4

Contingent Valuation of Health

Donald Kenkel, Mark Berger, and Glenn Blomquist

4.1. Introduction

One approach to valuing a nonmarket good is to conduct a survey and ask people directly what they would pay for the good, hypothetically assuming (contingent upon) the existence of a market for the good. This approach is termed the contingent valuation method and has been applied to a variety of nonmarket goods, including health.

The purpose of this chapter is to review the applications of the contingent valuation method to the problem of valuing health, including the new survey described more completely in Part 2. Since the goal is to find useful empirical evidence on the value of health, both methodological issues and actual results are discussed. Major methodological issues arising in contingent valuation are discussed in Section 4.2. Section 4.3 reviews empirical estimates of the value of health from six contingent valuation studies. Section 4.4 compares estimates of the value of health from contingent valuation studies with estimates based on the cost of illness approach discussed in Chapter 3. Section 4.5 concludes.

4.2. Major Issues in Contingent Valuation

Contingent valuation is an established research method for valuing nonmarket goods. Since it is a fairly flexible approach providing a conceptually correct and complete measure of willingness to pay, it has been applied to a wide variety of nonmarket goods, especially in the area of environmental economics. Studies have also compared the results to indirect market methods for valuing such goods. Many methodological issues concerning the contingent valuation method have been addressed as well. For comprehensive reviews of the literature, see Cummings, Brookshire, and Schulze (1986) and Mitchell and Carson (1989). In the brief discussion that follows, the focus is on the accuracy that can be expected for values from contingent studies.

Biases and Contingent Valuation

The basic reason contingent valuation results may be inaccurate is the possibility that the responses are biased away from the unobservable true maximum willingness to pay (or accept). Types of bias often mentioned include

hypothetical bias, strategic bias, starting point bias, vehicle bias, and information bias, though these categories can overlap.

Hypothetical bias and strategic bias can be understood as a dilemma for contingent valuation. On the one hand, if respondents believe the questions to be entirely hypothetical, they have little incentive to give accurate information concerning their maximum willingness to pay. On the other hand, if they see the exercise as playing an important role in future policy making, and not hypothetical, respondents may have incentives to strategically misrepresent their values.

Other biases stem from the structure of the contingent valuation questionnaire. If a bidding process is used that begins by asking whether the respondent is willing to pay a certain amount, respondents may view this figure as appropriate and so bids would be biased toward the starting point. An alternative questionnaire structure, the dichotomous choice contingent valuation, avoids the starting point bias. In a dichotomous choice questionnaire, respondents are presented with a policy and a randomly chosen policy price and asked to respond yes or no to a close-ended value elicitation question (Hoehn and Randall 1987).

A more general problem is the vehicle by which the contingent payment is made. If it is suggested that the payment will occur through a concrete vehicle such as an increase in taxes, respondents who dislike taxes may underreport their values or protest the exercise by giving zero bids.

Finally, the values reported by respondents in a contingent valuation experiment may be sensitive to the information provided them during the questioning, and even the order of questions asked may be important.

Various studies shed light on the importance of the possible biases to which the contingent valuation method may be subject. The fundamental problem—that contingent valuation is hypothetical—has been investigated by conducting experiments that include both hypothetical payments and actual cash payments. Bishop and Heberlein (1986) conducted surveys of hunters who had received free early season goose-hunting permits. For actual cash payments, the mean willingness to sell was $63, while for hypothetical payments the mean willingness to sell was $101. Mitchell and Carson (1986) dispute this finding: in a reanalysis of Bishop and Heberlein's data they find no statistically significant difference between the hypothetical and actual values. However, Bishop and Heberlein defend their original methodology and present preliminary results from a new survey that supports the finding that hypothetical bias exists. For a discussion of this debate, see the chapter by Bishop and Heberlein and the appendix by Carson and Mitchell in Cummings, Brookshire and Schulze (1986).

Other sources of bias can also be more or less directly tested by varying the starting point, payment vehicle, or information given or by changing the incentives for strategic behavior. Results to date are somewhat inconclusive,

though Cummings, Brookshire, and Schulze (1986) tend to minimize the importance of strategic bias and starting point bias, while noting that payment vehicle and information may be more important sources of bias. No strong consensus seems to have been reached in this area, and in particular a number of researchers believe starting point bias may be quite significant. For a discussion of the various studies' results that relate to these biases, see Cummings, Brookshire, and Schulze (1986, chap. 3).

In short, existing reviews of the contingent valuation method suggest that bias problems are not insurmountable and that careful design of the survey can minimize them in many cases. This points to the need to carefully consider the design of the survey that produces any contingent valuation results. Of particular concern are the trade-offs faced in survey design. For instance, it may be possible to reduce hypothetical bias by using more concrete payment and delivery vehicles, but only at the cost of increasing the chances of strategic behavior. The trade-offs chosen in designing a particular survey need to be explicitly recognized and discussed.

Willingness to Pay versus Willingness to Accept

Both the *willingness to pay* (how much a respondent is willing to pay to acquire a nonmarket good) and *willingness to accept* (how much compensation is required by the respondent to give it up) are theoretically definable measures of a person's valuation of a nonmarket good. Practical application in contingent valuation surveys, however, has made it evident that willingness to pay responses yield more realistic results under most survey conditions used until now. Cummings, Brookshire, and Schulze (1986), citing this evidence, include the use of willingness to pay questions as one of four reference operating conditions that form the basis of reliable contingent valuation.

Theoretically, valuations should be about the same for increases or decreases in a nonmarket good about a given base, especially for small changes. Willingness to accept valuations, however, have proved to be appreciably larger in most survey applications. This effect has been studied in the literature of both psychology and economics. Brookshire and Coursey (1987) report experiments in which subjects are given repeated questionnaires in one approach and participate in repeated sealed auctions in another approach. In both approaches, it is found that subjects appear to learn, and willingness to accept bids converge to willingness to pay levels.

The studies reviewed in this chapter and the original study reported in Part 2 of this book elicit willingness to pay measures.

Accuracy of Contingent Valuation

Aside from issues of survey design and bias, the basic question remains, however: in a properly designed contingent valuation study, how accurate

are the values reported? In a sense, the question is unanswerable since the true values are unobservable. Several types of evidence can suggest a range of accuracy.

First, as Tolley and Fabian (1988) point out, studies have found that contingent values are systematically related to income, availability of substitute goods, and other variables that economic theory suggests should be important. This implies that the contingent market is to some extent similar to an actual market and that the values reported are not random but are reasonable subjects for economic analysis.

Second, a number of studies have compared the contingent valuation method to alternative indirect market methods of valuing nonmarket goods. Cummings, Brookshire, and Schulze (1986) review these studies and stress that the results can not establish the accuracy of contingent valuation. But, "Assuming that, within the range of plus or minus 50%, value estimates derived from indirect market methods include 'true' valuations by individuals, these results suggest that CVM [contingent valuation method] values may yield 'accurate' estimates of value in cases where individuals have had some opportunity to make actual previous choices over that commodity in a market framework" (Cummings, Brookshire, and Schulze, p. 102).

Based on their comprehensive review of the methodology and practice of contingent valuation, Cummings et al. (1986) suggest a range of accuracy for carefully designed contingent valuation studies. (These suggestions are linked to a set of reference operating conditions that the study must meet for the accuracy range to apply.) At the least, "the method produces *order of magnitude* estimates—but we think one can argue that error ranges are much smaller" (Cummings, Brookshire, and Schulze, p. 233). At the best, "one might tentatively conclude that, given the current state of the art, the CVM is not likely to be more accurate than plus or minus 50 percent of the measured value" (p. 99). This plus-or-minus 50% range is a suggested reference accuracy, and though it is a somewhat arbitrary figure, it seems reasonable.

4.3. Applications of Contingent Valuation to Health

Scope of the Review

We turn to a critique of studies that use the contingent valuation method to value health symptoms related to air pollution. This section first briefly refers to a number of contingent valuation studies that have served to develop the method and demonstrate its ability to obtain estimates of willingness to pay, but they do not yield values directly comparable to those sought in this study, mainly because of the definition of the good being valued. The section then goes on to consider studies more narrowly focused on health

values. The first three studies mainly concern acute symptoms. The original motivation for these studies was to value symptoms linked to air pollution, but the symptoms valued are also of much more general interest. Loehman et al. (1979) used a mail questionnaire of the general public, while Dickie et al. and the new study reported in Part 2 used personal (telephone or household) surveys of the general public. The second set of studies concerns more serious, chronic illness. The earliest is Rowe and Chestnut (1984), which provides estimates of the value of a reduction in asthma symptoms, using personal interviews of a group of individuals suffering asthma. Viscusi, Magat, and Huber (1991) and Krupnick and Cropper (1992) both estimate the value of reductions in the risk of chronic bronchitis. Viscusi, Magat, and Huber use a sample from the general population. Krupnick and Cropper apply the same survey to a sample of people who are presumably more familiar with the "good" sold in the contingent market because they have a relative with chronic lung disease.

At the outset, the limited scope of this section should be explained. In line with the overall purpose of Part 1 of this book, the focus is on empirical estimates of the value of health. As a result, no attempt is made to report and review all of the findings of the studies in question. In particular, for our purposes the values of health are best summarized by simple statistics such as the median and mean values for the sample. Other statistical analyses, including the estimation of bid functions based on the contingent valuation responses, are not reviewed, though they are important parts of these studies. In addition, questions of methodology and survey design are only addressed in the context of evaluating the usefulness and accuracy of the value estimates produced.

Several important studies that use the contingent valuation method to value changes in air quality, including the health effects, are quite comprehensive: in two of them (Brookshire et al. 1979, Loehman, Boldt, and Chaiken 1981) respondents were asked separately about their values for the visibility and health effects of air pollution. The sum of these values may be the most meaningful estimate, and we have not attempted to use the health values because of a concern over the ability of respondents to disentangle the two values. The values of health alone may be overstated, reflecting part of the value of visibility, or understated if part of the value of health is included in the reported value of visibility.

A third study by Schulze et al. (1983) concentrates on health effects of ozone. Respondents were provided with descriptions of the health effects likely to result from air pollution levels and then asked for their values for a change in pollution levels. The descriptions are of the general form: for a given level of pollution, some people (or a certain percentage of people) experience these effects. Respondents might identify the general population risk as their own risk. So if they are told that 50% of people will experience

a symptom, they may view this as a 50% chance they will experience the symptom. Another interpretation is that the information provided helps remind the respondents of their experiences with air pollution. In this case, respondents will bid for a change based on their prior subjective probability estimates of experiencing a symptom given varying levels of pollution. Or, they may adjust their prior beliefs on the basis of the information given. In either case, the commodity the respondents are valuing is a change in risks (probabilities of symptoms) that is not strictly observable to the researcher, and for the purposes of this book these changes are not useful in attempting to establish individual health values.

Shechter et al. (1988) and Shechter and Kim (1991) employ a comparative approach to the valuation of health damages caused by air pollution in an urban environment. Both studies use data from the same sample of individuals to derive empirical results by means of different analytic methods. Contingent valuation is employed in these studies to give direct estimates of the health value of environmental improvement. Indirect estimates are obtained by deriving an expenditure function or a household production function in which the demand for housing services or medical services leads to a measure of the benefits of cleaner air. In these and other studies, Shechter and his colleagues introduce stress, anxiety, and personality variables in the theoretical and empirical investigations of the role of psychological factors in the demand for environmental quality (Zeidner and Shechter 1988; Shechter and Zeidner 1990).

The analysis of these studies is highly complementary to the approach of the present volume, although the empirical results are not directly comparable. The reason is that the dependent variable employed by Shechter and his colleagues is air quality rather than health status. The health status variables used as regressors (e.g., asthma, bronchitis) are grouped together, so it is not possible to infer the health value of a specific condition.

The studies reviewed up to this point help establish contingent valuation as a useful approach to valuing health and environmental quality but do not yield estimates of the value of health directly comparable to the approach of this book. We now turn to studies that provide such estimates.

Loehman et al. (1979)

STUDY DESIGN. The study by Loehman et al. (1979) concerns the benefits of controlling sulfur oxides in Florida. A mail contingent valuation survey was sent to 1,977 residents in the Tampa Bay area, resulting in 432 returns. Willingness to pay questions were asked about the following three groups of symptoms: shortness of breath/chest pains; coughing/sneezing; head congestion/eye/ear/throat irritations. Values were elicited for minor and severe symptom days, which were defined briefly. Respondents were asked to value 1 day, 7 days, and 90 days of relief. No mention was made of any

specific underlying disease, nor were causes such as air pollution mentioned. No specific delivery vehicle, such as a pill, was employed, and a simple, abstract payment vehicle—"tell us how much you would pay"—was chosen. The means of payment was a checklist, or payment card, ranging from $0 to $1,000 per year in 10 increments.

The Loehman et al. study design is similar to our seven-symptom survey described in Part 2. In both cases a pure health attribute approach was used. The Loehman et al. study carefully avoided the introduction of redundant information in its introductory letter, its symptom narrative, and in its delivery and payment vehicles. One difference between the design of our survey and the Loehman et al. survey is the large number (24) of similar willingness to pay questions of the latter survey. Our approach was to employ fewer questions on any survey instrument in order to avoid taxing the respondents' concentration and the extent of their information and preference review, a problem that might account for the relatively low return rate (22%) encountered by Loehman et al. It also could imply a reduction in the accuracy of their estimates of the value of health.

The major difference between Loehman et al. and the other contingent valuation studies reviewed below is that the Loehman et al. study used a mail questionnaire. The advantage to using this approach is that the lower cost per survey completed allows a larger sample size. There are several disadvantages. An obvious question is whether the respondents are representative of the general population. Loehman et al. test for this and find that the sample seems to be more or less representative, at least in terms of standard demographic characteristics.

Another problem with using a mail survey is that in a contingent valuation experiment there will be some protestors, or people who either refuse to participate in the contingent market or do not understand the nature of the exercise. In a personal interview, follow-up questions and interviewer comments can help identify respondents who are protestors. A mail questionnaire gives no indication of the identity of protestors, except for the bids themselves. Loehman et al. note the presence of bids from respondents who gave values of $1,000 (the highest amount on the payment card). These bids were statistically outlyers, and the respondents exhibited intransitivity of preferences. It seems reasonable that these respondents were protestors. However, it is also possible that these individuals simply had high values for health. The limited information from a mail questionnaire means this problem is difficult to resolve.

A final disadvantage of using a mail questionnaire is that a payment card is often used. Such a card lists the possible amounts people might be willing to pay, and the respondents choose among the different amounts. Designing a card that covers a wide range of low to high values and allows small but important differences between values to be reported is difficult. In addition, some have questioned whether such a card elicits *maximum* willingness to pay responses. Cummings, Brookshire, and Schulze (1986) suggest that, if a payment card is used, it should be followed with iterative bidding, but this is not feasible in the context of a mail questionnaire. These problems indicate that the values from the Loehman et al. study may be inaccurate, and in particular they may be underestimates of the maximum willingness to pay for health.

RESULTS. Table 4.1 lists the median and mean bids found by Loehman et al. All bids are expressed in terms of 1984 dollars, to insure comparability

TABLE 4.1. Contingent Values of Health from Loehman et al. (1979) (in $)

Symptom	Median Bid	Mean Bid
1 day of:		
Shortness of breath/chest pains:		
Mild	8	78
Severe	18	127
Coughing/sneezing:		
Mild	4	42
Severe	11	73
Head congestion, eye, ear, throat irritation:		
Mild	6	52
Severe	13	85
7 days of:		
Shortness of breath/chest pains:		
Mild	22	118
Severe	57	218
Coughing/sneezing:		
Mild	13	71
Severe	32	116
Head congestion, eye, ear, throat irritation:		
Mild	15	66
Severe	33	129
90 days of:		
Shortness of breath/chest pains:		
Mild	56	233
Severe	156	403
Coughing/sneezing:		
Mild	37	138
Severe	81	236
Head congestion, eye, ear, throat irritation:		
Mild	40	145
Severe	99	288

with other estimates of the value of health discussed in this book. The bids were adjusted using the consumer price index and were rounded to the nearest dollar.

The bids cover a fairly wide range. For one day of relief, the lowest median bid is $4 for mild coughing/sneezing, and the highest median bid is $18 for severe shortness of breath/chest pains. However, the mean bids for 1 day of symptoms are often an order of magnitude larger, ranging from $42 for mild coughing/sneezing to $127 for severe shortness of breath/chest pains. There is generally a smaller difference between median and mean bids for 7 days of relief and 90 days of relief.

The large difference between median and mean bids results from properties of the distribution of bids. As Loehman et al. describe it, the distribution is clearly not normal but includes a large number of relatively low bids, with a few bids in the upper tail of the distribution. These bids were for $1,000, the highest bid possible, and represent the possible protestors discussed above. The mean bids are much more sensitive to these outlyers than are the medians, and so the means are much larger than the medians.

In their analysis, Loehman et al. use only the median bids. One justification for this use is normative. They argue that the median is "indicative of majority voting since it indicates the bid which at least 50 percent of the population would agree to pay" (Loehman et al. 1979, p. 232). Though this majority voting criterion is certainly reasonable, it represents an alternative to the standard methodology of applied welfare economics, where programs are evaluated using the criterion of a potential Pareto improvement. Using this criterion, all individuals' values are given equal weight, including the very high values. It is possible that a program that represents a potential Pareto improvement would not be favored by over 50% of the population. Potentially, though, payments by gainers could compensate the losers by enough that all would favor (or at worst be indifferent to) the program. If this standard of applied welfare economics is accepted, the correct summary statistic is the mean, which puts equal weights on all, and not the median.

Loehman et al. also justify their use of median bids by noting that the median is less likely to be biased due to the outliers. V. K. Smith (personal communication 1987) explains how this problem could justify use of median bids even if the potential Pareto improvement criterion is accepted as relevant. If a distribution of individuals' true values of health in a population is known, the mean value is the correct summary statistic as explained above. Applying this reasoning to a distribution of values resulting from a contingent valuation experiment is not necessarily correct. To do so requires the assumption that all contingent valuation responses are judged as equally good estimates of each individual's willingness to pay. Arguments that have been made in the contingent valuation literature for the use of the median implicitly assume that not all responses to contingent valuation questions are equally good estimates of each individual's willingness to pay. In particular, there is a presumption that very large or very small responses are more likely to have large errors associated with them. Since the mean value is more affected than the median, the mean would be a less robust estimate of the "average person's" willingness to pay. In this case, if outliers are a problem, the median bid may be preferred.

Accepting the criterion of a potential Pareto improvement as the relevant welfare guideline, the choice of using median or mean values from a contingent valuation study depends upon the informational content assumed for different responses. Reporting median bids avoids overstating values due to the effect of very high bids, which may be inaccurate in the sense that they are not a true reflection of willingness to pay. At the same time, legitimately high bids are also given little weight. In addition, though the very high bids may be inaccurate, they probably do indicate that these individuals are actually willing to pay an amount higher than average. Finally, the argument is symmetric with respect to low bids. While very low bids probably do indicate that these individuals have lower than average willingness to pay values, the true values may not be as low as the values reported in the contingent valuation experiment.

To rigorously account for all of the considerations discussed above requires a model of how people respond to contingent valuation questions. In chapter 15 there are the beginnings of such a model, but it does not allow any definite conclusions to be made regarding the mean versus median question. In practice, both mean and median values are important pieces of evidence. Inferences of the informational content of very high and very low bids can be drawn from careful consideration of the study design and the distribution of bids found. For the Loehman et al. results, the problems inherent in a mail survey and the distribution of bids suggest that the high bids are not accurate reflections of willingness to pay. Thus the median may be a more robust summary statistic.

It is interesting to note the relationships between the bids for 1 day, 7 days, and 90 days of relief found by Loehman et al. Using mild coughing/sneezing as an example, the bid for 1 day is $4, while the bid for 7 days is $13, roughly three times as large. The bid for 90 days is $37, about nine times as large as the bid for 1 day. Roughly similar results are found for other median bids. For mean bids the ratios are even smaller; the bid for 7 days of relief from mild coughing/sneezing ($71) is less than twice the bid for 1 day ($42), and the bid for 90 days ($138) is only about three times the 1 day bid.

Two explanations for these relationships are possible. The marginal disutility from sickness (symptoms) could be diminishing rapidly, so that extra days of symptoms do not matter much and the individual is willing to pay increasingly less for relief from the symptoms. This does not seem plausible, especially since decreasing marginal disutility from sickness implies increasing marginal utility from health, which is not consistent with the assumptions of economic theory. A second possibility is that the respondents had trouble valuing large changes in health because these changes were outside of their experiences. That bids for unfamiliar commodities may be inaccurate has been suggested by users of the contingent valuation method (see Cummings, Brookshire, and Schulze 1986). This explanation seems to be more powerful in explaining why bids for 90 days of relief (an unfamiliar commodity to most people) are so small compared to the bids for 1 day of relief (a more familiar commodity within the range of most people's experiences). It is less powerful in explaining the ratio of bids for 1 day and 7 days of relief since both are probably familiar experiences to most people.

Dickie, et al. (1987)

STUDY DESIGN. Dickie et al. (1987) conducted a telephone survey of 221 residents of Glendora and Burbank, California. The most important contingent valuation results are for nine symptoms related to ozone pollution: sinus pain, cough, throat irritation, tight chest, could not breathe deep, pain on deep breath, out of breath easily, wheezing/whistling breath, and headache. The careful survey design clearly encouraged and guided respondents to think carefully about the health symptoms to be valued. Following a set of questions on standard socioeconomic measures, respondents were asked about their experience with a list of symptoms. Focusing on their most bothersome symptoms, respondents were asked about the frequency, duration, and severity of symptoms, as well as averting actions taken in response to the symptoms.

The next step was to determine their willingness to pay for 1 day of relief. Respondents were asked to value up to three symptoms. Only respondents who experienced a symptom were asked to value it. One hundred sixty-five respondents reported having had at least one symptom and so answered one or more contingent valuation questions. Respondents who had experienced more than three of the symptoms were asked about the three most bothersome. The number of respondents providing willingness to pay bids for the nine ozone symptoms varied from 11 for wheezing/whistling breath to 61 for headache.

Dickie et al. consider in detail the reliability of the contingent valuation method, and this is reflected in both the survey design and in the analysis of the results. As part of the survey, respondents were given an opportunity to revise their bids, after being presented information about their implied

total monthly bid for avoiding symptoms. There was particular concern about unrealistically high bids. Dickie et al. note that half of the symptoms distributions include at least one value of $5,000 or more to eliminate 1 day of symptoms. In two cases the implied monthly bids initially totaled $899,910, which exceeded not only the respondents' monthly incomes but even their annual incomes by substantial margins. Given the opportunity to revise their bids, the highest monthly total was in the range $501–$600.

While there is an obvious concern about the reliability of the initial bids, Cropper and Freeman (1991, p. 203) point out that the procedure used by Dickie et al. is also problematic. If average willingness to pay is less than marginal willingness to pay, the total monthly bid as calculated in the Dickie et al. survey exceeds total willingness to pay. Presenting information about the total monthly bid to the respondents may then be misleading and may cause the revised bids to be too low (i.e., to be below true willingness to pay).

RESULTS. Table 4.2 presents the main results of the Dickie et al. contingent valuation survey. The authors' attention to reliability issues is again reflected in the detailed presentation of results. The first part of Table 4.2 presents the median and mean initial bids, and the median and mean bids after respondents were given the chance to revise their bidding. The second part of Table 4.2 presents mean bids after the responses were subject to trimming and consistency checks, as described below. The range of values reported in Table 4.2 is extremely wide, making it imperative to develop some judgments as to which values are the most reliable.

As in other contingent valuation studies, a small number of bids in the upper tail of the distribution of bids have a pronounced influence of the sample means reported in Table 4.2. To explore this problem further, Dickie et al. subject the contingent valuation bids to trimming and consistency checks. In trimming, either 5% or 10% of bids from each tail of the distribution are arbitrarily eliminated. In the consistency checks, bids were excluded for a number of different reasons. First, bids were excluded if the implied total monthly bid to eliminate a symptom exceeded monthly household income. Additional checks excluded bids if the bid for symptom relief was inconsistent with other information provided in the survey. An example of an inconsistency is if the respondent bid a large amount (defined as over $100) for a symptom that he or she judged to be of low severity and was a symptom for which he or she took no averting action. Zero bids were similarly excluded if other information suggested the symptom was important to the respondent.

All three methods used to improve reliability—allowing respondents to revise bids, trimming the samples, and subjecting the bids to consistency checks—result in substantially lower means, as extremely high bids are removed from the sample. Using the initial bids, the mean value placed on a

TABLE 4.2. Contingent Values of Health from Dickie et al. (1987) (in $)

A. Initial and Revised Values

Symptom	Initial Value		Revised Value	
	Median	Mean	Median	Mean
Sinus pain or discomfort	3.50	239.50	1.00	4.53
Cough	1.00	355.10	.00	1.61
Throat irritation	3.00	15.00	1.00	3.74
Chest tightness	5.00	813.72	.80	3.55
Could not breathe deep	1.00	1,139.58	.00	2.78
Pain on deep breath	3.50	954.13	.00	4.67
Out of breath easily	.00	7.88	.00	1.94
Wheezing/whistling breath	2.00	54.36	.00	2.53
Headache	1.00	178.39	1.00	3.24

B. Trimming and Consistency Checks

Symptom	Trimmed Means (in $)		Consistency Check Means (in $)
	5%	10%	
Sinus pain or discomfort	10.98	9.42	13.15
Cough	9.96	8.65	11.46
Throat irritation	12.08	7.00	16.60
Chest tightness	449.74	18.11	19.60
Could not breathe deep	567.74	213.20	14.74
Pain on deep breath	376.21	22.25	26.70
Out of breath easily	5.43	2.16	6.91
Wheezing/whistling breath	15.78	15.78	11.50
Headache	12.42	8.33	18.80

day of relief is often above $100 and, in one case, is above $1,000. While there are still a few cases of means in excess of $100, after trimming most of the means are below $20. The highest mean value placed on a day of relief in the sample subjected to consistency checks is $26.70. The highest mean value in the sample of revised bids is $4.67.

The results just reviewed document the influence a few high bids have on the sample means for the value of symptoms. The substantial differences between medians and means also stem from this feature of the distribution of bids. The question as to whether the high bids are reliable indicators of individual willingness to pay is harder to resolve. The trimming procedure is obviously arbitrary, although Dickie et al. begin to develop a model of symptom value formation that could justify some procedure along these lines. The problem noted by Cropper and Freeman (1991)

applies to both the consistency checks and the bid revision process. A person could legitimately report a high marginal value on a day of relief but be unwilling or unable to pay 30 times that for a month of relief. Both of the other contingent valuation surveys of light symptoms reviewed in this section provide some evidence in support of this conjecture, although it is not clear why this relationship between marginal and average valuations should exist.

It should also be noted that the sample includes a number of unrealistically low values. Dickie et al. report that, for six of the nine symptoms, the modal bid for relief was zero. In the sample of revised bids, in five cases the median bid was zero, implying that half of the sample bid zero for relief. If respondents fully understand the nature of the contingent valuation survey, zero bids literally mean these respondents place no value on symptom relief. Any evidence that the symptoms caused discomfort or averting actions suggest a nonzero value. More realistically, zero bids are probably best interpreted as representing very low values that respondents have approximated as zero. Since replacing zeros with very low bids would not dramatically change the means reported, this problem is of less practical importance.

Given the problems noted, it is probably inappropriate to look for very precise estimates of the value of symptom relief from the Dickie et al. results reported in Table 4.2. The trimmed samples, the samples subjected to consistency checks, and the revised bids provide useful evidence on the lower bound to placed on the value of symptom relief: in the range from $2.00 to $5.00 a day. However, each of the methods used to improve reliability also systematically removed high bids from the sample. While the actual bids reported may be unrealistically high, the respondents making these bids probably placed a higher than average value on symptom relief. The procedures used by Dickie et al. therefore tend to create a downward bias in the estimates of the average value of symptom relief.

Original Contingent Valuation Study

STUDY DESIGN. Part 2 of this book contains a detailed description of the design of our original contingent valuation experiment and the considerations involved in this design. The experiment consists of four surveys valuing (1) 1 day of relief from seven light symptoms, such as coughing, and so on, (2) 30 days of relief from these same seven symptoms, (3) relief from mild and severe angina (chest pain) given that the respondent already suffered from 10 days of this symptom, and (4) relief from mild and severe angina given that the respondent already suffered from 20 days of this symptom. Separate surveys were used to keep the length of the survey at a level where reasoned responses could be reached, but respondents' patience and concentration were not overtaxed. A total of 199 interviews were completed,

roughly equally divided among the four types of surveys. The surveys were personal interviews of a randomly selected sample from Chicago and Denver.

Of the total of 199 completed surveys, 23 surveys were removed from the sample. Several criteria were used to determine which responses to remove. First, protestors who refused to give any bids were removed from the sample. Protestors are distinguished from those who wished to bid zero. Zero bidders were left in the sample on the grounds that the bids were felt to be legitimate. A second group excluded from the sample were those respondents who indicated that they would pay any amount for the improvement in health or exorbitantly high amounts (two or three times their yearly income). The last group of respondents removed from the sample were random bidders whose bids bore no logical relationship to each other. Interviewer comments were used in all cases to help identify individuals unwilling or unable to participate in the contingent market.

A great deal of care was taken in the creation of the contingent market. The contingent commodities were described to the respondents, and the structure of the survey encouraged respondents to think about the commodities before bidding began. A form of iterative bidding was used. Abstract payment vehicles and delivery vehicles were chosen to avoid protests and to avoid distracting respondents from giving reasoned values. Finally, interviewer comments and analysis of the bids were used to identify protestors.

For the two surveys concerning the seven light symptoms, the structure of the survey instrument first helps the respondent to recall his own experience with these common symptoms and then establishes a standardized hypothetical product (relief from symptoms) to be valued. As a result, the respondent should be familiar with the commodity of the contingent market, an important prerequisite to obtaining accurate value estimates.

The procedure described above could not be exactly followed for the two surveys concerning angina since most respondents had little or no experience with this symptom. Standard questions on health status help the respondent to begin to think about his or her health and its importance. The contingent valuation section begins with a general two-paragraph introduction that asks the respondent to imagine having mild or severe angina and includes a brief statement about the extent of angina in the United States. The actual contingent valuation includes a description by the interviewer of the specific symptoms to be valued, and a card summarizing of this description is then handed to the respondent. This approach to survey structure was used to minimize the problems associated with respondents being unfamiliar with angina. While the value estimates resulting may not be as accurate as for the more familiar seven symptoms, it is felt that most respondents did give reasoned bids.

RESULTS. Table 4.3 presents the values for symptoms from the four surveys. Part A. of Table 4.3 presents median and mean bids for relief from 1 additional day of seven individual light symptoms and two combinations of symptoms. Part B. of Table 4.3 presents the same statistics for relief from 30 additional days of the same individual and combined symptoms. Parts C. and D. of Table 4.3 present bids for relief from angina. The number of additional days of angina, the severity of the angina, and the endowment that respondents were asked to assume described their situation are varied to provide a range of values.

The median bids for relief from 1 additional day of the seven light symptoms range from $11 for relief from a day of coughing to $20 for headaches. Mean bids are roughly two to three times larger, ranging from $25.20 for a coughing day to $50.28 for relief from a day of nausea. Relief from combinations of three symptoms is more highly valued than relief from one symptom alone but is not the simple sum of the values of the individual symptoms. For instance, a day of cough, throat, and sinus symptoms combined is valued at $65.60 The sum of the bids for relief from these symptoms individually is $89.22.

The difference between the median bids and the mean bids is substantially less than that found for the Loehman et al. (1979) results. As described above, the excessively large bids resulting from respondents who explicitly or implicitly protested the contingent market were removed from our sample. This shows one advantage of the personal interview structure compared to mail surveys: interviewer comments can help identify protestors. Since all responses were subject to the editing process, and the distribution of bids shows the smaller impact of the largest bids, the mean seems to be the most robust summary statistic for this sample. In other words, the assumption seems justified that all responses, even the very large and very small bids, have roughly equivalent informational content.

For relief from 30 days of the seven light symptoms, the median bids range from $95 for 30 days of coughing to $135 for 30 days of sinus problems. Again, mean bids are usually about two or three times larger than the medians, ranging from $166.50 for 30 days of coughing to $488.20 for 30 days of headaches. The same relationship between the bids for combinations of symptoms and the sum of the bids for relief from the individual symptoms is found as in the 1-day survey. A combination of three symptoms is valued more than any one symptom alone, but not as much as the sum of the bids for the three individual symptoms.

Just as in the Loehman et al. (1979) results, a somewhat surprising relationship is found between the bids for different days of relief. The mean bids for 30 days of relief from the light symptoms are not 30 times larger than the mean bids for 1 day of relief. The 30-day bids are closer to 10 times the size of the 1-day bids. Though these bids result from two different

TABLE 4.3. Contingent Values of Health from Original Study (in $)

A. Survey 1

1 Additional Day of Symptom	Median Bid	Mean Bid
Coughing	11	25.20
Sinus congestion	14	35.05
Throat irritation	13	28.97
Eye irritation	12.50	27.73
Drowsiness	15	31.49
Headaches	20	40.10
Nausea	17.50	50.28
Cough, throat, and sinus congestion	30.50	65.60
Drowsiness, headaches and nausea	25	95.08

B. Survey 2

30 Additional Days of Symptom	Median Bid	Mean Bid
Coughing	35	146.50
Sinus congestion	135	255.62
Throat irritation	100	206.26
Eye irritation	100	235.53
[illegible]	150	317.98
[illegible]	112.50	259.20
[illegible]	60	196.42
[illegible]	200	421.30
[illegible]	200	499.58

C. Survey 3

Relief from Angina, Given Days of Endowment of Angina	Median Bid	Mean Bid
1 mild day:		
Given 1 mild day	53	66.08
Given 10 mild days	50	83.95
1 severe day:		
Given 1 severe day	100	123.59
Given 10 severe days	100	144.74
5 mild days:		
Given 10 mild days	55	96.18
5 severe days:		
Given 10 severe days	150	192.90
10 mild days:		
Given 10 mild days	100	154.36
10 severe days:		
Given 10 severe days	200	261.84

TABLE 4.3. *(continued)*

D. Survey 4

Relief from Angina, Given Days of Endowment of Angina	Median Bid	Mean Bid
1 mild day:		
Given 1 mild day	53	90.24
Given 20 mild days	40	99.05
1 severe day:		
Given 1 severe day	75	278.88
Given 20 severe days	60	208.78
10 mild days:		
Given 20 mild days	100	287.63
10 severe days:		
Given 20 severe days	125	506.25
20 mild days:		
Given 20 mild days	100	486.25
20 severe days:		
Given 20 severe days	200	844.38

samples of individuals, in terms of observable characteristics, the samples seemed similar. Another possible explanation is that the results reflect diminishing marginal disutility from sickness, but this explanation implies increasing marginal utility from health, which seems implausible. In addition, other results from these surveys support the more standard relationship of increasing marginal disutility from sickness. Finally, it could be argued that 30 days of sickness are a more unfamiliar commodity to most individuals, so they are undervaluing it. This possibility points to the continued need for a formal model of how respondents react to contingent valuation questions since it is not obvious why bids for an unfamiliar commodity would be systematically biased downward.

The third survey concerns the value of relief from angina (chest pain), given an endowment of up to 10 days of severe angina. Median bids range from $50 for relief from 1 mild day given an endowment of 10 mild days, to $200 for relief from 10 severe days given an endowment of 10 severe days. The mean bids are fairly close to the median bids, ranging from $66.08 for relief from 1 mild day given an endowment of 1 mild day to $261.84 for 10 severe days given an endowment of 10 severe days. For comparable endowments, median and mean bids for mild days are always less than bids for severe days, as would be expected. Comparing across endowments, it is generally true that relief from a given number of days of angina is valued more highly as the endowment increases. This is consistent with increasing marginal disutility of illness and is the expected relationship.

The fourth survey also concerns angina, but the endowment ranges up

to 20 days of mild and severe angina. Median bids range from $40 for relief from 1 mild day given an endowment of 20 mild days to $200 for relief from 20 severe days given an endowment of 20 severe days. Mean bids show a larger difference between the value of 1 day and 20 days of angina. The mean bid for relief from 1 mild day given an endowment of 1 mild day is $90.24, while the mean bid for 20 severe days given an endowment of 20 severe days is $844.38. Again, as expected, relief from severe days of angina are valued more highly than relief from mild days. However, comparing bids across endowments, the results do not always support that increasing the endowment increases the bid for a given number of days of relief. For example, the mean bid for relief from 1 severe day given an endowment of 1 severe day is $278.88, while the mean bid for relief from 1 severe day given an endowment of 20 severe days is only 208.78 This difference may not be highly significant. Closer examination of the bids reveals that some respondents bid a large amount to be completely free of angina while placing a small value on a day at the margin given a large endowment. Though this behavior is not consistent with increasing marginal disutility of illness, it is not necessarily irrational. Whether individuals with actual experience of angina would bid in this way is an interesting and open question.

It is possible to compare the results of the two surveys on angina in a few cases where identical commodities were valued by the different samples of individuals. The mean bid for relief from 1 mild day given an endowment of 1 mild day is $66.08 for survey 3 and somewhat larger for survey 4 at $90.24. A larger difference is found for the only other case in which the surveys are directly comparable. In survey 3, the mean bid for relief from 1 severe day given an endowment of 1 severe day is $123.59, while in survey 4 the mean bid is $278.88. This larger mean bid in survey 4 reflects the influence of a few very high bidders who bid a large amount to be completely free of angina. In fact, the median bid from survey 4 for relief from 1 severe day of angina given an endowment of 1 severe day ($75) is less than the median bid from survey 3 ($100). These results show the effect a few bids can have on the summary statistics and suggest that the values reported for relief from angina may not be highly accurate.

Rowe and Chestnut (1984)

STUDY DESIGN. The study by Rowe and Chestnut (1984) provides estimates of the value of a reduction in asthma days for people with asthma. The economic research supplemented research under way at the University of California, Los Angeles, School of Medicine concerning the effects of air pollution on asthmatics. The UCLA project included over 90 subjects from Glendora, California (in 1983); the general questionnaire that included the contingent valuation questions was completed by 64 adults and 18 parents of children under 16 years of age. Of this total sample of 82, there was only one refusal. After evaluation of the bids, including checking for protestors and other respondents whose bids were judged to be inaccurate on the bias of consistency checks, 65 bids were retained. The fact that asthmatics were sampled instead of the general population is arguably a strength, not a weakness, since people with asthma are a group likely to be affected by pollution who may value the change differently than the general population. Unfortunately, the sample was not chosen so as to be representative of asthmatics in general.

Contingent valuation bids were obtained by asking the respondents, "If federal, state, or local governments set up programs that could reduce pollens, dusts, air pollutants, and other factors throughout this area that might reduce your (and your household's) bad asthma days by half, but would cost you increased tax dollars, what would be the *maximum increase* in taxes each year that you and your household would be willing to pay and still support such a program?" A number of aspects of this contingent market deserve comment. First, the good or commodity being bid on is a reduction by half of the respondent's and his household's bad asthma days. Given the respondent's experience with asthma and the earlier questions in the questionnaire, it seems reasonable that the respondents understood the commodity and by this point in the experiment had prior valuation and choice experience with respect to consumption levels of it. The major drawback of this definition of the commodity is that it is different for each respondent. What constitutes a "bad asthma day" is subjective, and since the number of bad days varies across respondents, so does the number of bad days removed implied by the 50% reduction.

Second, it was made clear that the reduction in asthma days would be the result of a governmental program and paid for by an increase in taxes. That is, relatively concrete vehicles for the delivery of and payment for the good are used. Though this makes the contingent market more realistic, the added realism is purchased at the cost of increasing the possibility of problems such as strategic bias or protestors (either at the idea of increased taxes or the impossibility of such a program). In addition, experience in focus groups in Chicago showed that mentioning the environment as a cause of health condition seemed to distract the respondents from providing reasoned bids. This problem may not have existed for the asthma patients, however, since other results of the project showed that they had a good understanding and accurate perceptions of the effects of pollution on their conditions.

Third, an element of uncertainty is introduced into the market since it is stated that the program improving air quality "might" reduce bad days by half. This wording raises difficulties in interpreting the bids. Is one respondent bidding a small amount because the reduction in asthma days is not worth much to him or because his subjective probability that the program

will work is relatively low? The extensive analysis of the bids supports the former interpretation, but the issue can not be entirely resolved.

Two more general problems of the structure of the contingent market should be mentioned. First, there is the problem of the bidding format. The Rowe and Chestnut study used a payment card format. It was designed to eliminate some of the problems associated with this format; they note that problems may remain.

The second problem is the treatment of protest bids and extreme values (either zero bids or very large bids). The ideal is to retain all bids that reflect the true value, no matter how extreme, and to remove bids that do not. To be a useful bid, the respondent must be willing to participate in the contingent market and fully understand the nature of the exercise. Rowe and Chestnut carefully examine the zero bids and subject bids to a consistency check. This process necessarily involves some rather ad hoc procedures and is to a certain extent subjective. It would be interesting to know how sensitive the bid results are to the editing process. As mentioned earlier, this process results in 17 of 82 bids being rejected, or roughly 20%.

RESULTS. The results of the Rowe and Chestnut study relevant for this review can be very easily summarized. They found a mean bid for a 50% reduction in bad asthma days (for 65 observations) of $401 per year, with a standard deviation of $85. This is for an average number of bad days reduced equal to 19. Thus, on average a bad asthma day is worth about $21. Of course, this average value cannot in general be used to value a marginal change of 1 bad asthma day.

Viscusi, Magat, and Huber (1991)

STUDY DESIGN. Viscusi, Magat, and Huber (1991) extend contingent valuation methodology to measure willingness to pay for reductions in the risk of developing chronic bronchitis. Chronic bronchitis risk reduction differs substantially from the health improvements considered by the other contingent valuation studies reviewed in this chapter. It is a lifelong illness with much more serious respiratory symptoms than those considered by the light symptom contingent valuation studies (Loehman et al. 1979; Dickie et al. 1987; and the original contingent valuation study in Pt. 2). Since Viscusi, Magat, and Huber focus on the most severe form of chronic bronchitis, it is also a much more serious condition than the chronic asthma in Rowe and Chestnut (1984). In addition, where Rowe and Chestnut use a sample of asthmatics who were quite familiar with the symptoms to be valued, Viscusi, Magat, and Huber use a sample of healthy individuals. Finally, they measure the value of a reduction in the risk of chronic bronchitis rather than the certain changes considered in the other studies. In light of these differences, Viscusi, Magat, and Huber make a number of innovative changes in the design of the contingent valuation survey.

A central problem faced by Viscusi, Magat, and Huber is the difficulty of communicating chronic disease effects to potential sufferers. To address this problem, the contingent valuation questionnaire is administered via an interactive personal-computer program. The program includes several tests to determine whether the respondent understands the valuation task being asked of him or her. If the respondent fails the tests, the program provides additional information before proceeding with the questionnaire. Responses were also subject to a set of consistency checks, and subjects were excluded from the sample if their responses could not be used or indicated that they did not fully understand the valuation task. Viscusi, Magat, and Huber discuss in detail the consistency checks and the number of subjects who failed the various tests. This discussion should be very useful to future contingent valuation studies facing similar problems.

Viscusi, Magat, and Huber employ another innovative design feature: their approach measures the rates at which people are willing to trade off chronic bronchitis risk reduction in terms of the risk of an automobile fatality (risk-risk trade-off) as well as in dollars (risk-dollar trade-off). The approach is partly motivated by the suspicion that consumers may be more willing and able to specify the rates of trade-off of one risk with another. Viscusi, Magat, and Huber note that, when faced with risk-dollar trade-offs, some subjects cannot envision that they would voluntarily subject themselves to higher risks for a finite amount of money. The risk-risk form of the morbidity valuations can be converted into dollar values by placing dollar values on the fatality risks as in the empirical literature on the value of a statistical life.

The study uses a sample of 389 shoppers from a blue-collar mall in Greensboro, North Carolina. Three series of valuation questions were asked. The first series yielded the risk-risk trade-off, specifically, the rate of trade-off between chronic bronchitis risk reductions and increases in the risk of an automobile fatality. The second series yielded the risk-dollar trade-off. The third series yields a dollar measure of the value of reducing the risks of an automobile fatality. Using this measure, the risk-risk form of the valuation can be converted and compared to the risk-dollar responses.

RESULTS. Viscusi, Magat, and Huber develop a range of estimates of the dollar value of reducing the risk of chronic bronchitis. As is often done when valuing mortality risks, the results can be expressed as the dollar value of a statistical case of chronic bronchitis. This form is a convenient way to convey results about the dollar value of small changes in risk, but care must be taken when interpreting the results. Valuing a statistical case of chronic

bronchitis at, for example, $883,000 does not mean that an individual is willing to pay that sum to avoid a case of chronic bronchitis with certainty. Instead a large number of individuals are each willing to pay a much smaller amount to reduce the risk of chronic bronchitis. Summing over the individuals, they are willing to pay $883,000 for a risk reduction that in a statistical sense is expected to result in one less case of chronic bronchitis.

From the risk-dollar trade-off questions, the mean value of a statistical case of chronic bronchitis is $883,000, with an associated standard error of $114,000. As is the case in the other contingent valuation studies reviewed in this chapter, the distribution of responses is skewed so that the mean is substantially larger than the corresponding median value of $457,000.

Viscusi, Magat, and Huber also convert the risk-risk results into dollar values, using several different estimates of the value of a statistical life. If a statistical life is valued at $2 million, using the mean of the risk-risk results implies that a statistical case of chronic bronchitis is worth $1,360,000. Using the median of the risk-risk results, a statistical case of chronic bronchitis is worth $640,000. If the value of a statistical life is higher than $2 million, the implied value of a statistical life is higher than $2 million, the implied value of a statistical case of chronic bronchitis increases accordingly. If the value of a statistical life is $5 million, the implied values for a statistical case of chronic bronchitis are $3,400,000 (mean) and $1,600,000 (median).

Krupnick and Cropper (1992)

STUDY DESIGN. Krupnick and Cropper (1992) study whether people's familiarity with chronic lung disease affects their willingness to pay for risk reductions. They accomplish this by administering the survey designed by Viscusi, Magat, and Huber to a sample of people who have a relative with chronic lung disease. The sample of 189 subjects responded to newspaper ads in the Washington, D.C., area. Each subject had a relative with chronic lung disease but did not have a chronic respiratory condition himself or herself.

The design and use of the computer-interactive survey was identical to Viscusi, Magat, and Huber, with two exceptions. First, a set of questions was appended to the end of the survey, asking about the respondent's familiarity with his or her relative's disease and about the severity of that disease. Second, Krupnick and Cropper randomly assigned the respondents to two versions of the survey instrument. The second version was modified to elicit valuations of "a case of chronic respiratory disease like your relative's." In the first version, a case of chronic bronchitis was described as in the Viscusi, Magat, and Huber survey.

RESULTS. As in Viscusi, Magat, and Huber, the value of a statistical case of chronic bronchitis can be estimated directly based on responses to the risk-dollar trade-offs and inferred based on responses to the risk-risk trade-offs. From responses to the first version of the survey, based on the risk-dollar trade-offs the mean value of a statistical case of bronchitis is $2.08 million. The mean of the responses to the second version of the survey is $1.84 million. Based on a $2 million value of a statistical life, the mean risk-risk trade-off results imply values of a statistical case of bronchitis of $1.34 million from the first survey version and $1.76 million for the second version.

Krupnick and Cropper also analyze the results to determine if the respondents' familiarity with the disease significantly affected the reported valuations. They do this by comparing their results to those of Viscusi, Magat, and Huber. To control for differences in sample characteristics, they estimate regression models that predict responses to the trade-offs as functions of respondent characteristics. Krupnick and Cropper test the hypothesis that the predicted mean of responses from the estimated models are the same across samples, conditional on respondent characteristics. They find that for the risk-dollar trade-offs, the respondents more familiar with chronic lung conditions reveal a statistically significantly higher mean willingness to pay to reduce the risk of chronic bronchitis. No statistically significant difference is found for the risk-risk trade-offs. Based on these results and additional analysis, they suggest that the answers to risk-risk trade-offs are more stable than answers to risk-dollar trades.

4.4. Comparing Cost of Illness and Contingent Valuation

Introduction

The cost of illness (COI) approach and contingent valuation (CV) are two important methods that allow a dollar value to be placed on a change in morbidity or sickness. A direct comparison of values based on these methods is undertaken in this section. This comparison is especially interesting because the methods are in some sense complementary. The cost of illness approach, focusing on medical expenditures and forgone earnings, uses widely available data and straightforward empirical techniques, so it is generally accepted on a practical level by many health professionals. However, there is no strong theoretical basis for using COI values in benefit cost analysis. That is, there are serious questions whether a COI value associated with a given change in morbidity will be close to what an individual would be willing to pay for that change. In contrast, contingent valuation experiments can be designed to directly estimate what an individual would be willing to pay for a certain change in morbidity. So CV values are estimates of the conceptually correct benefit measures for benefit cost analysis under certainty. Unfortunately, the proper design of CV experiments is difficult and still controversial, and some economists tend to be skeptical of the

actual values given by individuals in a CV experiment. On a practical level, COI values are often judged superior to CV values, while on a theoretical or conceptual level, CV values are preferred.

Due to the perceived practical advantages of the cost of illness approach, recent theoretical work has investigated the relationship between COI values and an individual's true willingness to pay (WTP) for changes in morbidity. Harrington and Portney's (1987) theoretical analysis supports the conclusion that a COI value is a lower bound to the true WTP, for the certainty case. The more general model presented in Chapter 2 also implies that under plausible conditions, COI < WTP under certainty; the model also allows the analysis to be extended to the case of uncertainty.

Contingent valuation studies of the value of morbidity have considered changes in health status that occur with certainty. This seems justified since the costs of adding uncertainty seem large in light of the problems encountered in surveys that deal with concepts of uncertainty and the benefits of adding uncertainty in the context of nonserious morbidity may be small. In this section, only the relationship between willingness to pay and cost of illness for certain changes can be directly addressed.

The empirical evidence presented in this section is used to test the hypothesis that the cost of illness values are lower bounds to the true willingness to pay values. Values reported in CV experiments are used to represent the true WTP values for a change from being certainly sick to being certainly well. On the assumption that the CV values are reasonable proxies for the true WTP, the empirical results support the hypothesis that COI < WTP. Alternatively, the fact that this reasonable relationship holds between COI- and CV-reported WTP can be seen as additional evidence on the usefulness and reliability of contingent valuation methods.

Next, previous work comparing cost of illness and contingent valuation is reviewed. The results of a new contingent valuation experiment are presented to test the hypothesized relationship. The analysis is extended to a preliminary discussion of the relationship of COI and WTP values under certainty, and the amount an individual would be willing to pay for a change in health risks. No direct evidence is available on willingness to pay for morbidity risks, but the analysis suggests an approximation from the evidence on certainty values is possible.

Previous Work Comparing COI and CV

Two contingent valuation studies on the value of morbidity contain some evidence on the relationship between cost of illness values and CV values. The first study, reported in Loehman et al. (1979), estimated median willingness to pay bids for reductions in air pollution–related symptoms. They note that the bids "are probably low compared to out-of-pocket costs of illness" (p. 233). As an example, the income loss per day for a person with

an average income would be $65, while the highest median reported for 1 day of relief from severe symptoms (shortness of breath) is $10.92. Including the value of medical expenditures would cause COI to exceed the CV bid by a larger amount. The difference may be in part due to paid sick leave and medical insurance causing out-of-pocket expenses to be low. Another problem is the use of median CV bids. In order to avoid overstating WTP because of the influence of a few very large bids on the means, they instead used the much smaller medians. This might have resulted in an understatement of WTP, however, which might explain why the CV bids are small relative to reasonable COI values. At least, the median CV bids should be compared to median COI values. In any case, Loehman et al. do not collect the data that would allow a direct comparison of individuals' CV bids and their experienced or expected costs of illness. Thus, their results seem to be only a weak indication that WTP is less than COI; that is, this is weak evidence against the hypothesis that COI is a lower bound to WTP.

A second CV study, by Rowe and Chestnut (1984) on the value of asthma, is more suitable to a direct comparison of CV bids representing WTP and the cost of illness. The first body of evidence on WTP compared to COI is the respondents' rankings of the importance of the benefits they might receive from reduced asthma. Based on statistical analysis of the rankings, Rowe and Chestnut conclude that discomfort and effects on leisure and recreation activities, which are part of WTP but not part of COI, clearly ranked above medical costs and work lost, which are the only components of WTP that a COI value includes. So according to these rankings, COI estimates do not include the most important benefits of reduced morbidity. This indicates that WTP should therefore exceed COI.

The second body of evidence from the Rowe and Chestnut study is a comparison of the total WTP bid and a constructed COI value. This method reported yields a ratio of WTP/COI of 1.6, supporting the hypothesis that WTP is greater than COI. Other approaches to measuring this ratio examined in their larger study suggest a ratio as high as 3.7.

Unfortunately, the data collected do not include forgone earnings, so to construct the COI value Rowe and Chestnut had to assume that the earnings forgone were equal to the medical costs. The assumption is justified on the grounds that the respondents' rankings of the importance of forgone earnings and medical expenditures were nearly identical. The comparison of WTP to COI does not seem sensitive to any inaccuracies inherent in this assumption.

Another problem in the construction of the COI value is that it includes only variable medical expenditures, such as medicine or doctor visits. The asthmatics interviewed also had significant fixed cost expenses on onetime goods such as intermittent positive pressure breathing machines. From Rowe and Chestnut's (1984) table 1, the total (household) fixed cost expenses

were $713, compared to total (household) variable expenses per year of $528. Clearly, the entire sum of fixed costs expenditures should not be compared to the willingness to pay for an improvement in morbidity. However, since the improvement would change individuals' marginal decisions on the purchase of a onetime good, ideally some (unknown) portion of the fixed expenses would be included in a COI value. It does not seem likely that doing so would change the result that WTP is greater than COI.

In general, while the Rowe and Chestnut study is not the ideal test of the hypothesis that WTP exceeds COI, it does offer strong support of that relationship. The final caveat is that the study involved only a relatively small sample of individuals with a chronic condition, asthma, and may not be relevant for the general population.

Comparisons of COI and CV values from the Loehman et al. (1979) and Rowe and Chestnut (1984) studies are thus somewhat inconclusive. The first study contains very weak evidence against the hypothesis that WTP exceeds COI. The second study contains much stronger evidence that supports the hypothesis, but problems with the study may limit its applicability.

Comparing COI and CV—New Results[1]

The contingent valuation study described in detail in Part 2 of this book was designed to collect the necessary data for a direct comparison of CV willingness to pay bids for changes in health status with certainty and experienced cost of illness. Only the surveys on seven light symptoms (coughing spells, stuffed-up sinuses, throat congestion, itching eyes, drowsiness, headaches, and nausea) are used for this comparative analysis. The surveys on angina could not be used because few of the respondents had experience with angina and its related cost of illness.

The total sample of the seven light symptom surveys used in the analysis was 131, using door-to-door and mall-intercept interview methods. Out of this sample, nine observations were unusuable because they were incomplete. Because of the limited scope of the sample, we view this empirical study as illustrative.

Table 4.4 compares the mean WTP and private COI for each of the seven symptoms in the contingent valuation survey. The comparison is made among those who have experienced the symptom in the previous year, that is, those for whom we have COI data. The private COI calculated consistent with the prevailing measure in the COI literature. It is the expenditures on medicine and doctor visits less any insurance payments plus any

1. The material in this subsection originally appeared in M. Berger, G. Blomquist, D. Kenkel, and G. Tolley, "Valuing Changes in Health Risks: A Comparison of Alternative Measures," *Southern Economic Journal* 53, no. 4 (April 1987): 967–84. Reprinted with permission.

TABLE 4.4. Willingness to Pay and Private Cost of Illness Comparisons of Means

Symptom	Sample Size[a]	Mean Daily Willingness to Pay (in $)[b]	Mean Daily Private Costs of Illness (in $)[c]	*t*-Value[d]
Coughing spells	27	105.34	11.29	2.12*
Sinus congestion	43	38.84	6.79	2.22*
Throat congestion	24	43.93	14.27	1.59
Itching eyes	16	172.23	14.56	1.24
Heavy drowsiness	6	173.89	21.50	2.57*
Headaches	48	173.21	3.33	2.07*
Nausea	18	91.24	2.36	2.03*

[a] Only those experiencing the symptom are included.
[b] Willingness to pay to avoid one extra day of the symptom.
[c] Calculated as expenditures on doctor visits and medicine net of insurance reimbursements plus lost earnings, expressed on a daily basis.
[d] Test of the null hypothesis that willingness to pay is less than or equal to private costs of illness.
* Indicates hypothesis rejected at 0.05 level of significance in a one-tailed test.

lost earnings. Both the individual WTP and COI measures are expressed on a daily basis.[2]

Out of the entire sample of 122 individuals, the subsamples of those who had experienced the various symptoms in the previous year ranged in size from six for drowsiness to 48 for headaches. Within each of these subsamples, the mean WTP always exceeded the mean cost of illness. The last column of Table 4.4 indicates that in five of the seven cases, the differences were significant at the .05 level in a one-tailed test.

Another way to test the equality of the private COI and the WTP is through the use of a nonparametric sign test (see Hoel 1971, pp. 310–15). This type of test is less sensitive to extreme WTP or COI values than is the *t*-test. For the sign test, the 192 WTP-COI pairs across all seven symptoms are compared. In 174 cases, the WTP exceeds private COI. If the WTP-COI pairs had in fact come from the same distribution, we would expect that in only 96 cases would WTP exceed COI. We can then test whether 174 is significantly greater than 96 by using the binomial approximation to the normal distribution.[3] The resulting value of the test statistic is 11.26,

2. The contingent valuation experiments were conducted for both 1-day and 30-day changes in the experience of the various systems. Implicit in the normalization to 1-day changes is the assumption of constant marginal costs in the case of cost of illness and constant marginal utility in the case of willingness to pay.

3. The standard deviation for calculating the normal distribution test statistic is constructed under the null hypothesis that the WTP-COI pairs come from the same distribution. In this

which is significantly different from zero at a .001 level of significance, further adding to the empirical evidence that WTP exceeds COI.

There are two types of additional evidence that support the finding that WTP exceeds COI. First, we asked individuals to rank the reasons for their values for symptom relief. Focus group feedback led to development of a five-item list that covered most reasons. The reasons and the percentage of the 122 respondents who ranked the reason as the most important are comfort (67%), loss of work at home (6%), loss of work away from home (12%), loss of recreation (2%), reduction of medical expenses (11%), and other (2%). So, as in the Rowe and Chestnut (1984) study, the components of the value of health included in COI are ranked as less important than the components COI omits.

We also estimated simple ordinary least squares regressions of WTP on the private COI.[4] In each case the intercept is positive and, in most cases, is significantly different from zero. The slope term is never significantly different from zero. However, in the cases in which it approaches significance, it is positive. Thus, the regression results are consistent with the above finding that in general WTP exceeds COI, although there does not appear to be any strong tendency for the two to move together. This suggests that it is not possible to predict WTP based on COI. So while WTP/COI ratios could be computed based on the means reported in Table 4.4, yielding ratios of about 3 to over 50, the regression results suggest that these ratios are not particularly meaningful.

Implicit in our WTP-COI comparison is the assumption that the symptoms that people experienced in the previous year are the same as those that they are bidding on in the contingent valuation experiments. For the light symptoms included in the survey, the differences are rather inconsequential. When the samples are limited to those who reported that their symptoms were the same, not worse, or less severe than the contingent symptoms, the mean of WTP is still greater than the mean of COI for each symptom, and although the dollar differences are greater for four of the seven symptoms, only two of the t-values are significant at the .05 level due to the smaller sample sizes. The nonparametric sign test yielded a test statistic of 8.77, and the regression results are similar to those described above.[5]

case the probability that WTP > COI is 1/2, and the standard deviation for the binomial approximation to the normal distribution is $174 \times 1/2 \times 1/2 = 43.5$.

4. These and other results not reported in this chapter are available upon request.

5. A final piece of corroborating evidence is contained in the survey. Individuals were asked how much they would be willing to pay to avoid all the symptoms they had experienced in the previous year. Of the 46 individuals who did not experience symptoms in combinations with one another, 41 had WTP > COI, yielding a nonparametric sign test statistic of 5.3, which is highly significant. The mean WTP greatly exceeded the mean COI, and a simple regression yielded results similar to those described previously.

6. The value of time lost from market or nonmarket activity is measured by multiplying the number of days lost by the daily wage (hourly wage × 8). This reduces the sample somewhat since not everyone in the sample worked in the previous year and thus reported a wage rate. We also expanded the definition of cost of illness even further to include days of market and nonmarket activity "hindered." This cost of illness measure is the same as above except that it also includes the number of days hindered multiplied by one-half the daily wage. The means test, sign tests, and regressions were all recalculated, and the results are very similar to those described for the first expanded cost of illness measure.

Our empirical evidence suggests that the private COI, defined by excluding time lost from consumption, is less than WTP. Is it the exclusion of these time expenditures that is driving the result? In order to investigate this question, we use other information available from our contingent valuation survey to construct an expanded COI measure that can then be compared to the WTP values. This measure is the cost of medicine and doctor visits net of insurance reimbursements plus the value of time lost from any activity (e.g., market, work, school, work at home).[6] This increases the measured COI and is more compatible with theoretical models of COI. A comparison of the mean COI and WTP for the various symptoms indicates that WTP is greater than COI in six of seven cases (the exception is throat congestion), although the significance levels of the t-statistics are lower than before (they range from $-.165$ to 2.08). The nonparametric test produces a test statistics of 5.48, which is again significant at the .001 level, indicating WTP > COI. Regressions explaining WTP again produce positive (although smaller) and mostly significant constant terms and insignificant COI coefficients. So overall, the exclusion of lost consumption time does not appear to be the reason for our earlier result. Our empirical results are consistent with the hypothesis that consumer surplus exceeds the private COI, whether or not the value of lost consumption time is included. It should be noted, though, that our earlier measure, excluding the value of lost consumption time, is more consistent with that used in COI studies.

The next step is to generalize our results to the relationship between willingness to pay for a change in morbidity risks and the expected COI. From the theoretical model of Chapter 2, if exogenous change that lowers the probability of contracting an illness causes individuals to reduce their preventive expenditures (i.e., if dX/dE is negative), then willingness to pay for a change in risks exceeds expected consumer surplus (CS). This is true since individuals would also be willing to pay their preventive expenditure savings to avoid increases in health risks. While our survey contains no direct evidence on the sign of dX/dE, fortunately, it contains some indirect evidence. Individuals are asked whether they have made various defensive

expenditures for health reasons: whether they have purchased air conditioners, air purifiers, humidifiers for their home or car, or made other preventive expenditures. Nontrivial proportions of the full sample have made some type of preventive expenditure. But more interesting are the differences between those who have and have not experienced at least one of the seven light symptoms.[7] While the percentages of the two groups are almost equal for the purchase of humidifiers, those who have experienced at least one of the seven symptoms are more likely to have made expenditures in the other three categories than those who have not. The difference is most pronounced for air conditioners. No one in the group not experiencing any symptoms purchased on air conditioner for health reasons, but 19 of those having at least one of the seven symptoms did so.

What does this pattern of preventive expenditures tell us about the sign of dX/dE? The pattern is consistent with a negative dX/dE in the following way. Assume that those having experienced the symptoms also experience worse exogenous environmental conditions. This results in a higher probability of experiencing the symptom. In looking across the sample, we observe an increase in the quality of the environment ($dE > 0$) in moving from those who have experienced at least one of the symptoms to those who have not. The resulting change in preventive expenditures then appears to be negative. It should be stressed that the above explanation is only consistent with $dX/dE < 0$. The data in the survey do not allow for a strict test of hypothesis.

However, if it is true that $dX/dE < 0$, then our empirical results are also consistent with willingness to pay for a change in morbidity risks being greater than the expected COI. This allows us to make statements about our theoretical model with uncertainty from our empirical results, which by practical necessity are couched in terms of certainty, and yield only estimates of willingness to pay under certainty, in other words, an estimate of consumer surplus.

One final illustration will help show the usefulness of our empirical consumer surplus estimates. From the theoretical model, it is plausible that the

7. The proportions of the full sample having made various preventive expenditures, and the proportions among those who have and have not experienced at least one of the seven light symptoms, are as follows:

Preventive Expenditure	*Full Sample*	*No Symptoms*	*One or more Symptoms*
Air conditioner	.151	.000	.188
Air purifier	.110	.044	.126
Humidifier	.311	.318	.309
Other	.074	.056	.078

expected change in consumer surplus is a lower bound on willingness to pay for a change in health risks. Since the contingent valuation experiment measures CS, if we assume some value for the change in probabilities of becoming sick, we can estimate a lower bound for the value of the reduction of health risks. For example, in Table 4.4 we report that, among those having experienced coughing spells in the previous year, the mean CS for avoiding 1 extra day of cough with certainty is \$105.34. These individuals had on average approximately 48 days of coughing spells in the previous year. If we assume that the probability of having a coughing spell on any given day is constant throughout the year, the mean individual faces approximately a .13 probability of having a coughing spell each day. A lower-bound estimate of the willingness to pay for a 10% reduction in the risk of a coughing spell on any given day for the mean individual is simply $-\text{CS}\ dH/dE$ or $\$105.34 \times .013 = \$1.37$. The willingness to pay for a whole year's worth of 10% reductions is $\$1.37 \times 365 = \$500.05$. Lower bounds on the values of changes in the risks of the other symptoms can be similarly calculated. It should be stressed, however, that our lower-bound estimates, while useful for comparisons among approaches, should be used for policy purposes with caution. Our small sample is probably not representative of the entire U.S. population. In addition, it should be recalled that the contingent valuation experiment contained no direct evidence on the value of morbidity risks, and the lower-bound estimates depend upon the theoretical model used in chapter 2.

Conclusion about Comparisons

Our empirical work provides evidence on WTP and COI for seven light symptoms in the certainty case: coughing spells, stuffed-up sinuses, throat congestion, itching eyes, heavy drowsiness, headache, and nausea. The WTP values that are obtained are equivalent to consumer surpluses. The results suggest that WTP exceeds COI, but there is no strong indication that WTP and COI move together in any systematic fashion. Assuming that exogenous changes affecting health risks reduce preventive expenditures, our results also imply that the WTP for reduction in health risks that arises from our uncertainty-based model exceeds expected COI. We then provide an illustrative lower-bound estimate of the value of a change in health risks from our contingent valuation survey.

The results of the new empirical work thus tend to confirm Rowe and Chestnut's (1984) preliminary results that WTP exceeds COI. It should be noted that this relationship is also found in the experimental mail survey completed (see Chap. 13), but the results are for a very small sample. So there is a growing body of evidence that suggests that contingent valuation responses on WTP exceed COI, as predicted by several theoretical models. The major limitation is the small sample sizes of the studies.

4.5. Conclusions and Summary of Contingent Valuation

An assessment of the contingent valuation method suggests that with careful design the resulting value estimates may be fairly accurate. With this in mind, this chapter reviewed six studies that applied the contingent valuation method to the problem of valuing health: Loehman et al. (1979), Dickie et al. (1987), and the study described in Part 2 address the value of light symptoms; Rowe and Chestnut (1984), Viscusi, Magat, and Huber (1991), and Krupnick and Cropper (1992) address the value of asthma and bronchitis. Each of these studies seems to be carefully designed, though certain problems are noted. Future work could focus on two general problems: (i) the treatment of extremely large bids that are statistical outliers, and (ii) the relationship between bids for a marginal day of relief, compared to the average value per day of relief when a larger change in health is being contemplated.

While the health effects valued are not exactly the same, certain comparisons can be made between the results of four of the six studies. Each of the four studies implies a value for one day of respiratory symptoms, though not always of the same symptoms. From the Loehman et al. (1979) study, 1 day of coughing/sneezing has a mean value of $138 (mild day) or $236 (severe day). In terms of the initial bids, the mean value placed on a day of coughing from the Dickie et al. (1987) study is $355.10, but the mean falls to between $1.61 and $9.96 after alternative methods to improve reliability are applied to the sample of bids. Our study finds that relief from 1 day of coughing, throat, and sinus problems has a mean value of $65.60. The Rowe and Chestnut (1984) study implies that relief from 1 day of asthma symptoms is worth on average about $20.

These different values can be reconciled, to some extent. First, the Rowe and Chestnut value is not a value for a marginal day of relief but an average value for 1 day, given an average of 19 days of symptoms relieved. Thus, it is not really comparable to the other estimates. The Loehman et al. study is more directly comparable to Dickie et al. and our study. In general, somewhat different values result. But if we compare median bids across the studies, or compare mean bids across the studies, the values are closer. The range of values is narrowed further if, as argued above, the results from the Dickie et al. study are best viewed as lower-bound estimates.

5

Household Health Production, Property Values, and the Value of Health

Richard Clemmer, Donald Kenkel, Robert Ohsfeldt, and William Webb

5.1. Introduction

Although public policies and other external factors are important determinants of health, health remains at least partially under the control of the individual consumer. By analyzing the decisions consumers make when faced with trade-offs between health and other economic goods, it is sometimes possible to infer the value of health to the consumer. This chapter reviews two approaches used in empirical studies that concern such trade-offs: the household health production framework, and the hedonic analysis of housing markets. The main focus is on developing useful empirical estimates of the value of health based on these approaches.

In the household production framework, the individual is seen as producing the commodity health by combining his or her own time and effort with purchased goods such as medical care, diet, and so on. Some recent theoretical and empirical work has used this framework to derive expressions for what an individual would be willing to pay for an exogeneous improvement in health or health risks. The theoretical studies, such as the model developed in Chapter 2 and the references therein, investigate how the conceptually correct willingness to pay measure will be related to observable quantities, namely, the cost of illness and preventive expenditures (averting behavior). Since health is not the only commodity produced by the household, the approach can also be used to value environmental benefits more generally; see Smith (1991).

In keeping with the focus on empirical value estimates, Section 5.2 examines several empirical studies that use the household production approach to estimate willingness to pay for health directly. Cropper (1981) estimates willingness to pay for health risks related to an index of air pollutants. Gerking and Stanley (1986) estimate willingness to pay for health risks related to ozone exposure, and Dickie and Gerking (1991) develop estimates