Contingent Valuation of Multiple Health Programs : Major Improvements or Major Biases ?*

Stéphane Luchini[1,3]
Christel Protière[2]
Cam Donaldson[4]
Jean-Paul Moatti[2,5]

1) Groupe de Recherche en Economie Quantitative d'Aix-Marseille (GREQAM-CNRS), Marseilles, France

2) INSERM Research Unit 379 "Social Sciences Applied to Medical Innovation", Marseilles, France.

3) Institut d'Economie Publique (IDEP), Marseilles, France.

4) School of Population and Health Sciences and Business School (Economics), University of Newcastle upon Tyne, UK.

5) Université de la Méditerranée, Aix-Marseille II, France

Preliminary Version
CES HESG 2003

Abstract

Contingent valuation is the most flexible and controversial non-market valuation technique. To date, most CV surveys in health care have dealt with separate evaluations (SE) of single program. In publicly funded health care systems, where the opportunity cost of investing in one program is manifested in terms of benefits foregone in other programs, it might be more relevant to ask respondents about their valuation of competing candidates in one exercice. That is to say to perform a joint evaluation (JE), which consists of evaluating simultaneously different programs in a single survey. However, the use of JE explicitly raises the question of the potential interdependence in the individuals' valuation of different programs. In the EuroWill project, three health programs were simultaneously evaluated and gave us the opportunity to test for functional relationships between the dependent WTP variables using a simultaneous equation model. On the basis of our results, we show that an anchoring effect may occur. According to this, we propose some perspectives for further research about the use of JE of different programs for economic evaluation of health benefits.

Keywords: Health care, Willingness to Pay, Contingent Valuation, Joint Evaluation, Simultaneous Equations Model

JEL Classification: H4, I1

*__Corresponding author__: Christel Protière: INSERM U379, Institut Paoli Calmettes, 232 Bd Ste Marguerite, 13273 Marseille cedex 9, France. E-mail: protiere@marseille.inserm.fr. Tel: 33.4.91.22.38.18 (poste : 47 12).

1 Introduction

The contingent valuation (CV) method is the most flexible and controversial non-market valuation technique. The method relies on the direct elicitation of values from respondents by confronting them with a constructed or simulated market presented in a survey. The survey aspect of the method makes it the most flexible among the non-market empirical valuation techniques. CV has been extensively used for valuing public goods in the environmental field (Bateman and Willis 1999). Recent literature in health care economics has shown increasing interest in the use of CV for measuring willingness-to-pay (WTP) as the economic value of health benefits (Olsen and Smith 2001). To date, most CV surveys in health care have dealt with separate evaluations (SE) of single program or of alternative ways to deliver the same type of care. Therefore, the use of WTP data to inform public decision making about allocation of resources across different health programs would involve comparison of monetary estimations of consumer surplus obtained separately from different evaluations, which could be problematic in several respects. In publicly funded health care systems, where the opportunity cost of investing in one program is manifested in terms of benefits foregone in other programs, it might be more relevant to ask respondents about their valuation of competing candidates in one exercice (Olsen and Donaldson 1998). That is to say to perform a joint evaluation (JE), which consists of evaluating simultaneously different programs in a single survey.

It is claimed that JE provides the respondent with a well defined consumption set, over which preferences can be more easily expressed while such an explicit reference structure is missing when the assessment cognitive exercise concerns a single program (Bazerman, Moore, Tenbrunsel, Wade-Benzoni, and Blount 1999). However, the use of JE explicitly raises the issue of the potential interdependencies in individuals' valuations of different programs and, thus, the question of which econometric model best account for this. Researchers in environmental economics (Brookshire et al. 1979; Randall et al. 1981; Braeutigam and Noll 1984; Lave 1984) and experimental psychology (Tversky et al. 1990) have long been aware of such interdependence. Experimental economists have already shown how preferences for outcomes may change across joint versus separate evaluation of alternatives (Bazerman et al. 1999).

Our main assumption is that the interdependence between WTP values may be partly explained by an anchoring effect. Tversky and Kahneman (1974) first identified anchoring effect, stating that: «in many situations people make estimates by starting from an initial value that is adjusted to yield the final answer [...] Different starting points yield different estimates, which are biased toward the initial values». To date, most anchoring effects which have been observed in CV surveys have concerned starting point bias in iterative bidding games using dichotomous choice questions, the initial bid proposed to respondents strongly influencing their subsequent answers (Herriges and Shogren 1996; Green et al.

1998; O'Connor et al. 1999). In the context of multiple environmental programs valuation, Payne et al. (2000) have shown another kind of anchoring effect where the starting point seems to be spontaneously generated by respondents themselves: the first good valued in the sequence and the monetary response given to that good serves as an anchor or modulus for subsequent valuations. In their CV study, respondents evaluated five environmental programs in a randomized order using open-ended questions. Responses were analyzed with a panel data model. However, such an econometric model assumes that all WTP values for a given individual are exogenously determined by the same explanatory variables. For example, gender is supposed to have the same effect whatever the program being valued. It follows that specific program variables have to be introduced for all programs. Considering that all explanatory variables have the same impact on the WTP values for different programs seems however too restrictive. This is particularly true in the health field where health programs are addressed to specific populations.

The aim of this paper is to propose the use of a simultaneous equation models as a better way to account for potential interdependecies between WTP values. To date, few CV studies have evaluated more than one health program simultaneously (Olsen and Donaldson 1998; Olsen 1997; Donaldson et al. 1997; Miedzybrodzka et al. 1995). The EuroWill project, which was developed in six European countries to assess the feasibility of the CV method as a tool for measuring preferences of members of the general public about health care programs, was based on the JE of three different programs. The French part of the EuroWill survey gave us the opportunity to test the extent to which the reference structure that is exogenously provided to the respondent in a JE affects the estimates of WTP for each given program using a simultaneous equations model to uncover a functional relationship between the dependent WTP variables. On the basis of our results, we suggest further research about the use of JE of the benefits of different health programs. We also draw some conclusions about the appropriate use of CV for aiding decision making about the allocation of resources within publicly funded health cares systems.

2 Material and Methods

2.1 Data collection

The data used in this paper come from the French EuroWill survey carried out during May to July 1998, in a representative sample of the population of 18 years of age and over in South Eastern France. WTP questions were used to elicit respondents' preferences for three health programs covering heart Disease (HD), Breast Cancer (BC), and a Helicopter Ambulance (HA) service.

After explaining the purpose of the WTP exercise, the interviewers gave respondents

a description of the three health care programs they were going to evaluate. These options were described in terms of the proposed service, the population who would benefit from the new program and expected health outcomes with an without the programs. Table 1 presents a brief synthesis of programs description (See appendix A for a complete description).

The structure of the questionnaire was as follows. Respondents were first asked questions about their individual risk perceptions and family experience with each specific disease/situation addressed by each program. WTP for each three programs was then elicited always in the same order: HD, BC, and HA. Each Question about WTP was asked first in terms of whether the respondent would be willing to pay for the program and, if so, what would be his/her maximum amount. The questionnaire also included questions about reasons for being willing or not willing to pay and finally questions on respondents' demographic characteristics and self-perceived health status. This structure followed the basic format of the EuroWill questionnaire used in five other European countries.

WTP questions were asked in terms of additional contributions for health care. To avoid "protest zeros" resulting in part from an aversion to contributing to a publicly funded health care system, respondents who said "No" to paying in the form of increased contributions to the Sickness Fund of the French Social Security System, were asked if they would be willing to contribute in the form of a donation. If respondents were still not willing to pay, they were asked why, with responses categorized as follows: "Other programs are more valuable", "I can't afford it", "Not concerned", "Other public sector budgets should be cut", "Other taxpayers, better off, should pay", "Users should pay", "The health service should be more efficient", "I pay enough already", "I prefer other ways of paying", and "Other". According to the EuroWill protocol, respondents who felt into one of the first three categories were considered as having expressed a "true zero" WTP, while respondents who mentioned the other reasons were classified as "protestors". As usually done in CV studies, including EuroWill, protest responses were ultimately removed from WTP estimations.

2.2 Econometric model

An appropriate method to test the hypothesis that in a joint evaluation exercise, individual valuation of each program is affected by a common reference structure implicitly constructed by respondents (*i.e.* that expressed WTPs are interrelated), is to use a simultaneous equations model. The structural form of such model can take into account possible endogeneity biases and can be written as follows

$$WTP_i\Gamma = X_iB + U_i \tag{1}$$

where i is the individual index, WTP is a row vector of revealed WTPs of individual i such that

$$WTP = \begin{bmatrix} WTP_{1i} & WTP_{2i} & \dots & WTP_{Mi} \end{bmatrix}$$

where M is the number of programs evaluated in the protocol.[1] X_i is a row vector of explanatory variables. Γ is a $M \times M$ matrix of coefficients:

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{1M} \\ & & \vdots & \\ \gamma_{M1} & \gamma_{M2} & \dots & \gamma_{MM} \end{bmatrix} \tag{2}$$

with γ_{ij} being the coefficient associated with WTP of i^{th} program in the estimation equation of program j. The matrix B is a coefficient matrix where each column is the vector of coefficients for explanatory variables in a particular WTP equation:

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2M} \\ & & \vdots & \\ \beta_{M1} & \beta_{M2} & \dots & \beta_{MM} \end{bmatrix} \tag{3}$$

The structural disturbances are assumed to be randomly drawn from an $M-$variate distribution with

$$E[U_i] = 0 \qquad \text{and} \qquad E[U_i U_i^T] = \Sigma$$

with no inter-individual correlation. Hence, we assume that

$$E[U_i U_j^T] = 0 \qquad \forall i \neq j$$

Such a specification gives the opportunity to take into account the potential interdependencies between elicited WTPs: all equations are simultaneously used to estimate WTP for each program. Once the general model is explained, one is left with the identification problem. Indeed, without further information on the structural form, identification is impossible since the general model leads to an excess of M^2 parameters and the model is not estimable (Davidson and MacKinnon 1993). A way to solve the identification problem is to bring to attention our non-sample information. Such information that will be used in identifying the model will consist, for instance, of omitting variables from an equation which will lead to zeros being placed in the matrices B and Γ. This will decrease the total number of parameters to be estimated.

[1] It should be noted that, although it may seem challenging to consider M programs, in practice M may be small given that, at any one time, a decision maker may be concerned only with service increases or decrements at the margin Shackley and Donaldson (2000).

The sequential procedure used to elicit WTP values for each program and the fact that the sequence used was identical for all respondents (the first value being for the HD program, then for the BC program and finally for the HA program) allowed us to consider a special case of the general model where Γ is said to be triangular and is written as follows

$$\Gamma = \begin{bmatrix} 1 & \gamma_{12} & \dots & \gamma_{1M} \\ 0 & 1 & \dots & \gamma_{2M} \\ & & \vdots & \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{4}$$

This specification therefore corresponds to the assumption that the WTP for the first program is completely determined by the exogenous factors, while it affects WTP for the second program and WTP for the j^{th} program is affected by WTPs for all the previous ones. In addition to this assumption, we used a standard normalization in setting $\Gamma_{jj} = 1$ for all j. Each endogenous variable WTP_j has then a coefficient of unity in one and only one equation.

Moreover, additional nullity constraints on coefficients of B are needed to solve the identification problem since with a triangular matrix Γ, $M(M-1)/2$ undetermined values remain. Because some potential explanatory variables (for instance, individual risk perceptions and family experience with each specific disease) could be assumed to be only related to the specific corresponding program, these variables were not included in the WTP equations of the other programs. Hence, we will impose nullity constraints for some of the explanatory variables in each WTP equation and the corresponding parameters will be zero in the B matrix. For practical use, note that, as the number of programs jointly evaluated grows, the number of program specific explanatory variables has to increase too in order to solve the identification problem.

According to the EuroWill protocol, this simultaneous equation model will have to be applied to the subsample of respondents who gave WTP values to all three programs. However, the exclusion of protestors from WTP analysis may create potential misspecifications of WTP estimations (Haab 1999; Jorgensen et al. 1999). In order to control for potential sample selection bias (Heckman 1979), we add a single participation equation to the simultaneous equations model with the dependent variable defined as:

$$d_i = \begin{cases} 1 & \text{If the individual does not protest to all programs,} \\ 0 & \text{else.} \end{cases} \tag{5}$$

Hence, if $d_i = 1$ the WTP values for the M programs are observed and integrated in the simultaneous equation model. If $d_i = 0$, individual i is considered as a "protestor" and her/his answers to all M programs are left aside. The final econometric model is now

defined as

$$\begin{aligned} WTP_i^{\star}\Gamma &= X_i B + U_i \\ d_i^{\star} &= Z_i \alpha + v_i \\ d_i &= 1 \quad \text{if} \quad d_i^{\star} > 0; \quad d_i = 0 \quad \text{otherwise} \\ WTP_i^{\star} &= \left[WTP_{1i} \times d_i^{\star} \quad WTP_{2i} \times d_i^{\star} \quad \ldots \quad WTP_{Mi} \times d_i^{\star} \right] \end{aligned} \tag{6}$$

where the participation equation corresponds to a standard probit model assuming that error terms v_i are normally distributed.

In practice, the estimation method follows a procedure using four steps. In a first step, we estimate the probit participation equation in the total sample of respondents. In the second step, in the subsample of respondents who were non protestors, each WTP equation is explained in terms of the explanatory variables X_i and the disturbances U_i. This corresponds to the reduced form of the model. We estimate each equation of this reduced form and keep the fitted values. In the third step, we estimate the structural form of the simultaneous equation model. We use fitted endogenous variables obtained in the previous step instead of actual values in each WTP equation. As usual with simultaneous equation models, the fourth step consists of computing the estimate of the disturbance variance using the original variables rather than the predicted ones. For each step following the probit estimation of the participation equation, we control for selection bias introducing the inverse Mill ratio in each equation.

3 Empirical results

Characteristics of the total sample of 303 respondents who participated in the French EuroWill survey are described in Appendix B. Among the 303 respondents, a total of 163 individuals expressed a WTP for each of the three programs while the remaining 140 were classified as "protestors". Table 2 presents the final results of the probit estimation of the participation equation applied to the whole sample. After stepwise introduction in the probit model of all variables describing sample characteristics (Appendix B), only two factors were significant. First, respondents who ultimately expressed "some" or "many" difficulties in answering to the questionnaire were indeed more likely to have fully participated in the valuation exercise. This could be interpreted as people who give values for each program having thought more about the valuation exercise, as already found in other CV studies (Ryan and San Miguel 2000). Secondly, those who held private complementary health care insurance in addition to the publicly funded social security coverage were more likely to be protestors. Explanations of the second result could be that already having private protection either renders people to be less concerned within increase in health care supply as signals an "anti-public-sector" attitudes.

Among the 163 respondents whose WTP answers were available for the three programs, the proportions of respondents who expressed a null WTP (true zero), were respectively 21 (12.9%), 20 (12.3%) and 34 (20.9%) for HD, BC and HA. It must also be noted that among "true zeros", a higher proportion justified this choice with the reason that "other programs are more valuable" in the case of the HA program (14 (41.2%)), the rates for the HD (3 (14.3%)) and the BC (3 (15%)) programs being lower. Among these 163 respondents, mean WTP for the HA program (246 FF (+/- 292)) was significantly lower than for the HD program (346 FF (+/- 525) - $p<0.05$) and for the BC program (305 FF (+/-295)- $p<0.05$). High values of Spearman's correlation coefficients were obtained between WTP values for each program ($\rho_{hd,bc} = 0.687$, $\rho_{hd,ha} = 0.773$ and $\rho_{bc,ha} = 0.692$).

More than one third of the respondents (n=61; 37.4 %) gave similar WTP values for each of the three programs. Decreasing and increasing WTP across the sequence of HD, BC and HA programs were respectively observed for 42 (25.8 %) and 12 (7.4 %) respondents. For the remaining 48 (29.4 %) respondents, WTPs exhibited non monotonic trends (increase for BC then decrease for HA, decrease for BC then increase for HA).

Table 3 presents the results of the simultaneous equations model for the WTP for the three programs. Firstly, it must be noticed that the sample selection parameter is significant for the first WTP equation (HD program), suggesting that, in this case, parameter estimates in the sub-sample of non protestors would have been biased if we would not have controlled for sample selection through the probit estimation. Secondly, as predicted by conventional economic theory, level of income is the critical determinant of WTP for all three programs. However, while WTP for the HD program increases with income, for the two other programs, this relationship appears non monotonic , first increasing then decreasing beyond a certain threshold in the highest income category. It must however be remembered that the programs involved in our evaluation concern improvements in delivery of care following episodes of acute morbidity. Individuals in the higher social groups are well known to adhere to prevention and screening in highest proportions than the rest of the population (Jones 1994). Therefore, they may be less willing to contribute to publicly funded health programs potentially affecting the whole population. The apparently counter-intuitive result that respondents who express a higher risk perception for breast cancer were willing to pay less for the BC program (although this variable is only close to statistical significance at $p = 0.1$) may indeed depend on a similar kind of phenomenon: in the years previous to our study and in the geographic area where it took place, large screening programs for breast cancer were introduced; respondents who are the most sensitive to the risk of breast cancer may also be the ones who make a trade-off in favor of prevention of this disease rather than cure.

Thirdly, it must be noticed that no variable related to health status and experience with the disease in the family environment remain significant in the model. The only

exception is age for the HD program which is close to statistical significance, with younger respondents willing to pay more for this program. This result seems contradictory with the usual expectation of human capital models that the derived demand for curative medical care increases with age as health capital depreciates faster (Grossman 1972). However, this general rule may not apply to specific health care interventions: the expected value of heart operations may be greater when they are viewed as preventing premature death rather than prolonging life in older ages.

Finally, and more importantly, Table 3 confirms the hypothesis of functional interdependencies between WTPs of jointly evaluated programs: WTP of the previously evaluated program in the sequence of elicitation (WTP_{HD} for the BC program, WTP_{BC} for the HA program) significantly affects WTP of the next program. Moreover, this effect seems to increase for the last program (HA).

4 Discussion

Because most of previous CV surveys in heath care have dealt with the evaluation of only one single option (or alternative ways to deliver the same health care intervention), the methodological problems raised by the potential interdependence between respondents' evaluations of different programs in the same survey have rarely been discussed in the literature. Moreover, when JE of different programs has been previously performed in the field of health, WTP estimations were computed using independent estimations for each program which ignored potential interdependencies and may have led to major misspecifications (Luchini et al. 2003).

The EuroWill project was one of the first empirical attempts to carry out a joint CV of several health care programs within the same CV survey. Using a simultaneous equations model, our analysis demonstrates that the multiple program valuation context generally affects the value of a change in each particular good. Our result confirms those obtained in previous CV surveys for multiple environmental programs (Hoehn and Loomis 1993; Payne et al. 2000; Atakelty et al. 2000) although these studies used different survey designs: in the context of JE, the valuation of each program is affected by valuations of the other programs.

Of course, as in any other CV study, our analysis cannot escape the issue of the extent to which the elicitation procedure has directly influenced respondents' answers. It is already a well-established fact that WTP answers may differ depending on how many goods or programs respondents are asked to value (Bazerman et al. 1999) and how much information is given about each of them (Munro and Hanley 1999; ProtiÃ¨re, Donaldson, Luchini, Moatti, and Shackley 2003). One possibility, which has been extensively discussed in the

public economics theoretical literature (Green and Laffont 1979), is that the hypothetical character of the CV method does not provide incentives for true preference revelation and that respondents rather answer strategically in order to affect the outcome of the study. There is however little empirical evidence to support the hypothesis of systematic strategic bias in CV studies (Bohm 1972; Scherr and Babb 1975; Smith 1979; Milon 1989). Therefore, attention has rather been focused on various psychometric effects which could introduce biases in answers to CV survey questions (Mitchell and Carson 1989; Green et al. 1998). The context of JE of multiple programs is logically more suited than single evaluation (SE) to detection of some of the most important of these effects, such as sequencing effect, whereby the ordinal position of a program in a sequence affects its valuation (Payne et al. 2000), and anchoring, whereby of the first good and its associated WTP exert an influence on WTP amounts for the subsequent programs submitted to valuation, as shown by Payne et al. (2000) and also illustrated by our results.

The results of the EuroWill multiple valuation of health programs and the use of a simultaneous equations model made it possible to identify and quantify an iterative anchoring effect in which the active valuation of a prior good in a sequence directly affects the valuation of the subsequent one. Moreover, the explanatory variables selected and their associated coefficient differ between programs. An obvious limitation of the design of the French part of the EuroWill survey is that we have not randomly permutated the sequence order of the programs submitted to valuation. Therefore the identified anchoring effect cannot be strictly separated from a potential classical sequence effect. It must however be noted that evaluation of both effects simultaneously would have increased dramatically the number of parameters in the econometric model to the extent that identification may have become impossible. One of the other EuroWill studies tested for ordering founding that rankings, but not WTP values, where subject to such an effect (Stewart, O'Shea, Donaldson, and Shackley 2002).

Of course, the interdependence between WTP values for different health programs jointly evaluated in the same CV exercise which we found in the EuroWill study should not be viewed as surprising, given what we know about the psychology of scaling and valuation and the nature of the typical cognitive CV task. Psychologists have clearly shown that construing a preference task involves two interrelated steps for the individual respondents: first, deciding which features might matter, thereby providing structure, and, second, providing substance to that structure by deciding how to interpret potentially relevant features (Fischhoff 1991; Fischhoff et al. 1980; Payne et al. 1999). Such processes are a necessary and normal part of everyday life. In hypothetical exercises like CV surveys, problems arise if respondents invent more of a transaction than is effectively proposed, *i.e.* evaluate a different proposal than the one being offered (Fischhoff et al. 1999). If interdependence between WTP answers in JE is interpreted as the result of psychometric biases induced by the elicitation procedure, our results however suggest that these biases

will remain implicit and totally uncontrolled in the context of SE of programs. While the EuroWill study suggests caution in using monetary expressions of value from a CV survey as estimates of absolute economic values of health goods in cost-benefit calculations and/or inter-programs resources allocation, it also suggests that using CV data from separate evaluations may indeed maximize these risks of under or over estimations of «true» preferences of the population.

Moreover, JE can be seen as a way as a way to include direct and forceful reminders to respondents of the economic context in which they have to express their WTP values, providing them a relevant range of available health programs. Indeed, in the case of applications of CV to environmental goods, the National Oceanographic and Atmospheric Administration (NOAA) Panel (Arrow et al. 1993) has strongly pointed out the necessity to adopt study designs which are appropriate for reminding respondents of substitutes and budget constraints when performing a CV task.

The issue however remains open as to the extent to which the interdependencies observed in JE of multiple programs should not be interpreted only as psychometric "biases" but rather capture real economic relationships between preferences toward different health goods which can be substitutes or complements. Environmental economists have long been aware of interdependencies in CV of different programs and have often proposed an interpretation of these observed interdependencies in terms of substitution or complementary effects between environmental goods (Brookshire et al. 1979; Randall et al. 1981; Braeutigam and Noll 1984; Lave 1984; Hoehn and Randall 1989; Hoehn and Loomis 1993; Atakelty et al. 2000). As already shown in the environmental field, the EuroWill study suggests that failure to account for interdependence in valuing public programs may lead to systematic bias in conventional cost-benefit estimates. Our results may also be interpreted as the fact that JE of multiple health programs may help to ensure that respondents effectively consider substitutes and complements when valuing each good of interest. Of course, testing for substitution and/or complementary effects between the three programs would have necessitated a more complex study design than the one developed by the EuroWill protocol. Study designs proposed by Hoehn and Loomis (1993) and Atakelty et al. (2000), which allow respondents to choose any combination (including none) of the various programs and provide a series of repeated choices of various combinations of the programs, would have been more appropriate, although they may raise specific cognitive difficulties for respondents and ignore potential psychometric biases deriving from these difficulties. To decide whether JE of multiple health programs only makes more explicit major psychometric biases in the CV method or, in reverse, is better able to capture real economic behaviors, further research would need to adopt designs aimed at testing simultaneously for complementary and substitution effects between programs and psychometric effects related to the elicitation procedure.

5 Acknowledgments

The authors wish to thank Olivier Chanel, Russell Davidson and Thierry Magnac. The empirical part of this paper is based on the EuroWill project of the use of willingness to pay to assess community preferences for health care, funded by the European Commission. The views expressed in the paper are those of the authors.

References

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman (1993, January). Report of the NOAA panel on contingent valuation. *Technical Report*, **58**, p1601–1614.

Atakelty, H., W.L. Adamowicz, and P.C. Boxall (2000). Complements, substitues, budget constraints a,d valuation. *Environmental and Resource Economics*, **16**, 51–68.

Bateman, I.J. and K.G. Willis (Eds.) (1999). *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developping Countries*. Oxford University Press.

Bazerman, M.H., D.A. Moore, A.E. Tenbrunsel, K.A. Wade-Benzoni, and S. Blount (1999). Explaining how preferences change across joint versus separate evaluation. *Journal of Economic Behavior & Organization*, **39**, 41–58.

Bohm, P. (1972). Estimating demand for public goods : an experiment. *European Economic Review*, **3**, 111–130.

Braeutigam, R. and R. Noll (1984). The regulation of surface freight transportation: The welfare effects revisited. *Review of Economics and Statistics*, **66**.

Brookshire, D.S., R.C. d'Arge, W.D. Shulze, and M.D. Thayer (1979). Methods development for assessing air pollution control benefits. Vol. II, EPA-600/5-79-001b, National Tecnical Information Service, Spingfield, VA.

Davidson, R. and J.G. MacKinnon (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.

Donaldson, C., P. Shackley, and M. Abdalla (1997). Using willingness to pay to value close substitutes: carrier screening for cystic fibrosis revisited. *Health Economics*, **6**, 1.

Fischhoff, B. (1991). Value elicitation: is there anything in there? *American Psychologist*, **46**, 835–847.

Fischhoff, B., P. Slovic, and Lichtenstein (1980). Knowing what you want: measuring labile value. In T. Wallsten (Ed.), *Cognitive processes in choice and decision behavior*, pp. 117–141. NJ:Erlbaum: Hillsdale.

Fischhoff, B., N. Welch, and S. Frederick (1999). Construal processes in preference assesment. *Journal of risk and uncertainty*, **19**, 139–164.

Green, D., K.E. Jacowitz, D. Kahneman, and D. McFadden (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics*, **20**, 85–116.

Green, J. and J.J. Laffont (1979). *Incentives in public decision making*. Amsterdam: North Holland.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, **80**, 223–255.

Haab, T. (1999). Non-participation or misspecification? The impacts of nonparticipation on dichotomous choice contingent valuation. *Environmental Resource Economics*, **14**, 442–461.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, **47**, 153–161.

Herriges, J. and J. Shogren (1996). starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management*, **30**, 112–131.

Hoehn, J.P. and J.B. Loomis (1993). Substitution effects in the valuation of multiple environmental programs. *Journal of Environmental Economics and Management*, **25**, 56–75.

Hoehn, J.P. and A. Randall (1989). Too many proposals pass the benefit cost test. *American Economic Review*, **79**, 544–551.

Jones, L.J. (1994). *The social context of health and health work*. London: MacMillan Press Ltd.

Jorgensen, B.S., G.J. Syme, B.J. Bishop, and B.E. Nancarrow (1999). Protest responses in contingent valuation. *Environmental and Resource Economics*, **14**, 131–150.

Lave, L.B. (1984). Controlling contradictions among regulations. *American Economic Review*, **74**, 471–476.

Luchini, S., C. Protière, and J-P. Moatti (2003). Eliciting several willingnesses to pay in a single contingent valuation survey: Application to health care. *Health Economics*, **12**, 51–64.

Miedzybrodzka, Z., J. Semper, P. Shackley, M. Abdalla, and C. Donaldson (1995). Stepwise or couple antenatal carrier screening for cystic fibrosis?: women's preferences and willingness to pay. *Journal of Meddical Genetics*, **32**, 1995.

Milon, J.W. (1989). Contingent valuation experiments for strategic behavior. *Journal of environmental economics and management*, **17**, 293–308.

Mitchell, R.C. and T. Carson, R. (1989). *Using Surveys to Value Public Goods : The contingent Valuation Method*. Washington D.C.: Resources for the Future.

Munro, A. and N.D. Hanley (1999). Information, uncertainty, and contingent valuation. In I.J. Bateman and K.G. Willis (Eds.), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developping Countries*, pp. 258–279. Oxford University Press.

O'Connor, R.M., M. Johannesson, and P-O. Johansson (1999). Stated preferences, real behaviour and anchoring: some empirical evidence. *Environmental and Resource Economics*, **13**, 235–248.

Olsen, J.A. (1997). Aiding priority setting in health care: is there a role for the contingent valuation method? *Health Economics*, **6**, 603.

Olsen, J.A. and C. Donaldson (1998). Hearts and hips: Using willingness to pay to set priorities for public sector health care programmes. *Social Science and Medicine*, **46**, 1–12.

Olsen, J.A. and R.D. Smith (2001). Theory versus practice: a review of 'willingness to pay' in health and health care. *Health Economics*, **10**, 39–52.

Payne, J.W., J.R. Bettman, and D.A. Schkade (1999). Measuring constructed preferences: towards a building code. *Journal of risk and uncertainty*, **19**, 243–270.

Payne, J.W., D.A. Schkade, W.H. Desvouges, and C. Aultman (2000). Valuation of multiple environmental programs. *Journal of risk and uncertainty*, **21**, 95–115.

ProtiÃ¨re, C., C. Donaldson, S. Luchini, J-P Moatti, and P. Shackley (2003). The impact of information of non health attributes on willingness to pay for multiple health care programmes. *Social Science and Medicine*. Forthcoming.

Randall, A., J.P. Hoehn, and G. Tolley (1981). The structure of contingent markets: Some experimental results. Paper presented at the Annual Meeting of the American Economic Association, Washington, DC.

Ryan, M. and F. San Miguel (2000). Testing for consistency in willingness to pay experiments. *Journal of Economic Psychology*, **21**, 305–317.

Scherr, B.A. and E. M. Babb (1975). Pricing public goods: an experiment with two proposed pricing systems. *Public Choice*, **23**, 35–48.

Shackley, P. and C. Donaldson (2000). Willingness to pay for publicly-financed health care: How should we use the numbers? *Applied Economics*.

Smith, V.L. (1979). An experimental comparison of three public good decision mechanisms. *Scandinavian Journal of Economics*, **81**, 198–215.

Stewart, J., E. O'Shea, C. Donaldson, and P. Shackley (2002). Do ordering effects matter in willingness-to-pay studies of health care? *Journal of Health Economics*, **21**, 585–599.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 124–131.

Tversky, A., P. Slovic, and D. Kahneman (1990). The causes of preference reversals. *American Economic Review*, **80**, 203–217.

Table 1: Programs description (French EuroWill Study)

	Heart disease	Breast cancer	Helicopter ambulance
Proposed service	250 more heart operations on the 1,000 performed in the area	New treatment for poor diagnosis	Helicopter ambulance service better than "classic" ambulance
Concerned population	Men aged 60-70 years	Women aged 45-70 years	People having accident or disease
Outcomes	75% of men can live without pain and 25% with less pain. 1-2 additional years life Surgery mortality risk=2%	3 additional years survival for 3 women	15 life saved

Table 2: Participation equation to the WTP question for all three programs (probit model - French EuroWill study n=303)

Variable	Parameter estimate	P-value
Constant	0.436	0.0087
Difficulties with the questionnaire	0.353	0.0678
Having a private sickness complementary insurance	-0.495	0.0069

Table 3: Simultaneous equations model; WTP for three heath care programs (French EuroWill study n=163)

Variable	Program HD Parameter estimate (p-value)	Program BC Parameter estimate (p-value)	Program HA Parameter estimate (p-value)
Constant	806.0945772 (.0029)	5.420166628 (.9859)	96.48672059 (.7694)
Male	104.4527180 (.1877)	-88.56105145 (.2987)	35.60930400 (.7210)
Age	-4.084789922 (.1070)	-.8013178881 (.6927)	-1.692813353 (.3511)
Annual income between FF90,000 and FF132,000	258.5459127 (.0111)	140.1781293 (.0841)	112.0649365 (.0858)
Annual income between FF132,000 and FF240,000	287.8542732 (.0055)	231.8618088 (.0037)	142.0061469 (.0263)
Annual income more than FF240,000	564.6171427 (.0001)	214.4818857 (.0587)	61.54888751 (.5004)
More than 6 medical consultations during the last year	-139.6747017 (.2181)	149.6591864 (.1550)	-47.20392128 (.6848)
Professional certificate	126.8115113 (.3617)	-160.1225372 (.2241)	39.07725859 (.7737)
High school certificate at least	-88.85321083 (.3835)	-130.5203164 (.1093)	-14.60957053 (.8277)
Considering heath status a lot satisfactory	97.19885945 (.3437)	-117.0390449 (.2556)	51.51350546 (.6386)
Household having experience with the disease	63.31863344 (.4322)	-29.41544183 (.6414)	69.68012332 (.1667)
Considering own risk for having the disease equal to average	15.53465508 (.8780)	-41.10969110 (.6088)	37.53783866 (.7956)
Considering own risk for having the disease superior to average	147.4803666 (.2164)	-161.3153644 (.1073)	37.02768852 (.8213)
WTP for HD program		.7770648851 (.0495)	-.4023493843 (.4532)
WTP for BC program			.9907653835 (.0306)
Sample selection parameter	-541.3798073 (.0708)	99.73038138 (.6811)	-79.75532300 (.7408)

A Scenarios

A.1 Heart disease

250 more heart operations can be provided each year in addition to the 1000 which are currently done in " region PACA ". Most of the extra heart patients are men aged 60-70 years. They have chest pain and breathe heavily when strained. The operation will make 75% of the patients completely free from pain with less pain for the rest. Without the operation the patients are expected to live 8-10 years. With the operation they will on average live 1 to 2 years longer. The operation mortality risk is 2% (so 1 in 50 people die whilst being operated on).

A.2 Breast cancer

Each year, in " region PACA " 200 new cases of serious breast cancer are diagnosed. A new treatment more efficient than the current one could be provided. Most of the patients are women aged 45-70 years. The standard treatment permit to 27 women to live in average 4 years and only one year for the others 173 women. The new treatment could permit to 3 additional women to live in average 4 years instead of only one year.

A.3 Helicopter ambulance

The helicopter ambulance is an alternative to car ambulances. It carries a medical staff. The major advantage of the helicopter ambulance is that it provides a quicker transport to the hospital if there is an emergency (sudden illnesses or accidents). Each year in "region PACA" helicopter ambulances would have roughly 800 missions. A study has shown that helicopters sometimes save lives, but most people transported would have had the same probability of survival if a car ambulance were used. Then, we can expected that helicopter ambulances save 15 lives each year that would not be saved if car ambulances had been used.

B Sample characteristics

Subset 1: Socio-demographic factors

Name of variables	Description of variable	
SEX	Male	73 (44.8%)
AGE	Mean age (SD)	45.16 (± 17.66)
COUPLE	Living in couple	92 (56.4 %)
NBPERS	Number of persons living in the household	2.72 (± 1.3)
NBCHILD	Number of children under 16 years of age	0.66 (± 2.72)
DPT04	Living in the department *Alpes de Haute-Provence*	7 (4.3%)
DPT06	Living in the department *Alpes Maritimes*	27 (16.6%)
DPT13	Living in the department *Bouches du Rhone*	64 (39.3%)
DPT83	Living in the department *Var*	35 (21.5%)
DPT84	Living in the department *Vaucluse*	30 (18.4%)
STUDY	1: No level	27 (16.6%)
	2: Primary school	22 (13.5%)
	3: Professional certificate (CAP)	38 (23.3%)
	4: Secondary	29 (17.8%)
	5: High school certificate	30 (18.4%)
	6: University degree	17 (10.4%)
PROF	1: Farmer, Artisan, merchant	11 (6.8%)
	2: Senior manager	20 (12.3%)
	3: Intermediary profession	27 (16.6%)
	4: Employee	28 (17.2%)
	5: Worker	44 (27%)
	6: Non-working or retired	2 (20.2%)
INC1	Annual income less than FF90,000	61 (37.4%)
INC2	Annual income between FF90,000 and FF132,000	42 (25.8%)
INC3	Annual income between FF132,000 and FF240,000	39 (23.9%)
INC4	Annual income more than FF240,000	15 (9.2%)

Subset 2: Experience with health and health care

Name of variables	Description of variable	
MUT	Having a private sickness complementary insurance	120 (73.6%)
CONS0	No medical consultation during the last 12 months	23 (14.1%)
CONS1	One or more consultations during the last 12 months	140 (85.9%)
NONSAT	Considering health status not satisfactory at all	20 (14.8%)
SAT	Considering health status quite satisfactory	67 (41.1%)
PLUTSAT	Considering health status a lot satisfactory	37 (22.7%)
TRESAT	Considering health status totally satisfactory	35 (21.5%)
RPS	Considering own risk for having heart disease superior to average	37 (22.7%)
RPM	Considering own risk for having heart disease equal to average	80 (49.1%)
RPI	Considering own risk for having heart disease inferior to average	46 (28.2%)
RCS	Considering own risk for having cancer superior to average	40 (24.5%)
RCM	Considering own risk for having cancer equal to average	92 (56.4%)
RCI	Considering own risk for having cancer inferior to average	31 (19%)
RAS	Considering own risk for needing ambulance superior to average	41 (25.2%)
RAM	Considering own risk for needing ambulance equal to average	99 (60.7%)
RAI	Considering own risk for needing ambulance inferior to average	23 (14.1%)
EXPPON	Household already having experienced heart disease	70 (42.9%)
EXPSEI	Household already having experienced cancer	85 (52.1%)
EXPAMB	Household already having experienced ambulance service	89 (54.6%)