HELICOPTERS, HEARTS AND HIPS: USING WILLINGNESS TO PAY TO SET PRIORITIES FOR PUBLIC SECTOR HEALTH CARE PROGRAMMES

JAN ABEL OLSEN* and CAM DONALDSON**

*University of Tromsø, Department of Economics, NFH, 9037 Tromsø, Norway and **Health Economics Research Unit, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, Scotland

Abstract—The paper reports from a study that asked 150 interviewees their willingness to pay (WTP) in increased earmarked taxation for three different health care programmes: a helicopter ambulance service, more heart operations and more hip replacements. Reasons behind the stated WTP were asked for. Ordinary least squares regression analyses were used to analyse factors associated with WTP for each of the three programmes, and factors associated with the relative WTP for one programme compared with the total of the three. Comparisons were made of WTP for these programmes and the health outcome in terms of quality adjusted life years. © 1998 Published by Elsevier Science Ltd. All rights reserved

Key words—willingness to pay, priority setting, quality adjusted, life years, economic evaluation

INTRODUCTION

One of the greatest challenges for publicly financed health services is the setting of priorities among competing health care programmes. The most widely recommended criterion in the health economics literature is to apply some form of cost-effectiveness analysis and rank programmes according to lowest costs per unit of health outcome, where outcome is normally measured in terms of quality adjusted life years (QALYs). Although QALYs may be considered to be a better measure of health than many alternative measures, they account only in part for the preferences lying behind individuals' valuations of health care programmes. First, health gains are valued differently depending on the nature of the services which produce them (Eddy, 1991), e.g. the QALY outcome from rescue-type health care might be more highly valued than the same QALY outcome from other programmes. In addition, there are non-health benefits from health care use beyond the benefit derived from improved health outcome (Mooney, 1991; Mooney and Lange, 1993), e.g. information. Furthermore, there are equity concerns with regard to health care (Wagstaff, 1991), implying that the value of a sum of health gains depends upon its distribution.

The failure to accommodate for such preferences has led to increasing interest in willingness to pay (WTP) as an alternative way of valuing health care.† From a consumer sovereignty perspective, WTP is superior to QALYs in that no restrictions are imposed on which attributes of a health care programme people are allowed to value. However, the application of this valuation instrument is not without its difficulties. Much of the criticism against the use of WTP relates to the measurement biases inherent in the design of actual studies.

The aims of this paper are, first, to investigate whether what is in our view an improved design of a WTP study could prove fruitful to help set priorities among health care programmes. The paper reports on a study that elicited WTP values from members of the community with regard to three different public sector health care programmes. We are particularly concerned with the credibility of the results, i.e. the extent to which the absolute WTP figures appear reliable, and the relationships between respondents' WTP and their demographic characteristics and attitudes. Furthermore, emphasis is put on an investigation into why people would be willing to pay and why some would not be willing to pay.

The second aim of the paper is to compare the WTP values with the QALYs gained from each programme. In that way, it becomes clear if, and how, the two alternative valuation techniques result in different rankings of the programmes. Further, when the WTP per QALY differs between programmes, that could suggest their health outcomes are differently valued and/or differently influenced by characteristics other than their health enhancing capabilities.

In the following section, a summary of the state of the art of WTP as applied to health care is pre-
sented. Differences in focus and methodology between previous studies and the one presented in this paper are highlighted. In the third section, the method adopted for this study is outlined. The key results are then presented. The discussion in the fifth section focuses on methodological issues.

SOME RECEIVED WISDOM

Although quite a few studies have been carried out in the neighbouring subdisciplines of environmental and transport economics (Johansson, 1987; Jones-Lee, 1989), there are still a relatively small number of references to WTP for health or health care programmes. According to a count by Donaldson (1996), there are 28 studies of WTP for health or health care published in English language sources.* Advances have been made, but there are still some problems with these studies.

Most studies concentrate on the degree to which health care will improve the respondent's (or her/his family's) own health (Thomson, 1986; Johannesson, 1992). The implicit assumption is that there are no other benefits from health care. Further, it is common for preferences to be elicited from those who already are, or could be, users of the service being evaluated. This approach might be relevant in settings in which users pay for the full cost of health care (through charges or private insurance), and in which health care is viewed as an individual good. However, the approach becomes less relevant in settings in which the wider community pays, and in which health care is considered to have utility-bearing characteristics in the form of "caring externalities" (Culyer, 1971). If the mode of payment may affect one's stated valuation, there is a potential payment vehicle bias (Mitchell and Carson, 1988) in such WTP questions when applied to tax-funded health services.

In none of the 28 studies referred to above were WTP values elicited for more than one type of health care. The problem with doing such partial WTP studies is that, in a priority setting context, one would not have any idea of how the particular programme performs relative to other health care programmes. According to a count by Donaldson (1996), there are 28 studies of WTP for health or health care published in English language sources.* Advances have been made, but there are still some problems with these studies.

Most studies concentrate on the degree to which health care will improve the respondent's (or her/his family's) own health (Thomson, 1986; Johannesson, 1992). The implicit assumption is that there are no other benefits from health care. Further, it is common for preferences to be elicited from those who already are, or could be, users of the service being evaluated. This approach might be relevant in settings in which users pay for the full cost of health care (through charges or private insurance), and in which health care is viewed as an individual good. However, the approach becomes less relevant in settings in which the wider community pays, and in which health care is considered to have utility-bearing characteristics in the form of "caring externalities" (Culyer, 1971). If the mode of payment may affect one's stated valuation, there is a potential payment vehicle bias (Mitchell and Carson, 1988) in such WTP questions when applied to tax-funded health services.

In none of the 28 studies referred to above were WTP values elicited for more than one type of health care. The problem with doing such partial WTP studies is that, in a priority setting context, one would not have any idea of how the particular programme performs relative to other health care programmes on which the resources could be spent. Furthermore, partial studies may also create an "embedding problem" (see e.g. Kahneman and Knetsch, 1992), i.e. a tendency to place much the same value on a specific good (e.g. heart operations) as on a more broadly defined group of goods (the health care sector).†

The ways in which the study reported in this paper attempts to overcome the above problems are as follows: wider descriptions of the programmes were presented to respondents. These focused on: (1) the actual health care programme to which resources could be allocated, (2) a statement of patients' initial health problems, and health states after treatment, (3) duration of the improvement, and (4) characteristics of those benefiting in terms of age, sex and where they live.

Respondents are asked for their willingness to contribute in terms of extra earmarked taxation per annum. As such, we have a community programme focus, which is most relevant in societies where the health service is tax financed. Reasons for respondents' WTP were asked for, in order to facilitate an assessment of whether wider community benefits were taken into account. Three alternative programmes were presented, in order to make the respondents focus on their valuation of each particular programme rather than viewing an individual programme as one which represents the health care sector as a whole. Comparisons can then be made between the social valuations of alternative ways of allocating health care resources. More important, however, is the possibility of using this approach to gain increased understanding of which attributes of the different health care programmes are valued. This kind of priority setting study, looking at WTP for alternative programmes within one public sector, has (to our knowledge) not been done before. Only one study has made such broad comparisons, but not in terms of WTP (Jones-Lee and Loomes, 1994). Therefore, the study reported in this paper should be seen in many ways as hypothesis-generating rather than hypothesis-testing.

THE STUDY

The intention was to compare the value of spending resources on an existing helicopter ambulance service with the values of spending these resources on other health care programmes. The helicopter ambulance service was considered to be a particularly interesting candidate on which to carry out a WTP study. The service carries other assumed utility-bearing characteristics: (1) it is a "rescue service" which may reduce anxiety among people in rural areas; (2) although few people use or benefit directly from the service, it may have a considerable "option value" derived from knowing that the service is available should use in the future be required; and (3) it contributes to reducing regional inequalities in access. A thorough cost-effectiveness

*To our knowledge there are now four more studies in the literature.
†Of course, it is possible that such effects could arise in the context of cost-utility analysis when valuing a health state/outcome in isolation. For example, with the time tradeoff method, it may well be the case that the amount of time individuals are willing to give up to achieve separate gains in quality of life would, when added together, be greater than their lifespan!
study had made available unique outcome data on the helicopter ambulance operating in the north Norwegian county of Troms (Kristiansen, 1992).*

Two alternative health care programmes were presented, each having the same annual costs [10 million Norwegian kroner (£1 = NOK 10)] as the helicopter service. The criteria for choosing the alternatives were: first, it was likely that people were familiar with them, so brief descriptions were sufficient; and second, in comparing alternatives with a rescue-type programme such as the helicopter, we wanted one programme which produces health gains in the form of both life extension and improved quality of life, and one programme which produces improved quality of life only. The two programmes chosen were 80 more coronary artery bypass operations and 250 more hip replacements.

Rather than describing the programmes in individual probability terms, as advocated by Gafni (1991) and Morrison and Gyldmark (1992), they were framed as being to the benefit of the community, i.e. for people living in the region. Such a benefit can be seen as certain (due to the law of large numbers), a point made by O'Brien and Gafni (1996). The respondent could interpret her/his own probability implicitly in terms of the number of people benefiting out of the population. For the helicopter, respondents were informed how many people used the service and how many live in the target region. For the other two programmes, the relative increase in provision was emphasised. Since most people would have a fairly clear idea of how many live in Northern Norway, it was decided not to include an additional number on the cards describing the latter two programmes.

Study design

The study was carried out in October 1992. A member from each of 150 households in Troms county was interviewed; 75 in Tromsø, 25 in a smaller city and 50 in the rural area. The numbers were determined so as to be proportional to the sizes of the population living in these three different (types of) communities. There were no previous studies of this nature from which a sample size calculation could be made. People were selected in the following way: in the two cities, the telephone book was used to randomly select 15 and five streets, respectively, in each of which five people were interviewed. In the rural area five municipalities in different parts of the county were chosen, in each of which 10 people were interviewed. There, the interviewers started from two different locations, one in the municipality centre and one in a sparsely populated area.

From the starting address in each street/municipality the interviewers went along to each house (by ascending number) until they achieved the required quota of respondents. In total there were 135 who could not participate.† The usual excuse was that they were busy. Only one of the refusals expressed specific aversion to this type of health care evaluation. Six trained interviewers were used, and the average interview time was 35 minutes.

The introductory information for respondents focused on the proposed method of measuring individuals' valuations of health care programmes, i.e. in terms of the maximum amount one would be willing to contribute through increased taxation which would be earmarked to health care programmes. Respondents were also informed of the three programmes to be addressed in the interview. As the interview proceeded, the respondents were given a yellow card with a brief description of the helicopter ambulance service, and then a white card containing the same description with the payment options listed (see Appendix A). This is a common way of eliciting WTP values (Mitchell and Carson, 1988). The alternative of providing individuals with one WTP figure and asking for a “yes”/“no” response to whether they would be willing to pay that amount, as proposed by Arrow et al. (1993), requires a larger sample size than that used in this study.

After having circled the amount which was the maximum they would be willing to pay each year, respondents were asked, in open ended questions, to explain the reasons that they were willing to pay. Then the interviewer read out some precoded reasons which were thought to influence valuations. A visual scale on degree of importance for being willing to pay was handed to the respondents. They were then asked to rate the reasons on this scale, which was presented with equal spacing between the degrees of importance. To each of the four points on the scale, we attached the following values: 0 for “no”, 1 for “little”, 2 for “some” and 3 for “much” importance. The same procedure was used for the heart programme and after that the hip programme. For each programme a partial valuation was emphasised, WTP questions being accompanied by the statement “...if this was the only programme to be implemented”.

Before giving their valuations respondents were asked to compare the yellow cards and prioritise between them. Those who were not willing to pay anything in extra taxation, because of tax aversion, were asked whether they would be willing to give a

*The focus was on the marginal effects in terms of increased survival probability and improved quality of life that could be ascribed to the shorter transport time with the helicopter compared with the alternative ground ambulance. The mean time saving was 69 minutes for those who were transported.

†This “response rate” is encouraging, given that respondents were asked to let themselves be interviewed at the moment of approach which was during the period 5–10 PM.
Table 1. Willingness to pay per year (in NOK)

<table>
<thead>
<tr>
<th></th>
<th>Helicopter</th>
<th>80 more heart</th>
<th>250 more hip</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 143$</td>
<td>316</td>
<td>306</td>
<td>232</td>
</tr>
<tr>
<td>Mean</td>
<td>25</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>Median</td>
<td>200</td>
<td>200</td>
<td>150</td>
</tr>
<tr>
<td>Standard error</td>
<td>25</td>
<td>25</td>
<td>21</td>
</tr>
</tbody>
</table>

RESULTS

How much were people willing to pay?

Of the 150 completed interviews, seven were excluded because respondents would neither express any WTP figure nor would they prioritise. Some of them expressed reservations about the whole exercise (Table 1).*

There were no statistically significant differences between the mean WTP values for helicopter and hearts, but the value for hips was different from both helicopter and hearts (using Wilcoxon signed-rank test, $P < 0.0001$). Because we have fixed end points, and the intervals between the discrete payment options increased (from NOK 25 to 200) as the options themselves increased, normal distributions were not expected. Hence, for the following analysis, the three dependent variables, i.e. the WTP for each programme, were log-transformed.

Which factors influenced willingness to pay?

Ordinary least squares (OLS) regression analysis was used to test for factors associated with WTP. Table 2 gives a specification of the independent variables. The AGESQ variable was introduced because, for hearts and hips, the age variable was expected to have a curvilinear form with respect to WTP. The variable TIMEINT is introduced as an indicator of a possible "income effect". Although the partial valuation was emphasised, we would hypothesise that the longer the duration of the interview, the more time the respondent would have. Therefore, for hearts and hips, the age variable was included.

Most of these hypothesis are motivated by general utility theory; for example, women are more likely to be willing to pay for the hips programme because they are more likely to have their hips replaced. On education, it was thought that people educated to different levels would have different attitudes to health care priorities, although the direction of this effect is not clear a priori.

The same regression model was used to analyse WTP for each programme. This permits comparisons of coefficients across programmes.† Table 3 shows the results from the multiple regression analysis.

---

*Nine respondents were not willing to pay for any of the three programmes in any of the three modes of payments, but they were willing to prioritise. They are included.

†A simple correlation test did not give any indication of a possible multicollinearity between TIMEINT and the other independent variables.

‡It was hypothesised that people with children are willing to pay more for the helicopter. However, when a dummy for child(ren) under 16 in the household was introduced as an additional variable, it had no influence on WTP. Again, because the same model was preferred for all three programmes, this independent variable is not included.
Table 3. Factors influencing WTP (n = 143)

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Helicopter (log WTP)</th>
<th>Hearts (log WTP)</th>
<th>Hips (log WTP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.154*** (1.4672)</td>
<td>3.0776 (1.7140)</td>
<td>1.0710 (1.5995)</td>
</tr>
<tr>
<td>SEX</td>
<td>-0.8622** (0.2922)</td>
<td>-0.4894 (0.3431)</td>
<td>-1.2900*** (0.3185)</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.0178 (0.0567)</td>
<td>0.0333 (0.0663)</td>
<td>0.0703 (0.0618)</td>
</tr>
<tr>
<td>AGESQ</td>
<td>-0.0005 (0.0005)</td>
<td>-0.0068 (0.0006)</td>
<td>-0.0011* (0.0006)</td>
</tr>
<tr>
<td>INCPERS</td>
<td>0.3620 (0.2756)</td>
<td>-0.0724 (0.3220)</td>
<td>0.2732 (0.3004)</td>
</tr>
<tr>
<td>MUNICIP</td>
<td>-0.2923 (0.3328)</td>
<td>-0.5933 (0.3888)</td>
<td>-0.2364 (0.3629)</td>
</tr>
<tr>
<td>OWNHE</td>
<td>-0.1807 (0.3083)</td>
<td>-0.4433 (0.3601)</td>
<td>0.3277 (0.3361)</td>
</tr>
<tr>
<td>EDUC</td>
<td>-0.7735* (0.3201)</td>
<td>-0.5279 (0.3746)</td>
<td>-0.1388 (0.3490)</td>
</tr>
<tr>
<td>TIMEINT</td>
<td>0.0301 (0.0216)</td>
<td>0.0833*** (0.0252)</td>
<td>0.0801*** (0.0235)</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.241</td>
<td>0.144</td>
<td>0.205</td>
</tr>
<tr>
<td>Sigmal F</td>
<td>0.0000</td>
<td>0.0003</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Notes: B = coefficient, SE B = standard error of the coefficient. Two-tailed; one-tailed if according with a hypothesis: SEX hips, AGESQ hearts, TIMEINT hearts and hips.

* = \( P < 0.05 \);
** = \( P < 0.01 \);
*** = \( P < 0.001 \).

There is no evidence to support \( H_s^S \). For hips, women were willing to pay significantly more, which supports \( H_s^T \). The education level had a significant (negative) impact on WTP for the helicopter. There is not sufficient support for \( H_s^T \), \( H_s^P \), \( H_s^2 \) or \( H_s^3 \). When age was entered linearly, its coefficients were negative for all three programmes \( (P < 0.01) \). Although, when entered quadratically, the coefficients of AGE and AGESQ were not statistically significant any more, the important point is that, for hearts and hips, the sign for AGE switched from negative to positive (indicating \( H_s^T \)). The negative coefficients of AGESQ indicate that the WTP falls at older ages \( (H_s^{O3}) \).

As to the importance of the length of the interview, there is evidence in support of although \( H_s^{T3} \), \( H_s^{P3} \), \( H_s^{23} \) or \( H_s^{33} \). When age was entered linearly, its coefficients were negative for all three programmes \( (P < 0.01) \). Although, when entered quadratically, the coefficients of AGE and AGESQ were not statistically significant any more, the important point is that, for hearts and hips, the sign for AGE switched from negative to positive (indicating \( H_s^{T3} \)). The negative coefficients of AGESQ indicate that the WTP falls at older ages \( (H_s^{O3}) \).

As to the importance of the length of the interview, there is evidence in support of although \( H_s^{T3} \), \( H_s^{P3} \), \( H_s^{23} \) or \( H_s^{33} \). When age was entered linearly, its coefficients were negative for all three programmes \( (P < 0.01) \). Although, when entered quadratically, the coefficients of AGE and AGESQ were not statistically significant any more, the important point is that, for hearts and hips, the sign for AGE switched from negative to positive (indicating \( H_s^{T3} \)). The negative coefficients of AGESQ indicate that the WTP falls at older ages \( (H_s^{O3}) \).

While in Table 3 each of the three programmes is considered in isolation, the next model analyses what determines the relative preference of one programme.† As the dependent variables, the share of the WTP for the programme concerned over the total WTP for the three programmes was used, or, more precisely, the share of the sum of the partial WTP, i.e. a type of a “budget share”. Assuming partial valuations, this model will indicate how the relative WTP depends upon the background variables.§ The independent variables used in Table 4 are the same as those in Table 3 except that AGESQ is excluded.

Note that, for each of the independent variables in Table 4, the sum of the coefficients over the three programmes is zero. This is explained by the fact that the sum of the budget shares is always unity. In Table 4, it can be seen that men have a significantly higher relative WTP for hearts, which now lends support to \( H_s^S \).

Why were people willing to pay?

The open ended reasons. For each programme respondents were asked an open ended question about the reasons for their positive WTP. The stated reasons were grouped, allowing each respondent to give more than one reason. Assuming that these reasons reflect characteristics which are positively associated with the level of WTP, they were entered as dummies in a regression analysis together with INCPERS (i.e. adjusting for income differences). However, they did not explain anything, which suggests that the open ended responses were somewhat arbitrary. The structured precoded exercise proved to be more fruitful.
Jan Abel Olsen and Cam Donaldson

Table 4. Factors influencing WTP of each programme over the total WTP (n = 134)

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Helicopter (share WTP)</th>
<th>Dependent variable</th>
<th>Hearts (share WTP)</th>
<th>Hips (share WTP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B (SE B)</td>
<td>B (SE B)</td>
<td>B (SE B)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.8696***</td>
<td>0.1987</td>
<td>-0.0684</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1206)</td>
<td>(0.1010)</td>
<td>(0.1022)</td>
<td></td>
</tr>
<tr>
<td>SEX</td>
<td>0.0236</td>
<td>0.0546*</td>
<td>-0.0782**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0336)</td>
<td>(0.0281)</td>
<td>(0.0285)</td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>-0.0045**</td>
<td>0.0021*</td>
<td>0.0023*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td></td>
</tr>
<tr>
<td>INCipers</td>
<td>0.0043</td>
<td>-0.0606*</td>
<td>0.0562*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0298)</td>
<td>(0.0249)</td>
<td>(0.0252)</td>
<td></td>
</tr>
<tr>
<td>MUNICIP</td>
<td>0.0155</td>
<td>-0.0301</td>
<td>0.0146</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0379)</td>
<td>(0.0318)</td>
<td>(0.0321)</td>
<td></td>
</tr>
<tr>
<td>OWNHE</td>
<td>-0.0243</td>
<td>-0.0390</td>
<td>0.0633*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0359)</td>
<td>(0.0300)</td>
<td>(0.0304)</td>
<td></td>
</tr>
<tr>
<td>EDUC</td>
<td>-0.0543</td>
<td>-0.0060</td>
<td>0.0312</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0370)</td>
<td>(0.0310)</td>
<td>(0.0313)</td>
<td></td>
</tr>
<tr>
<td>TIMEINT</td>
<td>-0.0079**</td>
<td>0.0044*</td>
<td>0.0035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0026)</td>
<td>(0.0022)</td>
<td>(0.0022)</td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.109</td>
<td>0.123</td>
<td>0.138</td>
<td></td>
</tr>
<tr>
<td>Signif F</td>
<td>0.0029</td>
<td>0.0012</td>
<td>0.0005</td>
<td></td>
</tr>
</tbody>
</table>

Notes: B = coefficient, SE B = standard error of the coefficient. Two-tailed; one-tailed if according with a hypothesis: SEX hearts and hips, AGE hearts and hips.

* = P < 0.05; ** = P < 0.01; *** = P < 0.001.

The precoded reasons for WTP. The mean values for each of the reasons presented under each programme are given in Table 5.* There are open spaces in Table 5 because not all reasons were relevant to present for all programmes. The first two reasons for people's valuations can be interpreted as indicators for the importance of selfishness and altruism, respectively. The non-parametric Wilcoxon signed-rank test showed that ratings for altruism are significantly different from ratings for selfishness for helicopter (P < 0.0001) and for hips (P = 0.0209), but not for hearts.

The “option value” of the helicopter ambulance can also be considered as a selfish concern. The

*Within each programme, roughly 25% of the respondents had consistently given the highest score to all the listed reasons. Since this group appears to be less reflective on this exercise, there is an argument for excluding them in Table 5. However, although the mean values increase towards the upper end of the scale, their inclusion does not affect the ordinal ranking of the reasons under each programme.

†The helicopter ambulance is an existing service. People were asked to imagine that the money for the helicopter was to be transferred to other health services, and state how much they were willing to contribute to prevent it being closed down.

‡For each respondent, the mean value of the two reasons was used.

§Acknowledging that the distributions of the degree of importance attached to these reasons were far from normal (many claimed the reasons were of “much” importance), we examined the effect of entering these variables as dummies (1 for “much”, 0 otherwise). The results were close to those reported, but the explained variance was slightly less.

¶The interpretation of the B-coefficients of these attitude variables is as follows: for each unitary level of importance (from 0 to 3), the B-value gives the predicted percentage increase in WTP.

reason “don’t want a health service to close down”† can be interpreted as an indicator of “possession preference” (i.e. not wanting to give something up once it already exists), which is also known as the “endowment effect” (Kahneman et al., 1990). The selfish reason is strongest for the heart programme, but there is also much concern for “heart queues”. The “efficiency argument”, that some patients will return to work, got the lowest score. However, for the hip programme, the fact that “elderly will be able to manage themselves and will need less help from others” gained the highest score.

For each of the three programmes these precoded reasons from Table 5 were entered as independent variables in a regression analysis together with INCipers, and with the log-transformed WTP figure as the dependent variable. The major results can be summarised as follows. For the helicopter ambulance, the ratings for the “selfish” reason and the “option value” were merged into one “selfish” variable.† This came out with a positive influence on WTP (B = 0.29 and P < 0.01), however, slightly lower than the “altruistic” variable (B = 0.33 and P < 0.05). In a separate regression analysis with the four attitude variables entered together with the background variables of Table 2,§ the importance of altruism for WTP came out even stronger (B = 0.42 and P = 0.015).¶

For the heart programme, the altruistic reason came out with the highest positive sign, but it was not significant. For the hip programme, the “elderly will be able...” reason was the only significant reason for WTP (B = 0.52 and P = 0.001) when regressed together with the other three attitude variables and INCipers. The altruistic reason was not
significant. This "elderly will be able..." reason came out with the same $B$- and $P$-values in a separate regression analysis with the four attitude variables entered together with the background variables of Table 2.

**Why were people not willing to pay?**

Those who were not willing to pay earmarked taxes for the helicopter ambulance were asked whether they would be willing to pay in other ways. When faced with the two following programmes, this group were also asked which way they would prefer to pay. Among those with positive WTP, the vast majority were willing to pay in terms of taxes. However, the proportion differed from 73% among those willing to pay for helicopter, 83% for hearts and 88% for hips. The number of respondents who were unwilling to pay in any way were 17 for the helicopter, 22 for hearts and 23 for the hip programme. The stated reasons were grouped, again allowing each respondent to give more than one reason. The most widely expressed reason was: "the public sector has enough money to pay, so reallocation should be possible". Whilst this reason does not necessarily reflect zero valuations of the programmes, the other, more common, reasons are more indicative of zero valuations. Some simply said "can't afford to pay". For the helicopter ambulance, 11 held that "other health services are more important", while the straight selfish reason "I won't need it myself" was expressed by four. For the hearts programme, five said it is a "status activity which has got enough resources already", and three held that it is "self-inflicted lifestyle illness". A further three said that it was "inefficient" without further explanations. For the hips programme, four gave the reason that "it doesn't save lives".

**Comparing valuations with health outcomes**

One way of measuring the relative importance of characteristics of health care programmes which are not accounted for in terms of QALYs gained is to compare valuations from WTP with QALYs. The health outcomes as described on the three cards can be reduced to undiscounted QALYs (see the Appendix A). The outcome of the helicopter programme was five saved lives with a total of 150 life years (=150 QALYs†). The health outcome from the heart programme was estimated as 2.5 undiscounted QALYs per person $\times 80 = 200$ QALYs, while the outcome from the hip programme was estimated to be 4.5 undiscounted QALYs per person $\times 250 = 1125$ QALYs. As a second option, using conservative estimates by halving the quality enhancement part of the QALY gain, the outcome from the heart programme becomes 160 QALYs while that from the hip programme becomes 563 QALYs. Thirdly, discounting the first set of QALY estimates at a rate of 10% gives 47 QALYs for the helicopter, 100 QALYs for hearts and 563 QALYs for hips. Hence the ordering of the health effectiveness between the programmes is not sensitive to discounting or low estimates. The mean WTP per QALY gained for the different programmes under these three sets of assumptions can be seen from Table 6.

Although the relatively lower valuation of hips might be influenced by the ordering and the life saving focus of the first two programmes, the first line of Table 6 indicates the preferred ranking of the three programmes when based on WTP. Remembering that the costs of the three programmes were identical (NOK 10 million) (which respondents were informed about), the second line suggests a reverse ranking when based on QALY maximisation.

---

**Table 5. Mean scores on the precoded reasons for WTP scale on degree of importance (0-3)**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Helicopter</th>
<th>Hearts</th>
<th>Hips</th>
</tr>
</thead>
<tbody>
<tr>
<td>You/your family could need it</td>
<td>2.15</td>
<td>2.58</td>
<td>2.24</td>
</tr>
<tr>
<td>Other people could need it</td>
<td>2.78</td>
<td>2.67</td>
<td>2.50</td>
</tr>
<tr>
<td>Reassuring to know that it exists</td>
<td>2.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More equal access to health care regionally</td>
<td>2.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Don’t want a health service to close down</td>
<td>2.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Heart queues” and waiting lists</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some will be able to return to work</td>
<td></td>
<td>2.57</td>
<td>2.77</td>
</tr>
<tr>
<td>Elderly will be able to manage themselves and will need less help from others</td>
<td>2.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elderly people deserve better health</td>
<td></td>
<td></td>
<td>2.46</td>
</tr>
</tbody>
</table>

*There were not sufficient observations to enable comparisons of WTP across payment modes (voluntary donation, private insurance premium, and earmarked taxes).

†The cost-effectiveness study of the helicopter project showed that gains in terms of life years and QALYs were about equal (Kristiansen, 1992). Because of this the descriptions focused on healthy life years.
DISCUSSION

Some critics of the contingent valuation method (e.g. Diamond and Hausman, 1993) have been concerned with what they call "a clear mismatch between the hypothetical CV answers and the actual WTP", or, in other words, that respondents' altruistic attitude exceeds their altruistic action (see e.g. the findings obtained by Seip and Strand, 1992). Whilst it is rarely possible in CV studies to compare the stated WTP values with observed actions in order to assess the reliability of the answers, for the helicopter service we have some indication that the obtained WTP values in this study appear reliable. A comparison can be made between the mean WTP per household of NOK 316 for the helicopter ambulance with the membership fees of the Norwegian Air Ambulance, which is the foundation responsible for the operation of the helicopter. In 1993 the family membership fee was NOK 315, while that of a two-person household was NOK 240.*

Recently an expert panel in the U.S. gave some recommendations as to how a WTP study should be designed (Arrow et al., 1993). Although their report was published after our study was undertaken, to a large extent the views expressed by that panel correspond with our approach. In particular, a "conservative design" was used: (1) respondents were asked for their WTP rather than willingness to accept; (2) respondents were reminded of their budget constraint; (3) because three programmes were compared, there is less chance of an "embedding effect"; (4) the scale of the payment card had a reasonable upper limit, the midpoint being close to the actual cost per household; and (5) WTP per annum, rather than per month, was asked for. Given the above, we believe that the absolute WTP values reported in our study appear reasonable. Nevertheless, some outstanding issues remain to be discussed.

The partial valuation of each of the three programmes shows that the benefits exceed the costs in each case. While the helicopter is to the benefit of people in Troms county the other two programmes are for the entire region of Northern Norway, covering three times as many people. Strictly, a true comparison would require values elicited for all three programmes from a sample of people in the other two counties in Northern Norway.

Picking up "protest zeros"

In many WTP surveys a considerable number of respondents answer the questions about WTP with zero bids, i.e. they state they are not willing to pay. Given that respondents may still benefit from the programmes in question, the interpretation of the zero bids have evoked some concern in the literature (see e.g. Kidholm and Lankkilde, in press); are they "true zeros" or "protest zeros"?

Acknowledging that quite a few people appear to have "tax aversion", it was expected that some respondents would express zero bids because of the chosen mode of payment in this survey. Therefore, those who gave a zero bid to the first question were presented with follow-up questions on whether they would be willing to give a voluntary donation or, alternatively, pay for a private insurance premium. Of the 143 respondents included in the analysis, 51 were not willing to pay for the helicopter ambulance when faced with the WTP question in terms of increased taxation. However, when the alternative payment modes were offered, two thirds of them became willing to pay. Thus, a simple way of picking up some protest zeros in WTP surveys could be to include alternative modes of payment.

Publicity bias

The possibility of some biases in favour of the helicopter programme has to be acknowledged. During the months before interviewing there was much "turbulence" in the local media—particularly in Tromsø—as to the life saving importance of a new landing site for the helicopter ambulance.† This focus might have influenced the respondents to give higher valuations for the helicopter, although it was explicitly stated that this study was not concerned with the landing site. Although the mean WTP was higher in Tromsø than in the other city, in each city, the helicopter's mean value was equal to that of hearts, while the rural region valued the helicopter highest. This might indicate that the absolute WTP for the helicopter was influenced by media attention, but its relative performance in each city seems to have been unaffected. However, the figures do lend some support for the suggestion

---

*In Norway there are 800,000 contributing members out of a population of 4.1 million. Non-members, however, are also covered by the service.
†The doctors’ preferred choice would save 30 seconds compared to the cheaper alternative, and some of them had been very active in arguing for the life saving importance of this marginal time saving.

---

Table 6. Cost and mean annual WTP (in NOK) per QALY gained

<table>
<thead>
<tr>
<th></th>
<th>Helicopter ambulance</th>
<th>80 more heart operations</th>
<th>250 more hip operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean WTP</td>
<td>316</td>
<td>306</td>
<td>232</td>
</tr>
<tr>
<td>QALYs undiscounted</td>
<td>150</td>
<td>200</td>
<td>1125</td>
</tr>
<tr>
<td>Cost/QALY</td>
<td>67,000</td>
<td>50,000</td>
<td>8900</td>
</tr>
<tr>
<td>Mean WTP/QALY</td>
<td>2</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Mean WTP/QALY low est.</td>
<td>2</td>
<td>1.9</td>
<td>0.4</td>
</tr>
<tr>
<td>Mean WTP/QALY 10% disc</td>
<td>6.7</td>
<td>3.1</td>
<td>0.4</td>
</tr>
</tbody>
</table>
that the WTP for the first programme affects WTP for the subsequent ones.

Ordering bias

The lower valuation of hips could be influenced by an ordering effect. Although the partial nature of the valuations was explicitly stated, the interviewers got the impression that when being faced with the last of the three health care programmes, some respondents felt that they had contributed to two programmes (done their "fair share") and could not afford more (income effect), and/or they were tuned in on life saving which was not a part of the outcome of the hip programme. However, at the end of the exercise, an attempt was made to reduce the impact of an ordering effect by reading back the ranking of the programmes and their WTP figures. Respondents were then allowed to change the ranking and the figures.* Twenty took the opportunity and made changes.† However, given that such changes require extra time and a concession that one has changed one's mind during the interview, it may be that more people could have taken this opportunity.

The influence of the length of the interview (Table 3) indicates a general methodological lesson for WTP studies: that the stated WTP for programmes presented after the first one can depend on how much time the interviewee takes to accept and digest the partial nature of the WTP exercise. These results suggest an income effect, though it is small. Further experimental research is needed to disentangle the effects of ordering and income.

Explaining absolute and relative WTP

Compared with the adjusted $R^2$ obtained from WTP studies in the value of life literature (see Kidholm, 1992), the models used in our analysis perform well.

The precoded reasons for stating a positive WTP indicate that altruistic reasons are more important than selfish for all three programmes. One might doubt that people are truthful in giving such ratings. However, the regression analyses do suggest that people were willing to back up their claimed altruism with higher WTP.

Partial studies of WTP are generally limited in their ability to suggest anything with respect to how much the relative valuation of a particular programme is associated with which background variables. In this paper, the dependent variable, the share of total WTP for the three programmes, was introduced as an indicator for the relative preference of each programme compared to the total. To our knowledge, this is a new way of analysing such data. This approach provides new insights as compared with the partial models of Table 3. First, it supports the intuitive hypothesis that men have stronger relative preference for hearts and women for hips. Second, the relative preference for hearts and for hips increase with age. As this analysis was designed specifically for this study, it is important to test whether these results can be replicated. If so, this approach could prove fruitful for analysing relative preferences for health care programmes, and may hence be applicable in a priority setting context.

Distributional issues

As pointed out in footnote 1, it may seem contradictory to use WTP as a method for accommodating preferences with respect to equity. This is because WTP is often criticised for being associated with ability to pay.

It is important to establish two points here. First, non-monetary techniques for valuing health benefits also involve distributional assumptions, although this is not as explicit as in the case of WTP. For instance, with QALYs, it is assumed that one QALY is of equal worth to people no matter who they are or at what stage of life the QALY is gained. There is some tentative evidence to show that the assumption about one QALY being of equal value to everyone is erroneous. Williams (1988) has suggested, on the basis of survey work, that there may be some support amongst the population for discriminating in the distribution of marginal QALY gains in favour of the young, an idea which gained support even from the elderly members of the sample. However, similar to WTP, little has been done to progress beyond such results into specifying distributional weights.

Second, although one would expect an association between willingness and ability to pay, this does not, of itself, mean that results of all such studies are problematic in distributional terms. The issue is more subtle and depends on whether the preferences of those in lower income groups are different from those in higher income groups. If such preferences are similar, there is less of a problem. But if they are different, more care should be taken in interpretation of results. The challenge, then, is to be able to (at least) diagnose whether ability to pay is distorting results. In the study reported in this paper, we are limited to analysis of the distribution of WTP for each programme across income groups. If ability to pay is more of an influence on WTP for one programme relative to the other two, the distribution of total WTP across income groups for this programme will be of a different pattern from those of the other two. If the

---

*Some interviewees felt it was difficult to be sure about how much they could be willing to pay, but held that their valuations should be interpreted as their relative ranking between the three programmes.

†There were seven changes in WTP figures for the helicopter (five up and two down), 11 changes on hearts (eight up and three down) and 10 changes on hips (eight up and two down).
patterns of these distributions are similar, ability to pay can be said to have had little (or no) impact on relative valuations. Examination of WTP valuations across 11 income groups, ranging from “NOK 50,000 per annum or more” to “less than NOK 50,000 per annum”, revealed similar patterns (Donaldson, 1996). Thus, it would seem that the preferences of one (or a few) income groups did not distort the valuation of one programme relative to the others.

Community focus

Altruism might well have been influenced by a possible “community bias” in this study, which was particularly concerned with describing the programmes as public sector health care. Rather than consistently using the term “willing to pay”, we phrased the questions in terms of “willing to contribute”. This is a further way in which the design of this study complies with the recommendations of Arrow et al. (1993). The intention was to emphasise that we were concerned with the valuation of health care programmes that were meant for the whole community depending on need, and not with private insurance based health care only for those who had paid. As long as health services are publicly financed, we would argue that the correct framing is the one chosen.*

Following the suggestion of Gafni (1991) to frame health care programmes in terms of an individual insurance option, concern for others may not be included, although Gafni argues that such concern can be included (i.e. by asking how much a respondent would be willing to pay so that a programme were available to him/her and others). The notion of probability of use of services can still be incorporated into our framing though, but by describing health care actions in terms of how many people in the community are likely to benefit from them in any period of time. Given the size of the community is known or presented to the respondents, this is equivalent to providing people with a probability, but a community rather than an individual probability. Further research is required comparing these two ways of expressing probabilities.

WTP vs QALYs

The most striking thing from the comparisons of WTP and QALYs is the much lower valuation in terms of WTP for QALYs gained from hip replacements. This might reflect the attitude that “hip replacements don’t save lives”, i.e. that QALYs gained from life improvements are valued lower than those gained from life extending or life saving. Further, the higher valuation of heart operations might be influenced by a perception of the effectiveness of the programme existing before the marginal 80 would be added. Also, hearts charities and media have a tendency to exaggerate efficacy. The highest valued QALYs are those from the helicopter ambulance, which might reflect a preference for the “Rule of Rescue”. Life savings from this programme take place at a very critical risk level. Hence, this WTP study lends support to the view that life saving is more valued than the same QALY outcome from life extension and life improvement (see Nord, 1993).

Because the hearts and hips programmes were to the benefit of a wider community (Northern Norway with 420,000 inhabitants) than the helicopter programme (Troms county with its 145,000), this might well have influenced the respondents to place a higher value on the helicopter service. When adjusting the valuations for these differences in probabilities of own use, the ordering of hearts and helicopter changes, but still life improving QALYs are valued lower than life extending and life saving QALYs.

The type of comparisons as outlined in Table 6 could represent a fruitful approach to weighting the relative value of QALYs gained from various health care programmes. One measurement problem, though, is the evidence that WTP does not increase with the size of the good (see Arrow et al., 1993), which implies that it is the programme per se, rather than the number of patients treated by the programme, which is important for one’s WTP. While this is a general measurement problem for CV studies, it has particular relevance in the suggested context of using WTP to weight QALYs. We are certainly facing a problem if it is the case that respondents ascribe diminishing marginal WTP for additional patients treated, while the comparative QALY yardstick values additional patients at a constant marginal value. In other words, if the QALY denominator changes at a different speed from the WTP numerator, one should be cautious to generalise from the WTP/QALY comparisons in Table 6. Results expressed in this paper are therefore tentative. We certainly think that further experimental research on this measurement issue is required.

Acknowledgements—We acknowledge the helpful comments from Jostein Grytten, Ivar S. Kristiansen, Gavin Mooney and Erik Nord, as well as colleagues in the economics department in Tromsø and an anonymous referee. The usual disclaimer applies. This study has benefited from the financial support of the Norwegian Research Council (NRC) and the Scottish Office Department of Health (SODoH). However, the views expressed in this paper are those of the authors, not NRC or SODoH.

*It is worth noting that when attempting to recruit new members, The Norwegian Air Ambulance emphasises a community focus: “By joining the foundation The Norwegian Air Ambulance, you contribute to making life safer for a large part of the Norwegian population, yourself included”.

Jan Abel Olsen and Cam Donaldson
REFERENCES


APPENDIX A

The Descriptions Of The Health Care Programmes

The descriptions* were presented on separate cards (each separate description was presented on a single card using larger typescip).

*In framing the descriptions we wish to thank E. Nord, I.S. Kristiansen and O.H. Ferde, and from the University Hospital in Tromsø: Professor K. Rasmussen, Chief Consultants O. I. Solem and J. Thoner.

*In framing the descriptions we wish to thank E. Nord, I.S. Kristiansen and O.H. Ferde, and from the University Hospital in Tromsø: Professor K. Rasmussen, Chief Consultants O. I. Solem and J. Thoner.


**APPENDIX A**

The Descriptions Of The Health Care Programmes

The descriptions* were presented on separate cards (each separate description was presented on a single card using larger typescip).

*Helicopter Ambulance*

The helicopter ambulance is a supplement to car ambulances. It carries a doctor and provides quicker transport to the hospital.

The helicopter is primarily a benefit in connection with *acute illnesses and accidents* for the 80,000 people in Troms living outside Tromsø and Harstad.

The helicopter has roughly 260 missions each year. Studies have revealed that most people who have been flown would have had same probability of survival had a car ambulance been used. Doctors estimate that each year the helicopter will save *five human lives*, with a total of 150 life years, that would not be saved if car ambulance had been used.

*Heart Operations*

Eighty more heart operations can be provided each year in addition to the 300 which are currently done in Northern Norway.

Most heart patients are men aged 50–60 years. They have chest pain and breathe heavily when strained.

The operation will make 75% of the patients completely *free from pain* with less pain for the rest. Without the operation the patients are expected to live 8–10 years. With the operation they will on average *live one to two years longer*.

The operation mortality risk is 2% (so one in 50 people die whilst being operated on).

*Hip Replacements*

A total of 250 more hip replacements can be provided each year in addition to the 600 which are currently done in Northern Norway.

Most hip patients are women and on average 70 years old. They have problems in walking and have severely reduced physical capabilities.

The operation will make 95% of the patients completely *free from pain* with less pain for the rest. Without the operation the patients are expected to live 8–10 years. Withdraw the operation they will on average *live one to two years longer*.

The operation mortality risk is 2% (so one in 50 people die whilst being operated on).
The Payment Card*

<table>
<thead>
<tr>
<th>NOK</th>
<th>0</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>100</th>
<th>150</th>
<th>200</th>
<th>250</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>800</th>
<th>1000+</th>
</tr>
</thead>
</table>

Make a ✓ by each amount you are sure you would pay.

Make an X by each amount you are sure you would not pay.

Make a O around the amount which is the maximum you would be willing to pay each year.

The health status measurement literature does not provide index numbers of health states which would mirror the health gain of the hearts and hip patients as described on the cards used in this study. Therefore, some data from Torrance (1987) were used and adapted.

For the 80 patients who would benefit from a heart operation, their initial health status index was set at 0.85. After treatment, an index of 1 was set to reflect the health state for the 75% who would be free from pain, and 0.9 for the remaining 25% with less pain. The quality increment is described to last 8–10 years. Using nine years, each patient’s expected gain from the health enhancement part becomes:

\[(1 - 0.85)*0.75 + (0.9 - 0.85)*0.25\]*9 = 1.125 QALYs.

Furthermore, an additional life quantity increment is described to last one to two years. Using 1.5 years, each patient’s expected gain from the life extension part becomes:

\[(1*0.75 + 0.9*0.25)*1.5 = 1.4625\] QALYs.

Adding these two parts and adjusting for the described survival rate of 0.98 gives each patient’s expected QALY gain from the programme:

\[(1.125 + 1.4625)\] QALYs * 0.9825 QALYs.

For the 250 patients who would benefit from a hip replacement, their initial health status index was set at 0.7, and after treatment at 1. This gain is described to last for the remaining 15 years of patients’ lives. Each patient’s expected gain then becomes:

\[(1 - 0.7)*15 = 4.5\] QALYs.

Given that the QALY gains from the two programmes might be overestimated due to low initial health state indices and/or high indices after treatment, a simple sensitivity test is to halve the quality enhancement parts. That would give 2 QALYs per heart patient, and 2.25 QALYs per hip patient. Given the life extension part of a heart operation, a conservative estimate of the health enhancement does not affect the QALY gain very much.

*In framing the payment options, we acknowledge the helpful suggestions from Mike Jones-Lee.