Hypothetical versus real willingness to pay in the health care sector: results from a field experiment

Karen Blumenschein, Magnus Johannesson, Krista K. Yokoyama, Patricia R. Freeman

Abstract

We conducted a field experiment comparing hypothetical and real purchase decisions for a pharmacist provided asthma management program among 172 subjects with asthma. Subjects received either a dichotomous choice contingent valuation question or were given the opportunity to actually enroll in the program. Three different prices were used: US$ 15, 40, and 80. In the hypothetical group, 38% of subjects said that they would purchase the good at the stated price, but only 12% of subjects in the real group purchased the good (p = 0.000). We cannot, however, reject the null hypothesis that “definitely sure” hypothetical yes responses, as identified in a follow-up question, correspond to real yes responses. We conclude that the dichotomous choice contingent valuation method overestimates willingness to pay, but that it may be possible to correct for this overestimation by sorting out “definitely sure” yes responses. © 2001 Elsevier Science B.V. All rights reserved.

JEL classification: C93; D61; I10; Q26

Keywords: Willingness to pay; Contingent valuation; Field experiments; Calibration

1. Introduction

In the contingent valuation method, individuals are asked about their hypothetical willingness to pay for a defined good (Mitchell and Carson, 1989). In the most commonly used elicitation approach, the dichotomous (yes/no) choice approach, individuals accept or reject
only one price and opportunity to purchase the defined good. By varying the price in different subsamples of respondents, it is possible to derive the aggregate demand curve (Bishop and Heberlein, 1979). The hypothetical dichotomous choice question is often framed in terms of a vote on a referendum (Mitchell and Carson, 1989). The contingent valuation method was developed in environmental economics and has been widely used to estimate the willingness to pay for environmental changes (Hanemann, 1994). The method is also increasingly used in the health care field to estimate the willingness to pay for health care programs as a basis for cost-benefit analyses of health care programs (Diener et al., 1998).

Despite its widespread use, the contingent valuation method is highly controversial among economists (Diamond and Hausman, 1994; Hanemann, 1994; Portney, 1994). The nucleus of the controversy is the extent to which hypothetical choices in the contingent valuation method correspond to real economic choices. The extent to which hypothetical choices mimic real choices is an old controversy in economics. Wallis and Friedman (1942) criticized the use of hypothetical choices in experiments over 50 years ago, and this topic has been debated ever since (Kagel and Roth, 1995; Thaler, 1987).

The correspondence between hypothetical and real willingness to pay has been studied in a number of laboratory experiments. Generally, these experiments have focused on the dichotomous choice approach and have compared the fraction of hypothetical and real yes responses (i.e. acceptances to buy the good at the given price) between experimental groups. In most of these experiments, the proportion of hypothetical yes responses has significantly exceeded the proportion of real yes responses (Cummings et al., 1995, 1997; Loomis et al., 1997; Johannesson et al., 1998; Blumenschein et al., 1998; Cummings and Taylor, 1999).

The results of these experiments suggest that the dichotomous choice contingent valuation approach is associated with a general overestimation problem, which seriously undermines the usefulness of the method. This overestimation problem, often referred to as hypothetical bias, has motivated research into various calibration approaches for removing the bias. Cummings and Taylor (1999) recently tested a so-called “cheap talk” design for removing hypothetical bias. This approach includes an explicit discussion of the hypothetical bias problem in the contingent valuation question, to induce respondents to eliminate hypothetical bias from their answers. In a set of laboratory experiments involving four different public goods there was no significant difference between the proportion of real yes responses and the proportion of hypothetical yes responses with a cheap talk design (Cummings and Taylor, 1999).

An alternative approach to remove hypothetical bias that has yielded promising results is to use the degree of certainty in the hypothetical yes responses to sort out “false” yes responses. In two experiments Johannesson et al. (1998) and Blumenschein et al. (1998) tested the hypothesis that “definitely sure” yes responses, identified in a follow-up question, correspond to real yes responses. In the Johannesson et al. (1998) experiment, the “definitely sure” yes responses significantly underestimated the real yes responses whereas in the Blumenschein et al. (1998) experiment, the null hypothesis of no difference between “definitely sure” yes responses and real yes responses could not be rejected. In both of these experiments, data were also collected about the degree of certainty of the hypothetical yes responses.

---

1 Currently, the most common approach to economic evaluation in the health care field is cost-effectiveness analysis rather than cost-benefit analysis. See Johannesson (1996) for the conceptual differences between cost-effectiveness analysis and cost-benefit analysis.
responses on a scale between 0 (very unsure) and 10 (very sure). Utilizing the within-sample
data from these experiments (i.e. data where the subjects responded to a hypothetical valuation question followed by a real valuation question), Johannesson et al. (1999) estimated a calibration function based on the degree of the individual’s self-assessed certainty (on the 0–10 scale) and a variable representing the price level. This calibration function removed the hypothetical bias from both experiments (Johannesson et al., 1999).

A weakness of the experimental work carried out so far on hypothetical bias and different
calibration approaches is that the work has been carried out under rather artificial conditions
in the laboratory using mainly college students as subjects. Field experiments, conducted
in the ‘real world’ are a crucial next step in this area of research. In this paper, we report the
results of a field experiment of hypothetical versus real willingness to pay for a health care
good. The good valued in this experiment is a pharmacist provided asthma management
program. We examine whether dichotomous choice contingent valuation questions lead to
hypothetical bias for this good. In addition, we test if “definitely sure” hypothetical yes
responses correspond to real yes responses and if calibrated hypothetical yes responses
correspond to real yes responses (using the calibration function estimated by Johannesson
et al. (1999)). We find that the hypothetical yes responses significantly overestimate the
real yes responses (38 versus 12%), but that there is no significant difference between the
“definitely sure” hypothetical yes responses and the real yes responses (14 versus 12%).
Our results suggest that it may be possible to eliminate hypothetical bias in dichotomous
choice contingent valuation studies of health care by adding a simple follow-up question
about the certainty of the hypothetical response. The design of the experiment is described
below followed by the results. The paper ends with a discussion section.

2. Experimental design

The experiment involved two different experimental groups referred to as the “hypotheti-
cal group” and the “real group”. Subjects in the hypothetical group received a dichotomous
choice contingent valuation question about purchasing a pharmacist provided asthma man-
agement program, whereas subjects in the real group were given the opportunity to actually
purchase the program. Three different prices were used: US$ 15, 40, and 80. A pilot study
was carried out to determine the prices. The study had approval from the University of Ken-
tucky Medical Institutional Review Board. Details of the experiment are further described
below.

2.1. The good

The good used in our experiment was a pharmacist provided asthma management pro-
gram. This program was designed to assist asthma patients in attaining optimal management

2 An exception to this is the early field experiments reported in Bishop and Heberlein (1979) and Heberlein and
Bishop (1986) on the willingness to pay and willingness to accept for goose hunting and deer hunting permits.
These studies indicated that both hypothetical willingness to accept and hypothetical willingness to pay questions
were associated with an overestimation problem, although the results were much more pronounced for willingness
to accept.
of their asthma, thereby enhancing the patient’s life quality and decreasing their utilization of expensive health care services such as emergency department visits for asthma exacerbations. In the United States, the traditional pharmacist role of medication dispenser has evolved to include the provision of “cognitive services” such as disease state management. Cognitive services are “those services provided by a pharmacist to or for a patient or health care provider that are either judgmental or educational in nature” (Christensen et al., 1993). Currently, few pharmacists perform cognitive services in community settings, although the numbers are increasing (Carter et al., 1997; Smith et al., 1999). Leaders within the pharmacy profession feel that cognitive services represent the future direction of community pharmacy practice (Christensen and Hansen, 1999).

Pharmacist provided cognitive services, such as disease state management, have traditionally not been paid for by patients or included as benefits in health care plans. In the current competitive marketplace, some pharmacists are working to establish these services in order to create new revenue streams. A number of pharmacists across the US have developed asthma disease management programs (Pauley et al., 1995; Kelso et al., 1995; Smith et al., 1999), however, very few actually offer their services for sale to the general public. Some are offered only to Medicaid patients while others rely on grants for funding. The pharmacists participating in this study had never offered the asthma management service for purchase to the general public prior to this evaluation.

2.2. Subject recruitment

Subjects were recruited from 10 pharmacies in the state of Kentucky in the US. Five of these pharmacies were included in the real group and five pharmacies were included in the hypothetical group. The pharmacies in the real group agreed to deliver an asthma management program to any subject that purchased the good. All of the pharmacists in the real group had received extensive training on providing the asthma management program prior to the implementation of this study. The training was provided by American Pharmacy Services Corporation.

The pharmacists identified asthma patients at their site using the following algorithm: the computer system at the pharmacy was utilized to generate a list of patients ages 18 and older with asthma by running a report of the disease state field for asthma. Then, the computer system generated a second list of patients ages 18 and older who had received a prescription for an asthma medication in the past 6 months. Patients who were common to each of the above lists were considered candidates for subject recruitment. Potential subjects were contacted by phone. After confirming the asthma diagnosis, they were asked if they would be interested in participating in a scientific study that involved an interview of approximately 15–20 min. Subjects that agreed to participate in the study were given a mutually convenient appointment time for the interview, which was carried out at the pharmacy. To compensate for their time, each subject received US$ 25 upon completion of the survey. A trained interviewer conducted the interview sessions and all of the interviews took place between 1 October 1999 and 19 November 1999. In total, 173 asthma patients were interviewed (84 in the hypothetical group and 89 in the real group).
2.3. The questionnaires

In both groups, a written description of the pharmacist provided asthma management service was given to the subject (the description of the asthma management program is given in the Appendix). The interviewer read the description to the subject and responded to any questions, the subject had regarding the service. Then, the interviewer gave the subject, a written copy of the survey. The interviewer read the contingent valuation question to the subject and the subject marked his/her response on the survey form. In the hypothetical group, the interviewer also read the certainty questions to the subject, and again, the subject marked his/her response on the survey form.3

2.3.1. Real group

Subjects in the real group were given the opportunity to enroll in the asthma management program at their pharmacy. After receiving a description of the asthma management program, the following question was posed:

“You are being offered to purchase the asthma management service that was just described to you. If you choose to purchase the service, you will have to pay here and now with cash, check or credit card. Will you buy this service here and now at a price of US$ 15? Please circle your answer below”.

The price was varied between US$ 15, 40 and 80, but only one price was used at each pharmacy. The reason that we did not vary the price within a pharmacy was due to the fact that the pharmacists were unwilling to charge patients at the same pharmacy different prices. The low and the middle price was used at two pharmacies each whereas the high price was only used at one pharmacy.

2.3.2. Hypothetical group

Subjects in the hypothetical group received a hypothetical dichotomous choice contingent valuation question. The question consisted of a description of the asthma management program after which the following question was posed:

“Assume that you are being offered to purchase the asthma disease management service that was just described to you. Assume that if you choose to purchase the service, you will have to pay here and now with cash, check or credit card. Would you buy this service here and now at a price of US$ 15? Please circle your answer below”.

The price was randomly varied between US$ 15, 40 and 80 among subjects. To get the same proportion of subjects at each price as in the real group, 40% of subjects received US$ 15, 40% received US$ 40 and 20% received US$ 80. The dichotomous choice contingent valuation question was followed by a question in which the subjects who answered yes were asked if they were “probably sure” or “definitely sure” of their yes answer. This question was phrased in the following way:

3 All data and questionnaires are available from the authors upon request.
Are you probably sure or definitely sure that you would buy the asthma management service here and now if the price for the service was US$ 15? Please circle your answer below.

This question was followed by an additional question where the certainty of the hypothetical yes response was assessed on a visual analog scale between 0 (very unsure) and 10 (very sure). This question was phrased in the following way:

Mark with an “x” on the line below how sure you are that you would buy the service here and now at a price of US$ 15.

2.4. Statistical analysis

The data collected in the study was used to carry out three comparisons. First, the proportion of yes responses in the hypothetical group was compared with the proportion of yes responses in the real group. This tests for the existence of hypothetical bias. Second, the proportion of “definitely sure” yes responses in the hypothetical group was compared with the proportion of yes responses in the real group. This tests if a follow-up question that divides yes responses into “definitely sure” and “probably sure” yes responses can be used to calibrate hypothetical dichotomous choice responses. Finally, the proportion of hypothetical yes responses calibrated according to the calibration function estimated by Johannesson et al. (1999) was compared with the proportion of yes responses in the real group. This tests if a follow-up question about the degree of certainty of the yes response on a 0 (very unsure) to 10 (very sure) scale can be used to calibrate hypothetical dichotomous choice responses.

Both non-parametric and parametric methods were used to compare the proportion of yes responses between the groups. A contingency table Pearson chi-square test was used to compare the proportion of yes responses between the groups testing the null hypothesis of no difference (D’Agostino et al., 1988). The drawback of the non-parametric tests is that they do not control for any background variables that may differ between the groups. Therefore, we also evaluated our hypotheses using regression analysis controlling for the price level and a host of socio-economic and disease variables collected in the study. We tested if a hypothetical group/real group dummy variable was statistically significant in a probit regression equation (Greene, 1997). The price of the asthma management program was also included in the regression analysis to control for the price in the comparison of hypothetical and real responses. Including the price variable also provides a validity test of the study, since it tests if the demand curve is downward sloping (Smith, 1994). The following socio-economic variables were included in the regression equation: gender (1 = female), age, years of education, and personal income (US$ per year in pre-tax

---

In the calibration function estimated by Johannesson et al. (1999), the following probit regression equation is used to predict whether a hypothetical yes response is a real yes response or not: \[-5.81 + 0.62 \text{(value on the certainty scale)} + 2.28 \text{(the proportion of hypothetical yes responses at that price)}\]. This equation can be used to estimate the cut-off value on the certainty scale where a real yes response will be predicted. The cut-off is equal to: \[(5.81 - 2.28, \text{(the proportion of hypothetical yes responses at that price)})/0.62\].
Table 1
Background characteristics

<table>
<thead>
<tr>
<th></th>
<th>Hypothetical group</th>
<th>Real group</th>
<th>p-Value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Age</td>
<td>51.86</td>
<td>16.25</td>
<td>48.27</td>
</tr>
<tr>
<td>Proportion female</td>
<td>0.76</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>Education (years)</td>
<td>11.10</td>
<td>3.76</td>
<td>11.92</td>
</tr>
<tr>
<td>Personal income (US$ per year)</td>
<td>15417</td>
<td>22431</td>
<td>16966</td>
</tr>
<tr>
<td>Years with asthma</td>
<td>19.59</td>
<td>19.39</td>
<td>18.05</td>
</tr>
<tr>
<td>Asthma severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion with mild asthma</td>
<td>0.15</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Proportion with moderate asthma</td>
<td>0.54</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Proportion with severe asthma</td>
<td>0.31</td>
<td>0.37</td>
<td></td>
</tr>
</tbody>
</table>

We also included two variables for asthma: years with asthma and self-perceived asthma severity (divided into mild asthma, moderate asthma and severe asthma). These variables were included to control for any differences in the subject populations between the hypothetical group and the real group. We also tested for structural differences between the experimental groups. This was done by adding interaction terms between the hypothetical group dummy variable and all other variables. A likelihood ratio test was then carried out to test if this unrestricted model differed from the restricted model without the hypothetical group dummy variable and the interaction terms.

3. Results

3.1. Background characteristics

Table 1 provides the background characteristics for the two samples. On average, subjects were somewhat younger in the real group and the proportion of women was higher. The average income and education was also slightly higher in the real group. However, none of the differences in background characteristics between the groups were statistically significant, with the exception of the gender difference that was significant at the 10% level ($p = 0.080$).  

5 A categorical question was used to assess the annual personal income for the subjects. The following income categories were used: <US$10,000, 10,000–20,000, 20,001–30,000, 30,001–50,000, 50,001–100,000, >100,000. Based on the responses to this question, a continuous income measure was constructed by setting the income for each subject to the midpoint of the interval (US$125,000 was used as the midpoint for the highest income category).

6 All continuous variables were entered without any transformation in the probit regression. We also estimated all results with a logarithmic transformation of all continuous variables, however, this did not affect any of the conclusions reported below. The McFadden pseudo-$R^2$ for the estimated probit equations was slightly lower with a logarithmic transformation.

7 For continuous variables, a two-sided $t$-test was used to test for statistical differences between the groups. For the categorical variables, a contingency table Pearson chi-square test was used.
3.2. Hypothetical bias

Table 2 shows the proportion of yes responses in the experimental groups. In the hypothetical group, 59% said they would purchase the asthma management program at a price of US$ 15. At the US$ 40 and 80 prices, the proportion of hypothetical yes responses was 29% and 17%, respectively. The proportion of subjects that actually bought the asthma management program was substantially lower. At the low price (US$ 15) 16% bought the good and at the US$ 40 price 14% bought the good. At the price of US$ 80, no one bought the good. According to the non-parametric test, the null hypothesis of no difference in the proportion of yes responses between the hypothetical and the real group could be rejected at the US$ 15 price ($p = 0.000$), and for all prices combined ($p = 0.000$). At the US$ 80 price, the null hypothesis could be rejected at the 10% level ($p = 0.087$), whereas at the US$ 40 price the difference between the groups was not statistically significant at the 10% level.

We also estimated the mean willingness to pay in each experimental group using the non-parametric method developed by Kriström (1990). The mean willingness to pay was estimated to be US$ 29.23 in the hypothetical group and US$ 8.97 in the real group.\(^8\)

In the parametric tests we controlled for the price level and the background characteristics. The estimated probit regression equation is shown in Table 3. The hypothetical group dummy variable was highly significant rejecting the null hypothesis ($p = 0.000$). The price variable was also highly significant in accordance with a downward sloping demand curve. None of the background variables were statistically significant at the 10% level.\(^9\) The likelihood ratio test statistic for the test of structural differences was 20.352 (9 d.f.). This test was significant at the 5% level, consistent with the above tests.

\(^8\) In the estimation of mean willingness to pay, it was assumed that the maximum willingness to pay was equal to the highest price (US$ 80) used in the study, and that the proportion of subjects with zero willingness to pay was equal to the proportion of ‘no’ responses at the lowest price used in the study (US$ 15).

\(^9\) We also tested adding an interaction term between the hypothetical group dummy variable and the price variable allowing the slope of the aggregate demand functions to vary between the experimental groups. It has been argued that the slopes will differ for hypothetical and real willingness to pay data, due to a higher variance in hypothetical willingness to pay (Haab et al., 1999). The interaction term was, however, not significant (coefficient = $-0.003$, $t$-value = $-0.281$) and we could not reject the null hypothesis of equal slopes between the groups.
Table 3
Results of probit estimations of the effect of the experimental group on the probability of a yes response

<table>
<thead>
<tr>
<th>Definition of yes responses in the hypothetical group</th>
<th>All yes responses</th>
<th>“Definitely sure” yes responses</th>
<th>Calibrated yes responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t-Value</td>
<td>p-Value</td>
</tr>
<tr>
<td>Constant</td>
<td>−0.311</td>
<td>−0.404</td>
<td>0.686</td>
</tr>
<tr>
<td>Hypothetical group</td>
<td>0.922</td>
<td>3.841</td>
<td>0.000</td>
</tr>
<tr>
<td>Price</td>
<td>−0.019</td>
<td>−3.301</td>
<td>0.001</td>
</tr>
<tr>
<td>Age</td>
<td>−0.0005</td>
<td>−0.063</td>
<td>0.950</td>
</tr>
<tr>
<td>Female</td>
<td>0.032</td>
<td>0.111</td>
<td>0.911</td>
</tr>
<tr>
<td>Education</td>
<td>−0.035</td>
<td>−0.904</td>
<td>0.366</td>
</tr>
<tr>
<td>Personal income</td>
<td>0.000004</td>
<td>0.745</td>
<td>0.456</td>
</tr>
<tr>
<td>Years with asthma</td>
<td>−0.001</td>
<td>−0.141</td>
<td>0.888</td>
</tr>
<tr>
<td>Moderate asthma</td>
<td>0.155</td>
<td>0.489</td>
<td>0.625</td>
</tr>
<tr>
<td>Severe asthma</td>
<td>0.112</td>
<td>0.320</td>
<td>0.749</td>
</tr>
</tbody>
</table>

\( \alpha \) 173 173 173
Chi-square (p-value) 30.024 (0.000) 16.350 (0.060) 19.944 (0.018)
Log-likelihood −81.997 −59.633 −68.243
McFadden pseudo-\( R^2 \) 0.155 0.121 0.127
Individual prediction (%) 79.191 86.705 84.971
3.3. Calibration of hypothetical yes responses

The first calibration method tested was that “definitely sure” hypothetical yes responses corresponded to real yes responses. A proportion of “definitely sure” yes responses in the hypothetical group is shown in Table 2. The proportion of “definitely sure” yes responses was 28% at the US$ 15 price, 9% at the US$ 40 price and 0% at the US$ 80 price. The overall proportion of “definitely sure” yes responses was 14%, which was very close to the 12% yes responses in the real group. According to the non-parametric test, the null hypothesis of no difference in the proportion of “definitely sure” hypothetical yes responses and the proportion of real yes responses could not be rejected at the 10% level at any of the prices or for all the prices combined. The mean willingness to pay based on the “definitely sure” yes responses was estimated to be US$ 10.59, which was close to the mean willingness to pay in the real group.

The estimated probit regression equation for the comparison of “definitely sure” hypothetical yes responses and real yes responses is shown in Table 3. The hypothetical group dummy variable was not significant at the 10% level (p = 0.925). The price variable was highly significant in accordance with a downward sloping demand curve. None of the background variables were statistically significant at the 10% level. The likelihood ratio test statistic for the test of structural differences was 9.213 (9 d.f.). This test was not significant at the 10% level, consistent with the above tests. Thus, we cannot reject the null hypothesis of no difference between “definitely sure” yes responses and real yes responses in any of the tests.

The second calibration method tested was to use the calibration function estimated by Johannesson et al. (1999) based on the scale assessing the degree of self-expressed certainty in the hypothetical yes response. The proportion of calibrated yes responses in the hypothetical group is shown in Table 2. The proportion of calibrated yes responses was 41% at the US$ 15 price, 9% at the US$ 40 price and 11% at the US$ 80 price. The overall proportion of calibrated yes responses was 21%, compared to the 12% yes in the real group. According to the non-parametric test the null hypothesis of no difference in the proportion of calibrated hypothetical yes responses and the proportion of real yes responses could not be rejected at the 10% level at the US$ 40 price, the US$ 80 price or for all the prices combined. At the US$ 15 price the null hypothesis could, however, be rejected at the 5% level (p = 0.024). The mean willingness to pay based on the calibrated yes responses was estimated to US$ 16.20, compared to US$ 8.97 in the real group.

The estimated probit regression equation for the comparison of calibrated yes responses and real yes responses is shown in Table 3. The hypothetical group dummy variable was not significant at the 10% level (p = 0.180), and the null hypothesis could thus not be

---

10 We also tested adding an interaction term between the hypothetical group dummy variable and the price variable allowing the slope of the aggregate demand functions to vary between the experimental groups (Haab et al., 1999). The interaction term was, however, not significant (coefficient = −0.018, t-value = −1.012) and we could not reject the null hypothesis of equal slopes between the groups.

11 The calibration function predicts real yes responses if the degree of certainty on the 0 (very unsure) to 10 (very sure) scale exceeds 7.05 at the US$ 15 price, 7.99 at the US$ 40 price and 8.39 at the US$ 80 price.
rejected. As in the previous regression equations, the price variable was highly significant in accordance with a downward sloping demand curve. The education variable was significant at the 10% level with a negative sign, indicating that a higher education decreases the demand for the asthma management program. None of the other background variables were statistically significant at the 10% level. The likelihood ratio test statistic for the test of structural differences was 9.219 (9 d.f.). This test was not significant at the 10% level. Thus, we cannot reject the null hypothesis of no difference between calibrated yes responses and real yes responses in the parametric tests.

4. Discussion

It is becoming increasingly common to carry out contingent valuation studies within the health care sector. The ability of the method to correctly estimate the willingness to pay for a health care program is, however, unclear. We carried out a field experiment directly comparing responses to a dichotomous choice contingent valuation question with real purchase decisions. As noted in the introduction to this paper, several laboratory experiments have shown that dichotomous choice contingent valuation questions overestimate willingness to pay (Cummings et al., 1995, 1997; Loomis et al., 1997; Johannesson et al., 1998; Blumenschein et al., 1998; Cummings and Taylor, 1999).

These laboratory experiments have several potential limitations when it comes to extrapolating the results to contingent valuation studies of health care. They have been concerned with common private goods (e.g. a box of chocolates or a calculator) or environmental goods, and the validity of the method may differ between goods. The laboratory experiments have been carried out under rather artificial conditions involving mainly college students. Furthermore, the “stakes” involved have been small, varying between US$ 1 and 10.

Our field experiment of a pharmacist provided asthma management program therefore provides new important evidence compared to the previous laboratory experiments. According to our results, the proportion of hypothetical yes responses significantly overestimated the proportion of real yes responses. The overall proportion of hypothetical yes responses (38%) was more than three times greater than the proportion of real yes responses (12%). This result is consistent with the results of the laboratory experiments, and suggests that the dichotomous choice contingent valuation approach is associated with a general overestimation problem that is robust across goods and variations in the experimental design. This result is also consistent with results showing that dichotomous choice questions lead

---

12 We also ran the probit regression equation for just the US$ 15 price to test if the difference at this price, which was significant according to the non-parametric test in Table 2, would still be significant after controlling for the background characteristics. In this regression equation, the hypothetical dummy variable was significant at the 10% level (coefficient = 0.677, t-value = 1.835), rather than at the 5% level as in the non-parametric test.

13 We also tested adding an interaction term between the hypothetical group dummy variable and the price variable allowing the slope of the aggregate demand functions to vary between the experimental groups (Haab et al., 1999). The interaction term was, however, not significant (coefficient = −0.003, t-value = −0.231) and we could not reject the null hypothesis of equal slopes between the groups.

14 An exception to this is the experiment by Loomis et al. (1997) where the price varied between US$ 2 and 120.
to overestimation even for objective estimation tasks, e.g. the height of the tallest redwood in California (Green et al., 1998).

We also tested two ways of calibrating dichotomous choice contingent valuation questions. Both approaches are based on using the degree of certainty in responses to calibrate the responses. In the first approach only “definitely sure” hypothetical yes responses, identified in a follow-up question, were counted as yes responses. The overall proportion of “definitely sure” yes responses was 14%, which was very close to the proportion of real yes responses of 12%. We could not reject the null hypothesis of no difference between “definitely sure” yes responses and real yes responses for any of the statistical tests. This result is consistent with the findings by Blumenschein et al. (1998), where a close correspondence between “definitely sure” hypothetical yes responses and real yes responses was identified in a laboratory experiment involving a pair of sunglasses. In another laboratory experiment involving a box of chocolates, Johannesson et al. (1998) found that the proportion of “definitely sure” yes responses significantly underestimated the proportion of real yes responses. That study was, however, carried out in another country (Sweden) and language, and the follow-up question to distinguish between “probably sure” and “definitely sure” yes responses may have been interpreted in different ways in the two countries.¹⁵

The second calibration approach that we tested used the calibration function estimated by Johannesson et al. (1999) to calibrate the yes responses in our experiment. That calibration approach differs from the approach described above in that the degree of certainty of hypothetical yes responses is measured on a continuous scale, and that it also takes into account the price level. The overall proportion of calibrated yes responses was 21%, which is higher than the proportion of real yes responses of 12%. We could not statistically reject the null hypothesis of no difference between calibrated yes responses and real yes responses ($p = 0.180$ in the probit equation controlling for the price level and the background characteristics). If the results for the US$ 15 price were analyzed separately there was, however, a significant difference between calibrated yes responses and real yes responses. The statistical analysis could not clearly differentiate between the two calibration approaches, but in terms of magnitudes, the calibration function worked much less well than the simple “probably sure/definitely sure” distinction.

Our results can be compared to the results of the study by Champ et al. (1997), which compared hypothetical dichotomous choice questions about donating a specified amount to a public good with actual donations to the public good. They also assessed the certainty of the hypothetical donation responses on a 1–10 scale from very uncertain to very certain. They found that hypothetical donations significantly exceeded real donations, but that there was no significant difference if only subjects that were very certain of their yes response (i.e. 10 on the scale) were counted as real yes responses. Thus, the results are similar to our findings, however, there are some important differences between the studies that limit their comparability. We compared real and hypothetical questions about willingness to pay for a private

---

¹⁵ The wording in the follow-up question to distinguish between “probably sure” and “definitely sure” responses also differed slightly in the study by Johannesson et al. (1998). Rather than using the terms “probably” and “definitely” the Swedish terms “ganska” and “helt” were used. According to the dictionary “ganska” is translated to: fairly, rather, quite, pretty and “helt” is translated to: entirely, completely, absolutely, totally, wholly, altogether.
good. Champ et al. (1997) on the other hand compared real and hypothetical questions about voluntary donations to a public good, where real donations were interpreted as a lower bound on the willingness to pay for the public good. It is possible that the calibration of hypothetical questions differs between voluntary donations and willingness to pay, and between private and public goods. We nevertheless tested the same calibration as used by Champ et al. (1997) on our experimental data, i.e. only hypothetical yes responses with 10 on the certainty scale were counted as real yes responses. This led to 13% calibrated hypothetical yes responses, which did not differ significantly from the proportion of real yes responses of 12%.

Other approaches to calibrating dichotomous choice contingent valuation questions have been suggested in the literature (Blackburn et al., 1994; Cummings and Taylor, 1999). One approach that has yielded promising results is the so-called “cheap talk” design for removing hypothetical bias tested by Cummings and Taylor (1999). With this approach, an explicit discussion of the hypothetical bias problem is included in the contingent valuation question, to induce respondents to eliminate hypothetical bias from their answers. An advantage of the calibration approaches tested in our study compared to the “cheap talk” design is that they are easier to incorporate in contingent valuation surveys (since they only involve a short follow-up question about the certainty of responses). Cummings and Taylor (1999) used a rather long script to explicitly discuss hypothetical bias, and the authors noted that this may be infeasible for telephone applications of the contingent valuation method. Further work is needed to test if the same results can be achieved using a shorter “cheap talk” script. It would also be very interesting to carry out experimental studies to compare the “cheap talk” design with “definitely sure” yes responses and real yes responses to test which, if any, approach is to be preferred for calibration purposes. It is also possible of course that the same calibration cannot be used for different goods, e.g. environmental goods and health care goods. However, as noted above, the general pattern of overestimation of willingness to pay with the dichotomous choice contingent valuation approach has so far been robust across goods and variations in the experimental design.

A potential concern with our findings is that the correspondence between “definitely sure” yes responses and real yes responses does not validate our calibration method in general, but is just a coincidence. Only counting “definitely sure” yes responses as yes responses will clearly lower the incidence of yes answers and by coincidence it may fall so much that it equals the proportion of real purchases in the real group. To address this concern it is necessary to establish a link between the degree of certainty in hypothetical yes responses and real yes responses. This can be studied if the same individuals first respond to a hypothetical contingent valuation question followed by a certainty question and then face a real purchase decision (i.e. a within subject design). This has been done in previous experiments which have shown that the degree of certainty in hypothetical yes responses is a very strong predictor of whether a hypothetical yes response

---

16 Note, however, that the scale used by Champ et al. (1997) differs somewhat from the scale used in this study. Champ et al. (1997) used a scale from 1 to 10 whereas we used a scale from 0 to 10.
In our regression analysis we also tested if the probability of a yes response was related to the price level and a host of background variables. The price variable was highly significant with a negative sign, consistent with a downward sloping demand curve. The results indicated that the willingness to pay decreased with education; this is logical if knowledge about asthma is related to the level of general education, since the good valued was an asthma management program intended to educate the subjects about their disease and to assist them in achieving optimal treatment. A negative effect of education on the demand for health care is also consistent with Grossman’s demand for health model, which predicts that education increases the productivity of health production and decreases the demand for health care (Grossman, 1972). The income variable had a positive sign, but was not significant. Normally we expect the demand for health care to increase with income and a positive income elasticity is considered an important test of the validity of contingent valuation studies. In this case, the expected a priori sign of income is, however, unclear. This is because participation in the asthma management program involved a significant amount of time for the subjects (1.5–2 h over a 3 month period), and the opportunity cost of time can be expected to increase with income.

For the health care sector, it is also interesting to note the similarity between contingent valuation surveys and the methods used to estimate quality weights used to construct quality-adjusted life-years (QALYs) (see Torrance (1986) for an overview of these methods). Both approaches rely on respondents’ choices in hypothetical situations and QALY weights may therefore also be subject to hypothetical bias. A drawback of the QALY approach is that comparisons with real choices are very difficult.

In conclusion, our study indicates that the dichotomous choice contingent valuation method overestimates the willingness to pay for health care programs. This implies that the validity of existing contingent valuation estimates for health care programs is questionable and the values should be used with great caution. On the positive side, our results suggest that the overestimation can be corrected by sorting out “definitely sure” yes responses as identified in a simple follow-up question. Further work is needed to establish the validity of this calibration approach.

---

17 In the data on sunglasses collected by Blumenschein et al. (1998) and Johannesson et al. (1999) there were 35 hypothetical yes responses in the within subject comparisons. Of the 23 subjects who answered yes to the hypothetical question and were “probably sure”, 3 answered yes to the real question. Of the 12 subjects who answered yes to the hypothetical question and were “definitely sure”, 10 answered yes to the real question. The follow-up question dividing hypothetical yes responses into “definitely sure” and “probably sure” yes responses thus correctly predicted 85% (30/35) of the real choices. As shown in Johannesson et al. (1999) the degree of certainty in a hypothetical yes response on the 0–10 scale used in the calibration function also correctly predicted about 85% of the real yes responses from within subject comparisons.

18 Estimating the effect of the income variable was complicated by the fact that there was relatively little variation in the income measure. This is due to the categorical nature of the question used to assess personal income. As many as 59% of the subjects fell into the lowest income category (less than US$ 10,000 per year).
Acknowledgements

We thank Jonathan Gruber and an anonymous referee for helpful comments. This research was funded by a Drug Information Association Research Grant.

Appendix A. Description of the asthma management program

Managing your asthma correctly is an important effort you can make to ensure that you lead a normal and healthy life. Taking your medications exactly as prescribed and knowing how to use your inhaler and peak flow meter correctly are vital to minimizing asthma flare-ups.

Numerous studies have shown that when pharmacists or other health care professionals provide educational programs focused on helping patients with asthma understand and manage their condition better, many significant benefits are observed. These benefits include decreased visits to the emergency department; decreased admission to the hospital due to asthma flare-ups; decreased days of school or work missed due to asthma symptoms; improvement in asthma symptoms, and improvement in overall quality of life (i.e. patients feel better overall).

This asthma management program will be offered at this pharmacy with your regular pharmacist providing the service. The program consists of 3 monthly visits to the pharmacy, for a total of 1.5–2 hours spent with the pharmacist. The pharmacist that will provide the service will contact you to set up the monthly appointments. With your permission, the pharmacist will work with your prescriber to ensure your asthma is being properly managed. Following each visit, a letter will be sent to your prescriber summarizing what was discussed during your visit. The pharmacist may make recommendations to your prescriber as necessary to try and improve the management of your asthma.

The first visit will last approximately 45 min. The pharmacist will begin by taking a medication history as well as a history of your asthma and related conditions. You will fill out a short survey and a short true/false quiz about asthma which you will then discuss with your pharmacist. The pharmacist will provide an overview of the asthma disease process, including how asthma affects breathing, why people get asthma, and what types of asthma there are. Use of a peak flow meter will be demonstrated and the importance of accurate measurement and documentation of peak flow rate will be explained. Also, the importance of accurately recording your asthma symptoms and how symptoms are related to the severity of asthma will be explained. The pharmacist will review the proper use of your inhalers and assess the technique by which you are currently using your inhalers. Suggestions to improve the way you use your inhalers to ensure that you get maximum benefit from the medications will be made when appropriate. You will be given information on how your specific medications work in asthma and the proper times to use each medication. You and your pharmacist will set goals. You will be able to call your pharmacist if you experience any worsening of asthma symptoms or to ask any questions related to the information discussed at the visit.

Your second visit will last about 30 min. At this visit you will answer questions about your asthma symptoms and other problems that you may have experienced since the previous
visit. You will review the goals set by you and your pharmacist, and discuss with the pharmacist your current medications, with emphasis on any changes that may have occurred in your medication regimen since the previous visit. The pharmacist will check your peak flow diary and explain the results to you. Your pharmacist will create a self-management “action” plan based on your peak flow results and will explain how to use the plan to minimize the severity of your asthma flare-ups. If applicable, smoking cessation education will be provided. The progress made toward achieving the goals you set for yourself will be assessed and goals for the next visit will be set. Finally, the technique by which you are using your inhalers and peak flow meter will be reassessed.

The third visit will also last approximately 30 min. You will answer a questionnaire on asthma symptoms experienced since the last visit. All of your medications will be reviewed. The pharmacist will review your peak flow diary and use of the self-management “action” plan. The pharmacist will discuss environmental triggers that may cause a flare-up of asthma symptoms and will help you identify ways to avoid these triggers. If applicable, there will be a review of smoking cessation education. Progress toward attainment of the goals set previously will be evaluated and goals for the future will be set. Finally, the technique by which you are using your inhalers and peak flow meter will be reassessed.

These monthly visits can continue at your discretion, however, the initial fee you pay today will only cover the first three visits that I have just described to you.

References