

Decision makers' and scientists' opinion about contingent valuation and choice experiments for measuring willingness to pay in health care: Results from a survey in Germany

Oliver H. Günther, Hans-Helmut König

University of Leipzig

Objectives: Assessment of willingness to pay (WTP) by contingent valuation (CV) and choice experiments (CE) is increasingly performed in economic evaluation of health care. However, the question of whether the methods for measuring WTP are acceptable to decision makers and scientists has remained largely unacknowledged. The aim of this study was to learn more about decision makers' and scientists' opinion concerning these methods.

Methods: An expert group developed a questionnaire consisting of key items that may influence the opinion about CV and CE according to the constructs "attitude toward behavior," "subjective norm," and "behavioral intention" as defined by the Theory of Reasoned Action by Ajzen and Fishbein. In a survey, seventy-seven decision makers representing key institutions in the German healthcare system and forty-two scientists in health economics completed the questionnaire.

Results: Scientists and decision makers in particular did not show a high intention to use methods for measuring WTP. Skepticism regarding precision of the methods and subjects' capability to imagine paying an amount of money for a certain health commodity were stated along with the assertion that the hypothetical decision-making scenario was rather a distant reality. Nevertheless, the majority of scientists and decision makers did not state rejection of the methods.

Conclusions: Increasing the probability of using methods for measuring WTP in health care, the hypothetical scenarios should be made more realistic and payment vehicles should be used to help patients relate payment to a real health benefit. Moreover, an intensive discussion on the potential usefulness of CV/CE without excluding ethical concerns in comparison to existing alternatives has to be resumed.

Keywords: Willingness to pay, Choice experiments, Contingent valuation, Cost-benefit analysis

Economic evaluation of healthcare programs aims to aid decision makers with their difficult choices in allocating healthcare resources, setting priorities, and implementing health

policy. Drummond et al. (7) define economic evaluation as the "comparative analysis of alternative courses of action in terms of both their costs and consequences." Whereas costs are always measured in monetary terms, consequences can be analyzed in three different ways, reflecting the different types of economic evaluation.

This study was funded by the German Federal Ministry of Education and Research (grant number 01ZZ0106).

First: in cost-effectiveness analysis (CEA), consequences are assessed in terms of the immediate effects on health. These are usually clinically defined units appropriate to the area of study, such as “life-years saved,” “cases detected,” and so on.

Second: in cost-utility analysis (CUA), consequences are assessed in more generic terms such as quality-adjusted life-years (QALYs). QALYs are supposed to simultaneously capture gains from reduced morbidity and reduced mortality and to integrate these into a single measure. They are calculated by weighting the duration of health states with an index score of health-related quality of life ranging from 0 for death to 1 for perfect health.

Third: in cost-benefit analysis (CBA), consequences are measured in the same way as costs, i.e., in monetary terms. The challenging question with CBA in health care is obvious: How can we measure or convert benefits from health programs into monetary values? Two frequently used approaches for the monetary valuation of benefits are the contingent valuation (CV) method and the choice experiments (CE). Both are preference-based methods, which infer the value (i.e., willingness to pay) of healthcare programs by analyzing the stated choices of respondents from hypothetical prespecified choice sets.

In a CV study, participants are presented with a scenario representing an improvement over the current state. Participants are then directly asked to indicate their maximum willingness to pay for this improvement either as an open-ended estimate or by some form of forced choice. The value indicated by each respondent is assumed to reflect the value of the program for the individual (15).

A typical CE application extends the “single shot” CV approach by presenting a respondent with a choice between two or more alternatives, each described by the relevant levels of their attributes. This choice process is then iterated so as to build up a set of choices for each respondent. Repeating this process with a sample group permits the efficient collection of a substantial data set concerning underlying preferences. If price is included as an attribute, it is possible to indirectly estimate willingness to pay for the provision of a specified good (21).

Within the framework of CBA, there has been an intensive discussion, especially in health economics, about the attractiveness of CV and, more recently, about CE. One of the attractive features of CE is that—in comparison with CV—they provide the decision maker with greater information concerning preferences for a wide range of provision permutations. CE have not yet been exposed to the volume of explicit research that has been applied to the CV method. As shown by O’Brien and Gafni (17) and by Lancsar and Savage (16), both CV and CE are directly grounded in welfare economic theory. This finding helps decision makers and scientists reveal the net economic impact of any health program. Moreover, the analysis is not restricted to comparing programs within health care. In particular, choices presented

in CE seem to reflect what people do in real life situations: Mandy Ryan and colleagues (20;22;23) revealed in some of their studies that CE is able to model nondemanding behavior, and account for context effects, and further results indicated that, even when subjects were presented with choices for unfamiliar goods, they developed a consistent pattern of preferences.

Despite these advantages, CBA using CV or CE has only rarely been performed in the economic evaluation of health care in comparison to CEA and CUA. There has been concern that measuring WTP by CV or CE may pose a cognitive task too difficult for many respondents and that results may be biased and lack precision (2;4).

The purpose of the present study was to highlight the issue that, despite these advantages of CV and CE within the framework of CBA, their application in economic evaluation is rather rare. In particular, this study aimed to understand the opinions held by decision makers’ and scientists’ regarding the methodology used to elicit WTP. An additional aim was to predict the possible factors influencing future usage by these professionals. The following aspects were analyzed in detail: (i) What opinion do decision makers and scientists in health care express concerning methods for measuring WTP? (ii) Which factors influence their opinion about the use of WTP measurement in health care? (iii) How strong is the influence of such factors?

METHODS

The study was conducted in two consecutive steps: In the first step, an expert group developed a questionnaire consisting of key items that might influence decision makers’ and scientists’ opinion about CV and CE for measuring WTP. In the second step, the questionnaire was used in a survey of decision makers in the German healthcare system and scientists working in the field of health economics in Germany.

The Theory of Reasoned Action by Ajzen and Fishbein (1) served as the conceptual framework guiding the development of the questionnaire. It was designed to explain and predict behavior with a small number of variables. The basic assumption underlying this theory is that people do things for more or less rational reasons and can be summarized as follows:

- Behavior is determined by the “behavioral intention.”
- “Behavioral intention” is determined by “attitudes” and “subjective norms,” and the relative importance of each depends on the relative size of weights assigned to them.
- “Attitudes” are determined by their behavioral beliefs (beliefs about the consequences of a behavior) and evaluations of those beliefs.
- “Subjective norms” are determined by normative beliefs (perception of significant referents’ belief of the behavior) and motivations to comply.

Thus, the Theory of Reasoned Action provides a simple framework to analyze behavior. It has been used widely in behavioral science and could easily be applied to the present study: According to the Theory of Reasoned Action, the best predictor of decision makers' and scientists' use of CV or CE is their behavioral intention that comprises two major factors: the belief that the use of CV or CE is right or wrong (attitude) and the belief regarding social pressure to either use or not use CV or CE (subjective norm). Therefore, in further analysis, "behavioral intention" is seen as the dependent variable explained by specific attitudes and subjective norms about measuring WTP with CV or CE.

Development of Questionnaire

The group of experts involved in the development of the questionnaire consisted of sixteen scientists from psychology, economics, and mathematics, whose research work had a focus on measuring preferences, and of four decision makers from various medical insurers. Their participation was voluntary.

At the first stage of the questionnaire's development, key notions concerning methods for measuring WTP were collected by the experts during individual brainstorming sessions. The underlying idea was to identify all factors that might influence the opinion of decision makers and scientists about methods for measuring WTP. At the next stage, the frequency of the mentioned various notions was counted. Notions mentioned by two or more experts were used in the questionnaire.

The key notions identified were put into questions and assigned to the constructs "attitude toward behavior," "subjective norm," and "behavioral intention" of the Theory of Reasoned Action. The experts checked the questions for ambiguous or unclear wording, which may have influenced the semantic content of the questionnaire concerning the construct to be measured. For this purpose, probing questions were used (5). Moreover, the experts were asked to assess the level of difficulty of each question using a five-point Likert scale. Consequently, questions with lack of discriminative ability or ambiguous wording were removed. The final questionnaire consisted of eighteen questions (items), beginning with a simple and short description of CV and CE as methods for measuring WTP.

"Attitude toward behavior" was assessed by seven questions on the respondent's opinion regarding: the validity and reliability of methods for measuring WTP, the required methodical knowledge of the investigator and appropriateness, practical aspects and costs of measuring WTP. "Subjective norm" was assessed by two questions: the first question referred to talking with colleagues about methods for measuring WTP, the second question assessed the respondent's degree of rejection of methods measuring WTP in health care. "Behavioral intention" was assessed by five questions referring to the intention to use WTP measurement in health

care and to promote the respective methods. Each item of "attitude toward behavior," "subjective norm," and "behavioral intention" was measured by a four-point semantic differential scale, ranging from 1 (negative) to 4 (positive). Additionally, four questions were included to assess the frequency of use of WTP measurement in the past, the level of knowledge about methods for measuring WTP, decision makers' and scientists' years of experience in the health system, and their age.

Survey

For the survey, a sample of 221 subjects consisting of 163 decision makers in the German healthcare system and 58 scientists working in the field of health economics were selected. To identify the most important decision makers in the German healthcare system, one has to be aware of its organizational structure (8). It has a decentralized organization, characterized by federalism and delegation to nongovernmental corporatist bodies who are, in turn, the main players in the social health insurance system. Hence, all players involved are organized at the federal as well as at the state level. There are sixteen states ("Länder") within each of which there are the physicians' (and dentists') associations on the providers' side and the sickness funds and their associations on the purchasers' side.

Decision makers were selected based on a database provided by the German Federal Ministry of Health, which contained a list of all purchasers and providers of the statutory health insurance (SHI) program (3). Decision-makers' names and addresses were identified by a search in the Internet viewing the organizational charts of sickness funds, the Hospital Federation, the Association of SHI-Accredited Physicians, and Ministries of Health on federal and state levels. In detail, the sample contained seventeen state secretaries representing the government level, fifty-nine chairmen of the Hospital Federation and the Association of SHI-Accredited Physicians on the provider side, and eighty-seven chairmen of medical funds on the purchaser side. Scientists were selected based on a listing provided by the German Coordinating Agency for Public Health (Deutsche Koordinierungsstelle für Gesundheitswissenschaften) from which experts in the field of health economics were identified (6).

Questionnaires were sent to the subjects by mail in June 2004, asking them to answer the questionnaire personally. A reminder was sent out after 14 days. All subjects who did not return the questionnaire after the announced deadline of 1 month received a fax, asking for possible reasons for nonparticipation. Four possible reasons were presented in form of a forced choice task: (i) I found the questionnaire incomprehensible; (ii) I refuse to answer questions about WTP in health care; (iii) I refuse to answer any questionnaire; (iv) I have no time to answer the questionnaire.

Statistical Methods

Descriptive statistics such as mean score and standard deviation (SD) were used to give an overview of the decision

makers' and scientists' opinion on the various aspects of methods for measuring WTP assessed by individual questions. Differences in means were tested by Student's *t*-test.

To assess the tendency of the respondents' opinion, the four-point semantic differential scale was dichotomized. Answers scoring 1 or 2 were considered to reflect a rather negative opinion, and answers scoring 3 or 4 were considered to reflect a rather positive opinion; the proportion of negative and positive opinions was calculated for each item.

The constructs "attitude toward behavior," "subjective norm," and "behavioral intention" were formed by calculating the mean score over all items representing one construct. To assess internal consistency of items measuring the constructs, Cronbach's alpha was calculated. Cronbach's alpha measures how well a set of items records a single one-dimensional latent construct, ranging from 1 for perfect internal consistency to 0 for no consistency at all (11).

Pearson's correlation coefficient was calculated to assess the correlation between the constructs "behavioral intention," "subjective norm," "attitude toward behavior" as well as sociodemographic variables and expertise (years of professional experience, level of knowledge about methods for measuring WTP, age, and a dummy variable for being a decision maker or scientists). Significant correlations greater than .5 were viewed as large, .5–.3 as moderate, and anything smaller than .3 as small (11). Ordinary least square regression analysis was used to estimate quantitative functional relationships between "behavioral intention" as the dependent variable and "attitude toward behavior," "subjective norm," as well as sociodemographic variables and expertise as independent variables (11). A normal probability plot was used to check for normal distribution of residuals in regression analysis. For statistical testing, the level of significance was set at $\alpha = .05$.

RESULTS

Of the 221 subjects contacted, 138 (62 percent) responded to the questionnaire. Two of the respondents found the questionnaire incomprehensible and, thus, did not complete it. Seven respondents stated that they had no time to complete the questionnaire, and ten respondents indicated that they would not participate in a survey about WTP in health care. In total, 119 questionnaires were completed (77 by decision makers and 42 by scientists), equaling an overall response rate of 54 percent (decision makers 47 percent, scientists 72 percent). The proportions reported in the following refer to the 119 scientists and decision makers who completed the questionnaire.

Sociodemographic Variables and Expertise

The mean age of responding decision makers (48.7 years, SD 7.7) and scientists (50.2 years, SD 11.0) was very similar. Of all respondents, 75.6 percent reported more than

10 years of professional experience. The level of knowledge of methods for measuring WTP was higher among scientists: 80.5 percent of the scientists compared with 45.5 percent of the decision makers indicated a level of knowledge of 3 or more on a five-point Likert scale, with 1 reflecting low and 5 reflecting high level of knowledge. Four decision makers and eight scientists reported that they had already conducted measurements of WTP in health care.

Attitude Toward Methods for Measuring WTP

A total of 96.1 percent of the decision makers (mean score, 3.53; SD .660) and 97.6 percent of the scientists (mean score, 3.74; SD .497) believed that investigators who conduct measurements of WTP must have a high level of methodological and statistical knowledge (Figure 1). There were 54.8 percent of the decision makers (mean score, 2.36; SD .634) and 47.4 percent of the scientists (mean score, 2.50; SD .647) who thought that the hypothetical scenarios presented to subjects when measuring WTP resembled decision-making situations in a more distant than realistic way.

A total of 53.3 percent of the decision makers (mean score, 2.33; SD .741) and 52.4 percent of the scientists (mean score, 2.38; SD .660) believed that subjects would not be capable of imagining payment of a defined sum of money for a specific healthcare service. A total of 74.7 percent of the decision makers (mean score, 2.21; SD .501) and 81.0 percent of the scientists (mean score, 2.12; SD .504) thought that the costs of measuring WTP were rather high. A total of 74.7 percent of the decision makers (mean score, 2.17; SD .554) and 70.7 percent of the scientists (mean score, 2.27; SD .593) thought that the methods for measuring WTP were rather imprecise.

Nevertheless, 46.7 percent of the decision makers (mean score, 2.39; SD .634) and 65.0 percent of the scientists (mean score, 2.65; SD .662) believed that CE were an appropriate method for supporting decisions about the allocation of collective resources in health care; these mean scores were significantly different ($p = .04$). CV was considered an appropriate method by 48.7 percent of the decision makers (mean score, 2.34; SD .758) and 51.2 percent of the scientists (mean score, 2.46; SD .745).

Subjective Norm

A total of 32.9 percent of the decision makers (mean score, 2.21; SD .789) and 69.0 percent of the scientists (mean score, 2.95; SD .882) reported talking with colleagues about measuring WTP in health care; these mean scores were significantly different ($p < .001$; Figure 2). Only 22.1 percent of the decision makers (mean score, 2.90; SD .771) and 27.5 percent of the scientists (mean score, 3.18; SD 1.010) stated rejection of the methods for measuring WTP.

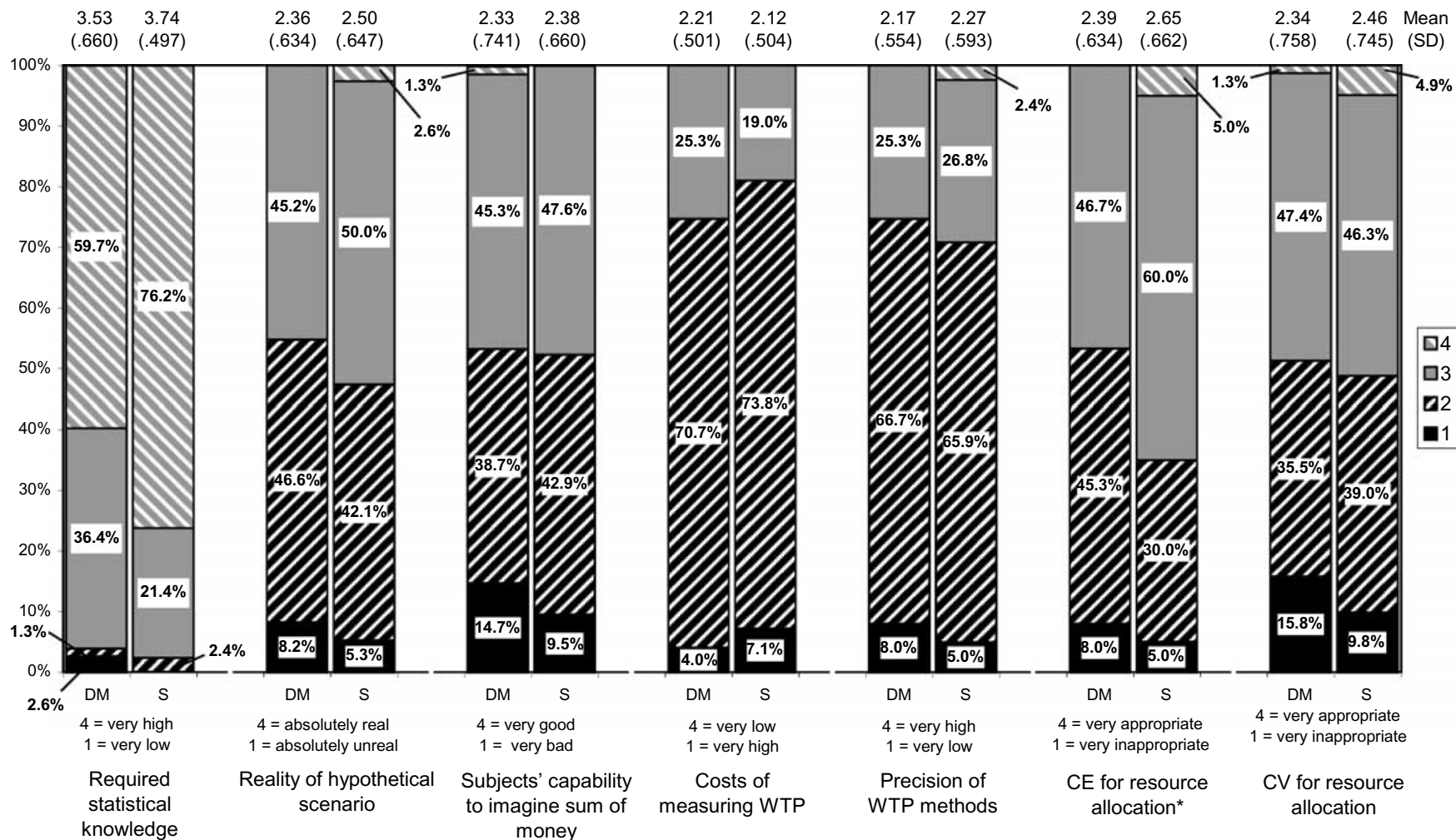


Figure 1. Frequency distribution and mean score of responses to items of the construct “attitude toward behavior” (DM, Decision Makers; S, Scientists). An asterisks indicates that a mean score was significantly different between decision makers and scientists ($p < .05$). Due to missing values for some variables, the sample size was $73 \leq n \leq 77$ for decision makers and $38 \leq n \leq 42$ for scientists.

Opinion about measuring willingness to pay

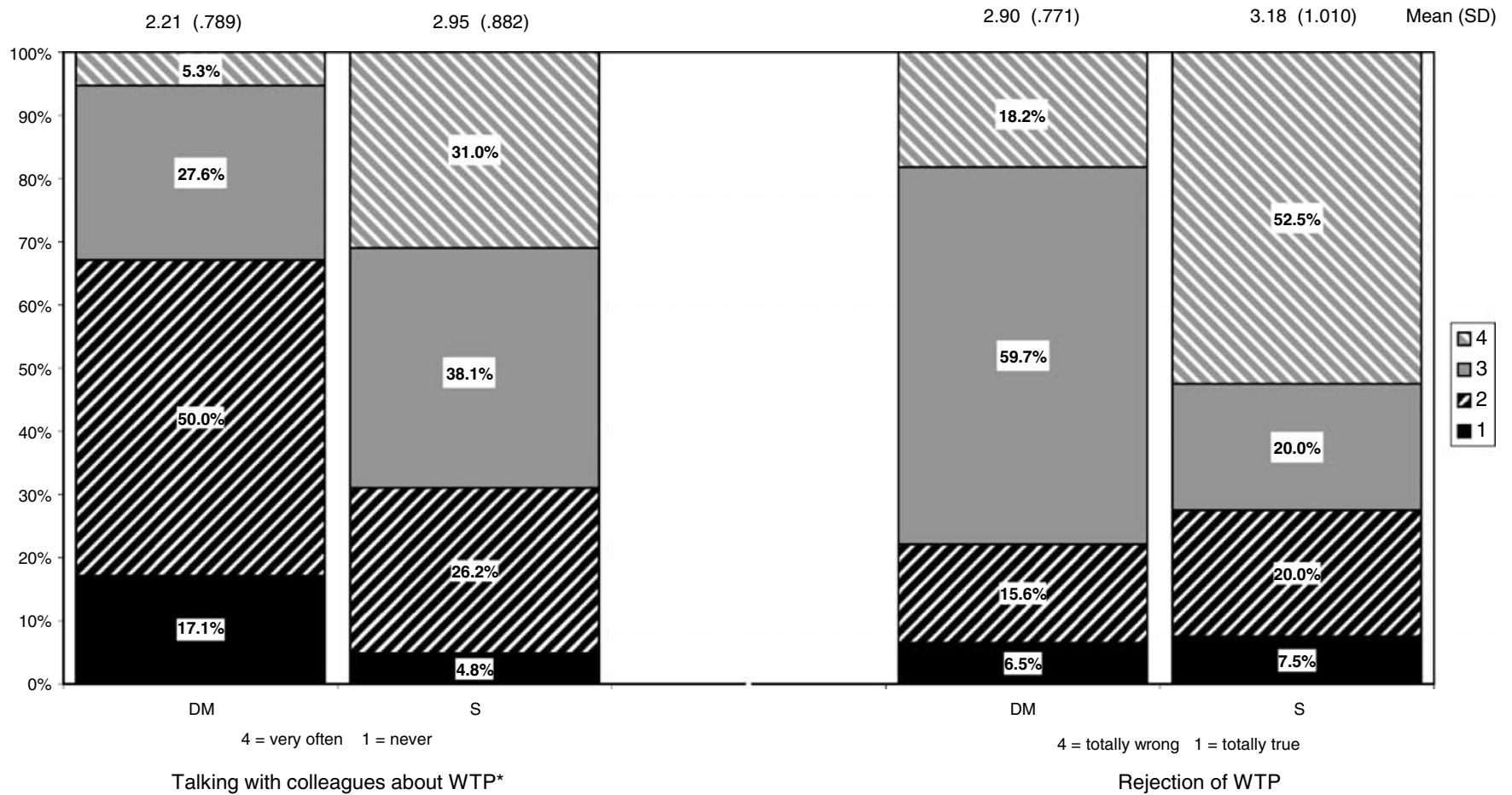


Figure 2. Frequency distribution and mean score of responses to items of the construct “subjective norm” (DM, Decision Makers; S, Scientists). An asterisks indicates that a mean score was significantly different between decision makers and scientists ($p < .05$). There were one missing value of the variable “Talking with colleagues about WTP” for decision makers ($n = 76$) and two missing values of the variable “Rejection of WTP” for scientists ($n = 40$).

Table 1. Correlation Analysis of the Constructs of the Theory of Reasoned Action and Sociodemographic Variables (Total Sample)^a

Correlation	Behavioral intention	Subjective norm	Attitude toward behavior	Working experience (yr)	Level of knowledge for measuring WTP	Profession (0 = decision maker; 1 = scientist)	Age (yr)
Behavioral intention	1	.639 ^b	.288 ^b	-.098	.362 ^b	.222 ^c	-.220 ^c
Subjective norm		1	.467 ^b	-.052	.375 ^b	.324 ^b	-.093
Attitude toward behavior			1	.023	.041	.066	-.030
Working experience (yr)				1	-.016	-.062	.408 ^b
Level of knowledge for measuring WTP					1	.362 ^b	-.247 ^b
Profession (0 = decision maker; 1 = scientist)						1	-.075
Age (yr)							1

^a Due to missing values for some variables, the sample size used for calculations was $103 \leq n \leq 119$.

^b $p < .01$.

^c $p < .05$.

WTP, willingness to pay.

Behavioral Intention

A total of 19.7 percent of the decision makers (mean score, 1.83; SD .737) and 31.0 percent of the scientists (mean score, 2.17; SD .908) stated their intention to use methods for measuring WTP; these mean scores were significantly different ($p = .03$; Figure 3). The intention to promote the establishment of the methods for measuring WTP in health economics was stated by 17.3 percent of the decision makers (mean score, 1.81; SD .711) and 41.5 percent of the scientists (mean score, 2.29; SD 1.031); these mean scores were significantly different ($p = .01$).

A total of 13.3 percent of the decision makers (mean score, 1.65; SD .707) and 38.1 percent of the scientists (mean score, 2.19; SD 1.018) intended to further optimize the methods for measuring WTP; these mean scores were significantly different ($p < .01$). The intention to use measurement of WTP for supporting decisions about the allocation of collective resources in health care was stated by 32.9 percent of the decision makers (mean score, 2.24; SD .798) and 31.7 percent of the scientists (mean score, 2.17; SD .972). There were 46.1 percent of the decision makers (mean score, 2.41; SD .786) and 73.8 percent of the scientists (mean score, 2.79; SD .976) who reported the intention to learn more about the methods for measuring WTP; these mean scores were significantly different ($p = .02$).

Correlation Analysis of the Constructs and Sociodemographic Variables

The internal consistency of the constructs “behavioral intention,” “subjective norm,” and “attitude toward behavior” were satisfactory with a Cronbach’s alpha of .88, .56, and .59, respectively. For all subjects, the correlation matrix showed a significant correlation between all constructs (Table 1). “Behavioral intention” was more strongly correlated with “subjective norm” ($r = .639$; $p < .01$) than with “attitude toward behavior” ($r = .288$; $p < .01$). Correlation between “subjective norm” and “attitude toward behavior” was moderate ($r = .467$; $p < .01$). Furthermore, behavioral intention was moderately correlated with level of knowledge ($r = .362$; $p < .01$). Moderate correlation was also found between subjective norm and level of knowledge ($r = .375$; $p < .01$) as well as being a scientist ($r = .324$; $p < .01$).

Regression Analysis Explaining Behavioral Intention

Multiple regression analysis showed that only “subjective norm” ($\beta = .537$; $p < .01$) had a significant influence on “behavioral intention” (Table 2). The regression model was able to explain more than 50 percent of the variance of the dependent variable ($R^2 = .51$). A normal probability plot

Table 2. Results of Regression Analysis Using “Behavioral Intention” as the Dependent Variable (Total Sample)^a

Independent variables	Beta	Standard error	T value	<i>p</i> value
Intercept	1.123	.506	2.22	.029
Subjective norm	.537	.085	6.32	<.001
Attitude toward behavior	-.014	.150	-.09	.927
Working experience (yr)	-.021	.061	-.35	.728
Level of knowledge for measuring WTP	.074	.049	1.50	.137
Profession (0 = decision maker; 1 = scientist)	.075	.112	.67	.503
Age (yr)	-.009	.006	-1.52	.131

^a Due to missing values for some variables, the sample size used for the calculations was $n = 97$; $R^2 = .51$.

WTP, willingness to pay.

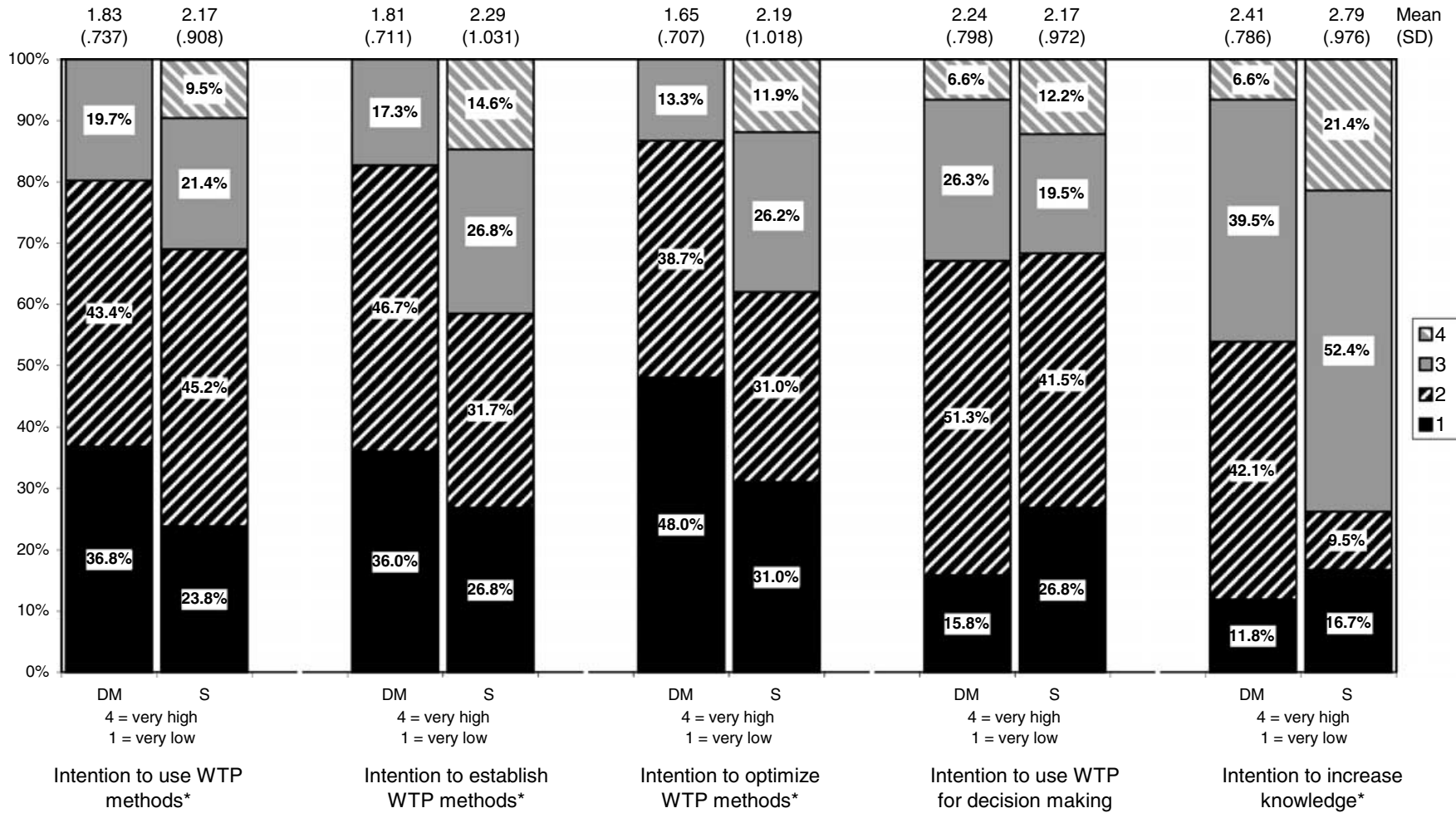


Figure 3. Frequency distribution and mean score of responses to items of the construct “behavioral intention” (DM, Decision Makers; S, Scientists). An asterisks indicates that a mean score was significantly different between decision makers and scientists ($p < .05$). Due to missing values for some variables, the sample size was $75 \leq n \leq 76$ for decision makers and $41 \leq n \leq 42$ for scientists.

indicated that residuals in regression analysis were normally distributed.

DISCUSSION

The survey showed that both decision makers and scientists believed that investigators measuring WTP must have a high level of methodological and statistical knowledge. In addition, they expected methods for measuring WTP to be rather cost-intensive and imprecise. Neither decision makers nor scientists made a clear commitment as to whether the hypothetical scenarios presented to subjects came close to real decision-making situations. Moreover, both groups were rather skeptical whether subjects were capable of imagining that they should pay a certain amount of money for a specific healthcare service. Decision makers were rather noncommittal as to whether methods for measuring WTP were either inappropriate or appropriate to support decisions about allocating collective resources in health care. However, the majority of scientists considered CE in particular to be an appropriate method for resource allocation. Furthermore, decision makers reported to talk significantly less about methods for measuring WTP with colleagues than scientists did. Yet, both groups revealed a rather low level of rejection of methods for measuring WTP. The majority of both decision makers and scientists stated that they did not intend to use, optimize, or establish methods for measuring WTP. However, almost half of the decision makers and more than half of the scientists showed the intention to learn more about methods for measuring WTP.

These present results revealed that especially decision makers did not have a high intention to use methods for measuring WTP on the one hand but that they might consider them as a support when allocating resources, although skepticism regarding several issues concerning validity and costs were stated on the other hand. That finding might support the statement given by Olsen and Smith (18) that there seems to be a “mismatch between the theoretical glory of WTP and the usefulness for public health policy.”

As Hoffmann and Schulenburg (14) reported, 46 percent of decision makers in Germany based their decision on some sort of evaluation, mostly conducted by themselves, rather than on formal economic evaluation concepts. Our study indicated a low level of practical experience, but a relatively high (self-) reported level of knowledge about the methods of measuring WTP of decision makers. However, the latter could be susceptible to social desirability, which leads to an over-reported level of knowledge (9). As a result, one may assume that decision makers reported spontaneous attitudes rather than consolidated ones. This assumption means they might be amenable to change their attitude, provided they receive persuasive systematic scientific information about methods for measuring WTP.

It would be appropriate to conceptualize the use of methods for measuring WTP within the framework of CBA as one

type of economic evaluation. There is some evidence in the literature that most decision makers in Germany prefer to refine CEA/CUA and particularly the cost per QALY analysis than to use methods for measuring WTP within CBA (10;24). The reason given for this reluctance is that very few measurements of WTP have been conducted (and published) in the field of health economics in Germany so far. Moreover, considering that 89 percent of the population were covered by SHI, which directly pays the doctor's costs, as well as remedies, drugs, appliances, hospital treatment, and preventative health care, it is not remarkable that the majority of decision makers and scientists were rather skeptical whether subjects are able to make a monetary trade-off for goods in health care (7).

This study suggests that the dissemination of methods for measuring WTP may be dependent on the “subjective norm” perceived by decision makers and scientists. Supported by the results of regression analysis, it seems that the more decision makers and scientists talked with their colleagues about methods for measuring WTP and the less rejection of the methods they expressed, the stronger the intention was to use the methods for measuring WTP. According to Ajzen and Fishbein (1) “subjective norm” is “. . . a specific behavioral prescription attributed to a generalized social agent.” This statement means that peer group opinion as perceived by individuals who approve or disapprove the methods for measuring WTP is an important factor influencing the “behavioral intention.” Because it is presumed that most of the decision makers and scientists have only limited knowledge and practical experience, it seems important to them how their social environment values the methods for measuring WTP. Personal attitude, therefore, might be replaced by social influence. As indicated by the opposition of a small number of decision makers refusing to answer questions about WTP, the provision of health care is an emotive issue and studies of measuring WTP are often viewed by the public (or important reference persons) as being somehow supportive of policies aimed at removing the provision of state-supplied health services and may find echoes in ethical concerns.

One general limitation of this study is that CV and CE were bundled together into “methods for measuring WTP;” thereby not allowing differentiation between all the “pros” and “cons” of CE and CV. Although a small pool of CE/CV comparisons in agricultural economics report higher WTP values with CE than CV derived values (13), there is, to our knowledge, only one study of a direct CE/CV comparison in health economics reporting no significant differences in WTP estimates (19). In view of this result and to keep the used questionnaire as short as possible (to enable an increase in response rate), we bundled CV and CE as methods for measuring WTP. The only exception allowing differentiation between CV and CE, was one question about the appropriateness of resource allocation. Here, scientists preferred CE to CV, whereas decision makers were rather noncommittal as to whether methods for measuring WTP were either

inappropriate or appropriate to support decisions about allocating collective resources in health care. Thus, it could be that scientists wish to overcome certain anomalies in CV by using CE (12).

The response rate to this questionnaire almost certainly was increased due to its brevity and the reminders that were issued (nearly 10 percent of decision makers who did not return the questionnaire cited lack of time). Although brevity was achieved at the cost of eliciting additional “motivational values”—as normally used by the “Theory of Reasoned Action” for determining future use—this must be considered in the light of the predictive power of the probably higher response rate. However, despite these efforts, the overall non-response rate was still 47 percent, being much higher among decision makers than scientist. One reason for nonresponding may have been ethical concerns, as indicated by ten decision makers who refused to participate in any survey about WTP in health care. Thus, nonparticipation may be motivated due to the perception that measuring WTP may support arguments aimed at excluding services from the service package of the SHI program.

CONCLUSIONS

Although currently the majority of decision makers and scientists does not intend to use, optimize, or establish methods for measuring WTP, most of them do not reject these methods and many are willing to learn more about them. To increase the likelihood of using these methods, decision-making scenarios should be made more realistic, for example, by using qualitative methods to identify salient attributes of hypothetical decisions. Adequate payment vehicles (such as percentage of income) should be used to help patients relate payment to a health benefit. In addition, the complex statistical model used to calculate WTP should be broken down into accessible parts, and various methods for measuring WTP should be compared to test accuracy. Finally, taking into account the strong influence of “subjective norm,” discussion not excluding ethical concerns within peer groups should be encouraged to increase the level of acceptance regarding WTP measures.

POLICY IMPLICATIONS

Given the great share of scientists and decision makers not rejecting methods for measuring WTP and being willing to learn more about the methods, the likelihood of using these methods for decision making may be increased by promoting the development of more realistic decision scenarios and adequate payment vehicles as well as encouraging discussion of methodological and ethical concerns within peer groups.

CONTACT INFORMATION

Oliver H. Günther, Psychologist (oliver.guenther@medizin.uni-leipzig.de), Research Fellow, **Hans-Helmut**

König, MD, MPH (hans-helmut.koenig@medizin.uni-leipzig.de), Professor, Health Economics Research Unit, Department of Psychiatry, University of Leipzig, Johannisallee 20, D-04317 Leipzig, Germany

REFERENCES

1. Ajzen I, Fishbein M. *Understanding attitudes and predicting social behavior*. New Jersey: Prentice-Hall; 1980.
2. Baron J, Greene J. Determinants of insensitivity to quantity on valuation of public goods: Contribution, warm glow, budget constraints, availability and prominence. *J Exp Psychol*. 1996;2:106-125.
3. Bundesministerium für Gesundheit und soziale Sicherung. *Gesetzliche Krankenversicherung*. 2004. Available at: <http://www.bmgs.bund.de/deu/gra/datenbanken/gkv/4504.cfm>. Accessed August 2005.
4. Cookson R. Willingness to pay methods in health care: A sceptical view. *Health Econ*. 2003;12:891-894.
5. Converse JM, Presser S. *Survey questions: Handcrafting the standardized questionnaire*. Beverly Hills: Sage; 1986.
6. Deutsche Koordinierungsstelle für Gesundheitswissenschaften. *Who is who in public health*. 2004. Available at: <http://www.medsoz.uni-freiburg.de/dkgw/whoiswho/recherche.htm>. Accessed August 2005.
7. Drummond M, Jonsson B, Rutten F. The role of economic evaluation in the pricing and reimbursement of medicines. *Health Policy*. 1997;40:199-215.
8. European Observatory on Health Care. *Health care systems in transition: Germany 2000*. 2000. Available at: <http://www.euro.who.int/document/e68952.pdf>. Accessed August 2005.
9. Fisher RJ. Social desirability bias and the validity of indirect questioning. *J Consumer Res*. 1993;20:303-315.
10. Gress S, Niebuhr D, Rothgang H, Wasem J. Criteria and procedures for determining benefit packages in health care: a comparative perspective. *Health Policy*. 2005;73:78-91.
11. Hair JF, Tatham R, Anderson R, Black W. *Multivariate data analysis*. New Jersey: Prentice-Hall; 1988.
12. Hanley N. Cost-benefit analysis and environmental policy making. *Environ Plan C*. 2001;19:103-118.
13. Hanley N, Ryan M, Wright R. Estimating the monetary value of health care: Lessons from environmental economics. *Health Econ*. 2003;12:3-16.
14. Hoffmann C, Graf von der Schulenburg JM. The influence of economic evaluation studies on decision making. A European survey. The EUROMET group. *Health Policy*. 2000;52:179-192.
15. Klose T. The contingent valuation method in health care. *Health Policy*. 1999;47:97-123.
16. Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: Inconsistency between current methods and random utility and welfare theory. *Health Econ*. 2004;13:901-907.
17. O'Brien B, Gafni A. When do the “dollars” make sense? Toward a conceptual framework for contingent valuation studies in health care. *Med Decis Making*. 1996;16:288-299.
18. Olsen JA, Smith RD. Theory versus practice: A review of ‘willingness-to-pay’ in health and health care. *Health Econ*. 2001;10:39-52.

19. Ryan M. A comparison of stated preference methods for estimating monetary values. *Health Econ.* 2004;13:291-296.
20. Ryan M, Amaya-Amaya M. 'Threats' to and hopes for estimating benefits. *Health Econ.* 2005;14:609-619.
21. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: Current practice and future research reflections. *Appl Health Econ Health Policy.* 2003;2:55-64.
22. Ryan M, San Miguel F. Revisiting the axiom of completeness in health care. *Health Econ.* 2003;12:295-307.
23. Ryan M, Skatun D. Modelling non-demanders in choice experiments. *Health Econ.* 2004;13:397-402.
24. Wasem J, Hessel F, Kerim-Sade, C. Methods of comparative economic evaluations of therapies and for rational allocation of resources across sectors of health care systems—introduction, advantages, risks. *Psychiatr Prax.* 2001;28:S12-S20.