Valuing morbidity: An integration of the willingness-to-pay and health-status index literatures

F. Reed Johnson a,*, Erin E. Fries a, H. Spencer Banzhaf b

a Triangle Economic Research, 1000 Park Forty Plaza, Suite 200, Durham, NC 27713, USA
b Duke University, Durham, NC, USA

Received 1 February 1996; revised 1 December 1996

Abstract

Placing dollar values on human health has long been a controversial aspect of policy analysis and remains difficult given the relatively small number of morbidity-valuation studies available. By combining both the economic and health literature, this paper offers an alternative approach to morbidity valuation and provides estimates for a wide range of short-term health conditions. © 1997 Elsevier Science B.V.

1. Introduction

Several recent applications in the fields of health policy and environmental policy have evaluated the cost-effectiveness of programs or treatments that reduce morbidity. (See, for example, Sintonen and Allander, 1990; Johannesson, 1995; Edelson et al., 1990; Schulman et al., 1991; van Hout et al., 1993.) Other researchers have attempted to weigh the costs of particular therapies or drug treatments against their benefits (Weisbrod, 1981; Johannesson et al., 1991). At a time when increasing health-care costs are a serious concern, the health field has been forced to closely examine benefits and costs associated with reduced...
morbidity. In addition, regulatory agencies are increasingly requiring industries such as electric utilities to account for the social costs, including health effects, that their activities may cause (Freeman et al., 1992). Increased interest in benefit–cost analysis (BCA) for use in the fields of health policy and health, safety, and environmental regulation has created a need for the monetary valuation of the benefits of improved health.

Although most research has focused on cost-effectiveness ¹, there are several reasons why BCA may be preferred by economists. Perhaps the most important reason is the fact that BCA is firmly grounded in welfare economics; that is, it requires that individuals make trade-offs between income and other goods, such as health. Therefore, BCA can provide information on the value of one program versus another. In contrast, cost–effectiveness analysis (CEA) provides a cost per health-outcome unit and, therefore, gives information only on the technical efficiency of one program relative to another program producing the same outcome. It does not provide a basis for comparing the net benefits of different outcomes or between health outcomes and alternative uses of scarce resources ². If an outcome is cost-effective but that outcome is not highly valued, the fact that its cost per outcome is low is not particularly relevant ³. A second reason why BCA may be preferred to CEA is the fact that some improved health outcomes, particularly those that are short-term, may not be captured well by measures that are commonly used in CEA. For example, many conditions resulting from a small increase in the level of air pollution are short-term or simply worsen conditions that already exist. In such cases, measures like quality-adjusted life-years may not be appropriate. Finally, BCA accounts for the fact that preferences vary across individuals, and across health outcomes, while CEA assumes that preferences for given outcomes are the same across individuals (e.g., all individuals wish to maximize life-years saved). However, preferences for health may not be uniform across all individuals.

Despite the fact that BCA may be preferred to CEA for some applications, the literature providing monetary health values is deficient in both breadth and quality

¹ A review of the health services research literature by Elixhauser et al. (1993) indicated that from 1985 to 1990, there were 724 cost-effectiveness studies and 401 benefit–cost studies.

² In terms of basic microeconomic theory, BCA provides the efficient level of production of health outcomes by providing information on marginal costs and marginal benefits. CEA, however, only gives information on the average costs of producing health outcomes; it cannot indicate which outcomes should be chosen or at what level the outcomes should be produced.

³ In a study that evaluated 500 different life-saving health interventions, Tengs et al. (1995) found that the cost of certain interventions exceeded hundreds of millions of dollars per life-year saved. In most of these cases, it is likely that the benefits of the intervention did not come close to approaching the costs. However, we cannot state that definitively, because benefits were not measured. Furthermore, certain outcomes may not be beneficial; a recent study by Desvousges et al. (1996) shows that respondents' WTP for increased longevity actually was negative for some very low quality-of-life health states.
of coverage. While good theoretical papers continue to appear, most applied results are relatively old. Until new studies appear, however, economists must use these results when evaluating health and environmental policy. When required to value a given health condition, economists typically have used a 'best-study' approach. This procedure requires selecting a single study based on such criteria as its age, sample size, methodology, and validity (e.g., Rowe et al., 1995; Harrison et al., 1993). However, this approach has at least three drawbacks. First, by valuing health conditions on a best-study basis, researchers are unable to provide estimates for conditions that have not been studied. Second, they must discard potentially useful information from other available studies. Finally, researchers must make difficult and subjective choices among studies, especially when no single study clearly is better than another.

For these reasons, a statistical synthesis, or meta-analysis, of the morbidity-valuation literature is warranted. By synthesizing information from several studies, meta-analysis can improve statistical precision while basing its results on a whole body of evidence rather than on just one study. A properly performed meta-analysis thereby allows researchers to have more confidence in their conclusions. Although meta-analysis has been commonly used in the health sciences in the last several decades, the technique remains relatively rare in economics. (For examples, see Smith and Huang, 1993; Smith and Osborne, 1995; Boyle et al., 1994; Smith and Yoshiaki, 1990.)

In this study we pool almost all of the available morbidity WTP estimates for short-term conditions, then link them to the health-state index literature through the quality of well-being (QWB) index, which provides a measure of each health condition's perceived severity. This approach allows us to pool values from different studies and for different health conditions based on differences in severity. Thus, we use all the available information to estimate the value for a specific effect, such as a cough-day, not just the information from any one study or even just the cough-day values from multiple studies. Furthermore, because any health condition can be given a score on a health-status index within the range of our data, we can predict the WTP for any short-term health condition with specified characteristics.

*Some researchers, such as Fisher (1984), have favored COI studies simply because of the lack of existing WTP estimates. Other researchers have found the limited amount of available data frustrating when conducting transfer studies. For example, after reviewing available WTP studies, Krupnick and Kopp (1988) simply assigned arbitrary bids for low, medium, and high estimates rather than using the WTP data.


6 Mauskopf and French (1991) suggested a conceptually similar approach for using health-status indexes to extrapolate WTP from a single condition to other health conditions.
In the following discussion, we first review the morbidity-valuation studies for various short-term health conditions, including a description of various data problems in these studies. We next discuss our use of the QWB index. We then propose a conceptual framework for the utility of health. This is followed by the empirical results from our examination of the relationship between WTP and QWB using panel models. Finally, we discuss potential applications and offer conclusions from our study.

2. Morbidity-valuation studies

To estimate benefits for use in BCA, economists have estimated the monetary value of avoiding or reducing health conditions in a number of ways, including cost-of-illness (COI) studies and WTP studies. COI studies measure the health-care and other pecuniary costs associated with a given condition, while WTP studies measure people's actual willingness to sacrifice consumption of other goods and services in return for improved health. Because traditional economic theory indicates that WTP studies are the theoretically correct measure of the social value of health, we use only WTP studies in this analysis.

In practice, economists primarily have used two methods for measuring people's WTP to avoid or reduce health conditions: averting-behavior studies and contingent-valuation studies. In studies of averting behavior, economists estimate people's implicit value for health by examining how much they are willing to sacrifice in financial expenditure and time for goods, such as an air conditioner or water filter, that minimize the risk of experiencing health conditions. Although averting-behavior studies measure willingness to pay, the approach is complicated by the fact that most averting behaviors produce joint products—air conditioners, for example, may produce both better health and comfort. Moreover, use of these studies in our analysis would require information about the efficacy of these goods in producing health-information that may be limited. For these reasons, we have not included averting-behavior studies.

Contingent valuation (CV) measures stated, rather than revealed, preferences. CV methods use surveys to ask people what they would be willing to pay for some nonmarket commodity, such as improved health. When eliciting WTP for health states, CV surveys ask respondents to value a given health outcome in a hypothetical context. CV has been used to elicit values for both morbidity and mortality (Tolley et al., 1986; Loehman et al., 1979; Dickie et al., 1988; Chestnut et al., 1988; Rowe and Chestnut, 1985; Gerking et al., 1988). Although CV has become

---

widely accepted in recent years, the method clearly has some limitations. Critics have cited many potential problems. Nonetheless, CV surveys are based on sound theoretical concepts and may result in valid and reliable WTP values if conducted properly.

We reviewed the available WTP studies and identified five CV surveys of short-term morbidity conditions: Loehman et al. (1979), Tolley et al. (1986), Dickie et al. (1987, 1988), Rowe and Chestnut (1985), and Chestnut et al. (1988). Each study estimated values for multiple symptoms. Four of these studies elicited respondents’ WTP to reduce or avoid common health effects associated with air pollution such as cough, wheezing, asthma attacks, headaches, and chest discomfort. In addition, Tolley et al. (1986) and Chestnut et al. (1988) elicited WTP to avoid angina attacks. These five CV studies, with a total of 53 data points for all health conditions, provide a basis for estimating the empirical relationship between stated WTP and health-status scores.

3. Data and survey-design concerns

Although the morbidity-valuation studies used in this analysis are WTP studies that follow theoretic principles, we nonetheless have several concerns about these data. For example, some use elicitation formats that may induce bias. Several also have small sample sizes and/or low response rates. In addition, because most of the CV health studies ask respondents WTP questions about several different conditions, the results from these studies may be affected by intra-study anchoring or sequence effects. Furthermore, Dickie et al. (1987, 1988) used a revision procedure that has been criticized (e.g., Cropper and Freeman, 1991; Krupnick and Kopp, 1988).

Some of our concerns involve theoretical considerations. For example, some WTP studies included in our analysis ask for the value of avoiding an increase in a health condition, while others ask for the value of reducing a health condition. Assuming that the marginal utility of health diminishes as conditions approach

---

9 A smaller version of this study was published by Berger et al. (1987).
10 Although there are other quality WTP studies, even of short-term symptoms, many of these do not specify the exact number of days being reduced or avoided, but instead present more relative quantities such as chronic hypertension (Johannesson et al., 1993; and Johannesson et al., 1991). To control for temporal differences among studies, we must know the number of days alleviated.
11 For example, iterative bidding games may cause ‘‘yea-saying’’ on the part of respondents. Payment cards, in which respondents choose their WTP from a given range, may also induce bias by suggesting a range of values.
perfect health, WTP should vary depending on whether the respondent is valuing the avoidance or the reduction of a condition relative to a common reference point\textsuperscript{12}. However, the difference between avoiding additional days and reducing the number of current days may not be large for short-term conditions that do not result in large changes in overall health status.

Similarly, although most morbidity-valuation studies collected some information on respondents' baseline health, they generally did not collect adequate information on each respondent's current health endowment (e.g., average number of days per year that the respondent experiences a given condition). This information is important, as a person who has 10 cough-days a year probably views a one-day reduction or avoidance in cough-days differently than a person who has 30 cough-days a year \textsuperscript{13}. While most morbidity-valuation studies typically ask respondents only to consider their current health status, Tolley et al. (1986) clearly specify a hypothetical baseline for those respondents who answer angina-valuation questions. However, the same study does not specify a baseline condition for milder conditions, making it difficult to know exactly what change in health respondents were valuing. Similarly, other studies in this analysis do not use hypothetical baselines and do not collect detailed information on respondents' current health endowment.

Related to the problems with defining baseline health is the concern that morbidity-valuation studies do not always clearly describe the health condition. Respondents should be told the exact nature and specific symptoms they will experience, the severity of discomfort associated with the symptoms, etc. Although Tolley et al. (1986) described each condition in detail before the valuation question, other studies in this analysis simply asked respondents for their WTP for a 'mild' or 'typical' episode of a named condition. Although most conditions included in this analysis are common ailments, in some cases respondents may never have experienced the condition or may have assumed varying attributes for a typical episode \textsuperscript{14}.

While these concerns are important, we again note that there are very few morbidity WTP studies available and none is without flaws. Given these concerns, policymakers would be right to hesitate before using any single study, a further motive for our research. Ideally, we would control for data and methodological

\textsuperscript{12} Over 75 percent of the estimates included in this analysis — those from Tolley et al. (1986), Loehman et al. (1979), and Chestnut et al. (1988) — measure WTP for the avoidance, rather than the reduction, of a given condition.

\textsuperscript{13} In a study of angina, Brien et al. (1994) found that the mean bid for a one-day reduction from mild angina was higher for respondents endowed with ten mild days than respondents endowed with one mild day. Tolley et al. (1986) found similar evidence.

\textsuperscript{14} The Loehman et al. (1979) study and the Tolley et al. (1986) study asked respondents to value conditions regardless of whether they had experienced them.
effects in our meta-analysis. Unfortunately, it is not possible to disentangle the indicated concerns from other study-specific effects in such a small sample.

4. The QWB health-state index

Health-state indexes emerged out of the need to evaluate both the increased length and quality of life achievable by alternative health-care programs. The indexes are based on the idea that both objective factors, such as behavior or motor function, and subjective factors, such as people’s ability to fulfill the roles and expectations they have for themselves, affect people’s well-being. The indexes place the objective and subjective components of health along a continuum of functional health states, and then collapse them into a single, univariate scale that measures different levels of well-being (see Fanshel and Bush, 1970 and Torrance, 1986 for further discussion). In this way, health-state indexes can provide a quantitative measure of people’s relative preferences for various health states.

In this analysis, we use the QWB health-state index. The QWB index is a general health index that rates health status on a scale from 0 to 1, where 0 represents death and 1 represents perfect health. This index comes from a large body of literature based on a model first proposed by Fanshel and Bush in 1970. Over the last two decades, the index has been operationalized, updated, and revised several times. In 1991, Oregon Health Services (Oregon Health Services Commission, 1991) estimated a new set of index weights.

We chose the QWB index for a number of reasons. First, it is one of the few indexes that provides a fairly straightforward yet very comprehensive measure of health. Second, the utility weights for the index are based on a large sample of the general population, the appropriate sample for general policy applications. Finally, the QWB index has been used in several health-program evaluation studies (Wu et al., 1990; Oregon Health Services Commission, 1991). Other general health-state indexes include the Sickness Impact Profile (Bergner, 1993), the EUROQOL (Rosser and Sintonen, 1993), and the 15D Index (Sintonen and Pekurinen, 1993). Although these indexes are also prominent, they are extremely detailed and require complex assessments to measure health status. The QWB index allows for a reasonable level of detail, while providing a complete measure of health.

Although the standard gamble, measuring choice under uncertainty, is the elicitation technique most directly related to the expected utility of health, the rating scale used to elicit preferences for the QWB index may also provide a measure of the utility for health. According to Torrance (1987), standard gamble and rating-scale exercises had similar results after the appropriate power curves were used to calibrate measurements.

We use the QWB index to provide a standardized unit of health. However, it is not a complete utility function. As our conceptual framework later in the paper indicates, other variables, such as income and demographics, also determine the health-state utility.

Studies include Bush et al. (1972, 1973); Patrick et al. (1973); Kaplan et al. (1976, 1978); Kaplan and Bush (1982); Kaplan and Ernst (1983); Kaplan and Anderson (1988, 1990); Kaplan et al. (1993); Kaplan et al. (1995).
The QWB index measures health in four dimensions: three 'function states' — mobility (MOB), physical activity (PHY), and social activity (SOC) — and the most severe symptom/problem complex (SYM). Because it includes dimensions for both function levels and symptom/problem complexes, the index incorporates both illness and dysfunction into one index score. Two major studies have estimated weights for the index's four dimensions from surveys of the general population: Kaplan et al. (1993) and the Oregon Health Services Commission (1991). We use the Kaplan et al. weights for this analysis 19.

Although health-state indexes offer the most conceptually appealing approach for evaluating policies that have an impact on human health, some researchers are concerned about their reliability, stability, and validity (Mehrez and Gafni, 1989; Torrance, 1986). While recognizing these concerns, we note that the QWB index has been extensively evaluated by health researchers. (See Kaplan et al., 1978; Froberg and Kane, 1989; Kaplan and Ernst, 1983 for a discussion of these issues.) Although some criticisms of health-state indexes may limit their usefulness in certain contexts, our analysis indicates that the QWB index is a useful construct for rating the short-term conditions with which we are concerned.

5. The data

Table 1 lists the available WTP values from the existing short-term morbidity-valuation studies in 1993 dollars 20. These WTP values come from the five CV studies discussed above (Chestnut et al., 1988; Dickie et al., 1987, 1988; Loehman et al., 1979; Rowe and Chestnut, 1985; Tolley et al., 1986). We applied a 5-percent alpha trim to both ends of the distribution of WTP values from Loehman et al. (1979), as the authors did not adjust for outliers or protest bids, opting instead to report median values 21. In the case of the Dickie et al. (1987, 1988) study, we use a restricted sample of the unrevised bids (with inconsistent responses removed) as a compromise between the extraordinarily high preliminary bids and the revised bids.

To link these WTP values to health states, we assigned a QWB score (using the Kaplan weights listed in the appendix) to each health condition listed in Table 1 using the four dimensions of the index discussed previously 22. We categorized

---

19 For a detailed discussion of each set of weights, see Johnson et al. (1996).
20 Despite the fact that WTP studies of chronic morbidity exist (such as Johannesson et al., 1991, 1993; Thompson et al., 1984; Viscusi et al., 1991; Krupnick and Cropper, 1992), our analysis intentionally includes only short-term CV studies valuing relatively similar durations (i.e., no condition in our analysis exceeds 90 days).
21 We thank Edna Loehman for providing the original data from the study.
22 Because we cannot construct QWB scores with complete accuracy, the use of QWB score as an independent variable in our regression model may introduce an errors-in-variables problem. The direction of this type of bias is indeterminate (Greene, 1993).
each condition based on the descriptions in each CV study. For example, Tolley et al. (1986) described ‘cough’ to respondents as follows: ‘You will cough about twice an hour in spells that last 10 to 20 seconds. You will feel the cough in your chest, but it is not severe enough to make you red in the face’ (vol. 3, p. 107). We considered this health condition relatively mild with no effect on function level. Therefore, we computed the following QWB score:

\[
\text{QWB} = 1 + \text{MOB} + \text{PHY} + \text{SOC} + \text{SYM} \\
= 1 + (-0.000) + (-0.000) + (-0.000) + (-0.257) \\
= 0.743 \\
\]

We repeated this procedure for all valuation estimates listed in Table 1. In cases where respondents received no description of the health condition, we assumed a moderate case and calculated QWB scores based on a review of major medical guides and the expert opinion of several health professionals.

6. A conceptual framework of utility for health

Before presenting the empirical results of this meta-analysis, we first present a simple conceptual framework of private utility for health. We assume that each individual has preferences with conventional properties that can be described as follows:

\[
U = U[H, X(H)] , \quad H = f(\sum_i \sum_j h_{ij}, D) \\
\]

where \(H\) indicates the quality and duration of health status; \(h_{ij}\) is an index weight for health attribute \(i\), level \(j\); \(D\) is the duration of the health status; and \(X\) is all other goods. Utility then is a function of health and all other goods, and the utility of all other goods is dependent on health.

Suppose health outcomes are determined exogenously. Maximizing Eq. (2)
<table>
<thead>
<tr>
<th>Study</th>
<th>Health condition</th>
<th>Number of days</th>
<th>1993 $ WTP</th>
<th>WTP/day</th>
<th>Standard error</th>
<th>Sample size</th>
<th>QWB (Kaplan)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chesnut et al. (1988)</td>
<td>Angina</td>
<td>2</td>
<td>216.00</td>
<td>108.00</td>
<td>-</td>
<td>22</td>
<td>0.641</td>
</tr>
<tr>
<td>Chesnut et al. (1988)</td>
<td>Angina</td>
<td>1</td>
<td>131.00</td>
<td>131.00</td>
<td>-</td>
<td>22</td>
<td>0.641</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Throat congestion</td>
<td>1</td>
<td>21.75</td>
<td>21.75</td>
<td>7.77</td>
<td>25</td>
<td>0.830</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Cough</td>
<td>1</td>
<td>15.01</td>
<td>15.01</td>
<td>4.44</td>
<td>26</td>
<td>0.743</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Sinus congestion</td>
<td>1</td>
<td>17.23</td>
<td>17.23</td>
<td>4.13</td>
<td>41</td>
<td>0.830</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Wheezing</td>
<td>1</td>
<td>15.07</td>
<td>15.07</td>
<td>9.08</td>
<td>8</td>
<td>0.743</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Shortness of breath</td>
<td>1</td>
<td>9.05</td>
<td>9.05</td>
<td>5.84</td>
<td>11</td>
<td>0.743</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Pain on deep inspiration</td>
<td>1</td>
<td>34.98</td>
<td>34.98</td>
<td>13.72</td>
<td>10</td>
<td>0.701</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Chest tightness</td>
<td>1</td>
<td>25.68</td>
<td>25.68</td>
<td>6.79</td>
<td>15</td>
<td>0.701</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Cannot breathe deeply</td>
<td>1</td>
<td>19.31</td>
<td>19.31</td>
<td>7.98</td>
<td>19</td>
<td>0.743</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Headache</td>
<td>1</td>
<td>24.63</td>
<td>24.63</td>
<td>6.02</td>
<td>49</td>
<td>0.756</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Eye irritation</td>
<td>1</td>
<td>19.86</td>
<td>19.86</td>
<td>8.38</td>
<td>34</td>
<td>0.770</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Runny nose</td>
<td>1</td>
<td>12.80</td>
<td>12.80</td>
<td>6.53</td>
<td>13</td>
<td>0.830</td>
</tr>
<tr>
<td>Dickie et al. (1988)</td>
<td>Chest congestion</td>
<td>1</td>
<td>3.14</td>
<td>3.14</td>
<td>1.70</td>
<td>10</td>
<td>0.701</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Cough/sneeze (mild)</td>
<td>1</td>
<td>12.11</td>
<td>12.11</td>
<td>22.39</td>
<td>356</td>
<td>0.743</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Cough/sneeze (severe)</td>
<td>1</td>
<td>33.43</td>
<td>33.43</td>
<td>55.75</td>
<td>356</td>
<td>0.682</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Cough/sneeze (mild)</td>
<td>7</td>
<td>39.92</td>
<td>5.70</td>
<td>65.80</td>
<td>356</td>
<td>0.743</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Cough/sneeze (severe)</td>
<td>7</td>
<td>87.35</td>
<td>12.48</td>
<td>134.12</td>
<td>356</td>
<td>0.682</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Cough/sneeze (mild)</td>
<td>90</td>
<td>106.29</td>
<td>1.18</td>
<td>146.61</td>
<td>356</td>
<td>0.743</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Cough/sneeze (severe)</td>
<td>90</td>
<td>237.51</td>
<td>2.64</td>
<td>411.84</td>
<td>356</td>
<td>0.682</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Shortness of breath (mild)</td>
<td>1</td>
<td>34.86</td>
<td>34.86</td>
<td>79.01</td>
<td>356</td>
<td>0.743</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Shortness of breath (severe)</td>
<td>1</td>
<td>70.16</td>
<td>70.16</td>
<td>173.96</td>
<td>356</td>
<td>0.622</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Shortness of breath (mild)</td>
<td>7</td>
<td>85.72</td>
<td>12.25</td>
<td>142.07</td>
<td>356</td>
<td>0.743</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Shortness of breath (severe)</td>
<td>7</td>
<td>209.72</td>
<td>29.96</td>
<td>415.76</td>
<td>356</td>
<td>0.622</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Shortness of breath (mild)</td>
<td>90</td>
<td>233.72</td>
<td>2.60</td>
<td>483.36</td>
<td>356</td>
<td>0.743</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Shortness of breath (severe)</td>
<td>90</td>
<td>493.16</td>
<td>5.48</td>
<td>735.09</td>
<td>356</td>
<td>0.622</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Head congestion (mild)</td>
<td>1</td>
<td>19.84</td>
<td>19.84</td>
<td>39.53</td>
<td>356</td>
<td>0.756</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Head congestion (severe)</td>
<td>1</td>
<td>46.66</td>
<td>46.66</td>
<td>86.14</td>
<td>356</td>
<td>0.695</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Head congestion (mild)</td>
<td>7</td>
<td>41.90</td>
<td>5.99</td>
<td>64.21</td>
<td>356</td>
<td>0.756</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Head congestion (severe)</td>
<td>7</td>
<td>90.48</td>
<td>12.93</td>
<td>134.48</td>
<td>356</td>
<td>0.695</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Head congestion (mild)</td>
<td>90</td>
<td>115.12</td>
<td>1.28</td>
<td>163.50</td>
<td>356</td>
<td>0.756</td>
</tr>
<tr>
<td>Loehman et al. (1979)</td>
<td>Head congestion (severe)</td>
<td>90</td>
<td>317.02</td>
<td>3.52</td>
<td>554.16</td>
<td>356</td>
<td>0.695</td>
</tr>
</tbody>
</table>
Table 1 (continued)
WTP values for various health conditions

<table>
<thead>
<tr>
<th>Study</th>
<th>Health condition</th>
<th>Number of days</th>
<th>1993 $</th>
<th>WTP/Day</th>
<th>Standard Error</th>
<th>Sample Size</th>
<th>QWB (Kaplan)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rowe and Chestnut (1985)</td>
<td>Asthma day</td>
<td>9.5</td>
<td>578.33</td>
<td>60.88</td>
<td>122.57</td>
<td>65</td>
<td>0.683</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Throat congestion</td>
<td>1</td>
<td>40.04</td>
<td>40.04</td>
<td>4.41</td>
<td>176</td>
<td>0.830</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Throat congestion</td>
<td>30</td>
<td>285.05</td>
<td>9.50</td>
<td>29.68</td>
<td>176</td>
<td>0.830</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Drowsiness</td>
<td>1</td>
<td>43.41</td>
<td>43.41</td>
<td>4.76</td>
<td>176</td>
<td>0.680</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Drowsiness</td>
<td>30</td>
<td>439.45</td>
<td>14.65</td>
<td>61.77</td>
<td>176</td>
<td>0.680</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Angina (mild)</td>
<td>1</td>
<td>88.22</td>
<td>88.22</td>
<td>74.33</td>
<td>176</td>
<td>0.701</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Angina (mild)</td>
<td>10</td>
<td>206.07</td>
<td>20.61</td>
<td>220.56</td>
<td>176</td>
<td>0.701</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Angina (mild)</td>
<td>20</td>
<td>649.14</td>
<td>32.46</td>
<td>1,233.41</td>
<td>176</td>
<td>0.701</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Angina (severe)</td>
<td>1</td>
<td>164.99</td>
<td>164.99</td>
<td>135.68</td>
<td>176</td>
<td>0.580</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Angina (severe)</td>
<td>10</td>
<td>349.56</td>
<td>34.96</td>
<td>326.81</td>
<td>176</td>
<td>0.580</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Angina (severe)</td>
<td>20</td>
<td>1,127.25</td>
<td>56.36</td>
<td>2,149.27</td>
<td>176</td>
<td>0.580</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Cough</td>
<td>1</td>
<td>34.83</td>
<td>34.83</td>
<td>4.03</td>
<td>176</td>
<td>0.743</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Cough</td>
<td>30</td>
<td>230.11</td>
<td>7.67</td>
<td>23.99</td>
<td>176</td>
<td>0.743</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Nausea</td>
<td>1</td>
<td>69.49</td>
<td>69.49</td>
<td>10.67</td>
<td>176</td>
<td>0.649</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Nausea</td>
<td>30</td>
<td>257.08</td>
<td>8.57</td>
<td>26.67</td>
<td>176</td>
<td>0.649</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Headache</td>
<td>1</td>
<td>55.42</td>
<td>55.42</td>
<td>6.38</td>
<td>176</td>
<td>0.695</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Headache</td>
<td>30</td>
<td>674.69</td>
<td>22.49</td>
<td>86.80</td>
<td>176</td>
<td>0.695</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Eye irritation</td>
<td>1</td>
<td>38.32</td>
<td>38.32</td>
<td>3.44</td>
<td>176</td>
<td>0.770</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Eye irritation</td>
<td>30</td>
<td>325.50</td>
<td>10.85</td>
<td>47.81</td>
<td>176</td>
<td>0.770</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Sinus congestion</td>
<td>1</td>
<td>48.44</td>
<td>48.44</td>
<td>4.41</td>
<td>176</td>
<td>0.769</td>
</tr>
<tr>
<td>Tolley et al. (1986)</td>
<td>Sinus congestion</td>
<td>30</td>
<td>367.09</td>
<td>12.24</td>
<td>37.48</td>
<td>176</td>
<td>0.769</td>
</tr>
</tbody>
</table>

Notes: Each health condition was assigned a QWB score based on how the condition was described in its study. Note, then, that the same condition may have several different QWB scores in the table above. Both Loehman and Tolley arbitrarily assign the categories mild and severe to health conditions. Loehman’s original sample was trimmed by 5 percent at each end of the distribution to account for outliers and protest bids.

subject to income $Y$ and a price vector $P$ maps $U$ into a corresponding indirect utility function:

$$V = V(P, Y, H)$$

(3)

Holding prices and income constant, changes in utility are determined by changes in health status. Monetary equivalents of a welfare change induced by a health improvement $\Delta H = H' - H_o$, where $H'$ is an improved health state over $H_o$, can be measured by eliciting the maximum willingness to pay to obtain $+\Delta H$. In this case, WTP is a measure of compensating surplus, CS:

$$V(P, Y, H_o) = V(P, Y - CS, H')$$

(4)

As Eq. (4) indicates, an individual is indifferent between the change in health status (from $H_o$ to $H'$) and the change in income ($Y$ to $[Y - CS]$). Eq. (4) applies only to health improvements. However, three of the five included WTP studies
measure willingness to pay to avoid additional days of a condition. This is a measure of equivalent, rather than compensating, surplus for $\Delta H = H' - H_0$, where $H' < H_0$. Although we acknowledge that these measures differ conceptually, we cannot control for such biases across studies, given the large number of study-specific characteristics that vary across studies.

Over the range of health improvements we are valuing (which are relatively short-term conditions), we assume that the marginal utility of income is constant. Constant marginal utility of income implies that $\frac{\partial U}{\partial X}/\frac{\partial X}{\partial H} = 0$. This condition probably holds only for relatively mild symptoms or symptoms of short duration; therefore, we assume a constant marginal utility of income for this analysis. However, we note that large changes in health could affect individuals' marginal utility of income as well as their income level (if, for example, they were unable to work). In such a case, the changing marginal utility of income would make the relationship between health and income ambiguous, depending on the marginal utility of each, and health might have to be modeled with several state-dependent utility functions.

This framework has implications for the functional form of WTP in our analysis. WTP for the reduction or avoidance of a health condition should increase with the severity of the condition. Furthermore, WTP should be positively related to the duration of that reduction or avoidance. In our analysis, these variables are measured as QWB and DAYS, respectively. Furthermore, we expect to observe evidence of diminishing marginal utility associated with health improvements in WTP values. Specifically, WTP to avoid a health decrement should increase at an increasing rate as the conditions move further from perfect health (i.e., become more severe).

While theory can similarly shed some light on the second derivative of the DAYS variable, the exact expectation depends on how we interpret this variable. If DAYS measures reductions in currently experienced sick-days, diminishing marginal utility of health would imply that WTP should increase at a decreasing rate with DAYS. If DAYS measures additional sick-days to be avoided, the same principle would imply that WTP should increase at an increasing rate with DAYS. We have raised this concern above in our discussion of data concerns, and note again that such second-order conditions may be of less importance in the small range of health conditions explored here.

In addition to severity and duration, income and such other demographic information as age should influence the WTP to reduce or avoid a health condition. Furthermore, detailed information on the baseline health status of the CV respondents also might be helpful in determining WTP. Unfortunately,

---

26 In a recent paper, Hanemann (1996) finds differences between surplus measures for changes of quality and surplus measures for price changes. See also Randall and Stoll (1980) for related conceptual work.
consistent measures of such variables were not reported in the morbidity-valuation studies. While large differences in demographics and the baseline health of samples across studies potentially could affect results, the use of the panel estimators in the regressions (discussed later) should capture systematic differences across studies. The direction of the bias from the omission of demographic variables in our empirical model is indeterminate.

A final theoretic concern involves the fact that WTP for a change in duration should equal $0 when QWB equals 1, or perfect health. Although QWB scores equal to 1 are outside the range of our data, we expect the willingness to pay for very mild health conditions to be relatively small.

7. Empirical results

Fig. 1 shows a loess fit of the data for one-day health effects. Although the loess fit shows that WTP is relatively insensitive to QWB scores of mild, one-day conditions between 0.75 to 0.85, it clearly indicates that a strong inverse relationship exists for QWB scores of 0.75 and below. The results from several different OLS model specifications (not reported here) confirmed the results of the loess fit. The OLS models show a strong negative correlation between WTP and QWB score, and a strong positive correlation between WTP and DAYS. The coefficient on QWB was not sensitive to the functional form of the DAYS variable.

Given the relatively weak relationship between QWB and the WTP for conditions with QWB scores greater than 0.75, we investigated these data further by estimating the model on different subsamples of observations. For example, removing the six most severe observations from the data set (three Lochman

---

27 Typically, bid functions from CV studies regress WTP as a function of demographic variables, while including little information on the commodity itself. Conversely, our empirical model regresses WTP as a function of the characteristics of the commodity (i.e., health status), while necessarily having to omit demographic variables.

28 The direction of omitted-variable bias depends on whether omitted variables are positively or negatively correlated with included variables. Any such correlations with demographics could be complex and the net effects indeterminate.

29 Although theory indicates that WTP should equal $0 when QWB score is 1, it is inadvisable to artificially constrain the intercept to 0. This artificial constraint could change the model’s slope completely and largely affect the model’s estimates, resulting in an inaccurate representation of the data.

30 We performed the Belsley et al. (1980) outlier analysis to determine if any observations are exerting an undue influence on the regression. This method generally only eliminated only a few outliers, depending on the regression model. We recalculated the regressions excluding these outliers and found that the resulting models possessed the same predictive power and the same magnitude of the coefficients. Because the trimmed samples did not improve our models, we chose to use the full sample ($N = 53$) in our final regressions.
observations and three Tolley observations), we find that QWB score is significant in only one of three panel models for the double-log form. When we remove the three most severe observations, we find a significant relationship between WTP and QWB in all three models. Finally, when we remove approximately 10 percent of the mildest conditions and 10 percent of the most severe conditions from the data, we observe a significant relationship between WTP and QWB score in two of the three panel models.

Trimming the data in this way indicates potential sensitivity to particular observations. Nevertheless, the 53 observations constitute the complete population of available estimates meeting our criteria. Thus it is appropriate to include all the observations if the meta-analysis is to be interpreted as a synthesis of all available information.

Generally, we found that milder conditions were more sensitive to duration, as measured by DAYS, while more severe conditions were more sensitive to severity, as measured by QWB score. Because different conditions are more or less related to WTP at different points on the QWB scale in this analysis, the results admittedly are affected by which studies are included. These results might suggest the use of state-dependent utility functions; such functions would account for differing utilities for income, health, and other goods at different levels of health. Given the limited number of observations in this analysis, however, we modeled these data as a single utility function.

Although the OLS models generally are consistent with expectations, they do not account for study-specific correlations among observations. For example, the Loehman et al. (1979) study contributes 18 of the 53 observations to the data set. We expect that one study's observations may differ systematically from other studies' observations for a number of different reasons. We, therefore, investigated three types of panel estimators: a fixed-effects model (FEM), a random-effects
model (REM), and a separate-variances model (SVM). The FEM assumes that a constant term varies across studies, but the error term is common across all studies. The REM assumes the intercept is a random variable with a study-specific disturbance term. The SVM is a heteroscedastic model that estimates a separate variance for each study using maximum-likelihood estimation.

Table 2 reports the results of the panel models for various model specifications. The FEM and REM models for each model specification produce similar coefficients and t-statistics for the independent variables. Generally, the coefficients for DAYS and Ln(DAYS) in the SVM are comparable to those of the FEM and REM, but the coefficients for QWB are much different. In all models, the linear specification does not provide a very good fit of the data. However, the double-log specification produces large adjusted $R^2$ values for the FEM and REM indicating a relatively good fit. The likelihood values likewise indicate that the double-log specification provides a good fit in the SVM.

The variance estimates in the SVM merit some discussion. These parameters indicate that the estimated variance for the Rowe and Chestnut study is larger than the other studies. However, there is only one observation from the Rowe and Chestnut study, and for two of the three SVM specifications, the estimate of $\sigma^2$ is statistically insignificant. Conversely, the relatively small variance estimate for the Loehman study is quite significant in all three SVM specifications. In two of the three SVM specifications, the Loehman study has the lowest $\sigma^2$. There are several reasons why the Loehman study may have produced WTP estimates with less variability. First, the Loehman study — the oldest study used in our meta-analysis — elicited WTP for relatively similar conditions. Furthermore, this study asked each respondent to provide WTP values for a series of health conditions with little explanation of each condition. In addition, these values were elicited using a payment-card format. These factors may have contributed to the small variation in the Loehman estimates.

---

31 A Lagrange multiplier test indicated that individual error components exist in the data and Hausman's chi-square test indicated that the REM offers the better representation of intra-study correlations. These tests both produced strong results: the Lagrange multiplier test had a chi-square of 566.89 ($p = 0.00$), while the Hausman test had a chi-square of 0.19 ($p = 0.91$).

32 All five study constants are included in the FEM (see Table 2). Because the Rowe and Chestnut study contributes only one observation to the meta-analysis, inclusion of a constant for this study means that the model will predict this observation with perfect accuracy, which increases the adjusted $R^2$ for this model. Although this inflated $R^2$ affects the comparison of fits between the FEM and REM, we felt less comfortable with the alternative of arbitrarily combining the Rowe and Chestnut study with one of the other studies in our meta-analysis and thereby only having four study constants.

33 The payment-card format elicits WTP for a given commodity by asking respondents to choose one of several dollar amounts listed. This technique may bias WTP estimates by suggesting a certain range of values (Mitchell and Carson, 1989).
In addition to having a good statistical fit, the panel models meet theoretical expectations. In all models, WTP increases with both severity and duration of the illness. Moreover, the nonlinear models indicate that WTP increases at an increasing rate as severity increases. Alternatively, they indicate that people will pay less and less on the margin for improvements closer to perfect health, as implied by the diminishing marginal utility of health. The double-log models have the additional advantage of interacting QWB and DAYS, and the results indicate that people will pay more to decrease the duration of a given illness by one day when that one day is for a more severe illness. As noted above, we did not have clear prior expectations for the second derivative of WTP with respect to DAYS. The data indicate that people pay more for reductions in duration at a decreasing rate.
8. Selecting a model for valuation purposes

Selection of a specific model for policy evaluation purposes depends on both theoretical and empirical considerations. The double-log specification meets theoretical concerns discussed previously in the conceptual framework section, while also providing a good fit of the data (i.e., adjusted $R^2$ values around 0.70 for the FEM and REM). Fig. 2 shows the residuals of the double-log specification for each model.

The residual plots for all three panel models — the FEM, the REM, and the SVM — are fairly similar. However, although all three models provide a reasonably good fit of the data, the FEM and REM both often overpredict WTP, particularly for those conditions found at either end of the QWB range. Conversely, the SVM has relatively small residuals for mild and severe QWB scores, but larger residuals for intermediate QWB scores.

Table 3 illustrates these differences for a selection of health effects. The FEM predicts a relatively narrow distribution of WTP values across QWB scores. Although the REM predicts a wider WTP distribution, it predicts relatively high WTP values for relatively mild conditions, such as mild cough. Finally, the SVM predicts a relatively broad range of WTP values for different QWB scores. Specifically, a very mild condition, such as mild cough, is associated with a relatively low WTP, while a more severe condition, such as severe asthma, has a WTP value almost three times that of mild cough. A wide range of WTP values is consistent with our expectations. It is worth noting, however, that although the point estimates differ across the panel models, overlapping confidence intervals for these estimates indicate that several of the point estimates are not statistically significantly different from one another.

Meta-analysis does not free the analyst of the necessity of making judgments. We have strong theoretical reasons to think that WTP approaches $0$ as the QWB score approaches perfect health. However, conditions near perfect health are outside the range of our data. Nevertheless, we prefer models that are consistent with low WTP predictions for very mild health conditions. Furthermore, we prefer models that predict a relatively wide range of WTP estimates for mild versus severe conditions. We also prefer models that show a good fit of the data and do well at predicting WTP for both single- and multiple-day reductions in health.

---

34 To estimate WTP for the FEM, we used the weighted average of all study constants as the intercept term. Also, note that Fig. 2 actually shows the difference between the predicted and actual WTP values, instead of the difference between the predicted and actual Ln(WTP).

35 Fig. 2 also shows that all three models consistently underpredict WTP for certain conditions. In cases where residuals are $200$ or more, the panel models are typically underpredicting WTP for conditions lasting 20 days or more. Because we are most interested in predicting WTP for mild conditions of short duration, we are not concerned about these underpredictions.
conditions. Because the SVM meets all of these criteria, we selected this panel estimator as our final model.

We list our WTP estimates from the SVM along with estimates from Tolley et al. (1986), Loehman et al. (1979), and Dickie et al. (1987, 1988) in Table 4. Several of the estimates from Tolley, Loehman, and Dickie fall within the confidence intervals of our estimated values. For purposes of comparison, we also have included other estimates from Alberini et al. (1994), Tolley et al. (1994), and Dickie et al. (1987). Specifically, we include WTP for a day of cold from study of Alberini et al. While the estimates for avoiding a day of cold are in the range of our cough and shortness of breath values, we note that this comparison must be made cautiously, given that the Alberini values represent the preferences of Taiwanese respondents. These preferences could potentially be much different from those found in the US. The Tolley et al. (1994) estimates, which the authors describe as ‘state-of-the-art’, are based on values from four different approaches: COI, CV, the health-production approach, and the QALY approach. Clearly these estimates are higher than those predicted from our analysis. Finally, we include averting-behavior estimates from Dickie et al. (1987) in the last column of Table

Table 3
Predicted WTP for 1-day avoidance of mild cough and severe asthma (1993 $) for the double-log specification

<table>
<thead>
<tr>
<th>MODEL</th>
<th>MILD COUGH (QWB = .743)</th>
<th>SEVERE ASTHMA (QWB = .622)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fem</td>
<td>$33 ($25-$43)</td>
<td>$61 ($41-$91)</td>
</tr>
<tr>
<td>rem</td>
<td>$46 ($35-$59)</td>
<td>$87 ($58-$130)</td>
</tr>
<tr>
<td>svm</td>
<td>$24 ($19-$31)</td>
<td>$74 ($50-$109)</td>
</tr>
</tbody>
</table>

Note: 90-percent confidence intervals are listed in parentheses.
Table 4
WTP values for 1-day avoidance of various health conditions (1993 $)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough (mild)</td>
<td>$24 ($19-31)</td>
<td>$34.83</td>
<td>$12.11</td>
<td>$15.01</td>
<td></td>
<td></td>
<td>$4-$19</td>
</tr>
<tr>
<td>Cough (severe)</td>
<td>$43 ($32-$56)</td>
<td>-</td>
<td>$33.43</td>
<td>-</td>
<td></td>
<td></td>
<td>$27 (low estimate)</td>
</tr>
<tr>
<td>Headache (mild)</td>
<td>$22 ($17-$28)</td>
<td>-</td>
<td>-</td>
<td>$24.63</td>
<td>$24.63</td>
<td></td>
<td>$26 ²</td>
</tr>
<tr>
<td>Headache (severe)</td>
<td>$38 ($29-$49)</td>
<td>$55.42</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$154 (high estimate)</td>
<td></td>
</tr>
<tr>
<td>Shortness of breath (mild)</td>
<td>$24 ($19-$31)</td>
<td>-</td>
<td>$34.86</td>
<td>$9.05</td>
<td></td>
<td></td>
<td>$27-$138</td>
</tr>
<tr>
<td>Shortness of breath (severe)</td>
<td>$74 ($50-$109)</td>
<td>-</td>
<td>$70.16</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye irritation</td>
<td>$19 ($14-25)</td>
<td>$38.32</td>
<td>-</td>
<td>$19.86</td>
<td></td>
<td></td>
<td>$27-$138</td>
</tr>
<tr>
<td>Cold</td>
<td>$24-$45³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ This column reflects averting-behavior estimates, based on a sample of persons with normal respiratory function.
² Dickie et al. did not specify whether this averting-behavior estimate reflected a mild or severe headache.
³ These are WTP estimates for 1-day of cold, with an average of 2.2 symptoms and restricted activities.

Note: 90-percent confidence intervals are listed in parentheses.
These values are lower than our estimated WTP values. It is unclear whether this relationship is to be expected. According to Dickie et al. (1987), averting-behavior studies overestimate WTP. However, it certainly could be the case that the cost of a given product may not reflect the maximum WTP to avoid or reduce a condition. Furthermore, averting-behavior studies have other complications stemming from the discrete nature of averting-behavior choices.

For the purpose of predicting confidence intervals for WTP estimates from SVM, we include the variance/covariance matrix in Table 5.

9. An illustrative application

One potential application of this model is to use the results to estimate the economic damages associated with the health effects of air pollution. We illustrate such an application with epidemiological evidence from Schwartz et al. (1988) and with data from Desvousges et al. (1995). According to these studies, a 5-percent decrease in average carbon monoxide concentrations in Hennepin County, Min-

There are a few other averting-behavior studies of health available, but unfortunately are not appropriate for comparison purposes here. For example, Cropper (1981) makes assumptions to simplify WTP estimation and subsequently shows that the averting-behavior costs equal COI, which are assumed to be directed related to earnings. Clearly this study measures values for health in a completely different way than is reflected in our valuation estimates. Gerking and Stanley (1986) measure the value of the reduction in ambient levels of certain air pollutants. However, they do not directly estimate WTP for health resulting from the reduction, which would provide an appropriate comparison. Finally, Dickie and Gerking (1991) use a household production approach to measure daily WTP to relieve respiratory symptoms and reduce air pollution. Unfortunately, the study calculates WTP for the group of symptoms as a whole. Therefore we cannot obtain WTP results from that study to compare with our results. Generally, however, values from averting behaviors were lower than our estimates.

Minnesota (Minneapolis) would result in about 48,000 fewer cases of headache per year, among other potential benefits. Assuming these generally are mild cases, our analysis would value the benefit of this reduction at about $1,056,000 annually, with a 90-percent confidence interval between $816,000 and $1,344,000.

10. Conclusions and implications

We have estimated a general valuation function combining a meta-analysis of morbidity-valuation studies and the QWB health-status index. The index provides a means of estimating a general valuation function for any short-term health condition that can be assigned a QWB score, regardless of whether the specific condition has been valued previously in an empirical study. The results illustrate some advantages of meta-analysis in contrast with conventional literature reviews, especially for constructing consensus on economic values for benefit-cost applications. Instead of choosing WTP estimates from a single ‘best study’, this approach allows us to integrate results from most of the available valuation studies and develop point estimates and confidence intervals for a range of health conditions.

Although the available valuation studies are subject to criticism, this meta-analysis nevertheless shows that the WTP estimates appear to be logically and statistically consistent. As theory would indicate, WTP for the reduction or avoidance of a health condition increases at an increasing rate as the health status worsens and increases as duration lengthens. Furthermore, WTP increases more for a given duration change at more severe health states. Our recommended heteroscedastic model, the SVM, presents a good fit to the available data and allows us to establish WTP for a number of health conditions with more confidence than we could using only a single ‘best study’.

Although this analysis provides a statistical synthesis of the evidence from the existing morbidity literature, there is a need for future research in several areas. Foremost in priority is the need for additional high-quality morbidity-valuation studies. Considering the fact that morbidity valuation is a potentially important component of public-policy analysis, there are relatively few morbidity-valuation studies that could pass current experimental-design and econometric standards. Newer techniques such as conjoint analysis could also be used to estimate hedonic values for health attributes. We do not consider this research to be an alternative to new empirical estimates; rather, it provides a tool for making the most of the ones currently available.

Acknowledgements

We would like to thank Bill Desvousges and Kerry Smith for their helpful comments on an earlier draft.
References


