of illness measure is a lower bound to willingness to pay as revealed by contingent valuation.

Chapter 5 draws implications for the value of health from studies of the household production of health. While relevant work is limited, several studies are reviewed that yield illustrative empirical estimates of the value of acute morbidity.

Chapter 6 discusses an approach widely used by health professionals: the quality-adjusted life year, or *qualy*, approach. The chapter describes the qualy approach and compares it to the economic approach used in this book. A major goal of the chapter is to incorporate useful results from qualy analysis into the framework of this book, enlarging available estimates of health values.

2

Framework for Valuing Health Risks

Mark Berger, Glenn Blomquist, Donald Kenkel, and George Tolley

2.1. Introduction

In this chapter we develop a model of health investment that yields a general expression for the value of changes in risk to human health. The preference-based values of morbidity risks and mortality risks are ex ante dollar equivalents of changes in expected utility associated with risk changes. The values of changes in morbidity risks and mortality risks are related to two alternative measures, costs of illness and preventive expenditures, which are thought to be lower bounds on the value of risk reductions. We demonstrate that these alternative measures are not even special cases of the more general measure and that the size relationships among the three measures are complex. Also, we derive the relationship between willingness to pay for risk changes and the consumer surpluses associated with health changes that occur with certainty.

This chapter begins with a review of several approaches to valuing changes in risks that are currently in use. The model of health risk behavior is developed, and implications for benefit estimation from the model are discussed.

2.2. Approaches to Valuing Health Risks

Cost of Illness

The traditional approach to measuring the benefits of improved health is based on avoidance of disease damages. The damage avoidance approach, which is the form used by health professionals and some health economists, is also referred to as the *cost of illness approach* or sometimes the *earnings expenditure approach.* The cost of illness approach relies heavily on the idea that people are producers, that is, human machines. Outlays for health services are seen as investments that improve people as productive agents and

This material was originally published in M. C. Berger, G. C. Blomquist, D. Kenkel, and G. S. Tolley, "Valuing Changes in Health Risks: A Comparison of Alternative Measures," *Southern Economic Journal* 53 (April 1987): 967–84. Reprinted with permission.

yield a continuing return in the future. The yield for improvements in health is the labor product created plus any savings in health care expenditures due to any reduction in disease (see Mushkin 1962, pp. 130 and 136). The costs of health degradation are the damages caused by the disease (or accident). The health expenditures made, the value of the resources used in supplying health care, are referred to as the *direct cost of illness.* The loss of labor earnings due to sickness and premature death, the value of the lost product of labor, is referred to as the *indirect cost of illness.* The value of health improvements is the sum of the reductions in direct and indirect costs of illness, that is, the damages that will be avoided. Studies employing the cost of illness approach include Weisbrod (1971), Cooper and Rice (1976), and Mushkin (1979).

Several deficiencies in the cost of illness approach are recognized: (1) the indirect costs are zero for retirees, full-time homemakers, and other people who do not work in the market; (2) an arbitrary decision must be made about forgone consumption expenditures, that is, gross or net labor earnings; (3) individuals are viewed as having no control over their health or health care expenditures; and (4) there is little basis in economic theory for the use of the costs of illness in benefit-cost analysis. An attempt has been made by Landefeld and Seskin (1982) to reformulate costs of illness values to more closely approximate a theoretically correct measure, but their study primarily focuses on externalities, and an approach more closely tied to individual optimization seems more appropriate. Chapter 3 examines in much greater detail the cost of illness approach as a possible source of estimates of the benefits of health risk reduction.

Willingness to Pay in Contingent Markets

The absence of a market for health as such has prompted consideration of direct questioning techniques to elicit willingness to pay for changes in health risks. Through a survey interview or laboratory experiment a hypothetical market is established, and individuals are asked to purchase changes in health directly contingent upon the existence of the market. Contingent valuation of mortality risks was pioneered by Acton (1973) in his study of heart attack treatment and has been used by Loehman et al. (1979) to value morbidity related to air pollution. Currently, there is renewed interest in direct questioning because it yields conceptually correct values of health risk that are difficult to estimate using other techniques.

Contingent valuation is considered in detail in Part 2. Empirical results applied to the value of morbidity are reviewed in Chapter 4.

Household Production of Health and Preventive Expenditures

While the cost of illness approach concentrates on damages or costs following the onset of illness, individuals can and do incur costs in efforts to prevent illness from ever occurring. In Grossman's (1972) model of consumption and production of the commodity "good health," individuals combine purchased goods such as medical care and their own time to produce health capital. Willingness to pay is the value of healthy time and is the sum of two terms: (1) the increment in labor earnings that is possible, and (2) the monetary value of the gain in utility associated with better health. Thus, the household production model gives a conceptual foundation for the relevance of labor earnings (indirect costs) for morbidity, but it also implies that a preference-based value will depend on the costs of producing health (preventive expenditures) and a utility, or consumption, value. An example of the household production approach is Cropper's (1981) micro study of the effect of air pollution on days lost from work due to illness. To value the health changes, she multiplies the wage rate by a factor derived from a specific production function.[1] This and other studies using the household production approach are discussed in Chapter 5.

The recognition that health is partly endogenous has also spawned the idea that health improvements permit a reduction in preventive expenditures and that the savings of preventive expenditures is the value of the health improvement. This general approach has been suggested as a way to measure the benefits of reducing pollution where the expenditures prevent not only damages to human health but also damages to property and so forth. Courant and Porter (1981) characterize the literature as having reached a limited consensus that such expenditures represent a lower bound to the total costs of pollution, a conclusion they dispute.

Smith and Desvousges (1985) find that households make adjustments to reduce the risk of exposure to hazardous wastes through drinking water. In their sample of households in suburban Boston, nearly 30% purchased bottled water regularly to avoid hazardous wastes, while smaller fractions installed water filters and attended public meetings as ways to reduce the risks. This study provides important evidence that averting or preventive behavior in response to pollution risks can be significant. However, the relation between preventive expenditures and the benefits of improved health has received too little attention. We explore this relationship.

Willingness to Pay in Implicit Markets

One implication of household production models of health is that individuals will make expenditures of money and time to improve their health and reduce risks to their health. By observing people's behavior in well-developed

1. Cropper (1981) obtains estimates of valuation of health changes only under very specific assumptions. Gerking and Stanley (1986) do so more generally, estimating the value of a change in health as the cost of preventive activity times an estimated ratio of marginal products of inputs in the health production function.

markets for ordinary goods and services, values can be derived for health, which is not traded explicitly. Much of this type of evidence comes from the labor market in the form of estimates of compensating wage differentials for jobs with extraordinarily high risks to health and survival. Most of the studies focus on implicit values of changes in the risk of a fatal accident.

Consumption activity also can involve exchanges between health and safety and other desirables. Estimates of willingness to pay have been made based on analyses of residential housing site choice, automobile seat belt use, speed of travel on highways, and cigarette consumption.[2] This work, like that in the labor market, has focused on mortality risk. Inherent in this methodology of estimating implicit values of health risks is that individuals know and perceive differences in health risks associated with various jobs and consumption activity and that they can choose among various alternatives.

When investigating how workers and consumers make choices regarding risks to health, it is important to recognize that the utility individuals derive from consumption depends upon their state of health. Fatal risks pose the most extreme case, where one state of the world corresponds to the individual's death. In this case, expected utility depends upon consumption if alive, and the satisfaction from leaving bequests if death occurs. More generally, economists use the theory of state-dependent utility as applied to variations in health status by Zeckhauser (1970) and Arrow (1974). Viscusi (1979) was the first to apply the state-dependent approach to estimate the value of fatal risks faced by workers. Viscusi and Evans (1990) estimated state-dependent utility functions for the case of nonfatal job injuries and provided the first empirical evidence that the marginal utility of consumption is lower in the ill-health state. Below we use the state-dependent approach to model both mortality and morbidity risks.

A General Framework for Valuation

At this point there appear to be two disparate approaches to valuation of health and risks: cost of illness, perhaps inclusive of preventive expenditures, and willingness to pay. Research has proceeded using one approach or the other, but only limited effort has been made to compare and reconcile the approaches. A paper by Harrington and Portney (1987) is noteworthy in that they show that for morbidity, under certain conditions, the cost of illness values will be a lower bound on the theoretically preferred willingness to pay values. Below we develop an eclectic model with endogenous

2. For a review of labor market studies, see Smith (1979). For a comprehensive survey of the literature on willingness to pay and fatality risks, see Blomquist (1982) and Fisher, Chestnut, and Violette (1989).

health risks and derive the preference based values for changes in health risks. The model considers morbidity and mortality and allows the probabilities of various health states and survival to be influenced by preventive activity and exogenous factors such as environmental quality. Terms for preventive expenditures and costs of illness in the benefit expression are identified for purposes of comparison with the conceptually correct willingness to pay. The model provides a framework for comparing values of health risks estimated using various techniques.

2.3. Human Health Risk Reduction Benefit Model

Assume a person's utility depends on the consumption of goods and services and the state of health. Utility may be expressed as

$$U = U(C, q), \tag{2.1}$$

where U is utility, C is consumption, and q is a vector of health characteristics.[3]

A person does not know with certainty, however, what his or her health will be or, for a given state of health, whether he will survive the period in question. In order to incorporate these uncertainties into the model, we specify the probability of health characteristics and probability of survival functions. The probability density function for health characteristics can be represented as

$$h(q; X, E), \tag{2.2}$$

where X is preventive expenditures and E is any exogenous shift variable, such as environmental change. The health characteristic probabilities are not immutable but, rather, are influenced by preventive measures chosen by the individual person and exogenous changes such as environmental improvement.

It is reasonable to assume that the healthier a person is, the greater are the chances of survival of a given period. In other words, probability of survival can be expressed as a function of health characteristics:

$$p = p(q), \tag{2.3}$$

where p is the probability of surviving the period.

3. Consumption, C, consists of both expenditures on market goods and services and on time, combined in fixed proportions. If the value of time is constant at the market wage rate, then consumption time expenditures are simply the product of the wage and the amount of time spent in consumption activities. Preventive expenditures (X) and costs of illness (Z) introduced below are also assumed to consist of expenditures on time and market goods combined in fixed proportions.

A final element of the model facilitates comparisons with the cost of illness approach for valuing health risk reductions. When in poor health, a person incurs cost such as medical expenditures and earnings lost due to days not worked. These costs will vary according to the degree of illness malfunction that occurs:

$$Z = f(q), \tag{2.4}$$

where Z is the cost incurred as a result of illness malfunctions. These expenditures reduce consumption and provide no utility on their own.[4]

In this framework, a person chooses preventive expenditures X in order to maximize the expected value of utility given the following income constraint:

$$M = C + X + Z, \tag{2.5}$$

where M is money income in the absence of any costs due to illness malfunctions.[5]

Preventive expenditures influence the expected value of utility in three ways: (1) X increases the probability of being in good health, therefore increasing utility if alive; (2) at the same time, increasing the probability of being in good health also increases the probability of being alive; (3) finally, by increasing the probability of being in good health, X expenditures decrease malfunction costs Z that can be expected, increasing the amount of income expected to remain for consumption. These benefits must be weighed against the direct loss in consumption made necessary by the preventive expenditures.[6]

More formally, the consumer's problem can be stated as

$$\max E(U) = \int_{-\infty}^{\infty} U(C, q)p(q)h(q; X, E)dq, \tag{2.6}$$

subject to the income constraint (2.5). Reexpressing the income constraint in terms of C and substituting it into (2.6), the consumer's problem becomes

4. Typically, the cost of illness approach only includes earnings lost or the value of time lost from work and excludes the value of time lost from consumption activities. Define $Z^* = Z - C_L$, where C_L is the value of time lost from consumption. In our empirical comparisons of the cost of illness and willingness to pay approaches in Section 4.5, we employ the more widely used Z^* definition of the cost of illness.

5. Money income, M, is the sum of nonlabor income and potential earnings. Assuming the wage rate is constant, potential earnings are simply the product of the wage rate and the total time in the period. The individual's problem can be expressed in terms of the choice of X, rather than its goods and time components, because of the fixed proportions assumption for X, C, and Z.

6. Just as with Z expenditures, X expenditures provide no utility directly by themselves.

$$\max E(U) = \int_{-\infty}^{\infty} U[M - X - f(q), q]p(q)h(q; X, E)dq, \tag{2.7}$$

where U, p, and h come from equations (2.1), (2.3), and (2.2), respectively.[7]

The integral in (2.7) gives utility under different health outcomes weighted by the probability of the various outcomes. Since utility always depends upon health, the situation could be described as a continuum of state-dependent utility functions, the possible states being the possible health outcomes. Different attitudes toward risk are allowed for through the shape of each state-dependent utility function. When utility is expressed as $U[M - X - F(q), q]$, it becomes apparent that preventive expenditures X directly reduce the amount of income left over for consumption. The term $p(q)$ in (2.7) adjusts utility by the probability of being alive. Assuming no utility if dead, $U[M - X - f(q), q]p(q)$ gives expected utility conditional on the state of health. A more extended analysis might consider utility of heirs as affected by bequest. The density function $h(q; X, E)$ weights expected utility by the probabilities of different states of health. The integration over health states thus gives expected utility for the period.

The model as described does not specify fully the mechanisms available to the individual to adjust to risk such as market insurance. The only opportunity the individual has is to make ex ante preventive expenditures X that change the probabilities of the different states. Another extension of this analysis could be to carefully describe what opportunities are available to the individual to adjust expenditures made in each state of the world. Though these opportunities could easily be made explicit in the present model, this section retains the simpler framework in order to make the comparisons between preventive expenditures, cost of illness, and willingness to pay for risk reductions more straightforward. However, in general, willingness to pay values are affected by the opportunities available to adjust to risk, so it is vital to note the simplified framework used.

The problem also becomes more tractable if a single health outcome measurable as a zero-one condition is considered. An example is the occurrence of a specified type of cancer as affected by environmental irritants. Another example is the occurrence of traffic accidents due to poor visibility brought on by air pollution, provided the major cost is associated with frequency of accidents, all having about the same expected severity, rather

7. Although the consumer's problem as expressed in equations (2.6) and (2.7) is single period in nature, it can be generalized to allow for multiperiod planning as has been done by Cropper (1977). In particular, suppose the probability density function, the probability of survival function, and the utility function all vary over time. Assuming an infinite planning horizon, the consumer's problem can be restated as max $E(U)$ = the integral from T to infinity of the integral from negative infinity to positive infinity of $U[M_t - X_t - F(q_t), q_t; t]p(q_t, t)h(q_t, X_t, E_t, t)dqdt$.

than the severity of an individual accident being importantly related to the degree of visibility. Tissue damage from contact with pollutants, such as liver damage, is another example as long as the principle effect is absence of unimpaired functioning rather than the degree of malfunctioning being associated with the degree of pollutant level.

A damage function, as might be the case for ozone, where the degree of discomfort rather than the presence or absence of discomfort is related to the level of pollution, requires a more extended analysis considering probabilities for more than two states of the world. Various degrees of symptoms along with their associated probability densities have to be considered, rather than just the presence or absence of symptoms. The integral in (2.7) would not simplify as it does in the case where there is only one malfunction state.

If health is a matter only of absence or presence of a deleterious condition, the probability density function $h(q; X, E)$ is discrete rather than continuous with probability concentrated at $q = 1$ for presence of the condition and $q = 0$ for absence of the condition:

$$\begin{aligned} h(q; X, E) &= H(X, E) && \text{if } q = 1, \\ h(q; X, E) &= (1 - H(X, E)) && \text{if } q = 0, \end{aligned} \tag{2.8}$$

where $H(X, E)$ is the probability of the absence of the condition.

In this case, the person decides at the beginning of the period what his or her preventive expenditures will be and then takes the resulting chance of what the health outcome will be for the period. A long planning period can be considered by letting consumption expenditures, illness costs, and preventive expenditures be average discounted present values, with the probabilities associated with survival and health status being averages of shorter-term probabilities, possibly allowing for cumulative exposure effects.

Because of the discreteness of q when health is a matter only of the absence or presence of a condition, the integral in (2.7) simplifies to a sum of two discrete states corresponding to $q = 0$ and $q = 1$. Using (2.8), the consumer's maximization problem is

$$\max E(U) = U_0 P_0 (1 - H) + U_1 P_1 H, \tag{2.9}$$

where

$U_0 = U(M - X, 0)$ is utility if free of the disease;
$U_1 = U(M - X - Z, 1)$ is utility with the disease;
$P_0 = p(0)$ is probability of survival if free of the disease;
$P_1 = p(1)$ is probability of survival with the disease; and
$H = H(X, E)$ is the probability of contracting the disease.

Equation (2.9) states that the expected utility to be maximized is the sum of utilities in the absence and the presence of the deleterious health condition, weighted by the probabilities of contracting and not contracting the disease and of surviving. As can be seen from the expressions for U_0 and U_1, utility depends both on the presence or absence of the disease; that is, there is state dependence. The income constraint has been substituted into the utility function just as in equation (2.7). In the discrete case, this constraint can be expressed as[8]

$$\begin{aligned} C &= (M - X) && \text{if } q = 0, \\ C &= (M - X - Z) && \text{if } q = 1. \end{aligned} \tag{2.10}$$

Differentiating equation (2.9) with respect to preventive expenditures X and setting the result equal to zero gives the first-order condition for a maximum

$$\begin{aligned} F &= U_0' P_0 (1 - H) - (U_1' P_1 H) - (U_0 P_0 H_X) + (U_1 P_1 H_X) \\ &= 0, \end{aligned} \tag{2.11}$$

where U_0' and U_1' are the marginal utilities of income when $q = 0$ and $q = 1$, respectively, and H_X is the change in the probability of contracting the disease resulting from an extra dollar spent on prevention. The first two terms give the decline in expected utility due to decreased consumption when an extra dollar is spent on defensive measures. The last two terms give the rise in expected utility due to decreased probability of contracting the disease as a result of the extra dollar spent on prevention. The first-order condition for a maximum is that the sacrifice of consumption given by the first two terms must just offset the gain from the reduced probability of contracting the disease given by the last two terms.

In order for the consumer to obtain a maximum, the second derivative of the expected utility function with respect to preventive expenditures must be less than or equal to zero. This second-order condition can be expressed as

$$\begin{aligned} \Delta &= U_0'' P_0 (1 - H) + (U_1'' P_1 H) - (U_0 P_0 H_{XX}) + (U_1 P_1 H_{XX}) \\ &\quad + 2H_X (U_0' P_0 - U_1' P_1) \le 0, \end{aligned} \tag{2.12}$$

where H_{XX} is the second partial derivative of $H(X, E)$ with respect to X, and U_0'' and U_1'' are the second derivatives of utility with respect to income when $q = 0$ and $q = 1$, respectively.

8. Note that, for any given individual, Z is fixed once the disease is contracted. In a more extended analysis, Z could be made to depend on other variables such as the price of medical care. Z could be made endogenous in the current framework if it were specified as a function of preventive expenditures.

2.4. Valuation of Changes in Risks to Human Health

Expressions for the marginal willingness to pay (WTP) for an exogenous reduction in health risks can be derived from this model. The totally differentiated expected utility function must be solved for the change in income that would be required to keep the expected utility constant when there is an exogenous change. The individual would be willing to pay the negative of this compensating variation for the exogenous improvements in health risks. Holding the expected utility constant by setting $dE(U) = 0$, we can solve for the WTP measure:[9]

$$-dM/dE = -[(U_0P_0 - U_1P_1)/m]H_E - \{1 + [(U_0P_0 - U_1P_1)/m]H_X\}dX/dE. \quad (2.13)$$

The numerator of the first term on the right-hand side is the difference in expected utility between being healthy and being ill. This is divided by

$$m = U_0'P_0(1 - H) + U_1'P_1H,$$

which is a weighted average of the expected marginal utility when healthy and the expected marginal utility when ill, with the weights being the probabilities of being healthy or ill. Thus m can be interpreted as the expected marginal utility of income.

So far, the analysis has neglected the fact that individuals choose the level of defensive expenditures so as to maximize expected utility. Rearranging the first-order condition given by equation (2.11) yields

$$(U_0P_0 - U_1P_1)/m = -1/H_X. \quad (2.14)$$

The left-hand side is familiar from the WTP expressions. As the dollar value of the difference in expected utilities between being healthy and ill, it can be interpreted as the marginal benefit of defensive expenditures that reduce the probability of illness. The right-hand side is the marginal cost of defensive expenditures.

Allowing the optimal choice of defensive expenditures as individuals adjust to the exogenous changes in health risks or the environment implies that equation (2.14) satisfies the first-order condition. Substituting the first-order condition as given by (2.14) into the WTP expression given in (2.13) results in

$$\begin{aligned} -dM/dE &= H_E/H_X + [-1 + (H_X/H_X)]dX/ \\ &= H_E/H_X. \end{aligned} \quad (2.15)$$

9. Terms involving the partial derivative of U with respect to q disappear since these terms are multiplied by dq, and $dq = 0$ since q is set at either zero or one. Similarly, recalling that the costs of illness Z are given by $Z = f(q)$, $dZ = f'(q)dq = 0$, since again $dq = 0$.

This simplification allows the WTP measure to be expressed, not in terms of the nonobservable utility function, but instead in terms of the health risk function H. In particular, equation (2.15) gives the WTP for a change in environment as a ratio of the marginal product of the environment in reducing health risks and the marginal product of preventive expenditures in reducing health risks. This result is very similar to the findings of others who suggest WTP for an environmental improvement can be expressed solely in terms of the production function (see Courant and Porter 1981; Needleman and Grossman 1983; Gerking and Stanley 1986; and Harrington and Portney 1987). One obvious difference is that, while in these models health is deterministically a function of the environment and defensive expenditures, in our model the probabilities of being healthy or ill are a function of these variables. Another difference is that our model considers mortality as well as morbidity.

Equation (2.15) is the basis for one approach to obtaining empirical estimates of willingness to pay. In principle, the health risk function $H(X, E)$ could be estimated, yielding the marginal products necessary to compute WTP. Gerking and Stanley (1986) use this strategy to estimate WTP for ozone reductions in a model with pure morbidity under certainty. (See Chap. 5 for a discussion of this and related studies.) However, Harrington and Portney (1987) and others emphasize the difficulties in correctly estimating a health or health risk production function.

The fundamental problem with the health production function approach is that it is hard to identify and measure all of the inputs that affect health. Harrington and Portney (1987) point out that typical epidemiological studies only explain a small fraction of the total variation in illness, suggesting that a number of important variables may have been omitted. Atkinson and Crocker (1992) explore the relative bias from omitted variables and measurement error when estimating health production functions. From the empirical example based on a widely used data set, they conclude that the production function estimates are likely to be the most sensitive to measurement error. Mullahy and Portney (1990) investigate an additional problem, the endogeneity of some health inputs. They use an instrumental variables approach to treat smoking as being endogenously determined, to estimate the effects of smoking and air quality on respiratory health. When estimating a health production function applicable to air pollution–induced morbidity, the health outcome is acute respiratory illness and not general health status. This could make the empirical estimation even more difficult since respiratory health is jointly produced with other aspects of health. Finally, equation (2.15) only holds as a marginal condition. Bockstael and McConnell (1983) show that it may also be very difficult to use the household production approach to estimate the value of nonmarginal changes. All of these problems indicate that the health production function approach to estimating WTP may be of limited usefulness. Below, other estimation strategies are investigated.

Equation (2.15) can be rewritten to allow for a more intuitive interpretation. Recalling that $H = H(X, E)$,

$$dH/dE = H_X(dX/dE) + H_E, \quad (2.16)$$

or, rearranging,

$$H_E = (dH/dE) - H_X(dX/dE). \quad (2.17)$$

Substituting this expression for the marginal product of the environment in reducing health risks into equation (2.15), we have

$$\begin{aligned} -dM/dE &= [(dH/dE) + H_X(dX/dE)](1/H_X) \\ &= (1/H_X)(dH/dE) - (dX/dE). \end{aligned} \quad (2.18)$$

Writing this benefit expression in terms of utility by using the left-hand side of the equation (2.15), we have

$$-dM/dE = -[(U_0P_0 - U_1P_1)/m](dH/dE) - (dX/dE). \quad (2.19)$$

This form of the benefit expression states that a person's WTP for an environmental improvement can be expressed as the sum of two terms. The first term is the dollar value of the expected difference in expected utilities between being healthy and ill multiplied by the change in health risks due to the change in the environment or other exogenous factor. The second term is the change in preventive expenditures resulting from the exogenous change.

Our model yields an expression for willingness to pay that is ex ante in nature, that is, before it is known whether the individual is sick. The value is that amount of income we have to take away from both states to keep expected utility constant. The value is defined by

$$\begin{aligned} U_0P_0(1 - H) + U_1P_1H - U(M - \hat{X} - dM/dE, 0)P_0(1 \\ - \hat{H}) - U(M - \hat{X} - Z - dM/dE, 1)P_0\hat{H} = 0, \end{aligned} \quad (2.20)$$

where the hat indicates the value of a variable after a change in E. In the context of uncertainty, our willingness to pay, $-dM/dE$, is similar to an option price (see Smith 1983) since it is a constant payment regardless of the state of nature that actually occurs. V. K. Smith (personal communication 1987) points out that, in the model described in this section, however, the framework in which individuals can purchase state contingent contracts is not fully specified, so it is difficult to restrict the payments to be constant across the states of nature. As explained earlier, the only opportunity for individuals to adjust to risk is the purchase of preventive expenditures. These features of the model mean that the willingness to pay measure,

$-dM/dE$, may not be consistent with conventional measures of option price. The measure is nevertheless a valid ex ante compensating variation for changes in risk.

Comparisons to Preventive Expenditures and Costs of Illness

It seems natural to assume that people will pay a positive amount for an environmental improvement. This means that to keep expected utility constant in the face of an exogenous improvement in the environment, an individual's income would have to be reduced; that is, $dM/dE < 0$, and positive willingness to pay is equal to $-dM/dE$. Inspection of the benefit expression given in equation (2.19) reveals that WTP could be positive if both terms, the utility value and the preventive expenditure value, are positive. Since the total derivatives, dH/dE and dX/dE, show how risk and expenditures change after optimizing behavior, however, the terms cannot be unambiguously signed. For the total derivatives, the general and plausible results and accompanying conditions are summarized in Table 2.1.

Preventive Expenditures

Consider the expenditure response of the individual to a change in the environment, dX/dE. Using the first-order condition, F, shown in equation (2.11) and the implicit function rule, it follows that

$$dX/dE = -F_EF_X = -F_E/\Delta, \quad (2.21)$$

where $\Delta < 0$ from the second-order condition given by equation (2.12). The sign of dX/dE then is the same as the sign of F_E. Differentiating F with respect to E we get

$$F_E = (U_0'P_0 - U_1'P_1)H_E - (U_0P_0 - U_1P_1)H_{EX}, \quad (2.22)$$

which cannot be signed unambiguously. The implication is that dX/dE need not be negative in that preventive expenditures could increase with an environmental improvement. Nonetheless, under plausible conditions dX/dE will be negative. If $H_{EX} > 0$, which is the case if H and E are substitutes, and if $(U_0P_0 - U_1P_1) > 0$, which is the case if expected utility when healthy exceeds the expected utility when sick, and if the difference between expected marginal utilities is small, then $F_E < 0$. If $F_E < 0$, then $dX/dE < 0$.

Change in Health Risk

The risk response to a change in the environment, dH/dE, depends in part on dX/dE, as can be seen from equation (2.16). The sign of dH/dE is negative if $dX/dE < 0$ and if H_E is larger in absolute value than H_XdX/dE; the sign of dH/dE is also negative if $dX/dE \geq 0$. In other words, the sign of dH/dE is negative except when $dX/dE < 0$ and, what seems to be unlikely, the direct effect (H_E) is less than the indirect effect (H_EdX/dE). While it is

TABLE 2.1. Comparative Statics of the Health Risk Model

	General Result	Plausible Results	Sufficient Conditions for Plausible Results
Preventive expenditures	$dX/dE \gtreqless 0$	$dX/dE < 0$	$H_{EX} > 0$ and $(U_0P_0 - U_1P_1) > 0$ and $(U_0'P_0 - U_1'P_1) \geq 0$
Morbidity risk	$dH/dE \gtreqless 0$	$dH/dE < 0$	$dX/dE < 0$ and $H_E > H_X dX/dE$ or $dX/dE \geq 0$
Willingness to pay and preventive expenditures (eq. 2.19)[a]	$-dM/dE \gtreqless dX/dE$	$-dM/dE > -dX/dE$	$dX/dE < 0$ and $dH/dE < 0$
Willingness to pay and cost of illness (eq. 2.23)	$-dM/dE \neq -Z(dH/dE)$	$-dM/dE \neq -Z(dH/dE)$[b]	Many exist
Willingness to pay and preventive expenditures—pure morbidity case (eq. 2.24)	$-dM/dE \gtreqless -dX/dE$	$-dM/dE > -dX/dE$	$dX/dE < 0$ and $dH/dE < 0$
Willingness to pay and costs of illness—pure morbidity case (eq. 2.24)	$-dM/dE \neq -Z(dH/dE)$	$-dM/dE > -Z(dH/dE)$	$dH/dE < 0$ and $dX/dE < 0$ and $U(C, 0) > U(C, 1)$ and $U(Z)\, m^{**} > Z$

[a] Willingness to pay is equal to $-dm/dE$.

[b] It is implausible that $-dM/dE = -Z(dH/dE)$. A set of sufficient conditions for this result is $dX/dE = 0$, U is not a function of q, $U(Z)\, m^* = Z$, and $P_0 = P_1 = 1$.

possible that the indirect effect can dominate even where there is evidence of counterproductive exogenous changes, alternative explanations are offered as being more plausible; for example, see Viscusi (1984*a*).

The upshot of this discussion is that, while the two terms in equation (2.19) taken together surely imply that a positive amount will be paid for an environmental improvement, it is not strictly true that the terms separately will each imply positive payments. It is the case, however, that the payments for reductions in risk and preventive expenditures will be positive under the plausible conditions that X and E are substitutes and the direct effect of E on H dominates the indirect effect through dX/dE. Under these conditions the willingness to pay for an environmental improvement is the sum of the utility value of the reduction in risk and the savings in preventive expenditures. Also under these conditions, the savings in preventive expenditures, dX/dE, is a lower bound on willingness to pay. If the conditions described above do not hold, then dX/dE is not necessarily a lower bound on WTP. Under no plausible conditions is dX/dE a special case of WTP.

COST OF ILLNESS. On the basis of the benefit expression, it is tempting to consider a value of exogenous improvement based solely on the costs of illness as special case of the general WTP measure. Indeed, there might appear to be conditions under which the expression approaches being a special case of WTP. For instance, if (1) defensive expenditures are nonexistent or unchanging, and if (2) health does not enter the utility function directly, the WTP expression shown in equation (2.19) collapses to the first term, and the difference in expected utilities between being healthy and ill only reflects the reduced level of consumption when ill due to the costs of illness incurred, Z. Even with these severe restrictions, however,

$$-Z\,dH/dE \neq \frac{U(M - X)P_0 - U(M - X - Z)P_1}{m^*}\, dH/dE, \quad (2.23)$$

where $m^* = U'[P_0(1 - H) + P_1 H]$. For Z to equal WTP, additional questionable restrictions are necessary. For example, sufficient conditions are that (3) the monetary value of the utility of consumption be equal to consumption expenditures, $Z = U(Z)/m^*$, and (4) the probability of survival be equal to one, $P_0 = P_1 = 1$ (see Table 2.1). In fact, there are no plausible assumptions that can be made to simplify the WTP measure to cost of illness. It is even less likely that WTP will equal Z^*, the more commonly used cost of illness measure that excludes the value of lost nonwork time.

MORBIDITY RISK. For the sake of brevity and because considerable attention has been given to mortality risk in previous articles, we focus on valuing changes in morbidity risks.[10] For the pure morbidity case, there is no possibility of death whether healthy or ill, so $P_0 = P_1 = 1$. The general WTP expression, equation (2.19), simplifies to

$$-dM/dE_{P_0 = P_1 = 1} = -\frac{U(M - X, 0) - U(M - X - Z, 1)}{m^{**}}\, dH/dE - dX/dE$$
$$= -\frac{U_0 - U_1}{m^{**}}\, dH/dE - dX/dE, \quad (2.24)$$

10. Although we concentrate on morbidity risk, we should note another implication of our model for the cost of illness approach. Typically cost of illness studies separately estimate the morbidity costs and the mortality costs and simply add them together (e.g., see Mushkin 1979, p. 385). From our model it is evident that willingness to pay for combined morbidity and mortality risks is not the sum of the willingness to pay for the special cases alone.

where $m^{*} = U_0'(1 + H) + U_1'H$, which is the expected marginal utility of consumption for the morbidity case.

The relationship between WTP and preventive expenditures is again, as in the case of morbidity and mortality, complex in that neither is unambiguously larger than the other. Again, however, under similar plausible conditions, dX/dE is a lower bound on WTP (see Table 2.1).

As in the case of morbidity and mortality, there is no reason to believe that WTP equals the savings in costs of illness, $-ZdH/dE$. Plausible conditions do exist, however, under which $-ZdH/dE$ is a lower bound on WTP. If $dH/dE < 0$ and $dX/dE < 0$, then WTP $> -ZdH/dE$ because ZdH/dE ignores the savings in preventive expenditures. One reason is that health enters directly in the utility function, and utility is enhanced by health; $U(C, 0) > U(C, 1)$. Another reason is that we expect the dollar value of utility lost due to losing Z dollars of consumption to costs of illness is less than Z. This relationship between the value of the utility of consumption and consumption expenditures, or labor earnings, has been explored in depth in the "value of life" literature. Conceptually it cannot be shown, strictly, what the empirical relationship should be (see Linnerooth 1979). Still, a representative theoretical conclusion is that the value of utility of consumption or earnings will "usually" exceed their dollar value; see Bergstrom (1982). Reviews by Blomquist (1981, 1982) and Chestnut and Violette (1984) of the estimates of the value of mortality risks are consistent with Bergstrom's conclusion. The implication for our case of morbidity is that $U(Z)/m^{**} > ZdH/dE$. This relationship, along with $U(C, 0) > U(C, 1)$, leads to WTP $> -ZdM/dE$. If also $dX/dE < 0$, then WTP exceeds $-ZdH/dE$ by a greater amount. So, while we cannot definitely conclude that cost of illness measures produce a lower bound for willingness to pay, the lower bound conclusion seems plausible. These results are summarized in Table 2.1.

Comparisons to Certainty Values of Morbidity

The willingness to pay expression in the pure morbidity case is shown in equation (2.24). The WTP holds expected utility constant in the face of an exogenous change in health risk. This can be compared to measures of certain changes in morbidity as follows.

Define consumer surplus (CS) as the dollar amount that holds utility constant in moving from the certainly sick to the certainly well state. For an irreplaceable commodity such as health this measure is what Cook and Graham (1977) call a "ransom." In terms of the model, CS is thus the difference between the utility in the healthy state and sick state $(U_0 - U_1)$ expressed in dollar terms by dividing by the marginal utility of income.

The expected consumer surplus associated with an exogenous change in the environment is the product of CS and the change in the probability of the certainly well state caused by the exogenous change:

$$\begin{aligned} \text{expected CS} &= -\text{CS } dH/dE. \\ &= \frac{-(U_0 - U_1)}{(\text{marginal utility of income})} dH/dE. \end{aligned} \tag{2.25}$$

Comparing equations (2.24) and (2.25), it is clear that the willingness to pay for changes in morbidity risks given by (2.24) is almost the expected value of consumer surplus, adjusted for changes in preventive expenditures. That is, equation (2.25) is almost the first term of equation (2.24). The only ambiguity in this comparison is that, in expressing the change in utility in dollar terms in equation (2.24), m^{**}, the *expected* marginal utility of income or money is used. Since m^{**} is a weighted average of marginal utilities when healthy and when ill, if we assume the marginal utilities are the same, the problem is resolved. In general, it is not clear when these two marginal utilities will be equal since differences in consumption levels and health status are involved. The relationship between the marginal utilities of income across states also depends upon the opportunities the individual has to adjust expenditures across states. For instance, with actuarially fair insurance available, the individual will equate marginal utilities across states, though this will not necessarily result in full insurance in the sense that levels of utility are equal across states (see Cook and Graham 1977). In any case, if the marginal utilities of income across states are close to each other, willingness to pay for a change in health risks is approximately equal to the expected value of consumer surplus, adjusted for changes in preventive expenditures.

Consumer surplus is what previous studies that address the pure morbidity case have measured in their valuation expressions since they have avoided the question of uncertainty. The empirical work in Parts 1 and 2 also makes use of consumer surplus. In particular, since it is difficult to appropriately incorporate uncertainty into the contingent valuation experiment, we measure consumer surpluses associated with certain changes in morbidity. However, we are able to approximate willingness to pay for risk changes by the expected value of these consumer surpluses as explained above.

2.5. Concluding Remarks

The main purpose of this chapter has been to compare preference-based willingness to pay measures for human health risk reduction with the main alternative approaches that are currently in use. After providing discussions of the various approaches, we construct an eclectic model from which we derive preference-based (WTP) values for changes in health risks, which are then compared with the alternative approaches. The model incorporates partly endogenous health, uncertainty, mortality, and morbidity. In fact, pure mortality and pure morbidity, to which previous studies have been confined, are considered as special cases of the more general framework.

In the general case, we find that the preference-based willingness to pay measure for reductions in health risks consists of two terms: a utility term, which reflects the cost of illness as well as other factors, and a term reflecting preventive expenditures. It does not follow, however, that benefit measures involving the cost of illness alone or preventive expenditures alone are special cases of our general willingness to pay measure. It is difficult or impossible to specify truly reasonable assumptions under which the willingness to pay measure collapses to a cost of illness measure or a preventive expenditures measure. Our emphasis is somewhat different from that of Harrington and Portney's (1987) in that their willingness to pay measure for a reduction in morbidity is reduced to the cost of illness measure under the assumptions that there are no preventive expenditures and that health does not enter the utility function directly.

Even the weaker result that the alternative benefit measures are lower bounds to the willingness to pay measure does not necessarily hold for our model. Without additional assumptions, we cannot establish any general comparisons between the three measures. We do find a set of plausible assumptions under which some comparisons of the alternative benefit measures can be made. First, it is necessary to assume that the environment and preventive expenditures are substitutes in reducing health risks. Second, the direct effects of a change in the environment on health risks must outweigh the indirect effects, so $H_E > (H_X)(dX/dE)$. Third, the marginal utilities of consumption when healthy and ill must be approximately the same.

If the above assumptions are made, for the special cases of pure mortality and pure morbidity, both the cost of illness and the preventive expenditures will plausibly be lower bounds to willingness to pay. The cost of illness approach understates the true willingness to pay for several reasons. First, it neglects the savings of preventive expenditures. Second, it does not allow for individuals to enjoy health directly; that is, it implies in our formulation that health q does not enter the utility function. Third, from the "value of life" literature it seems reasonable to conclude that the value of the utility of consumption will exceed consumption expenditures, so the utility lost due to expenditures lost resulting from cost of illness is greater than the cost of illness. It should be stressed that this result directly applies to the case of mortality but would seem to be plausible for morbidity as well.

Preventive expenditures also are likely to be a lower bound to willingness to pay. The preventive expenditures are not a complete measure of the benefits of health risk reduction to an individual because the individual en-

joys gains in expected utility as well as the savings of expenditures. Our model does not suggest any necessary relationship between the cost of illness and preventive expenditures measures.

One additional result is that the benefit of an exogenous change that improves both mortality and morbidity risks is not the simple sum of the benefits of morbidity risk reduction and the benefits of morbidity risk reduction.

Our results come from a model of individual maximizing behavior that considers the private costs and benefits. Thus, our results cannot be immediately generalized to social costs and benefits. However, we are able to draw some conclusions. For instance, we find in the case of pure mortality that private WTP and private cost of illness are unrelated since the latter does not matter to an individual if he dies. Only if we were to build in bequests, or impose some constraint on the amount of debts that could be left at death, would cost of illness enter the pure mortality framework. But we know costs of illness are not necessarily zero for society. So society's willingness to pay for a reduction in mortality risk may exceed the willingness to pay of the individual.

Empirical research on mortality risks has tended to confirm the prediction that benefit measures based on cost of illness will be lower bounds to benefit measures based on a willingness to pay approach. Further empirical work is needed to substantiate or refute the theoretical result that for morbidity the cost of illness will be smaller than the willingness to pay. Work along these lines is reviewed in Chapter 4. In addition, future empirical work could shed some light on the case where both mortality and morbidity risks are present. Data that contain contingent value estimates of willingness to pay, estimates of direct and indirect costs of illness, and preventive expenditures could be highly useful. These data would enable us to further investigate the questions examined in this chapter.