as age, gender and severity of illness) when comparing the effectiveness of alternative treatments,
may not be a good basis for collecting data for economic evaluations. There are very few health care interventions that work identically for all individuals. Recognizing individual differences is a very important part of conducting an economic evaluation because individual preferences greatly determine outcomes. For example, a medication that is more effective than another may not be better if people choose not to take it (because they do not like the side effects). In addition to differences in preferences, programs can also be inefficient for medical or physiological reasons. Instead of there being a single outcome for all, there is a distribution of possible outcomes which, as we shall see, means that for some individuals the health care intervention do not work.

1.1.3 Components of an Economic Evaluation

Health care programs take inputs (labor, capital, and so on) and transform them into outputs. In order to aggregate the inputs, one usually values the labor and capital using market prices to produce a measure of all the resources taken up in the health care intervention. This aggregate input measure is called ‘costs’ and is in monetary units. The outputs of the evaluation can come in different forms and these are the ‘consequences’ of the intervention. Note that, following Drummond et al. (1987), although inputs also have consequences, we use consequences in this text to mean output consequences; we simply refer to input consequences as ‘costs’. The most obvious output consequence to consider is what the health care industry immediately works with, such as a diagnostic test outcome or an operation successfully completed. These outputs are called ‘effects’ and expressed in natural units (such as a percentage detection or completion ratio). A broader measure of effects relies on ‘utilities’ (that is, estimates of the satisfaction of the effects) and the output unit is called a ‘quality adjusted life year’ (QALY) which measures the length of life that a person is expected to live and the satisfaction for that period that the person can expect. Lastly, the output can be expressed in the same monetary units as the costs, in which case the consequences are now called ‘benefits’.

In the health care field (see Drummond et al., 1987) the costs and benefits are disaggregated into three categories: direct, indirect and intangible. ‘Direct’ means directly related to the health care industry (the doctors, the hospitals and the patients). So physician and nursing expenses, and any hospital cost savings, are called direct costs and direct benefits respectively. ‘Indirect’ refers to inputs and outputs that pass outside the health care industry. The main measure of these indirect effects is via earnings forgone or enhanced due to treatment, as the earnings reflect the value of production lost to, or gained by, the rest of society. ‘Intangible’ refers to the pain and suffering that are caused or alleviated by a health care intervention.

Corresponding to the costs and consequences just identified, there are four types of economic evaluation that incorporate some or all of the components. All four evaluation types make use of costs. The type that only uses costs is called a cost minimization study. The other three types differ according to what kind of consequence they incorporate along with the costs. A cost-effectiveness analysis uses the effects, a cost–utility analysis uses the utilities, and a cost–benefit analysis uses the benefits. This book is devoted to explaining, analyzing and developing these four types of evaluation, a process that we begin in the next section.
1.2 TYPES OF ECONOMIC EVALUATION

As we have just seen, there are four main types of economic evaluation that exist in the health care literature as recognized and understood by economists. The four methods are: cost–benefit analysis (CBA), cost-effectiveness analysis (CEA), cost–utility analysis (CEA) and cost minimization (CM). However, most of those who carry out the health care evaluations are non-economists. They actually do use one of the four types of evaluation, but they label their work differently. We need to be aware that many studies are called CBAs when in fact they correspond to a different evaluation category.

The distinguishing characteristic of a CBA is that it places a monetary value on the consequences. Forming ‘benefits’ means that the consequences are in the same units as ‘costs’. One can tell whether the benefits are greater than the costs and thus know whether the expenditure is worthwhile. The main organizational principle of this book is that the three other types of economic evaluation can be best understood as special cases of CBA. We are now going to give a brief summary of the four evaluation types using this principle. We start with an outline of CBA and then introduce the other three methods in terms that make their relationship to CBA clear from the outset.

1.2.1 Cost–Benefit Analysis

Consider the simplest type of health care intervention (or treatment), that is, taking an aspirin, which we designate by the subscript 1. This medication leads to advantages and disadvantages. When the advantages and disadvantages are measured in monetary terms, we will call them benefits $B_1$ and costs $C_1$. The aspirin would be worth buying (make a person better off) if the amount of the benefits exceeded that of the costs:

\[ B_1 > C_1 \]  

Equation (1.1) presents the basic cost–benefit criterion. It is this relation that determines whether a health care expenditure should be approved or not.

The CBA approach can be thought of as dealing with the final result of a health care intervention (consumer or patient satisfaction). In the process of arriving at the end result, there is an intermediate stage, which involves transforming the treatment from an input into an output that can be represented as $E_1$. Here $E$ is called the effect of the intervention.

In the aspirin example the effect might be pain (headache) relief. Benefit estimation can then be thought to have occurred in two steps: first there was an effect of an intervention, and then a monetary value was assigned to it. Let the second step be interpreted to be ‘pricing’ the effect and denoted by $P_1$. The two steps combine to construct the benefit measure in the form of a product of the effect times the price, that is: $B_1 = P_1 \times E_1$. In which case, equation (1.1) can be rewritten as:

\[ P_1 \times E_1 > C_1 \]  

This criterion can be expressed in an equivalent way by dividing both sides of equation (1.2) by $C_1$. We then obtain the requirement that the benefit–cost ratio must exceed unity:
The CBA criterion has an alternative formulation when there is a financial budget constraint, which limits how much costs can be expended. The logic is that if there is a budget constraint, then using funds for one purpose precludes their use for another. Let the alternative health care intervention, which could be something unrelated to providing headache relief, such as a diagnostic test for cancer, involve benefits $B_2$ (equal to $P_2E_2$) and costs $C_2$. Then it is not sufficient that the benefit–cost ratio for intervention 1 exceeds unity; it must also exceed the benefit–cost ratio of the alternative intervention:

$$\frac{P_1E_1}{C_1} > \frac{P_2E_2}{C_2} \quad (1.4)$$

Criterion (1.4) ensures that if one spends on intervention 1, then one receives more benefit per dollar spent on costs than with the alternative use of the funds. (For a fuller discussion of using benefit–cost ratios instead of net-benefits as CBA criteria in the context of budget constraints, see Brent (1998a, chapter 2) and Vinod (1988).)

### 1.2.2 Cost-Effectiveness Analysis

In the health care evaluation field there has been a general unease with the step in CBA which entails the pricing of the effects, that is, with setting $P_1$ and $P_2$. The most popular approach is CEA, which tries to continue without the prices and work on the consequences side only with effects $E_1$ and $E_2$. In the classical version of CEA, there is envisaged to be a budget constraint, which means that treatments cannot be considered in isolation. However, treatment effects must be the same kind across alternatives. For example, one can compare either different ways of providing headache relief, or different ways of diagnosing cancer, but one cannot consider both effects at the same time. This means that the only difference between $E_1$ and $E_2$ is that $E_1$ (the effect coming from intervention 1) is at a different level of the same effect as $E_2$ (the effect coming from intervention 2). The aim now is to choose the intervention that supplies a unit of effect at lowest cost. Under a CEA, treatment 1 would be more cost-effective than treatment 2 if:

$$\frac{C_1}{E_1} < \frac{C_2}{E_2} \quad (1.5)$$

Rather than requiring the lowest cost for a given effect, one can instead try to achieve the most effect per dollar of cost. By inverting equation (1.5), we can see that the CEA is equivalent to requiring that:

$$\frac{E_1}{C_1} > \frac{E_2}{C_2} \quad (1.6)$$

Now compare equation (1.6) with (1.4). The only difference involves the inclusion of prices in the CBA criterion. Thus, if the prices of the effects are considered the same, which is valid in a CEA since it is a common effect that one is comparing with the two
treatments, we can set $P_1 = P_2 = P$ in equation (1.4). Dividing both sides of this equation by $P$, we obtain the CEA criterion (1.6).

Clearly, in principle, CEA can be regarded as a special case of a CBA. The requirements, however, are very stringent. The effect must be exactly the same for the treatments being compared. This means that not only must there be a single effect that is common to the treatments; the quality of the effect also must be identical across treatments in order for the value, or price, of the effect to be the same.

1.2.3 Cost–Utility Analysis

Note that, in a CEA, the only difference between $E_1$ and $E_2$ is that the quantity of the effect is different in the two treatments. It is not the effect itself that is different. If we want to compare entirely different effects (as with headache pain relief and the precision of a diagnostic test) and we do not want to use prices explicitly, then one needs to have a process that enables all effects to be converted to a common unit.

In a CUA, the common unit is a quality adjusted life year (a QALY). The idea is that any health care intervention either enables one to live longer (that is, have more life years) or to live healthier (that is, have a better quality of life). Thus, the aspirin by relieving pain adds to the quality of life, and the more accurate diagnostic test would (if it leads to preventative action that reduces the chances of dying) add to the quantity of life. The essence of a QALY then is entailed in finding out how much quantity of life someone is willing to give up in order to live a higher quality of life (as is the case when deciding whether to have life-threatening surgery that would remove an obstacle to living normally).

With the QALY as the common effect, the evaluation exercise returns to the CEA framework. A CUA is therefore the special case of a CEA where the effect $E$ is measured by a QALY, and the following criterion is used to replace equation (1.6):

$$\frac{QALY_1}{C_1} > \frac{QALY_2}{C_2}$$

(1.7)

Since a CEA is a special case of a CBA, and a CUA is a special kind of CEA, it means that a CUA is a restricted CBA, where the restrictions are that to result in (1.7) from equation (1.4) we have $E = \text{QALY}$ for each treatment, in addition to $P_1 = P_2 = P$ (where $P$ now relates to the price of a QALY).

Just as with a CEA, a CUA differs from a CBA in practice, not in principle. A CBA relies on the fact that individuals are used to purchasing goods on a market at a price. Individuals are familiar with the process of trading off valuations in terms of how much money they are willing to give up to receive more of a good. On the other hand, individuals are not used to buying a QALY and are therefore unfamiliar with the process of trying to obtain a QALY at lowest cost. Deriving meaningful estimates of QALYs is therefore at the heart of the CUA evaluation exercise.

1.2.4 Cost Minimization

In a CM, consequences play no part in the evaluation. The ‘trick’ therefore is to make disappear both parts of the benefit term, $P$ and $E$. This does not mean setting $P \times E = 0$,
for then criterion (1.4) would be totally undefined. Rather it means that ‘somehow’ \( P \times E = 1 \). With this specification of benefits for all treatments, equation (1.4) becomes:

\[
\frac{1}{C_1} > \frac{1}{C_2}
\]

Equation (1.8) is equivalent to requiring:

\[
C_1 < C2
\]

There is no unique way to satisfy the condition \( P \times E = 1 \). For example, one could assume \( P = 1 / E \) (although I know of no one who actually does employ this assumption). What would be logical is that one treat a CM as a special case of a CEA. The restriction then would be that all effects are the same, \( E_1 = E_2 = E \), and that the common effect would be standardized as a single unit, that is, \( E = 1 \). This restriction, together with the assumption of an equal valuation per unit of effect \( (P_1 = P_2 = P) \), would be sufficient to reduce equation (1.4) to (1.8) and ensure that a CM is also a special case of a CBA.

If it is difficult in practice to justify the requirements of a CEA, it is going to be even more difficult to justify a CM. The quantity of an effect must be exactly the same for all treatments; but also the quality of an effect must be the same. Unless consequences are identical across treatments, a CM would not constitute a valid evaluation of treatments.

### 1.3 CEA AS CBA: USING A THRESHOLD VALUE FOR PRICE

Now that we have a clear idea about what distinguishes the four main types of economic evaluation in the health care field, we are in a position to understand how CEA attempts to side-step the problem of determining whether an intervention is worthwhile. As we now explain, CEA uses a threshold value for the price of the effect, which we label \( P_T \). This use of a threshold value is a particular use of a general technique that is often used in health care evaluations to estimate hard-to-estimate values, a technique which we will be covering many times in the book.

Consider two alternatives, 1 and 2. In the current context alternatives 1 and 2 are mutually exclusive, which means that if we choose option 1 then we will not be adopting 2. To set the stage for the application that we are envisaging in this section, that is, where the National Health Service (NHS) in the UK is involved with the evaluation, we can think of 1 as being a new drug to treat cancer and 2 as being an existing drug for the same purpose that is currently being funded by the NHS. We begin with the CBA criterion \( B - C \) and then set the net benefits of the two interventions to be equal:

\[
B_1 - C_1 = B_2 - C_2
\]

We can bring all the benefits onto the left side of equation (1.10) and include all the costs on the right side:

\[
B_1 - B_2 = C_1 - C_2
\]
As before we will split the benefits up into the product of a price and an effect. This time the common price \( P \) part of \( P \times E \) is going to be \( P_T \). Equation (1.11) then becomes:

\[
P_T E_1 - P_T E_2 = C_1 - C_2
\]

or, as \( P_T \) is common to both effects:

\[
P_T (E_1 - E_2) = C_1 - C_2 \quad (1.12)
\]

Finally, divide both sides of equation (1.12) by \( E_1 - E_2 \):

\[
P_T = \frac{C_1 - C_2}{E_1 - E_2} \quad (1.13)
\]

On the right side of equation (1.13) is the incremental cost-effectiveness ratio (ICER). It measures how much extra effect is obtained by the new drug for the extra cost of the new drug. On the left side of equation (1.1) is the threshold price, that is, what the minimum amount one is willing to pay has to be to ensure that the new drug is worthwhile. Note that \( P_T \) was derived from the very first equation (1.11), which assumed that there was no difference in the net benefits of the two drugs. So if \( P_T \) were higher than this amount, there actually would be a difference and the net benefits of drug1 would be higher than 2. So threshold values are always those values that would tip the balance in the evaluation outcome from neutral to positive or negative.

The best-known institution that has adopted the evaluation criterion expressed by equation (1.13) is the UK’s National Institute for Health and Clinical Excellence (NICE). In NICE (2008) it outlines its evaluation guidelines. The institute was established in 1999 and is a UK government funded body that responds to requests by the Department of Health for guidance when evaluating new and existing health technologies. A technology includes:

- pharmaceuticals;
- medical devices;
- diagnostic techniques;
- surgical procedures;
- other therapeutic technologies; and
- health promotion activities.

The outcome measure that NICE uses is the QALY. Thus its version of equation (1.13) is essentially:

\[
P_T = \frac{C_1 - C_2}{QALY_1 - QALY_2} \quad (1.14)
\]

In this context \( P_T \) is the price of a QALY and its role is to convert a CUA into a CBA. The NICE has for over seven years used two values for \( P_T \), that is, £20 000 and £30 000. If the ICER is less than £20 000 the technology will normally be recommended. If the ICER is above £30 000 then the technology is questionable. But if the ICER is between £20 000 and £30 000 the recommendation is not so clear.
1.4 THE BASICS OF COST–BENEFIT ANALYSIS

Since CBA is the primary evaluation method, that the other types of evaluation only approximate, it is necessary to outline some of the fundamentals of the approach before we see it in action throughout the book. As we shall see in the applications section, there is a widespread (and increasing) reluctance to use CBA for health care evaluations. Outlining the fundamentals of CBA is especially important in order to dispel false perceptions of the weaknesses of CBA.

1.4.1 Why Cost–Benefit Analysis?

There are two parts to the answer to the question ‘Why CBA?’ The first involves explaining why, in general, CBA should be used rather than some other method of allocating resources. The second involves explaining why, in the health care context, one should use CBA rather than the other types of economic evaluation. We consider each part in turn.

CBA is at the core of any public policy course. Roughly, half of economics is trying to explain how the economy operates, and the other half is concerned with trying to alter how the economy operates. It is this second half that constitutes the public policy domain. Very few people would be interested in economics as a subject if there were not perceived to be some need to intervene in the economy. Policies to eliminate inflation, unemployment and poverty are all driven by this perceived need.

It is in this context of government intervention that CBA is most useful, for it provides a consistent framework for deciding when interventions are desirable or not. Considering whether to introduce a tariff reduction, a labor subsidy, an agricultural price support scheme, an environmental regulation, an interest rate change, or supply new funds for AIDS research are all matters for a CBA. When areas of economics, both within and outside the field of health care, do not employ the principles of CBA they engage in ‘bad’ public policy, in that their findings are based on incomplete analyses of what is socially desirable.

What, then, are the alternatives to using CBA? One could allow the market mechanism to allocate resources without any government intervention. But markets do have their imperfections and fairness/equity is not a market objective, even though in the health care field fairness is probably considered at least as important as efficiency. On the other hand, one could leave it to the political system to allocate resources. This has the drawback of government imperfections, in that voter preferences may be made secondary to political agent preferences. Even without government imperfections, it is not yet feasible to have a national vote every time a government intervention is to take place.

A CBA is designed to be a better reflection of what is socially desirable than market forces or direct government determination. As we shall see, CBA incorporates the efficiency rationale behind markets and makes adjustments for situations where market failure exists. It is efficiency that non-economists often ignore when in the health care field evaluators violate fundamental CBA principles. In addition, CBA can incorporate distributional considerations, which markets ignore and the health care field considers an essential ingredient in an evaluation.

In the health care context, CBA should be used rather than the other types of economic evaluation because it is the only method that can tell whether an intervention
is worthwhile. Even the most cost-effective intervention may not actually be socially desirable, especially when we compare it with alternative uses of funds. It is not sufficient that we concentrate only on the health care sector. Our evaluation framework must ensure that health care programs are evaluated on a consistent basis with those in education, nutrition, the environment and social security. The objective is that all public funds be spent wisely, not just those allocated (perhaps arbitrarily) to the health care sector.

The logic of this need to keep our perspective as general as possible comes out clearly even if health is the only social priority. For there are many different ways of achieving good health from areas not considered part of the traditional health care sector. For example, the high rate of female literacy in Kerala in India is a main reason why that state (with only average levels of income per head) has one of the lowest rates of infant mortality and one of the highest rates of life expectancy.

1.4.2 The Social Perspective in Cost–Benefit Analysis

The perspective in CBA is a very broad one as it embraces the effects on everyone in society. A social evaluation does not consider just the parties directly involved with an intervention, that is, the firms (the hospitals and the physicians) and the consumers (the patients as clients). It also covers those indirectly affected, including the family members of the patients and even the general taxpayer. Strictly, then, CBA should be called ‘social’ CBA to recognize the all-inclusive nature of the evaluation. However, this usage is not widespread either within or outside the health care field. So we shall just refer to the analysis as CBA, leaving the social connotation implicit.

It is interesting that in the health care field it is considered good practice to make the perspective explicit at the outset of a study. That is, it is thought necessary to specify whose perspective the study is from. Is it from the perspective of the hospital, the client, the government taxpayer or whatever? Although it is true that an economic evaluation has usefulness from each and every one of these perspectives, it is only the social perspective that is important for social decision-making. A study that ignores the costs for the families of patients with psychiatric problems is not very useful for making choices about the best place to house the seriously mentally ill. Consequently, one should always adopt the social perspective in an economic evaluation if one wishes to try to influence social decisions.

Even though the social perspective is primary, other perspectives are not irrelevant. In a mixed economy, where the government makes decisions recognizing its interaction with the private sector, it is important to know whether the social outcome is in accord with outcomes from a more narrow perspective. If it is socially worthwhile for people to be inoculated for TB, but it is not worthwhile from an individual’s point of view (as the benefits to the non-inoculated population who have a lower chance of contracting TB are ignored), then there could be an ‘incentive compatibility problem’ – which is to say that the socially desirable outcome will not in this case be chosen by the individual. Therefore some government incentive must be given to induce individuals to adopt the socially optimal outcome if they do not think it is in their best interests. (Chapter 4 explains this further.)
1.4.3 Efficiency Costs and Benefits

Whether an effect of a health care service is a ‘benefit’ or a ‘cost’ depends on what is the purpose of the expenditure. For example, a reduction in the number of persons born stemming from a population control program may be considered an advantage if one is trying to ensure that existing food supplies enable the most people to survive; but it would constitute a disadvantage if one were trying to build a large (human) army for defense purposes.

In welfare economics (the theoretical base for all policy economics) the objective is usually assumed to be to maximize (aggregate) consumer satisfaction (also called utility). It is this sense of maximizing satisfaction (something to be made more precise in later chapters) that is meant by saying that programs are efficient. People get satisfaction if they are willing to pay for something. With resources (income) limited, purchasing health care precludes the purchase of other goods and services, which also give satisfaction. There would be no point in people purchasing the health care if it did not make them better off. Because it is the individuals themselves who are to decide how to spend their income, and in the process make themselves better off, the basic value judgment behind the welfare economic approach to valuation entails the assumption of ‘consumer sovereignty’. That is, the individual is assumed to be the best judge of his/her own welfare.

If one does accept the assumption of consumer sovereignty, there is no good reason not to accept willingness to pay (WTP) as a measure of benefits, and therefore no good reason not to use CBA to evaluate health care expenditures. This has to be borne in mind when one considers the criticism of CBA that it is based on the questionable use of earnings to measure health benefits. The use of earnings (what was labelled ‘indirect benefits’ in section 1.1.3) is a part of ‘traditional’ CBA. The economic philosophy behind this valuation method is embodied in the ‘human capital’ approach. A person’s earnings are meant to reflect a person’s productivity. A health intervention by restoring a person’s productivity thereby provides a benefit to society.

The complaint has been that the inclusion of earnings biases programs in favor of those who work and earn, against those that affect children, housewives and the elderly. The traditional CBA approach thus is thought to be inequitable. However, the main criticism of the human capital approach by economists is that this approach ignores the preferences of the individual him or herself, and clearly does not fit in with the usual welfare economic base behind CBA, which is based on an individual’s WTP. Thus, one can be an opponent of the human capital approach, yet still be an advocate of modern (efficiency-based) CBA.

The equity issue is also something that is not outside CBA, even though it may not be a part of traditional CBA. When one uses WTP, one can weight effects according to their social significance. Thus, if one person is in poverty, their WTP can be given a premium so that their lack of ‘ability to pay’ can be allowed for. The real problems of CBA are therefore: (a) dealing with cases when consumer sovereignty does not apply (as with programs geared to the severely mentally ill); (b) trying to measure WTP when market valuations of WTP do not exist or are greatly distorted (which is often the case in the health care field); and (c) trying to obtain meaningful measures of the equity weights. The human capital approach is the ‘fall-back position’ that one should rely on only when other approaches cannot be applied. Traditional CBA is not necessarily ‘best practice’.
1.5 APPLICATIONS

Despite the superiority of CBA, it is not the method of choice in the health care evaluation field. We begin our applications with a recent survey that finds that non-CBA evaluations predominate and that their share is increasing over time. We use this survey to give an overview of health care evaluations in practice. Then we present two studies typical of those in the health care evaluation field. In the first we consider an evaluation of gallstones procedures using a CUA, and in the second we present a CEA of alternative sites for hypertension treatment. The emphasis in both of these studies will be on allowing for the side effects of health care interventions. We also use both these studies to reinforce the point that only a CBA can tell whether a treatment or intervention can be accepted or not. The fourth application gives an account of exactly how the $47 million estimated cost for the sixth stool guaiac protocol was obtained. We close with the application of the Millennium Village Project which goes into detail about how to know when it is necessary to carry out a full-blown effectiveness evaluation and when to settle for something simpler/cheaper.

1.5.1 A Survey of Health Care Evaluation Practice

In this case study we provide an overview of the health care evaluation field using a Canadian survey of 153 studies carried out by Tarride et al. (2009). Specifically, we report which category of economic evaluation is most used in practice, what data sources were used and who did the funding. In this way the reader can get a feel for the scope of the health care evaluation field and obtain an idea of what applications to expect later in this book. To set the stage we first give a brief summary of what types of evaluation took place in earlier periods using the survey by Elixhauser et al. (1993).

From 1966 to 1978, Elixhauser et al. claimed that CBAs in the literature were roughly in equal numbers as CEAs and CUAs combined (CMs were omitted as they did not include effects as well as costs). This parity then changed with CEAs and CUAs becoming the evaluation techniques of choice. During 1979–85, the share of CEAs and CUAs rose to 58.9 percent, and increased even further to 64.5 percent for 1985–90. The authors of the survey suggested that the controversy involved with putting a monetary value on life and limb that takes place in a CBA contributed to the relative appeal of CEA and CUA.

As we can see from Table 1.1, based on the survey by Tarride et al. (2009), in more recent times (2001 to 2006), the trend of evaluations moving away from CBA has continued. Owing to the technological breakthroughs that have occurred in the development of pharmaceutical drugs, the breakdown is done by drug and nondrug evaluations (which were approximately in equal numbers in the survey). Irrespective of whether drug or nondrug evaluations were being carried out, only 3–4 percent of the total involved CBAs. The majority of evaluations took the form of a CEA.

This finding that CBAs are nowadays relatively little used (though in absolute terms they have increased exponentially) is not just restricted to Canada. Walter et al. (2007) report that, of 66 CEA studies published in four main health economics journals since 2004, all used just the ICER (and none used the threshold value for a price of an effect which section 1.3 showed was necessary to convert a CEA into a CBA).
Table 1.1 also reports that about 80 percent of the studies disclosed their source of funding. The main funding source was industry which funded two-thirds (67 percent) of evaluations of drugs and 21 percent of nondrug evaluations. Given these funding sources, one should not be surprised that 63 percent of the studies used a government or third-party payer perspective and only 23 percent used a societal perspective.

The types of evaluation did not differ much between drug and nondrug evaluations. Differences were evident however in terms of funding (industry funded mostly drug evaluations) and by data source (more trials and less expert opinion in drug evaluations). The authors claimed that although not directly comparable, there was a lot of consistency between the pattern of evaluations in Canada and those elsewhere. For example, 51 percent of the Canadian studies used effectiveness data from a single source, and it was 66 percent in an UK NHS survey; and both the Canadian studies and those in Italy collected 50 percent of their data on evidence from randomized control trials.

A point which the survey wished to emphasize was the extent to which the evidence used in the evaluations was Canadian. Less than 50 percent of the studies indicated whether Canadian patients were included in the measures of effectiveness or if resource use was only Canadian. In a sense, this fact shows how globalized health care evaluations have now become. Many different countries are undertaking evaluations and these

<table>
<thead>
<tr>
<th>Type of evaluation (N = 153)</th>
<th>Total</th>
<th>Drug</th>
<th>Nondrug</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBA</td>
<td>3%</td>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td>CEA</td>
<td>58%</td>
<td>59%</td>
<td>58%</td>
</tr>
<tr>
<td>CUA</td>
<td>18%</td>
<td>19%</td>
<td>18%</td>
</tr>
<tr>
<td>CM</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>CEA and CUA</td>
<td>16%</td>
<td>16%</td>
<td>15%</td>
</tr>
<tr>
<td>CEA and CBA</td>
<td>1%</td>
<td>0%</td>
<td>1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data source (N = 273)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials</td>
<td>51%</td>
<td>62%</td>
<td>42%</td>
</tr>
<tr>
<td>Administrative database</td>
<td>5%</td>
<td>7%</td>
<td>3%</td>
</tr>
<tr>
<td>Literature review</td>
<td>11%</td>
<td>4%</td>
<td>16%</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>10%</td>
<td>13%</td>
<td>9%</td>
</tr>
<tr>
<td>Expert opinion</td>
<td>8%</td>
<td>2%</td>
<td>12%</td>
</tr>
<tr>
<td>Other</td>
<td>15%</td>
<td>12%</td>
<td>18%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Funding (N = 153)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic/granting agency</td>
<td>13%</td>
<td>10%</td>
<td>16%</td>
</tr>
<tr>
<td>Disease-related agency</td>
<td>10%</td>
<td>5%</td>
<td>14%</td>
</tr>
<tr>
<td>Government</td>
<td>10%</td>
<td>2%</td>
<td>16%</td>
</tr>
<tr>
<td>Industry</td>
<td>43%</td>
<td>67%</td>
<td>21%</td>
</tr>
<tr>
<td>No funding</td>
<td>1%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Other</td>
<td>3%</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>Funding undisclosed</td>
<td>21%</td>
<td>13%</td>
<td>27%</td>
</tr>
</tbody>
</table>

Source: Constructed by the author based on Tarride et al. (2009), tables 1 and 2.
countries are both using information from elsewhere and supplying information and results that other countries are using. But the concern of Tarride et al. is that there are potential differences in patient characteristics and health care systems between Canada and elsewhere. Another way of expressing this concern is that health care evaluations should respect consumer sovereignty and, if Canadian preferences are not being consulted, how can their preferences be respected?

We close this discussion of health care practice by covering some of Drummond’s (1981) observations comparing and contrasting economic evaluations in the health care field with those in other sectors. Drummond made his points in the context of evaluations prior to 1980, but they seem to be just as valid today.

In some respects, evaluations in the health care field are most suitable for the standard type of economic appraisal. Economic evaluations rely on ‘partial equilibrium analysis’, assuming that income and all other prices are held constant. Strictly, then, only ‘small’ projects are appropriately evaluated by economic appraisals. In the health care field as a whole this has definitely been the case. Evaluations have been applied to limited and tightly defined problems, such as whether hernias should be treated by inpatient or outpatient surgery, or whether one should screen for a particular disease.

A major difference between health applications and those elsewhere is over the selection of the type of intervention chosen for evaluation. In other fields, evaluations have been of large capital projects (for example, airports). In the health care field the applications are of procedures (treatments) and not capital expenditures, such as hospitals. (For an exception, see the final case study in this chapter where multiple large capital expenditures are involved.) This has advantages and disadvantages.

The advantages are that:

- Decisions over procedures are at the heart of resource allocation in the medical field and so it makes sense to focus on activities that are of most concern to those involved in the field.
- Consequently, the health care field has an edge over other areas that have neglected to evaluate procedures, for example in the transportation field, where there is a need to evaluate whether standby flights are worthwhile.

The disadvantages of concentrating evaluations on procedures are:

- Health evaluators have had to deal with the difficult issue of measuring outputs. A dam provides electricity. But what precisely does a test for breast cancer generate?
- Important issues have been missed, for example, with decisions over where to site hospitals.
- There is no direct link between the evaluation and the making of expenditure decisions. For example, it has been shown that kidney transplants are more cost-effective than any form of dialysis. The issue is, who is the client for this information? The question of transplants and kidney treatment as a whole is dealt with by health planners (hospital administrators, insurance companies, health maintenance organizations and public officials) who take a broad view, and would need a lot wider evaluation than just knowing about a particular set of procedures.
- The narrow focus has led to the neglect of other health producing measures, such
as changes in lifestyles and government agricultural policy, which may affect health just as much as the medical procedures.

1.5.2 Evaluating Gallstone Treatments

Whether to include indirect costs (forgone earnings due to the patient’s time being taken up by treatment) is a controversial issue. As we remarked earlier, there are those who consider its inclusion as inequitable. Other opponents stress its conflict with WTP principles. Even if one does endorse the inclusion of indirect costs, there are a whole host of measurement issues involved with estimating the extent to which society’s output will actually fall when a person cannot show up for work. In a CUA of gallstone treatments by Cook et al. (1994), they focus on establishing the quantitative importance of including indirect costs.

Indirect costs impact the choice of gallstone treatment in the following way. The standard treatment for gallstones was removing the gallbladder itself (called a cholecystectomy). This causes considerable post-operative sickness (pain, diarrhea and nausea) and a lengthy recovery period (hence a lengthy period away from work). As an alternative to this ‘open’ surgery, one could have the gallstone removed by a laparoscopic cholecystectomy (a laparotomy is a surgical incision into any part of the abdominal wall), which is a minimal access surgery where the post-operative morbidity (sickness) is greatly reduced and patients can resume normal duties within 2–3 days. Finally, there is now available a way of fragmenting the gallstones by external shock waves, called an extracorporeal shock wave lithotripsy or ESWL (lithotripter is Greek for ‘stone crusher’). With an ESWL being non-invasive, post-treatment sickness is minor and the patient can leave the hospital on the same day as treatment. The work disruption is thus least with an ESWL.

A CUA was a more appropriate evaluation technique than a CEA because there did not exist a single outcome to use as ‘the’ effect. A QALY covers the post-treatment sickness as a quality of life (QoL) issue, and the chance of dying from the operation as a quantity of life dimension. In the gallstone case, the chance of dying from surgery was thought to be 1 in 1000. The QALY outcomes therefore depended mainly on the QoL component. With life expectancies from gallstone patients set at 25 years, 0.001 times 25 produces a QALY amount of 0.025, that is, the quantity of life years element was about 9 days.

Living (having time) with severe pain was judged by the patients to be worth 89 percent of time without pain. So an 11 percent reduction in QoL was assigned to any life years spent in pain. Having severe diarrhea had a QoL loss of 19 percent. For the three symptoms combined that were associated with time after open cholecystectomy, the QoL loss was 56 percent.

Table 1.2 (based on Tables 3 and 4 of Cook et al. 1994) presents the results for the three kinds of gallstone treatment. Cook et al. used two different perspectives for estimating the QALY’s. To simplify matters, we record only the approach that had the higher QALY magnitudes.

Open cholecystectomy has higher costs and lower outcomes (a larger QALY loss) than laparoscopic cholecystectomy and so it cannot ever be more cost-effective than the non-open form of surgery, no matter how one values a QALY. The additional total cost ($1114) of ESWL over laparoscopic cholecystectomy (that is, $5536 – $4422) divided
Cost–benefit analysis and health care evaluation

by the additional effect (lower loss) of 0.005 QAL Ys (or 0.1210 – 0.1205) produces an estimated incremental cost of $2 228 000 per QAL Y gained from ESWL. If one omits the indirect costs, the incremental cost per QAL Y from ESWL would be 62 percent higher at $3 616 000. Cook et al. therefore found that omitting or including indirect costs makes a big difference to the gallstone treatment outcomes.

There are two aspects of the gallstone treatment study that we wish to highlight. On one hand, we see the strength of a CUA in that it quite naturally includes treatment side effects in the determination of outcomes. As we shall see in the next case study, trying to measure side effects in monetary terms for use in a CBA is very difficult. On the other hand, we see the inherent weakness of any CUA study. Cook et al. (1994) found that the cost per QAL Y gained from ESWL was $2 228 000 when indirect costs were included (and as high as $3 616 000 when indirect costs were omitted). At these costs per QAL Y, Cook et al. (1994) doubt that it will be found worthwhile buying. While this judgment is undoubtedly correct, with a CUA we need to know exactly what cut-off point would make a QAL Y worth buying. The authors suggest that a cost per QAL Y of $13 573 might be worthwhile (which would be obtained if we omit laparoscopic cholecystectomy from consideration and look at ESWL over open surgery). But how can one know this for sure? Only a CBA can tell whether a treatment is worth undertaking.

Whether to include indirect costs can be resolved as follows. In a traditional CBA, restoring one’s earnings ability is included as a benefit (indirect) and forgoing earnings in order to undergo treatment is included as a cost (indirect). So, clearly, ignoring indirect costs would not be valid in a traditional CBA because negative earnings are as important to record as positive earnings. Would it be valid to omit such costs from a CUA where the consequence is not measured in earnings?

Relative to a traditional CBA, a CUA can be interpreted to be omitting indirect benefits from the denominator of the cost-effectiveness ratio, but including indirect costs in the numerator. Omitting indirect benefits would be valid in a CUA because the consequence side has an all-inclusive outcome measure in the form of a QAL Y. Quality adjusted time can be used for any purpose one likes, including earning income. Using time for work is not an additional consequence. There would then be no question of trying to add earnings measured in dollars to the QAL Ys. Indirect costs, on the other hand, are not subsumed in the input side by some non-monetary index and so should be included on the denominator of a CUA.

Table 1.2 Costs and effects for alternative gallstone treatments

<table>
<thead>
<tr>
<th>Cost variables and QALYs</th>
<th>Open cholecystectomy</th>
<th>Laparoscopic cholecystectomy</th>
<th>ESWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital cost</td>
<td>$3366</td>
<td>$2581</td>
<td>$4007</td>
</tr>
<tr>
<td>Patient cost</td>
<td>$992</td>
<td>$395</td>
<td>$254</td>
</tr>
<tr>
<td>Cost of conversion</td>
<td>$4358</td>
<td>$3154</td>
<td>$4962</td>
</tr>
<tr>
<td>Indirect cost</td>
<td>$2564</td>
<td>$1268</td>
<td>$574</td>
</tr>
<tr>
<td>Total cost</td>
<td>$6922</td>
<td>$4422</td>
<td>$5536</td>
</tr>
<tr>
<td>Effect (QALY loss)</td>
<td>0.1650</td>
<td>0.1210</td>
<td>0.1205</td>
</tr>
</tbody>
</table>

Source: Cook et al. (1994).
1.5.3 A Worksite Hypertension Program

One of the intangible effects of taking medication is that there are side effects to the treatment. The medications for hypertension (high blood pressure) particularly had this drawback as people were refusing to take the required medication. Low compliance and high treatment dropout rates were mainly responsible for the fact that, in the 1980s, only 30 percent of the US population had their blood pressure under good control. If one were going to conduct a CBA of hypertension treatments, an important part of the analysis would then involve putting monetary valuations on the side effects of the medications. But little work actually has been done in this direction.

As an alternative to trying to find precise estimates of these intangible costs of treatment, one could try to side-step the issue by focusing on compliance as a separate treatment program and trying to achieve greater compliance at least cost. This in effect has been the approach adopted by Logan et al. (1981) who examined whether the greater convenience of worksite treatment programs would increase compliance. The relevance of Logan et al.’s CEA study for CBA is that if one can find a way to make the side effects issue ‘go away’, a standard CBA that ignored compliance effects would be more valid. In effect, by providing a worksite location for treatment that is more convenient, one is lowering the disutility of the side effects (in the sense that the side effects occur in a more comfortable setting and hence become more tolerable) and thereby increasing compliance with the medication regime.

Table 1.3 (based on Tables 3 and 4 of Logan et al. 1981) shows the costs and effects of worksite care (WS), where a person is treated by a nurse at a person’s worksite, compared to regular care (RC), where a person has to make an office visit to a physician. The effect of the treatment is recorded as a mm Hg (millimetre of mercury) reduction in the diastolic blood pressure (BP) of the patient. The direct costs are called here the ‘health system costs’. As we can see in Table 1.3, the direct costs are higher for the worksite program, but the patient costs are lower. When we add the two categories of cost we find that total costs are $31.53 higher for the WS program. For this additional cost, the WS program obtains a 5.6 mm Hg additional reduction in blood pressure (which means that a 1 mm Hg reduction in BP can be purchased for $5.63).

However, as pointed out by Birch (2004), the suggestion that a decision-maker could purchase a 1 mm Hg reduction in BP for $5.63 requires strong qualification. This per unit result, like all CEAs and CUAs, assumes perfect divisibility and (see Chapter 2) constant returns to scale. $5.63 is an average result which may not be representative if

<table>
<thead>
<tr>
<th>Variable</th>
<th>Worksite care (WS)</th>
<th>Regular care (RC)</th>
<th>WS – RC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health system costs</td>
<td>$197.36</td>
<td>$129.33</td>
<td>$68.03</td>
</tr>
<tr>
<td>Patient costs</td>
<td>$45.50</td>
<td>$82.00</td>
<td>($36.50)</td>
</tr>
<tr>
<td>Total costs</td>
<td>$242.86</td>
<td>$211.33</td>
<td>$31.53</td>
</tr>
<tr>
<td>Effect (mm Hg)</td>
<td>12.10</td>
<td>6.50</td>
<td>5.60</td>
</tr>
</tbody>
</table>

Source: Logan et al. (1981).
the distribution of results is skewed. Moreover, this average result for WS was obtained from a program covering 232 individuals costing in total $56,344,893. If this sum is not available to a decision-maker, then the $5.63 cost per 1 mm Hg reduction in BP has no relevance.

The hypertension study highlights a number of principles concerning how to carry out an incremental analysis for a CEA (and these principles will be presented in Part III). However, we are more concerned here with discussing the study’s relevance to the CBA principles established in this chapter. The most important point to note with a CEA is that it does not establish whether any procedure is worthwhile. Logan et al. (1981) conclude that if conventional treatment of hypertension (RC) is considered worthwhile, then it is more cost-effective to replace RC with WS treatment. We know that, under specific assumptions stated above, a 1 mm Hg reduction in BP can be purchased for $5.63 in the WS program. But we do not know whether this is a price that individuals (or society) are willing to pay. Moreover, even though the RC program is less cost-effective than the WS program (the C/E ratio is lower), the RC program could still be worthwhile if it is thought that $32.51 ($211.33/6.5) is worth paying for a 1 mm Hg reduction in BP.

Worth mentioning are the different results from alternative perspectives shown in the study. Logan et al. (1981) were careful to distinguish the total costs that are incurred by society as a whole from those related to the persons being treated for hypertension. In this case, costs for society moved in the opposite direction from those for the patients. The total cost increase of $31.53 from having the WS program consisted of a $68.03 increase in health system costs and a $36.50 reduction in patient costs. Clearly, patients will be more enthusiastic supporters of WS programs than the rest of society.

The fact that patients had the $36.50 saving from the WS program is vital information in understanding how a CEA can supplement a CBA. Recall that the problem was how to incorporate into an evaluation the side effects of the hypertension medication. The CEA comparison of WS with RC can be interpreted as a way of minimizing the costs for overcoming the side effects. Patients in the WS program received greater convenience. Not having to make an office visit translated into a cash equivalent saving of $36.50. One could argue that patients’ behavior revealed their preferences over side effects and so produced an estimate of what the side effects were worth. That is, $36.50 was sufficient to overcome the adverse side effects of the medication, or else the patients would not have taken the medication, and there would not have been the BP reduction observed.

The result is that we are suggesting the use of a two-stage evaluation process. First we employ a CBA to evaluate treatment ignoring the side effects. Then we add to the cost side a sum of money to minimize the side effects, which entails finding the most cost-effective way to overcome those side effects. For the hypertension case, the most cost-effective way was to use the WS program, and the total amount of money to add to the costs was the $68.03 increase in health system costs for the program. The most informative way to view the $68.03 figure is to consider the second stage to be taking the $31.53 amount, reflecting the total additional resources required by the rest of society to set up the nurses at the worksite, and add to this the $36.50 to compensate the patients for taking the medication and putting up with the side effects.
1.5.4 The Sixth Stool Guaiac Protocol

The sixth stool guaiac protocol is a classic in the health care field as a case study in showing what happens when apparently sensible health regulations or procedures are imposed without first undertaking a thorough economic evaluation of the costs and effects of those procedures. Neuhauser and Lewicki published their work in 1975. Subsequently, there has been a re-evaluation of the case study (see Brown and Burrows, 1990; Neuhauser, 1990; Gatsonis, 1990). Our view is that a lot can be learned from this study even if everything was not as assumed by Neuhauser and Lewicki. We first explain exactly how such a large figure of $47 million per case detected was obtained. Then we explore some of the wider implications following from the case study.

In earlier work, D.H. Greegor had studied asymptomatic colonic cancer, which involved testing the stool for occult blood. Unfortunately, some of those with cancer were not detected by a single test. Greegor therefore recommended, and the American Cancer Society endorsed, a protocol of six sequential tests. If a positive outcome for a test were found, the existence of cancer would be confirmed by a subsequent barium-enema examination.

The cost of an individual test was very small. The first stool test was assumed to cost $4, and each subsequent test cost $1. The barium-enema procedure was considered to cost $100. Clearly, so it seemed in a CM-type framework, it would be cheaper to use the stool tests than to use the barium-enema procedure on everyone. The issue was, though, how many extra cases would be detected by the sixth test.

Neuhauser and Lewicki used Greegor’s results, which were based on a screening of 278 people, and extrapolated them for a hypothetical study of 10,000 people. Thus, because two cases were confirmed out of the 278 cases, there would be 72 cases in a population of 10,000. They treated each one of Greegor’s tests as an independent event with a 91.67 percent chance of detecting cancer. After one test, with 91.67 percent of the cases detected, there would be 8.33 percent of the cases undetected. Test two would have a 91.67 percent chance of detecting the remaining cases, which means there would be an additional 7.64 percent cases detected (that is, 0.9167 times 8.33 percent). And so on. What this means is that of 72 cases of colonic cancer, 66 cases would be detected after one test, and 71 cases after two tests. Clearly, there were very few cases left undetected (precisely one) after two tests. By undertaking test six, only 0.0003 extra cases were detected. The cases detected and the corresponding costs are indicated below in Table 1.4 (based on their table 2). (The ‘incremental’ cost is the cost of one more test, while ‘marginal’ cost is the cost per case detected from one more test.)

The final column is what interests us. It shows that for an extra 0.0003 of a case detected, the extra cost involved with the sixth test was $13,190. In other words, an extra case detected by the sixth test was $47 million ($13,190/0.0003). Note that if the occult blood testing was bypassed and the barium used as the screening procedure for 10,000 persons, the total cost would be $1 million (10,000 times $100). The (marginal and average) cost would have been $13,900 per cancer case detected. Thus, the sixth stool guaiac protocol was not even cost minimizing (it was $710 higher than the barium procedure).

One general issue concerning the guaiac protocol that is often found in health care evaluations is the importance of assuming that treatments are divisible and that a
Table 1.4 Detection rates for successive stool guaiac tests

<table>
<thead>
<tr>
<th>Number of tests</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>69.9469</td>
<td>71.4424</td>
<td>71.9004</td>
<td>71.9385</td>
<td>71.9417</td>
<td>71.9420</td>
</tr>
<tr>
<td>Incremental gain</td>
<td>65.9469</td>
<td>5.4956</td>
<td>0.4580</td>
<td>0.0382</td>
<td>0.0032</td>
<td>0.0003</td>
</tr>
<tr>
<td>Total detected</td>
<td>77511</td>
<td>107690</td>
<td>130199</td>
<td>148116</td>
<td>163141</td>
<td>176331</td>
</tr>
<tr>
<td>Incremental cost</td>
<td>77511</td>
<td>30179</td>
<td>22509</td>
<td>17917</td>
<td>15024</td>
<td>13190</td>
</tr>
<tr>
<td>Marginal cost</td>
<td>1175</td>
<td>5492</td>
<td>49150</td>
<td>469534</td>
<td>4724695</td>
<td>47107214</td>
</tr>
<tr>
<td>Average cost</td>
<td>1175</td>
<td>1507</td>
<td>1810</td>
<td>2059</td>
<td>2268</td>
<td>2451</td>
</tr>
</tbody>
</table>

*Source:* Based on Neuhauser and Lewicki (1975).
proportional relationship concerning outputs and costs applies. One needs to be aware that $47 million was not actually at stake and could have been devoted elsewhere. The extra cost of the sixth test was only $13,190. The point was that for this $13,190 very little was obtained (0.0003 of a case detected). If one purchases cases at this rate, and if there were this proportional relation between cost and cases, then one would end up paying $47 million if one stopped when one purchased exactly one case. But, actually, only $13,190 would have been saved if the protocol had not been introduced. CEAs and CUAs are especially vulnerable to relying on this proportionality assumption because their outcome measures are always expressed in ratio form.

In terms of the specifics of the case study, the main message (emphasized by Gatsonis) is the need to obtain precise estimates of clinical and economic outcomes when making economic evaluations. The Neuhauser and Lewicki study was based on working with the assumption that only two cases of colonic cancer were found in the study population. However, these were just the proven cases. Some of the people who had negative test results could have had the disease; Gatsonis (1990) thought that the possibility that further cancers may have existed among them is not small. If instead of two cases there were three cases, all the numbers would change dramatically and the sixth test would not have been nearly so costly. One of the real problems therefore was in the Greegor data (concerning the prevalence of the disease) which Neuhauser and Lewicki used to carry out their evaluation.

The evaluation of the sixth stool guaiac protocol had a widespread impact. Not only did the study greatly affect the whole field of clinical decision-making, but it also changed actual policy. As pointed out by Getzen (1997), the study (and its aftermath) caused the American Cancer Society to amend its recommendations such that routine stool guaiac screening is now recommended only for individuals over age 50, or whose family history makes them at high risk for colorectal cancer.

1.5.5 The Effectiveness of the Millennium Villages Project

When should one carry out a full blown evaluation of an intervention and when should one use an evaluation framework that is simpler/cheaper? Clemens and Demombynes (2010) present some general advantages and disadvantages and then go into an application which highlights the answer in a specific setting.

The main advantages of carrying out a rigorous evaluation are increased transparency about what are the main factors that led to the results and avoiding the waste of scarce resources on ineffective projects. The disadvantages are that: inordinate attention is placed on easily quantifiable outcomes; the rigorous evaluation could take up too much time when a policy decision needs to be made much sooner; and resources are taken up in evaluation that could be used to scale up projects that are already known to work well.

The case study that Clemens and Demombynes (2010) focused on, to come up with a concrete example as to when a rigorous evaluation was necessary in practice, related to the Millennium Villages Project (MVP). This was a large experimental intervention that aimed to generate economic development in 14 village clusters across Africa. The defining characteristic was to provide a broad package of intense and expensive interventions in a poor, rural region over a number of years. The particular package of interventions included:
local infrastructure, such as roads, electricity and communications;
- improvements to agriculture technology such as fertilizers and new seed varieties;
- improvements in human capital involving the construction and staffing of schools;
- improvements to public health such as the creation of free clinics and improvements to water and sanitation facilities; and
- increased financial access in terms of the introduction of microloans and micro savings instruments.

Unfortunately the past history of integrated rural development (IRD) projects was not good. Clemens and Demombynes (2010) give examples relating to China, India and Kenya which were generally agreed to be unsuccessful, largely because any benefits that resulted were not sustained. The record was so bad that the World Bank decided to discontinue IRD projects.

However, proponents of the MVP argued that this time things would be different. There was now a 5- to 10-year commitment and not just 2–3 years as with typical IRD projects. Also the MVP had targets that were quantitative and time-bound and relied more on decentralization to local governments. Clearly a prima facie case for the need for a rigorous evaluation exists given the enormous differences in expectations of those who support the MVP and those who reject MVP type interventions.

To understand the thrust of Clemens and Demombynes's argument one needs an outcome measure. The MVP specified that under-5 child mortality is to be the primary indicator for the evaluation. But data on child mortality rates were not available for Clemens and Demombynes to analyze. In their analysis they looked at a number of different measures, such as access to improved sanitation, mobile phone ownership and human immunodeficiency virus (HIV) testing. We will use just one measure, bednet usage, that is, the percentage of children under 5 who slept under an ITN (insecticide-treated net) last night. This measure is appropriate because bednet usage was a major reason why child mortality fell so dramatically in Africa between 2005 and 2010. To simplify, we will focus just on bednet usage in Kenya as this was the country in Africa with the largest reduction in child mortality and where bednet usage was found to be the main significant determinant – see Demombynes and Trommlerova (2012).

Now we need a method for deciding whether the MVP was effective or not. The MVP organizers stated that project impact was to be judged on the basis of a before-and-after comparison. Since this is an evaluation method that is often used in health care evaluations we will start off analyzing results with this method. As we shall see, this method is simple to understand and use, but it is not a rigorous evaluation design. The first row of numbers in Table 1.5 gives the results for this before-and-after comparison for Kenya. The rise in bednet usage was dramatic in the MVP rising from 10 percent usage to 62 percent, a rise of 52 percentage points.

The trouble with the before-and-after comparison is that we do not know for sure that the after results are purely due to the MVP and not something else, such as the general rise in incomes throughout Kenya. To obtain some idea of what would have happened at the MVP site in the absence of the actual MVP intervention one could use the bednet usage experience of the region in which the MVP was situated. The second line of figures in Table 1.5 shows that in the MVP region bednet usage also rose rapidly from 28 percent to 59 percent, a difference of 31 percentage points. So the difference the MVP interven-
tion made was 52 percent minus 31 percent, that is, 21 percentage points (obtained by subtracting the second row total from the first row total). This latter figure is called the ‘difference in difference’ (DID) estimator and this better depicts what the MVP really contributed.

An alternative way of seeing what the DID figure is achieving is to consider a second way of calculating it, that is, by subtracting the second column total from the first column total and this is shown in the last row of figures in Table 1.5. One still gets the 31 percentage points difference. But this time we can see that most of this difference comes about because the initial difference was so large. In 2008, the MVP intervention was only 3 percentage points better than the MVP region. It looks as though not much was achieved, but the MVP intervention started off 28 percentage points lower and so 31 percentage points was the real turn around, not just 3 percentage points.

In a completely rigorous evaluation, one would still want to adjust the DID estimate for any pre-existing differences in the people who live in the MVP region and those in the MVP site. Formally, one needs to form a control group that has the same characteristics as those in the MVP site yet do not get exposed to the MVP package of interventions. We explore this idea in greater depth in Chapter 6. At this stage we only need to point out that generating a control group costs money and so comes at a price. Clemens and Demombynes (2010) estimate (conservatively) that it would cost $240 000 to establish a proper control group for this evaluation.

To get a feel for whether it would be worthwhile to invest $240 000 to obtain a more rigorous evaluation, we will assume that the MVP before-and-after results are what a simple evaluation would generate, while a rigorous evaluation with a control group would produce exactly the DID results. Now assume that the 52 percentage points improvement given by the simple evaluation would just break even, such that the net benefits would be zero. As the costs were $1.5 million per village in the Clemens and Demombynes (2010) analysis, this makes the benefits also worth $1.5 million. The rigorous evaluation on the other hand showed that the benefits were only 44 percent as large (that is, 21 percent over 52 percent). As 0.44 of $1.5 million is $0.66 million, the outcome should have been a loss of $1.5 million – $0.66 million, or $0.84 million. Doing a rigorous evaluation would clearly have been worthwhile seeing that for the cost of $240 000 one would have avoided the $840 000 loss, a saving of $600 000.

The message from this case study is that the simple before and after evaluation that is often used could easily be misleading (obtaining a positive result when it really is negative). A rigorous evaluation could save a lot of money especially if the intervention one is evaluating (like the MVP intervention) involves investing a lot of money (millions of

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MVP</td>
<td>0.10</td>
<td>0.62</td>
<td>0.52</td>
</tr>
<tr>
<td>MVP region</td>
<td>0.28</td>
<td>0.59</td>
<td>0.31</td>
</tr>
<tr>
<td>Difference (MVP – MVP region)</td>
<td>−0.18</td>
<td>0.03</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Source: Constructed by the author based on Clemens and Demombynes (2010, table 1).
dollars or more). If one wants to scale up the MVP as its proponents intend, one needs to discover any possible ineffectiveness early on by doing a rigorous evaluation and so avoid any large potential losses.

1.6 FINAL SECTION

All chapters in this book will contain a problems section, a chapter summary, and a section indicating where some of the unresolved issues will be covered in later chapters.

1.6.1 Problems

The four main methods for making an economic evaluation in health care have been introduced in this chapter. The first set of problems that follow require the reader to select and assemble from a set of categories of costs and outcomes the necessary ingredients to conduct each of the four kinds of evaluation.

The evaluation being considered in Table 1.6 is Boyle et al.’s (1983) study of neonatal intensive care in Canada. (All monetary figures cited are in Canadian dollars.) The provision of neonatal intensive care involves increased current capital expenditures (to control the respiratory, nutritional and environmental circumstances of the baby) in order to increase a baby’s future survival chances. The costs and consequences for babies with birth weight 1000–1499 g are listed in Table 1.6 (all figures are undiscounted).

1(a). Undertake a cost minimization comparison of neonatal intensive care (before versus with intensive care) from: (a) the hospital’s perspective and (b) society’s perspective.

1(b). Undertake a CEA comparison of neonatal intensive care from society’s perspective. (Hint: more than one category of consequence could be used to form the CEA comparison.)

Table 1.6 Evaluation of neonatal intensive care treatment

<table>
<thead>
<tr>
<th>Cost or consequence</th>
<th>Before intensive care</th>
<th>With intensive care</th>
<th>Incremental effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cost per additional survivor (to hospital discharge)</td>
<td>$5400</td>
<td>$14 200</td>
<td>$8 800</td>
</tr>
<tr>
<td>2. Cost per additional survivor (to death)</td>
<td>$92 500</td>
<td>$100 100</td>
<td>$7 600</td>
</tr>
<tr>
<td>3. Survival rate (to hospital discharge)</td>
<td>62.4%</td>
<td>77.2%</td>
<td>14.8%</td>
</tr>
<tr>
<td>4. Survival time (per live birth):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Life-years</td>
<td>38.8</td>
<td>47.7</td>
<td>8.9</td>
</tr>
<tr>
<td>b. QALYs</td>
<td>27.4</td>
<td>36.0</td>
<td>8.6</td>
</tr>
<tr>
<td>5. Earnings</td>
<td>$122 200</td>
<td>$154 000</td>
<td>$32 000</td>
</tr>
</tbody>
</table>

Source: Based on Boyle et al. (1983, tables 1 and 2).
1(c). Undertake a CUA comparison of neonatal intensive care from society’s perspective. (Hint: only one category of consequence can be used to form the CUA comparison.)

1(d). Undertake a traditional CBA comparison (that is, use the human capital approach) to evaluate neonatal intensive care from society’s perspective. (Hint: only one category of consequence can be used to form the CBA comparison.)

The second problem set reinforces an understanding of how, in the MVP case study examined in section 1.5.5, we took what was essentially a CEA and converted it into a CBA.

2(a). Explain all the steps (and assumptions used) to calculate the $1.5 million estimate of the benefits of what a rigorous evaluation of the MVP would generate.

2(b). Given your answer to (2a), what would you call the method used to generate this benefits estimate? (Hint: re-read section 1.3.)

1.6.2 Summary

This chapter started with a definition of a health care evaluation that stressed the need to consider alternatives. It was argued that one needed to carry out evaluations in order to ensure that funds were put to their best use. We then introduced the four main types of economic evaluation that comprise the health care evaluation field. In explaining the logic of each type of evaluation, we first set up CBA as the primary evaluation method, and then showed that the other types of evaluation could be considered to be special cases of a CBA. It was the fact that CBA measured both inputs and outputs in common, monetary terms that made CBA the only method that can actually tell whether a health care intervention should, or should not, be undertaken.

Although there is this unity from a theoretical point of view, the focus in all cases was on identifying the practical circumstances necessary to validate a particular method. This focus will be the defining characteristic of the book. That is, we are trying to develop an understanding of the strengths and weaknesses of trying to apply each of the evaluation types. With this in mind, we explained how it was necessary to generate a price for an effect if one were going to use the incremental cost-effectiveness ratio to decide whether interventions are worthwhile. We took this opportunity to introduce readers to the UK’s NICE program and to the use of the threshold method, two topics that we will be covering many times in the rest of the book.

In the outline of the basics of CBA, we identified consumer sovereignty as the main value judgment. For those applications where it is sensible to assume that the individual is the best judge of his or her welfare, modern CBA based on WTP is the appropriate evaluation framework. How practical WTP is as an approach to estimating benefits is a separate issue and one that will be explored in depth in this book. The modern approach has to be contrasted with the traditional approach to CBA based on the concept of human capital. Traditional CBA, which uses earnings to measure the benefits, has the merit that data on production are readily available. This largely accounts for its popularity in the health care field. But it is not based on firm welfare economic principles that depend on
individual preferences for their validity. Also the human capital approach has its critics on equity grounds.

The applications section revealed the strong preference in the practice of health care evaluations for non-CBA types of evaluation. Thus the applications covered the three main alternatives to CBA: CUA, CEA and CM. The CUA study of gallstone treatments highlighted the strength of CUA to be able to include routinely in its evaluations quality of life issues, such as pain and suffering; something which CBA often fails to incorporate. In the CEA of hypertension treatment locations, we showed how a CEA could be used in a back-up role to a CBA to help it include treatment side effects. However, in each of the first two applications, we emphasized the inherent weakness of using any non-CBA type of evaluation. One could not say whether any of the treatment alternatives were worthwhile. The final application, which was conducted in a CM-type framework, returned to the theme introduced at the beginning of the chapter. That is, why it is necessary to explicitly carry out a health care evaluation. Without a formal evaluation, one could be wasting dollars that could be used elsewhere to save lives or otherwise make society much better off.

1.6.3 Looking Ahead

With only the introductory chapter (Part I) covered at this stage, looking ahead involves providing a guide to the rest of the book. Each of the four evaluation types will be analyzed in depth. Part II deals with CM, CEA forms Part III, CUA is in Part IV and Part V is on CBA. Part VI summarizes the book.

Since all four evaluation types depend on costs, and this is all there is in a CM, we start, in Chapter 2, with defining costs from an economic perspective and examining the rationale of CM as a type of economic evaluation. Chapter 3 explains some of the problems of measuring costs and examines whether hospital and physician charges can be used. The concept of costs is then broadened in Chapters 4 and 5 to include the costs on the rest of society, whether they are people not directly involved with treatment (generating so called ‘external costs’) or general taxpayers.

Chapters 6 and 7 are on CEA. They analyze its relationship to CBA and cover the controversial issues involved with discounting effects. Chapter 8 starts the part on CUA and examines its link to CBA. Chapter 9 describes the main instruments used to estimate the utilities, and Chapter 10 examines other alternatives to a QALY in an equity context. The two different kinds of CBA, traditional and modern, are discussed in Chapters 11 and 12. Chapter 13 explores how to include equity considerations in CBA.

In the final chapter we take just one disease, HIV/AIDS (without treatment the human immunodeficiency virus HIV leads to the acquired immunodeficiency syndrome AIDS), and show how all the main methods covered in this book can be used to evaluate interventions in this one area. This chapter serves the purpose of being a summary of the whole book as the strengths and weaknesses of all the main evaluation methods presented throughout the book are brought together in one place and so can be readily compared.