

Ahlert, Marlis; Breyer, Friedrich; Schwettmann, Lars

Working Paper

How you ask is what you get: Willingness-to-pay for a QALY in Germany

DIW Discussion Papers, No. 1384

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Ahlert, Marlis; Breyer, Friedrich; Schwettmann, Lars (2014) : How you ask is what you get: Willingness-to-pay for a QALY in Germany, DIW Discussion Papers, No. 1384, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/97490>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

1384

Discussion Papers

Deutsches Institut für Wirtschaftsforschung

2014

How You Ask Is What You Get: Willingness-to-Pay for a QALY in Germany

Marlis Ahlert, Friedrich Breyer and Lars Schwettmann

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2014

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<http://www.diw.de>

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
<http://ideas.repec.org/s/diw/diwwpp.html>
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

How You Ask is What You Get: Willingness-to-Pay for a QALY in Germany

Marlies Ahlert, Friedrich Breyer and Lars Schwettmann

Abstract

We report results of a survey of a representative sample of the German population in which respondents were asked for their willingness-to-pay (WTP) for either an extension of their life or an improvement in their health corresponding to a gain of one quality-adjusted life year (QALY). While one version of the survey exactly copied the questionnaire used in the EuroVaQ project (Pennington et al. 2014), in other versions the wording and the survey technique were modified. The findings show first that Germans have no higher WTP for health gains than other Europeans. Second, the technique of posing the questions plays an important role when respondents are asked to imagine being in hypothetical situations. This clearly refers to the wording of the questions and the survey setting (personal or online interview). But even simple design elements such as offering an explicit option to say “No” right away greatly affect the answers, as does the position in the questionnaire that a particular question is given. This shows that in any attempts to base health care rationing decisions on the WTP of the population, where the latter is to be elicited using surveys, great care must be taken in designing the questionnaires.

JEL-Code: I18.

Keywords: willingness to pay, QALY, survey, Germany, EuroVaQ.

This version: May 20, 2014.

The research reported here was funded by the Deutsche Forschungsgemeinschaft (DFG) under the research grants AH 179/3 and BR 740/16. We are grateful to the EMB of the EuroVaQ Project for the provision of the original questionnaire and to Cam Donaldson and Sue Bell for their steady support in the phase of creating the German version. Statistical support by Maria Goehring and Clara Wolff and valuable comments by Hartmut Kliemt, Normann Lorenz, Robert Nuscheler, Mark Pennington, Winfried Pohlmeier and Joachim Winter are gratefully acknowledged.

Authors' Affiliations: Ahlert, Schwettmann: Martin-Luther-Universität Halle-Wittenberg;
Breyer: Universität Konstanz and DIW Berlin.

Corresponding Author: Prof. Dr. Friedrich Breyer, Fachbereich Wirtschaftswissenschaften, Universität Konstanz, Fach D 135, D-78457 Konstanz, Phone (+49-7531) 88-2568, Fax -4135, e-mail: friedrich.breyer@uni-konstanz.de

1. Introduction

Health care systems in developed countries are facing tremendous financing problems, given the rapid medical progress and limited resources from public or semi-public funds such as payroll taxes. In particular with respect to decisions on financing new and innovative health care technologies, every health care system must find rational methodologies to assess value for money. The procedures currently in place differ widely even between countries with similar GDP per capita. E.g. in England and Wales, the National Health Service provides a particular treatment to the population if its "costs per QALY gained" does not exceed a certain threshold, which lies between 20,000 and 30,000 GBP (NICE 2007, p.54), and the assessments are provided in a transparent process by the National Institute for Health and Clinical Excellence (NICE). In contrast to this prototype of open and explicit rationing, the covering decisions for German Social Health Insurance ("Gesetzliche Krankenversicherung, GKV") are made on a case-by-case basis by a decision-making body called "Gemeinsamer Bundesausschuss" (G-BA) with no obvious or transparent decision criteria.

No matter how explicitly and openly the decisions are taken, it is justified to require that they somehow reflect the preferences of the population which is affected by them both as potential recipients of medical services (patients) and as payers of taxes or social insurance contributions. Therefore, it would be desirable to know what value citizens place on the gains in health and life expectancy that can be achieved with the respective medical treatments. Thinking about such gains in terms of quality adjusted life years (QALYs), it is thus desirable to estimate a monetary value that members of society place on additional QALYs.

There is a literature on the monetary valuation of health, which falls into two categories: in the value-of-a statistical-life (VSL) literature, the object to be valued is (the avoidance of) a small risk of immediate death. This case is especially relevant in fields involving fatal hazards such as traffic, dangerous occupations or accidents such as fires. The other category tries to assess the monetary value of an additional (healthy) life year, a QALY, or some gain in health status over a period of time. It is obvious that gains (or avoidance of losses) of (more) healthy lifetime are the typical target of medical services so that this second branch is of greater relevance in the economics of health care. For a survey of the state of knowledge in this literature see Donaldson et al. (2010, p.11f.).

Over the last years, a group of experienced health economists from nine European countries and Palestine has tried to elicit the "monetary value of a QALY" in a research project called "EuroVaQ" (European Value of a QALY), which was funded by the Commission of the European Union under the Sixth Framework Programme. The main methodology consisted in online-surveys of approximately 4,000 persons in each of the participating countries that were conducted in late 2009 and early 2010 (Donaldson et al. 2010, Robinson et al. 2013, Pennington et al. 2014).

The survey questionnaire came in two versions. In each version a different approach was used for framing the hypothetical decision situations, and each approach was used for approximately one-half of the sample in each country. The first approach (“chained approach”) is based on the assumption that respondents are rational expected-utility maximizers. It uses a series of standard-gamble or time-tradeoff questions to translate the WTP for a small and everyday health gain described by comparing two vectors of EQ5-D health states into a WTP for a fraction (.1 or .05) of one QALY. One notable feature of this approach is that at least one of the standard-gamble or time-tradeoff questions which had to be answered in each of these series involved the option of immediate death with a very small probability.

In contrast, the “direct approach” tries to describe the gain of a QALY to the respondents without actually using the word. First a visual analogue scale, called “health thermometer”, is used on which 0 marks “death” and 100 “perfect health”, and respondents are asked to rate their own health on this scale. Then, using visual means, health gains and losses that last for a certain number of years are denoted as rectangles in a diagram in which time is measured along the horizontal and health along the vertical axis. Finally respondents are asked for their WTP to avoid a health loss of x points on the scale that lasts for $100/x$ years.¹

The results of the surveys using the direct approach show that mean and median responses differ considerably from country to country but also across survey questions (Pennington et al. 2014). The main message of the authors is that mean willingness-to-pay (WTP) values for a 1-QALY gain in the form of life extension is considerably higher than for an equal gain in quality of life, so that one may conclude that “a QALY is not a QALY”. What is equally striking is the fact that the elicited WTP values are very low. When, as is usual in online surveys, the top 1% answers are “trimmed”, the mean WTP-per-QALY values range between 6,000 and 21,000 USD (or roughly 4,000 to 14,000 GBP), and the medians are still lower (between 150 and 2,200 USD or between 100 and 1,500 GBP). The latter values are more than an order of magnitude smaller than the much debated threshold of 20-30,000 GBP for “cost-per-QALY gained” used by NICE in England and Wales in their funding recommendations. Moreover, between 20 and 45 per cent of respondents even expressed a WTP of zero for substantial gains in length and quality of life. These numbers throw considerable doubt on the validity of the results.

The largest EU member country, Germany, was not among the countries in which the survey was conducted. This fact alone would have suggested conducting a similar research in Germany in order to gain an understanding of the patterns which govern the differences in monetary valuation of health gains across Europe. In addition, using cost considerations in health care rationing is known to be particularly unpopular in Germany (Breyer 2013), which may suggest that WTP for health gains is especially high in this country.

Furthermore, such an endeavour could be used to examine whether the results of the original research, in particular the low median responses, may be due to either the online nature of the

¹ Some questions involved health gains of fractions of a QALY so that the 100 is replaced by some smaller number.

survey or a lack of understanding of the hypothetical nature of the questions by part of the respondents, i.e. in the difficulty of putting oneself in the position of someone who is sick *and* has to pay for health care himself. In addition, the large share of zeroes might be an artefact of the questionnaire structure because here a two-step procedure was used in which the respondents were first asked if they were willing to pay at all and only those who answered “yes” were asked for the amount of their WTP.²

For all of these reasons, the EuroVaQ study was extended between June 2012 and February 2014 to a representative sample of German citizens. As the chained approach uses alternatives involving small risk of death and it is well-known from the literature on the value of a statistical life (VSL) that many people have great difficulties in making consistent choices in such situations (Viscusi 1993, Hammitt and Graham 1999), we decided to administer only a questionnaire based on the direct approach. In addition we judged the direct questions to be easier to understand. Altogether, we used four different versions of the survey questionnaire and procedure:

Survey I: a direct translation of the English questionnaire³ was administered online to 1,501 respondents in June and July 2012.

Survey II: the questionnaire was modified so as to improve the understanding by the respondents of the hypothetical nature of the scenarios (on the details see Section 3) and was administered online to 1,500 respondents in February 2014.

Survey III: the questionnaire used in Survey II was administered in computer-assisted personal interviews to 507 persons in June and July 2012.

Survey IV: the questionnaire used in Survey II was modified by replacing the two-step procedure by a one-step one, i.e. by skipping the yes/no questions, and administered to 1,500 persons in February 2014.

In this paper we report the procedure and the results of the four German surveys on the value of a QALY (“GermanVaQ”). Section 2 briefly states the hypotheses to be tested with this study. In Section 3 we describe the survey questions and emphasize the differences between the survey versions. Section 4 contains a descriptive account of the results of the surveys, and in Section 5 we analyze the pattern of responses with respect to demographic and socio-economic determinants of WTP and the dependence of these monetary valuations on the present health status, size and timing of the expected health gain, and the ordering of situations in the survey. Finally, Section 6 concludes.

² In a recent study by Gyrd-Hansen et al. (2014) on framing effects in WTP questions, such a “binary response filter” resulted in higher frequencies of zero responses. The authors also detected a higher mean WTP among those respondents who were willing to pay a positive amount, but no effects on median values.

³ The translation was provided by the second author.

2. Theoretical hypotheses on WTP for a QALY

We first restate a number of hypotheses underlying already the original EuroVaQ study (*H1* to *H7*) before we formulate additional hypotheses which refer to the differences in survey design and between countries (*H8* to *H11*):

H1: WTP per QALY is higher when only fractions of a QALY can be gained (theoretically derived from the budget constraint and diminishing marginal utility of health).⁴

H2: WTP for a QALY decreases with the time span between payment and realization of the promised health gain (due to discounting).

H3: WTP for a QALY is higher when gains arise from extension of life rather than from health improvements for a given life span.

H4: WTP increases with age (because older people are more aware of the value of health).

H5: WTP for a QALY increases with the income available to the person (due to the budget constraint).

H6: WTP for life extension is higher for persons with a family than for singles (due to positive externalities).

H7: WTP for a QALY decreases with current health status (theoretically derived from diminishing marginal utility of health).

H8: Emphasizing the hypothetical nature of the survey reduces the share of protest-zeroes and increases mean and median WTP.

H9: Skipping the yes/no question reduces the share of zeroes.

H10: Administering the questionnaire in personal interviews reduces the variance of stated WTP.

H11: WTP for a QALY is higher in Germany than in most EuroVaQ countries.

3. The design of the questionnaires

As mentioned above, Survey I in the GermanVaQ study is a direct translation of the questionnaire used in UK into German. Therefore, the reader is referred to Donaldson et al. (2010, pp. 57ff.) for a detailed description.

In the introduction, the respondents are informed that the survey is part of a research project inquiring into the value placed by citizens on their own health and that their answers would be useful to inform governments in making decisions on allocating resources to and within the health care sector. Then the hypothetical nature of the questions is stressed and it is explained that “the amount you would be willing to pay for some treatments *if you had to* gives an indication of *how*

⁴ In the introduction of the questionnaire, respondents learned to interpret descriptions of health states as rectangles displaying the health and time dimension. Therefore, health gains are reflected by the area of the rectangle and we assume marginal utility of the size of these areas.

much you value health gains from those treatments compared with other things you might want to spend your money on”. The combination of these two pieces was designed to convey the message to respondents: If you value your health highly and if you think that the government should spend more on health care, then you should put large numbers in the answering boxes.

Surprisingly, in each of the questions of the corresponding version of the EuroVaQ survey, there was not only a sizeable fraction of people who expressed a WTP of zero for the corresponding health gain (of 1 QALY in most cases), but also, when asked to give reasons for their choice, many of these respondents ticked the box “I do value the treatment, but do not want to pay because the government should provide health care”. This clearly shows that they did not get the message mentioned above because they misunderstood either the purpose of the study or the hypothetical nature of the questions.

Therefore, in the alternative version of the GermanVaQ questionnaire used in Surveys II – IV, we added the following paragraph to the introduction:⁵ “Imagine that there are *no sickness funds* in Germany so that you have to pay no premiums or contributions for health insurance. Thus your net income is higher *by the respective amount* than it is in fact today. In return, you must pay *for every medical service out of your own pocket*. As you have known this for a long time, you have accumulated savings of one year’s income to be prepared for unexpected medical expenditures.”

In addition, in this version we reminded the respondents in every question for their WTP for a particular health gain that they should place themselves in this hypothetical situation. This was done with the following words:⁶ “Remember that we assume here that there is *no health insurance* and you have to pay all medical services yourself, if necessary from your savings that you have made for this purpose.” Finally, when asking for the specific amount, the question reads:⁷ “What would be the *highest* amount you would pay for this treatment today at age x, if there was no health insurance which pays for it and if you had the equivalent of one year’s income in your savings account?”

In the beginning, respondents were asked for their age, gender, occupation, region of residence (which of the 16 Länder), family status, number of children of different age groups, household size, household income, type of health insurance (public vs. private), expected life span and,

⁵ The German original reads: “Stellen Sie sich vor, es gebe in Deutschland *keine Krankenkassen* und Sie müssten daher auch keine Beiträge oder Prämien für eine Krankenversicherung zahlen. Ihr Nettoeinkommen wäre also *um den entsprechenden Betrag höher*, als es heute tatsächlich ist. Dafür müssten Sie *jede medizinische Behandlung selbst bezahlen*. Da Sie dies schon lange wissen, haben Sie ein Sparvermögen in Höhe eines Jahreseinkommens aufgebaut, um für unvorhergesehene Behandlungskosten gerüstet zu sein.”

⁶ The German original reads: “Denken Sie daran, dass wir hier unterstellen, dass es *keine Krankenversicherung* gibt und Sie alle medizinischen Behandlungen selbst bezahlen müssen, notfalls aus Ihrem dafür angesparten Vermögen.”

⁷ “Was wäre der HÖCHSTE Geldbetrag, den Sie HEUTE im Alter von x Jahren für diese Behandlung zahlen würden, wenn es keine Krankenversicherung gäbe, die dafür bezahlt, und wenn Sie ein Vermögen in Höhe eines Jahreseinkommens auf dem Sparkonto hätten?”

most importantly, current health as a point on the 0-100 health thermometer. Based on the latter two answers different exclusion criteria were applied throughout the questionnaire to make sure scenarios were applicable to individual participants. Respondents who stated a health level below 20 points or a life expectancy lower than two years were not included in the sample. Furthermore, if life expectancy was less than six years people were redirected to the “grey block” containing four scenarios which are still meaningful for them.⁸

As in the EuroVaQ questionnaire, there were altogether 13 different scenarios of which each respondent was presented 4 or 5. Table 1 contains an overview of the scenarios and gives information of the nature and timing of the health gain and the number of QALYs involved.

[insert Table 1 here]

For each of the scenarios, which consisted of the description of a health loss, which could be avoided by getting a medical treatment, it was first asked,

- a) *whether* the respondent was willing to pay something (only in Surveys I – III).
- b) If the answer to this question was “No”, the respondent was asked to give reasons why (Donaldson et al., 2010, p.55).
- c) If the answer to question a) was “Yes”, or in any case in Survey IV, the respondent was confronted with a screen in which different amounts (from 10 Euros to 300,000 Euros) appeared in a random order, which he was asked to allocate to one of three columns “willing to pay”, “not willing to pay” and “unsure”. In the end, he was presented the largest amount of the first and the smallest amount of the second column and was asked again what amount within this interval constituted his “maximum” WTP. Respondents who were not willing to pay at least 10 Euros could enter even smaller non-negative amounts and were asked to state their reasons if they entered a zero amount.

At the end of the questionnaire, respondents were confronted with all the answers they had given in the course of the survey and were given the opportunity to change their responses.

Survey III contained additional items that referred to extensions of the benefit package of the statutory health insurance. Hence, it was necessary to include only respondents with public insurance in the sample of Survey III. In contrast, participants of the three other surveys hold either private or public health insurance. As only privately insured individuals in Germany are used to pay for medical treatments (and are later reimbursed by their insurance company), the insurance status may have an effect on the evaluation of the hypothetical scenarios in the questionnaire. However, 90 per cent of the German population are members of public health insurance so that the bias introduced by this difference when comparing the surveys should not be too serious. Moreover, when we try to explain differences in WTP across respondents (see Section 5), insurance status will be used as one of the explanatory variables so that we can check the validity of this

⁸ The “grey block” includes scenarios D, E, I and J (see Donaldson et al., 2010, p.61-62). For several questions further criteria based on life expectancy or health state were applied which led to either an exclusion or a redirection to the “grey block”.

conjecture. In other dimensions, gender, age, and region (Länder), the German sample was created so as to ensure representativeness for the German population as a whole, although Survey III only included respondents with public health insurance. This goal was broadly achieved. However, people over 59 years are underrepresented, while especially respondents aged between 30 and 44 are overrepresented. Similar to many samples of the EuroVaQ countries (Donaldson et al., 2010, p.41 and Table 3.3) this is mainly due to an underrepresentation of elderly women. Furthermore, Survey I contained slightly more individuals from the East of Germany including Berlin (see Table 2 for sample characteristics).

[insert Table 2 here]

4. Descriptive Results

4.1 Zero willingness-to-pay

One striking result of the EuroVaQ study was that a sizable share of respondents (20 to 45%) expressed a zero WTP in any individual scenario (Pennington et al. 2014, Table I), and a similar result is found for the corresponding German survey (no. I). In the columns in the middle of Table 3 the corresponding shares for all versions are reported. The percentages vary tremendously across scenarios and, moreover, between surveys. In Survey I the highest proportion of zeroes (46.0%) is found in scenario I (a very small reduction of chances of dying in any given year amounting to one additional QALY at the end of life), which seems reasonable because in particular young respondents will heavily discount any life extension at the very end. In contrast, only 17.3% did not want to pay anything for a health gain of 25 points for 4 years in the near future (scenario A). On average over all scenarios, almost one-third (32.6%) of all respondents stated a WTP of zero. In Survey II, where the respondents were reminded that there was no health insurance, this share dropped only slightly to 29.2%, and when an interviewer was present it dropped further to 23.2%. The most striking difference, however, can be seen by comparing the results of Survey IV in this block with the other three. When respondents are asked directly for their WTP, skipping the introductory yes/no question, the share of zeroes drops to between 1.4 and 9.0%, with an average of just 3.0%. In line with the results of Gyrd-Hansen et al. (2014), this shows that the large share of zeroes in the other three surveys (and in the EuroVaQ study as well) is an artefact of including the yes/no question, which gives the number 0 a special prominence. Thus this finding is a striking support for *Hypothesis H9*.

[insert Table 3 here]

In Section 3 it has already been mentioned that some respondents did not get the basic message of the hypothetical situation where no government exists which may carry health care costs. This can be deduced from their statement that “the government should pay for health care”. In the EuroVaQ project respondents who stated only this reason were called “protestors”, and their share was about 7% (Pennington et al. 2014, Table I). From the numbers in the right block of Table 3 it can be seen that in our Survey I between 3.2% and 6.6% of respondents in each scenario can

be classified as “protestors”, with an average of 4.6%. This number drops considerably to around 2.5% in the Surveys II and III in which the respondent is reminded that “there is no health insurance” to pay for the treatment. Finally, in line with the total share of zeroes, also the share of protestors drops sharply in Survey IV to an average of .3 per cent.

4.2 Mean and median WTP for a QALY

In this section, we take a closer look at mean and median responses to the respective scenarios in all 4 surveys.⁹ While for normative purposes, e.g. for maximizing a utilitarian welfare function mean WTP is the appropriate measure, other reasons support focussing on the median values. As decisions on health care rationing are ultimately made by politicians seeking re-election in democratic countries relying on majority voting, mean preferences in the population should be less important than median preferences. Furthermore, the distribution of WTP responses is heavily right-skewed so that the mean is strongly dependent on a few observations in the right tail of the distribution while the median is much more robust to these outliers.

[insert Table 4 here]

Similarly to Pennington et al. (2014, Section 2.5) and omit both the “protestors” and the top 1% WTP responses in each scenario. For this “reduced sample”, Table 4 contains the maximum, mean and median responses to each of the 13 scenarios in all 4 surveys. As Survey III was conducted only among members of public health insurance (GKV), we present the respective data for the subset of GKV members for all surveys in Table 6 (in the Appendix). In scenarios D, E, M, O, and N, which refer to health gains of less than 1 QALY, all responses were divided by the respective fraction so that the numbers in the tables denote the implied “WTP for 1 QALY”, assuming strict proportionality of WTP. When calculating simple averages of the means and medians across scenarios, however, we distinguish between the subset of the first 8 scenarios, in which the WTP for a health gain of 1 QALY is directly elicited, and the last five in which only a fraction of a QALY is at stake. It is obvious from Tables 4 and 6 that higher mean and median amounts can be observed in the latter 5 scenarios (see the bottom part of both tables), which confirms *Hypothesis H1*.

Furthermore, some scenarios regard health gains in one year’s time, while others consider similar situations at the end of life. As expected (see *Hypothesis H2*), mean and median values are always higher in scenarios A, F and L compared to the corresponding cases described in B, G and I. We also conducted two-sided t-tests to compare mean values and median tests to investigate whether answers to the two scenarios in each pair differ in central tendency (see Siegel and Castellan, 1988). For all surveys, the tests reveal statistically significant differences between all but two pairs of scenarios.¹⁰ Hence, overall the WTP for a QALY depends on the timing of the gain.

⁹ Respondents had to state their amounts in Euros. However, as the surveys are from different years and in order to make results comparable to the EuroVaQ results, amounts have been converted to US\$ by using Purchasing Power Parity conversion rates for 2008 and by including inflation rates (IMF, 2013).

¹⁰ The two exceptions in Table 4 are: Mean values of scenarios A and B in Survey III ($p=0.061$) and median values of scenarios F and G in Survey IV ($p=0.555$).

Furthermore, in scenario A, respondents had to give one single value for a gain of one QALY, while in scenario P four amounts (over four years) had to be stated for a similar gain. Here, mean and median values are consistently higher in the latter case so that the payment procedure also seems to have an influence on the WTP.

In scenarios I, J, and L increases in longevity are considered while all other scenarios regard increases in quality of life (*Hypothesis H3*). From the mean values we detect some higher WTP for a one-QALY gain arising from life extensions, but especially median values do not show significant differences between increases in longevity and increases in quality of life.

Next we turn to the question whether Germans display a higher WTP for health gains than other Europeans (*Hypothesis H11*). To this end we compare the answers to our Survey I to the ones reported in Donaldson et al. (2010, p.94) for the 9 European countries in their study. In Figures 1 and 2, we thus compare the averages of the mean and median values, taken over the first 8 scenarios. Especially the figures for the median values show that Germany ranks well in the mid-field so that, surprisingly, *Hypothesis H11* is refuted.

[insert Figures 1 and 2 here]

Comparing the mean and median values in Table 4 between Surveys I and II, we find that in the latter, both means and medians are higher than in the former. As could be expected, the relative increase in the medians in Survey II compared to Survey I is even larger than in the means. Thus emphasizing the hypothetical nature of the questions more strongly leads to higher stated WTP values, which is in line with *Hypothesis H8*. This comparison also casts doubts on the validity of the elicited WTP in the EuroVaQ survey, which used the questionnaire without the additional reminders.

We next analyse whether the lower share of zeroes in Survey IV also affects the means and medians in the respective scenarios. If all respondents who would have chosen the “no” option in Survey II stated low positive numbers in Survey IV, one could imagine that at least the medians should be similar in these two surveys. This effect has been observed by Gyrd-Hansen et al. (2014). However, Table 4 clearly refutes this conjecture for the present study since the medians in Survey IV are almost consistently at least twice as high as in the corresponding scenarios in Survey II. Median tests reveal that respective values are always significantly higher in Survey IV. So the option of choosing “no” not only increases the share of zeroes but apparently affects at least the lower half of the distribution of responses. This again casts doubts on the validity of the results in the EuroVaQ survey where this option was offered.

Finally we examine whether the answering behaviour changes when the respondent is confronted with an interviewer. To this end we compare the results for Surveys II and III in Table 6, which is confined to publicly insured respondents. We first observe that mean and median responses are consistently higher in Survey III than in Survey II, although results of statistical tests on differences between respective values are mixed. If one accepts the general conclusion that the WTP values stated in these surveys are on the whole lower than the “true” WTP, then it follows

that values elicited in personal interviews are more valid than the ones from online surveys. However, *Hypothesis H10* is clearly rejected because in all scenarios the standard deviation of the responses in the personal interview exceeds the one in the otherwise equal online survey.

So far we have observed several variations between scenarios. Furthermore, differences between respondents with either public or private health insurance seem to exist if one compares results in Tables 4 and 6. In the next section we investigate the influence of different covariates by means of multivariate regression models.

5. Determinants of WTP for a QALY

In order to identify relevant determinants of WTP an appropriate regression procedure has to be selected. As revealed by the descriptive results presented in Section 4, a reasonable number of participants in each situation stated a WTP of zero. Facing similar results, Pennington et al. (2014) proceeded in two steps and used Logit regression to model the determinants of zero valuations, while they modelled positive WTP values using OLS regression after log transformation. In general, such a two-part procedure has the advantage of modelling two independent decisions, where all covariates are allowed to appear in all equations and, therefore, may display distinct effects on the different dependent variables considered. Hence, we applied a related two-part procedure called the “Cragg truncated normal hurdle model” (Cragg 1971; see also Greene 2012, Wooldridge 2010). More specifically, in the first part the probability of observing a positive WTP is predicted by a binary Probit model, while in the second part a truncated regression on the observations above zero WTP is estimated. While estimated coefficients can be interpreted in the same manner as OLS regression coefficients, a truncated model takes into account that the outcome variable is restricted to a truncated subset of its distribution, viz. positive amounts. Furthermore, as the distribution of positive responses is right-skewed, we chose log (WTP) as dependent variable in the second part.¹¹

We introduced two groups of covariates, individual characteristics and attributes of the scenarios. Descriptions of individual variables are given in Table 2. The selection of characteristics and their categories has been inspired partly by results of the EuroVaQ project. Explanatory variables were age in four age groups, monthly net household income in US\$ PPP for 2008 grouped into five brackets with identical thresholds for all surveys plus an additional indicator for “missing income”, three levels of education depending on the years at school and any type of university studies, gender, region (East Germany including Berlin versus West Germany) and self-rated health status (measured on a 0-100 visual analog scale) grouped into four brackets. Also, similarly to the EuroVaQ study a measure for size and composition of the household, called the OECD coefficient, has been introduced.¹² Finally, a dummy variable for having a full private health insurance was added.

¹¹ As a robustness check we also estimated OLS regression models at the second stage. The results convincingly confirmed the estimates of the truncated regression models.

¹² It distinguishes between single-person households and multi-person households, partitioned into couples with or

Based on the descriptions in Table 1, we identified three attributes of scenarios which entered the regressions as dummy variables. First, two factors denote whether a scenario is characterised by risky health gains (scenarios M, O and N) or certain gains of only a fraction of a QALY (scenarios D and E): Consequently the base category is a certain and full QALY. Secondly, life extensions (scenarios I, J and L) are contrasted with gains in quality of life. Third, gains at the end of life (scenarios B, G, and I) are separated from gains in one year's time. We also considered the order in which the questions had to be answered by including dummy variables for the position in the questionnaire. As some respondents had to answer 4 and others 5 questions, positions 4 and 5 were combined in one dummy. As we pooled the answers to Surveys I to III, we added dummy variables for the different survey versions.

The decision to state a zero WTP was taken separately in Surveys I, II and III, but not in Survey IV. This may lead to a sample selection bias at the second stage of the two-part model. Hence, at the first stage of the two-step procedure we estimated two binary Probit models on answers either from Surveys I-III or from Survey IV. A Wald test confirmed that the estimated coefficients in both models are indeed not the same. However, to investigate not only effects from characteristics of scenarios, but also from different survey methods, we pooled individual answers from the first three surveys and ran separate regressions for Survey IV. All samples contained several answers from each individual so that we allowed for clustering at the individual level by using respondents as primary sampling units (PSUs). Additionally, in each of the four surveys individuals have been sampled separately, so that they were divided into different strata.

[insert Table 5 here]

The results of the estimation of both parts are summarized in Table 5. As in the descriptive analysis above, "protestors" and the 1 per cent highest responses to each question are omitted. The statistical package STATA 12 was used to estimate all regression models. For the Probit models we report marginal effects rather than estimated coefficients, because they also reflect the extent of the effect observed. F-tests show that all models have more explanatory power compared to a model including only a constant. Apparently some independent variables display a statistically significant effect only on one of the two decisions considered. Hence, the decision to estimate a two-part model seems to be justified. Furthermore, we judge results from both samples as equally important. Finally, as expected the predicted probability to observe a positive WTP in Survey IV is very high, so that we will mainly focus on the truncated regression results for this sample. In the following we first consider characteristics of respondents and afterwards properties of the scenarios.

without children, sole-parent families and other private households with cohabitating members or families.

5.1 Effects of individual characteristics

The omitted age category was the 45 to 59 year group. In *Hypothesis H4* it has been stated that older people are supposed to be more aware of the value of health. This expectation is confirmed for the probability to state a positive WTP in Surveys I-III, which is lower in the two youngest age groups. In contrast, the amounts of stated WTP are significantly higher in the youngest age group than in the omitted group, whereas in Survey IV this effect cannot be observed.

Apparently, some very strong effects can be observed from the income dummy variables. In theoretical models this is expressed by the budget constraint (see *Hypothesis H5*). As the omitted category is the middle one, the coefficients clearly show that both the probability to state a positive WTP and the amount rise significantly and substantially with increasing income. Among the positive amounts, average WTP in the lowest (highest) income category is 76-83 per cent lower (28-41 per cent higher) than in the middle income group, depending on the survey. Unexpectedly, people who did not state their income had a lower propensity to state a positive WTP in Surveys I-III and also stated lower amounts in Survey IV.

Especially the estimated coefficients for the three middle income classes could be used to investigate the relation between WTP and household income in more detail for those individuals who stated a positive WTP. From the mid-points of the second, third and fourth income brackets (1,780, 2,450 and 3,290 US\$ PPP monthly, respectively, see Table 2) the relative differences from the lower and higher income bracket to the middle one can be calculated as -27.4% and $+34.3\%$. As the coefficients of income in the regression can be interpreted as percentage difference of WTP per QALY from the one in the middle income bracket, dividing them by the percentage differences in income yields the corresponding income elasticities. These values are quite substantial, when the second and third income bracket are compared (1.24 for Surveys I-III and .95 for Survey IV), and somewhat lower (.47 and .61, respectively) in going from the third to the fourth income bracket. Hence, the income elasticity of WTP per QALY appears to flatten out with higher income. In summary, similar to the EuroVaQ study, we detected a rather strong influence of income on WTP.

Holding income constant, individuals with a higher level of education displayed a significantly higher WTP, presumably because it needs some cognitive ability to evaluate health reductions described in the questionnaire and estimate monetary consequences. In contrast, the share of positive responses was not affected by educational status.

The current health status of a respondent may have a relevant effect on WTP for a QALY. According to our *Hypothesis H7* this is theoretically expressed by diminishing marginal utility of health. This hypothesis is hard to test because most respondents state rather high values for their own health on the 0-100 scale. Correspondingly, the evidence displayed in the regression results is rather weak. In Survey IV we can observe a higher probability to state a positive WTP in people with “good” health (70-89 points) as compared with the less healthy people (below 70 points),

whereas members of the highest health status group do not differ from the lowest group, which does not confirm *Hypothesis H7*.

Household size as measured by the OECD coefficient, which displays the number of family members or people in a household, affects only the amount of WTP in Surveys I-III significantly, but the sign is negative in both samples. This is in contrast to our *Hypothesis H6*. However, as we are holding family income constant, this may simply be an effect of the budget constraint: the larger the household the less can be spent for any single household member in case of illness.

The EuroVaQ group reports (Pennington et al. 2014, Section 3) that in some scenarios men were less likely to state a positive amount, but if they agreed to pay, they displayed a higher WTP. Our results of the two-part model confirm this observation. In Surveys I-III, women were significantly more likely (by almost 5 percentage points) than men to state a positive WTP. However, focussing only on those respondents who stated a positive value, men wanted to pay on average 21 per cent more than their female counterparts in Surveys I-III, and 17 per cent in Survey IV although the estimated coefficient is only significant at the 10 per cent level ($p=0.092$). In general, it seems that women are more aware of the necessity to pay something in order to improve their own health, but they are less willing to pay high amounts.

In Surveys I-III, East Germans wanted to pay about one-third less than West Germans. The corresponding value in Survey IV is 21 per cent ($p=0.077$). This effect cannot be due to income differences between both regions, as we are holding income constant. Note, too, that we have excluded all “protestors”, i.e. respondents who agreed to the statement that health expenditures should be financed by the government. Therefore, this kind of attitude cannot be the reason for the differences observed. There seems to be some additional reluctance to spend higher amounts for medical treatment in the “East”.

Finally, one could expect significant differences between respondents with private and public health insurance. As this is a specific characteristic of the German sample, corresponding results of the EuroVaQ group are not available. The estimated large and significantly positive coefficients in the second part of our model and in all surveys indicate that people with private insurance were willing to pay much more for a QALY, indeed. Again, this result holds although we have controlled for income effects. Probably the privately insured were more accustomed to pay for their health care and were also more aware of “true” costs.

To summarize this subsection, several individual characteristics of respondents were found to display a significant influence on both decisions regarded. Most of these observations are in accordance with the above stated hypotheses and with previous findings of the EuroVaQ group.

5.2 Effects of attributes of the scenarios

In Section 2 we have presented some hypotheses with respect to attributes of the scenarios. Furthermore, our descriptive results in Section 4 have already confirmed several of them. With the help of our regression results we are now able to make more robust statements.

Scenarios differed in the fraction of a QALY which could be gained by a medical treatment. In *Hypothesis H1* it was stated that WTP for a QALY is higher when only fractions of a QALY can be gained. Interestingly, the results of our two-part model suggest that the underlying decision process is probably more complex. In scenarios where only a fraction of a QALY could be gained (whether or not risk was involved) those who were willing to pay a positive amount indicated a value of a QALY which was considerably higher (on the order of twice as much or more) compared to those scenarios in which one QALY could be gained. This observation confirms the initial hypothesis.

In scenarios that involved the extension of life, significantly fewer respondents stated a positive WTP than in scenarios involving gains in the quality of life. We have already remarked in subsection 4.1 that scenario I, which describes a very small reduction of chances to dying in any given year, lead to the highest shares of zero WTP in Surveys I-III. However, we find opposing results for the positive amounts stated in the three life extension scenarios compared to health improvements. Amounts for life extensions are higher in Surveys I-III and, thereby, confirm our *Hypothesis H3*. In contrast, in Survey IV stated amounts for health improvements are significantly higher.¹³ Thus, the preceding yes/no question has had a considerable impact on the stated amounts of those respondents who initially agreed to pay at least something. Survey design might have had a considerable impact on the evaluation of different gains of one QALY.

Finally, some pairs of scenarios differed with respect to the time span between payment and health improvement. Payment is in most cases described to be now whereas the effect of treatment could materialize either in one year or at the end of life. The regression results reveal that participants in Surveys I-III were considerably less likely to pay a positive amount for a health gain that was said to take place at the end of their life rather than in one year's time, whereas in all surveys they stated significantly lower amounts (by almost one-half). These findings confirm *Hypothesis H2*.

5.3 Effects of the survey technique

The most striking result concerns the effect of the survey design on the probability to state a positive WTP: being confronted with a personal interviewer increases this probability by about 8 percentage points, while giving the choice of zero a much less prominent place in Survey IV significantly raises it by 23 percentage points compared to Surveys I-III. This latter result is derived from pooling all surveys and estimating a further Probit model including a dummy variable for Survey IV.

But even the position in the questionnaire has a strong effect, especially in Surveys I-III: in the first question, the share of positive amounts is highest, and it drops by a full 6 (8) percentage points for the second (fourth) question. Interestingly, the share is not lower for the third question,

¹³ Further regression results (not reported in Table 5) reveal that especially in scenario I amounts are rather low in Survey IV.

presumably because “shirkers” who chose the zero option to save time did not dare to do so three times in a row. In contrast, the amounts stated are significantly increasing in the course of answering in all surveys: in the second (third, fourth or fifth) question, the average WTP is 16-20 % (35-68%, 60-69%) higher than in the first, depending upon the survey. Apparently, when a respondent gets used to these difficult question, he or she is getting more courageous to state high amounts.

Finally, the amounts stated depend a lot on the wording of the questions: the simple trick of emphasizing and repeating the assumption that there is no health insurance which would pay for the treatment increases average amounts by a full 55 per cent, and presenting the respondent with a personal interviewer increases this gap to 114 per cent!

6. Conclusions

In this paper, we reported the results of a survey of a representative sample of the German population in which respondents were asked for their willingness-to-pay for either an extension of their life or an improvement in their health for a given period of time. While one version of the survey (Survey I) exactly copied the setting (online) and the questionnaire used in the EuroVaQ project, in the second version (Survey II) we took greater efforts to persuade test persons to imagine a situation in which there was no other source of funding for their health care than their own payment. A third version (Survey III) with the improved questionnaire was conducted as a computer-assisted personal interview, and in the last version (Survey IV), we skipped the question whether the person was willing to pay at all for the respective health gain.

The findings throw considerable light on the two main research questions of this endeavour: First, is Germany different?, and second, is the survey technique important for the results? Surprisingly, the first question must be answered in the negative: when the same questionnaire is used, the WTP values found in Germany are similar if not even lower than the ones in comparable other European countries. If these values are taken seriously, they throw doubt on the notion that people would give everything they have for living longer or being healthier.

Moreover, and perhaps more importantly, the technique of posing the questions plays an important role when respondents are asked to imagine being in hypothetical situations. This clearly refers to the wording of the questions and the survey setting (personal or online interview). But even simple design elements such as offering an explicit option to say “No” right away greatly affect the answers, as does the position in the questionnaire that a particular question is given.

It is difficult to determine the ‘best’ survey design out of the four alternatives regarded in this paper. We may interpret a “no” answer to the initial question in each scenario as a possibility to avoid a deeper consideration of the problem presented. Moreover, a “protest zero” indicates that the respondent has not understood the purpose of the whole questionnaire. By these measures, the extended clarification of the hypothetical character of the scenarios in Survey II has improved the quality compared to the baseline (Survey I). A further improvement was brought about by omitting the initial yes-no question (Survey IV). We thus think that the results are closer to the “true” WTP

of the entire sample in this survey than in Survey I. Nevertheless, it remains an open research question, which WTP would have been stated by those individuals, who have answered “No” to the initial question. The comparison between median values in Surveys II and IV has revealed that at least part of these respondents might be willing to pay a considerable amount. In this case, the initial binary question may have served as an opt-out option to circumvent more difficult or time-consuming questions.

Furthermore, we may even favour a personal interview (Survey III), which has still reduced the proportion of zero WTP in almost all scenarios compared to the otherwise similar Survey II. However, higher standard deviations of responses in Survey III can be related to the observation in survey research that there is no general result which shows that online surveys are less valid than personal interviews (see, e.g. Couper and Bosnjak, 2010).

For those who are convinced that health care rationing decisions should ultimately reflect the preferences of the affected citizens, this is essentially bad news. Not only is it impossible to determine a single “value of a QALY”, as Pennington et al. (2014) have demonstrated, but if the WTP for a specific type of a QALY is to be elicited from the population, great care must be taken in the design of the respective questionnaires when even seemingly innocuous design elements such as the order of the questions have such a great effect on the results. Whether even the best design is suitable to elicit the “true” WTP, remains still doubtful.

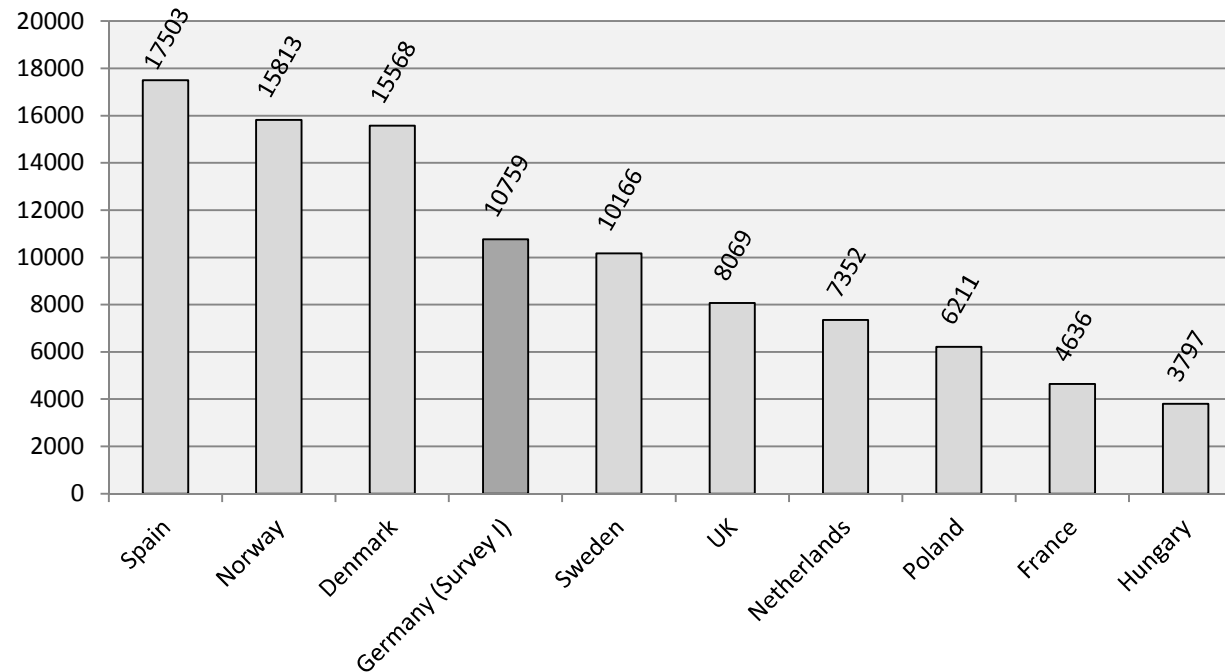
References

- Breyer, F. (2013), Implicit versus Explicit Rationing of Health Services, CESifo DICE Report 1/2013, 7-15.
- Bundesgesundheitsministerium (2014), Zahlen und Fakten zur Krankenversicherung. Received May 14, 2014 via <http://www.bmg.bund.de/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung/mitglieder-und-versicherte.html>.
- Couper, M.P. and Bosnjak, M. (2010), Internet Surveys, in: P.V. Marsden and J.D. Wright (eds.), *Handbook of Survey Research*, 2nd ed., Bingley: Emerald Group, 527-550.
- Cragg, J.G. (1971), Some Statistical Models for Limited Dependent Variables with Application to the Demand of Durable Goods, *Econometrica* 39, 829-844.
- Donaldson, C. et al. (2010), *European Value of a Quality Adjusted Life Year*, Final Publishable Report, Mimeo. http://research.ncl.ac.uk/eurovaq/EuroVaQ_Final_Publishable_Report_and_Appendices.pdf
- Greene, W.H. (2012), *Econometric Analysis*, 7th Edition, Boston: Pearson, Chapter 19.
- Gyrd-Hansen, D., Jensen, M.L., Kjaer, T. (2014), Framing the Willingness-to-Pay Question: Impact on Response Patterns and Mean Willingness to Pay, *Health Economics* 23, 550-563.
- Hammit, J. and Graham, J. (1999), Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?, *Journal of Risk and Uncertainty* 18, 33-62.

- IMF. International Monetary Fund (2013), World Economic Outlook Database; October 2013 Edition. Received 24 March, 2014 via <http://www.imf.org/external/pubs/ft/weo/2013/02/weo-data/weorept.aspx?pr.x=56&pr.y=1&sy=2008&ey=2014&scsm=1&ssd=1&sort=country&ds=.&br=1&c=134&s=PPPEX&grp=0&a=> and <http://www.imf.org/external/pubs/ft/weo/2013/02/weo-data/weorept.aspx?pr.x=75&pr.y=11&sy=2008&ey=2014&scsm=1&ssd=1&sort=country&ds=.&br=1&c=111&s=PCPI&grp=0&a=>
- National Institute for Health and Clinical Excellence [NICE] (2007), Guidelines Manual, <http://www.nice.org.uk/niceMedia/pdf/GuidelinesManualChapter8.pdf>
- Pennington, M., R. Baker, W. Brouwer, H. Mason, D. Gyrd Hansen, A. Robinson, C. Donaldson and the EuroVaQ Team (2014), Comparing WTP Values of Different Types of QALY Gain Elicited from the General Public, *Health Economics*, forthcoming.
- Robinson, A., D. Gyrd-Hansen, P. Bacon, R. Baker, M. Pennington, C. Donaldson and the EuroVaQ Team (2013), Estimating a WTP-based value of a QALY: The ‘chained’ approach, *Social Science and Medicine* 92, 92-104.
- Siegel, S. and Castellan, N.J. (1988), Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill.
- Statistisches Bundesamt (2014), GENESIS-Online Datenbank. Received May 14, 2014 via <http://www.destatis.de/genesis>
- Viscusi, W. (1993), The Value of Risks to Life and Health, *Journal of Economic Literature* 31, 1912-1946.
- Wooldridge, J.N. (2010), Econometric Analysis of Cross Section and Panel Data, 2nd Edition, Cambridge (Ma): MIT Press.

Tables and Figures

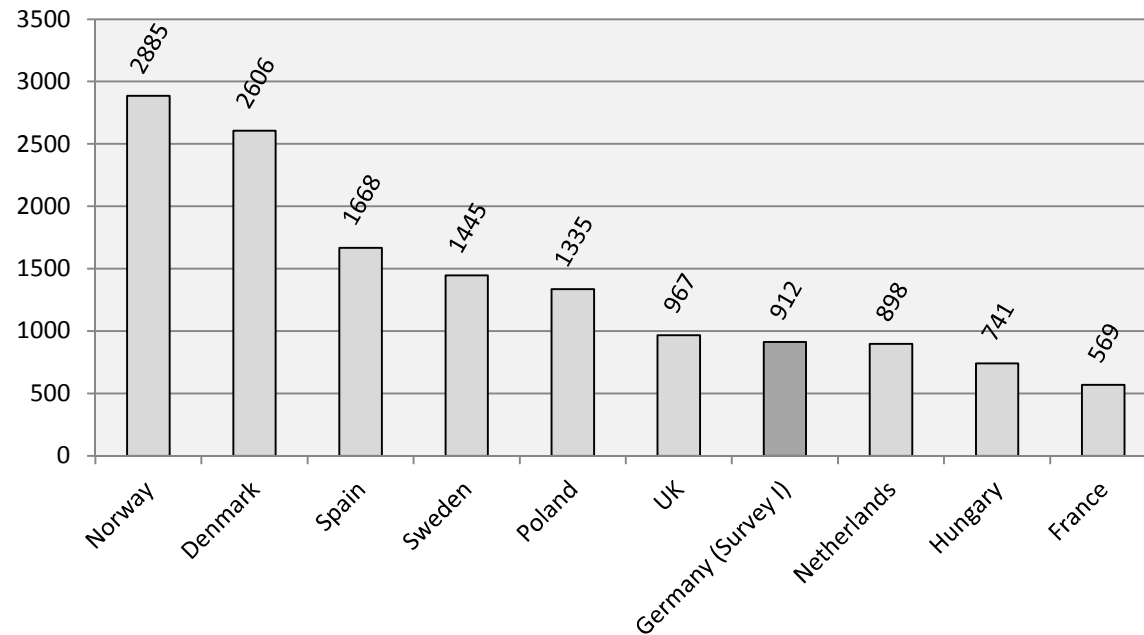
Figure 1: Overall Mean (Survey I, Scenarios A, B, F, G, I, J, L, P)
1% Trimmed ^{a)} (US\$ PPP 2008) ^{b)}



a) The trimming procedure for this data is similar to the one reported by Pennington et al. (2014) so that the top 1% of positive values are excluded in each scenario. In contrast, when regarding only the German data we trimmed the top 1% of all values except protesters.

b) Similar to the EuroVaQ project (Donaldson et al., 2010) all values are converted to US\$ using Purchasing Power Parity conversion rates for 2008 and regarding inflation rates (IMF, 2013).

Figure 2: Overall Median (Survey I, Scenarios A, B, F, G, I, J, L, P)
 1% Trimmed ^{a)} (US\$ PPP 2008)^{b)}



a) The trimming procedure for this data is similar to the one reported by Pennington et al. (2014) so that the top 1% of positive values are excluded for each scenario. In contrast, when regarding only the German data we trimmed the top 1% of all values except protesters.

b) Similar to the EuroVaQ project (Donaldson et al., 2010) all values are converted to US\$ using Purchasing Power Parity conversion rates for 2008 and regarding inflation rates (IMF, 2013).

Table 1: Overview of Decision Scenarios (see Donaldson et al. 2010, p.82)

Scenario	Health gain	Duration	QALY gain	When	Certainty/Risk
A	25 points	4 years	1	in 1 year's time	certainty
B	25 points	4 years	1	end of life	certainty
F	10 points	10 years	1	in 1 year's time	certainty
G	10 points	10 years	1	end of life	certainty
I	extra life	12+ months ^{b)}	1	end of life	certainty
J	No coma	12+ months ^{b)}	1	in 1 year's time	certainty
L	Extra life (terminal illness)	12+ months ^{b)}	1	in 1 year's time	certainty
P	25 points ^{a)}	4 years	1	in 1 year's time	certainty
D	25 points	1 year	0.25	in 1 year's time	certainty
E	10 points	1 year	0.1	in 1 year's time	certainty
M	25 points	4 years	0.1	in 1 year's time	10% risk
O	10 points	10 years	0.1	in 1 year's time	10% risk
N	25 points	4 years	0.05	in 1 year's time	5% risk

a) Scenario P differs from A in that in P the price is to be paid in 4 annual installments, but in A in one amount.

b) In scenarios I, J, and L 'additional life' is offered. The duration is adapted to the respondent's own health rating so that the gain at that health level amounted to one QALY (see also Donaldson et al. 2010, p.60).

Table 2: Characteristics of the Samples

Characteristic	Survey I (n=1,501)	Survey II (n=1,500)	Survey III (n=507)	Survey IV (n=1,500)	Germany (2012)
Age range (in years)					
18 to 29	17.6	19.1	17.0	19.1	16.8
30 to 44	29.0	27.1	25.4	27.1	22.9
45 to 59	25.8	29.5	26.6	29.5	28.3
60 or above	27.6	24.3	31.0	24.3	32.1
Net monthly household income: (in US\$ PPP for 2008)					
Lowest (<1,450)	19.5	17.2	25.4	15.4	
Low (1,450-2,109)	17.1	20.7	12.2	21.1	
Medium (2,110-2,789)	20.5	14.7	18.1	15.7	
High (2,790-3,789)	17.8	20.4	17.6	20.1	
Highest (≥3,790)	14.9	18.4	20.9	18.3	
Income not stated	11.5	10.1	10.8	10.6	
Educational level					
Low (up to ten years of schooling)	36.2	34.3	63.7	33.5	
Medium (additional three years of advanced education)	40.8	40.8	22.3	41.8	
High (any type of university study)	23.0	24.9	14.0	24.7	
Males (rather than females)	51.6	49.7	47.9	49.7	48.9
East Germany including Berlin (rather than West Germany)	26.8	20.3	22.1	20.3	19.8
Private health insurance (not public)	14.7	13.5	---	13.1	13.4
Own health (VAS, 20 to 100):					
Poor (20 to 69)	16.3	14.6	18.1	13.9	
Rather poor (70 to 79)	11.9	14.4	14.0	13.9	
Rather good (80 to 89)	24.0	23.8	24.7	23.5	
Very good (90 to 100)	47.9	47.2	43.2	48.7	
OECD coefficient: Mean value (standard deviation)	1.6576 (0.6154)	1.6487 (0.5450)	1.5542 (0.5227)	1.6743 (0.5676)	

Note: Given are percentage values except for the OECD coefficient. For Educational level four cases are missing in Survey I. For the OECD coefficient four, four, one and two cases in Surveys I to IV, respectively, are either missing or are ignored due to implausibly high values. Representative data for Germany: Statistisches Bundesamt (2014) and Bundesgesundheitsministerium (2014).

Table 3: Sample Sizes, Zero WTP and Protestors ^{a)}

	Total responses				Zero WTP (in % of total)				Protestors (in % of total)			
Survey	I	II	III	IV	I	II	III	IV	I	II	III	IV
Scenario												
A	762	736	246	733	17.3	14.1	8.9	1.4	3.7	2.0	2.0	0.3
B	372	367	121	367	40.1	37.9	25.6	2.2	4.6	1.6	1.7	0.5
F	735	714	231	715	18.8	20.6	19.0	1.4	4.6	2.8	2.2	0.1
G	349	349	110	365	35.0	36.1	38.2	1.4	3.2	2.0	3.6	0.3
I	748	760	263	760	46.0	37.9	39.2	5.0	4.4	2.6	1.9	0.1
J	744	757	261	758	30.2	27.5	18.4	3.6	5.1	3.2	3.4	0.3
L	727	740	252	741	33.6	32.6	22.2	9.0	4.4	3.4	1.2	0.3
P	376	369	122	367	29.0	22.0	9.8	2.5	6.6	2.4	3.3	0.5
D	380	371	124	377	30.0	27.0	14.5	1.6	5.5	6.7	4.0	0.5
E	392	389	131	392	36.2	37.3	31.3	1.5	5.4	4.1	4.6	0.3
M	325	367	124	367	37.5	26.4	25.0	3.3	4.0	2.5	2.4	0.0
O	361	363	118	364	28.0	22.6	17.8	2.7	4.2	1.1	1.7	0.3
N	325	367	124	367	41.8	37.1	31.5	3.8	3.7	1.9	2.4	0.3
Simple mean					32.6	29.2	23.2	3.0	4.6	2.6	2.5	0.3

a) “Protestors” are those respondents who only state that “government should pay for health care”.

Table 4: Willingness to Pay for 1 QALY: Descriptive Statistics – Reduced Samples ^{a)}

	Survey I			Survey II			Survey III			Survey IV		
Sce- nario	Max	Mean (S.E.)	Median	Max	Mean (S.E.)	Median	Max	Mean (S.E.)	Median	Max	Mean (S.E.)	Median
A	160,858	6,853 (650)	1,072	157,613	9,697 (725)	2,627	268,097	16,931 (2,163)	5,362	126,090	12,353 (720)	5,254
B	107,239	4,179 (665)	214	105,075	5,423 (644)	630	321,716	12,610 (3,166)	2,681	315,225	11,222 (1,349)	2,627
F	107,239	6,548 (597)	1,072	262,688	8,458 (804)	2,102	107,239	9,606 (1,251)	2,145	315,225	12,396 (922)	4,203
G	80,429	4,194 (551)	536	136,598	5,241 (735)	525	107,239	6,459 (1,666)	536	105,075	9,289 (825)	3,152
I	321,716	7,244 (1,100)	97	315,225	8,590 (1,046)	1,051	321,716	13,856 (2,844)	1,448	315,225	11,398 (1,092)	2,102
J	321,716	12,730 (1,411)	1,072	315,225	15,275 (1,464)	3,152	428,954	20,662 (3,250)	6,434	315,225	16,557 (1,214)	5,254
L	536,193	16,370 (1,842)	1,072	525,376	21,817 (2,256)	3,152	536,193	31,668 (4,837)	8,579	525,376	25,051 (2,192)	5,254
P	600,536	15,494 (2,700)	2,145	218,556	17,054 (1,648)	4,623	1,286,863	46,236 (12,320)	10,724	420,301	30,526 (3,144)	10,508
average	279,491	9,202	910	254,545	11,444	2,233	422,252	19,754	4,739	304,718	16,099	4,794
D	686,327	24,404 (3,860)	2,574	630,451	28,256 (3,649)	6,305	514,745	39,115 (8,164)	8,365	1,260,902	53,504 (6,466)	16,812
E	857,909	40,167 (5,852)	3,485	1,365,977	42,404 (5,814)	5,254	1,072,386	64,992 (15,933)	10,724	3,152,254	84,460 (12,241)	21,015
M	1,072,386	35,415 (6,362)	2,145	1,576,127	63,079 (9,388)	10,508	3,217,158	86,948 (27,389)	21,448	1,050,751	76,366 (8,193)	21,005
O	1,072,386	41,777 (6,666)	5,362	830,094	51,020 (5,422)	10,508	1,608,579	115,868 (22,031)	26,810	1,050,751	98,789 (9,574)	26,269
N	2,359,249	77,948 (15,804)	2,145	3,152,254	100,063 (16,031)	10,508	6,434,316	152,053 (55,194)	32,172	2,521,803	125,144 (14,229)	31,523
average	1,209,651	43,942	3,142	1,510,981	56,964	8,617	2,569,437	91,795	19,904	1,807,292	87,653	23,325

a) Samples are reduced by “protestors” and by the top 1% of answers in each scenario. Values reported are in US\$ to make them comparable to the EuroVaQ results (Donaldson et al., 2010). Euro amounts have been converted by using Purchasing Power Parity conversion rates for 2008 and by including inflation rates (IMF, 2013). To convert the values back to Euro amounts, numbers have to be multiplied by 0.9325 (surveys I and III from June/July 2012) or 0.9517 (surveys II and IV from February 2014).

Table 5: Regression Results (Reduced Samples)

Characteristics	Surveys I, II, and III				Survey IV			
	Probit: 1: WTP>0		Truncated (only WTP>0): ln(WTP)		Probit: 1: WTP>0		Truncated (only WTP>0): ln(WTP)	
	dy/dx	(S.E.)	Coef.	(S.E.)	dy/dx	(S.E.)	Coeff.	(S.E.)
Age 18 to 29	-.0473**	(.0175)	.3411**	(.0991)	-.0020	(.0081)	.0719	(.1461)
Age 30 to 44	-.0312*	(.0158)	.0906	(.0913)	.0024	(.0064)	.0811	(.1265)
Age over 59	.0089	(.0162)	.1268	(.0925)	.0063	(.0055)	.0108	(.1302)
Income Lowest	-.1279**	(.0216)	-.7625**	(.1176)	-.0299	(.0176)	-.8269**	(.1832)
Income Low	-.0755**	(.0202)	-.3405**	(.1073)	-.0219	(.0135)	-.2630	(.1525)
Income Higher	-.0103	(.0173)	.1627	(.0949)	.0093	(.0066)	.2067	(.1453)
Income Highest	.0167	(.0176)	.2829**	(.1006)	.0058	(.0085)	.4102**	(.1573)
Income missing	-.0995**	(.0245)	-.1030	(.1336)	-.0253	(.0170)	-.4605*	(.1879)
Education med.	.0141	(.0130)	.0900	(.0789)	.0075	(.0053)	.3581**	(.1147)
Education high	.0238	(.0157)	.3647**	(.0930)	-.0096	(.0076)	.4510**	(.1315)
Health 70-79	.0267	(.0205)	-.0806	(.1290)	.0159**	(.0045)	-.2882	(.2008)
Health 80-89	.0042	(.0189)	.0629	(.1140)	.0161**	(.0050)	.2252	(.1754)
Health 90-100	-.0026	(.0174)	.0640	(.1063)	.0091	(.0069)	.1185	(.1666)
Male	-.0450**	(.0114)	.2132**	(.0674)	-.0013	(.0051)	.1660	(.0985)
OECD coeff. (centered)	-.0049	(.0108)	-.1366*	(.0695)	-.0013	(.0043)	-.1000	(.0938)
East Germany	-.0238	(.0139)	-.3433**	(.0830)	-.0080	(.0069)	-.2085	(.1178)
Private insurance	.0217	(.0108)	.3175**	(.1005)	.0065	(.0069)	.2689*	(.1321)
QALY fraction (risk)	-.1473**	(.0210)	1.6971**	(.0998)	-.0134	(.0096)	1.2132**	(.1352)
QALY fraction (certain)	-.1272**	(.0202)	1.1037**	(.0929)	.0086	(.0054)	.8441**	(.1248)
Life extension	-.0889**	(.0135)	.2354**	(.0672)	-.0320**	(.0079)	-.3653**	(.0995)
End of life	-.1663**	(.0102)	-.4233**	(.0432)	-.0005	(.0029)	-.4543**	(.0480)
2nd scenario	-.0621**	(.0108)	.2050**	(.0382)	-.0000	(.0037)	.1650**	(.0411)
3rd scenario	-.0133	(.0144)	.3503**	(.0720)	-.0036	(.0053)	.6842**	(.0963)
4th/5th scenario	-.0796**	(.0113)	.6008**	(.0486)	-.0157**	(.0047)	.6950**	(.0671)
Survey II	.0157	(.0121)	.5546**	(.0730)	---	---	---	---
Survey III	.0815**	(.0155)	1.1370**	(.0996)	---	---	---	---
Constant	---	---	7.1491**	(.1395)	---	---	7.8112**	(.2011)
Sigma	---	---	1.9088**	(.0227)	---	---	1.9502**	(.0323)
y=Pr(WTP>0) (predict)	0.7456				0.9822			
# of strata	3		3		1		1	
# of PSUs	3,457		3,089		1,495		1,477	
# of obs.	14,738		10,814		6,588		6,384	
F-test	31.32**		144.96**		5.16**		119.99**	

Note: Willingness-to-pay for one QALY is given in US\$ PPP. Individual base case: Age 45 to 59, medium income, educational level low, health 20-69, female, West-Germany, public health insurance. Scenario base case: gain of one QALY with certainty, gain in one year's time, improved quality of life, first scenario in the questionnaire. In the Probit model, dy/dx refers to a discrete change of a dummy variable from 0 to 1 for all independent variables except for the OECD coefficient. OECD coefficients are centered on the respective mean. Levels of significance: * 5%, ** 1%.

Appendix: Table 6: Willingness to Pay for 1 QALY: Descriptive Statistics – Reduced Samples, only Public Insurance

	Survey I			Survey II			Survey III			Survey IV		
Sce- nario	Max	Mean (S.E.)	Median	Max	Mean (S.E.)	Median	Max	Mean (S.E.)	Median	Max	Mean (S.E.)	Median
A	107,239	5,778 (572)	858	157,613	8,766 (745)	2,102	268,097	16,931 (2,163)	5,362	105,075	11,321 (753)	4,203
B	64,343	2,955 (443)	214	84,060	5,115 (647)	525	321,716	12,610 (3,166)	2,681	315,225	10,555 (1,469)	2,627
F	193,029	6,111 (685)	858	210,150	7,601 (764)	2,102	107,239	9,606 (1,251)	2,145	315,225	11,853 (1,017)	3,678
G	85,791	3,555 (549)	536	157,613	5,360 (868)	525	107,239	6,459 (1,666)	536	105,075	8,733 (839)	3,152
I	214,477	6,003 (921)	64	315,225	6,534 (868)	525	321,716	13,856 (2,844)	1,448	315,225	11,192 (1,212)	2,102
J	321,716	11,304 (1,366)	858	315,225	14,520 (1,599)	2,627	428,954	20,662 (3,250)	6,434	315,225	15,518 (1,231)	5,254
L	428,954	14,052 (1,659)	1,072	525,376	18,009 (2,061)	2,102	536,193	31,668 (4,837)	8,579	525,376	24,649 (2,444)	5,254
P	214,477	10,720 (1,494)	2,145	218,556	15,704 (1,689)	4,203	1,286,863	46,236 (12,320)	10,724	420,301	29,413 (3,253)	8,826
average	203,753	7,560	826	247,977	10,201	1,839	422,252	19,754	4,739	302,091	15,404	4,387
D	686,327	18,307 (3,172)	2,574	630,451	29,400 (4,195)	5,044	514,745	39,115 (8,164)	8,365	1,260,902	52,174 (6,945)	14,711
E	643,432	28,872 (4,502)	3,217	630,451	39,067 (5,178)	3,152	1,072,386	64,992 (15,933)	10,724	3,152,254	84,249 (13,406)	21,015
M	1,072,386	33,140 (6,266)	2,145	1,576,127	67,122 (10,601)	10,508	3,217,158	86,948 (27,389)	21,448	945,676	67,533 (7,308)	15,761
O	750,670	32,638 (4,832)	5,362	830,094	52,027 (6,105)	10,508	1,608,579	115,868 (22,031)	26,810	1,050,751	91,477 (9,909)	21,015
N	2,359,249	77,724 (16,768)	2,145	3,152,254	104,928 (17,940)	6,305	6,434,316	152,053 (55,194)	32,172	1,681,202	109,646 (12,449)	31,523
average	1,102,413	38,136	3,089	1,363,875	58,509	7,103	2,569,437	91,795	19,904	1,618,157	81,016	20,805

a) Samples are reduced by “protestors” and by the top 1% of answers in each scenario. Values reported are in US\$ to make them comparable to the EuroVaQ results (Donaldson et al., 2010). Euro amounts have been converted by using Purchasing Power Parity conversion rates for 2008 and by including inflation rates (IMF, 2013). To convert the values back to Euro amounts, numbers have to be multiplied by 0.9325 (surveys I and III) or 0.9517 (surveys II and IV).