



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Health Economics 24 (2005) 990–996

JOURNAL OF
HEALTH
ECONOMICS

www.elsevier.com/locate/econbase

Implicit versus explicit ranking: On inferring ordinal preferences for health care programmes based on differences in willingness-to-pay

Jan Abel Olsen^{a,b,*}, Cam Donaldson^c, Phil Shackley^d

EuroWill Group¹

^a *Institute of Community Medicine, University of Tromsø, 9037 Tromsø, Norway*

^b *Health Economics Research Programme, University of Oslo, Norway*

^c *Centre for Health Services Research and Business School (Economics), University of Newcastle upon Tyne, UK*

^d *Centre for Health Services Research, School of Population and Health Sciences, University of Newcastle upon Tyne, UK*

Received 1 March 2004; received in revised form 1 March 2005; accepted 1 April 2005

Available online 12 May 2005

Abstract

The paper explores the merit of the willingness-to-pay (WTP) method as a way to elicit public preferences regarding health care priorities. The aim is to test the extent to which the implicit ranking inferred from the ordinal differences in WTP-values corresponds with respondents' explicit ranking of the same programmes. This issue of convergent validity is explored by face-to-face interviewing of population samples in six European countries—in total 1240 respondents. The most consistent result is the inconsistency of WTP and explicit ranking in all six countries. The convergent validity of WTP is low, particularly among those who did not state different WTP-values on the three programmes being considered.

© 2005 Elsevier B.V. All rights reserved.

JEL classification: C9; D7; I1

Keywords: Willingness-to-pay; Ranking preferences; Health care priorities

* Corresponding author. Tel.: +47 77 64 4832; fax: 47 77 64 4831.

E-mail address: jan.abel.olsen@ism.uit.no (J.A. Olsen).

¹ Listed in [Appendix A](#).

1. Introduction

There are principally two different contexts in which the willingness-to-pay (WTP)-method becomes useful for health care decision makers; to provide the benefit-measure in a partial cost–benefit analysis when *one* new programme is considered, and; to compare the relative values of alternative programmes when *several* programmes are competing for the same resources within a fixed health care budget. This paper is concerned with the latter context; in that, it explores the merit of the WTP-method as a way to elicit public preferences regarding health care priorities.

In publicly financed health services, where health care managers lack fiscal discretion to finance the programme in question by raising the revenues that people say they would be prepared to pay, the cost–benefit motivation for obtaining WTP-values becomes futile. Under such policy restrictions, the usefulness of the method lies in considering the differences in the obtained WTP-values as expressions of respondents' preferred ordinal ranking of alternative programme options.

While this way of inferring *implicit ranking* based on differences in partially stated WTP-values represents an indirect way of obtaining ranking preferences, a more *explicit* way is to present the alternative programmes and ask the respondent to compare them and give her preferred ranking of their importance. In theory, rational respondents would reveal consistent rankings between the two methods, but do they?

The aim of this paper is to test the extent to which the *implicit* ranking inferred from the ordinal differences in WTP-values corresponds with respondents' *explicit* ranking of the same programmes. This is an issue of consistency, or a test of *convergent validity*, which measures the degree to which concepts that should be related theoretically are interrelated in reality.

One reason why this issue has not yet received much attention in the literature is that nearly all applied WTP-studies in health care have been partial studies of one programme in isolation (Olsen and Smith, 2001; Smith, 2003). An exception was a study of three programmes that showed a substantial discrepancy between the explicit ranking and the ranking implied from the partial WTP-values (Olsen and Donaldson, 1998; Olsen, 1997). These findings raised doubts over the use of WTP in this health policy context. Thus, further investigation of convergent validity was called for, and the EuroWill project was established in part to address this issue (see Donaldson, 1999 and Donaldson and Shackley, 2003 for a presentation of the project). Up to now, only single (and at most, double-) country investigations of other important methodological issues addressed by EuroWill have been published (Stewart et al., 2002; Protière et al., 2004; Ryan et al., 2004; Olsen et al., 2004a,b). This note is the first to publish findings based on the population surveys in all six European countries in which this project took place, addressing the convergent validity issue across all of these countries.

2. Empirical evidence for convergent invalidity

The results reported below refer to representative population samples in which three different health care programmes were presented in face-to-face interviews. Respondents

were asked: “to rank the three programmes in terms of how important you think they are”. After this, *explicit ranking* (ER) exercise, they were asked partial WTP-valuations of the same programmes in the same order (independently of their earlier explicit ranking of the programmes); firstly, asking if respondents would be willing to pay, and, if so, how much was the maximum their household would be willing to contribute each year for each of the programmes.

In accordance with microeconomic theory, we would expect consistency between the ranking implied from WTP and their previously stated *explicit ranking*. When a respondent had ranked A before B, she would be expected to signal this preference by being willing to sacrifice *more* money (i.e. higher WTP) in order to get A. Thus, we would hypothesise that WTP shows convergent validity with explicit ranking.

The implied rankings from the WTP-answers depend on the level of differences in the values provided. If three different WTP-values were given, it means *strong ranking*, i.e. one programme was strictly preferred to a second that was strictly preferred to the third. If two equal values and one different, there is a *weak ranking*, i.e. respondents preferred one programme to the other two, between which they were indifferent, or, they were indifferent to the first two, both of which were preferred to the third. *No rankings* are of two kinds, either three identical positive WTP-values or three zeros.

Table 1 shows the numbers from each country. The first column shows the implied rankings based on the WTP-answers, where the different degrees of ranking is referred to as *strong*, *weak*, and *no*. The subsequent columns show which degree of ranking the respondents within each of these groups had stated on the *explicit ranking* question. Thus, each number in the WTP-column represents the corresponding row sum under ER. Complete convergent validity would be revealed if numbers were seen in the diagonal only, i.e. combination *strong* in WTP and *strong* in ER; combination *weak* in WTP and *weak* in ER; combination *no* in WTP and *no* in ER. As can be seen from the six country tables, the figures are concentrated in vertical lines under *strong*, particularly so for three countries.

First, the only consistent group of respondents appears to be those who stated *strong* in WTP, where nearly all had stated *strong* in ER. The inconsistency for the other groups could be conceived of as a ‘weakening up’ of respondents’ willingness to rank between the programmes as the preference elicitation exercise went from ER to WTP, in that there is a shift towards the left of the diagonals in Table 1, i.e. lower degree of ranking from WTP-responses. However, if we modify the complete convergent validity criterion of numbers in the diagonals only, we note from the box combinations [(*weak* row) and (*strong* column)] that many who should be expected to state three different WTP-values have still stated *two* different values, i.e. they used WTP to discriminate between two rather than three programmes.

Table 2 summarises the results across the six countries. First, the (unweighted) country average shows that only 23% stated *strong* rankings in their WTP-answers, while 73% had done so in their explicit ranking. Second, respondents are most reluctant to rank the programmes through WTP; the country average showed that 41% were indifferent to the three programmes. Third, an important type of inconsistency is the proportion of respondents who stated indifference (no ranking) through WTP, but who in the previous preference elicitation exercise had stated *strong* ranking. The reason for the

Table 1
Comparing the degree of explicit ranking (strong, weak, no) with the degree of ranking as implied from WTP

	WTP-ranking	Explicit ranking			
		Strong	Weak	No	Missing
Norway					
Strong	51	48			3
Weak	39	34			5
No – posWTP	47	39		1	7
No – zeros	25	21			4
Total	162	142		1	19
Denmark					
Strong	40	39			1
Weak	101	100			1
No – posWTP	83	78	1	2	2
No – zeros	87	75	1		11
Total	311	292	2	2	15
France					
Strong	80	79	1		
Weak	75	67	6	2	
No – posWTP	45	41	1	3	
No – zeros	101	86	11	4	
Total	301	271	19	9	
Portugal					
Strong	63	54	7	2	
Weak	92	71	16	5	
No – posWTP	19	13	3	3	
No – zeros	33	19	11	3	
Total	207	157	37	13	
Scotland					
Strong	38	25	10	3	
Weak	53	23	24	5	1
No – posWTP	52	14	16	22	
No – zeros	3	2	1		
Total	146	64	51	30	1
Ireland					
Strong	15	11	3	1	
Weak	34	21	6	7	
No – posWTP	46	16	5	25	
No – zeros	18	4	5	9	
Total	113	52	19	42	

lower proportion of inconsistent respondents in Scotland and Ireland can be ascribed to the much smaller groups who had given *strong* in the initial explicit ranking exercise. Lastly, the column ‘Convergent validity’ shows the percentages of respondents in the diagonals of Table 1.

Table 2

Country comparisons of WTP-ranking with explicit rankings (ER) (missing included)

	'Strong' in WTP (%)	'Strong' in ER (%)	'No' in WTP (%)	'No' in ER (%)	% of 'no' in WTP giving strong in ER	Convergent validity (%)
Norway	31	88	44	1	83	30
Denmark	13	94	55	1	90	13
France	27	90	49	3	87	31
Portugal	30	76	25	6	62	37
Scotland	26	44	38	21	29	49
Ireland	13	46	57	37	31	45
Country average	23	73	41	11	64	34

The most notable differences between countries lie in the explicit rankings, in that the Scottish and Irish were less prepared to discriminate. As compared with the other four countries, there were relatively less respondents to state strong ranking, and relatively more to state no ranking. Reasons for such inter country differences might reflect anything from cultural factors (i.e. *true* preferences differences) to differences in survey administration. It might seem that the Irish respondents have a stronger aversion in general to rank health care programmes. As for the Scottish survey, respondents were recruited in a shopping centre rather than in their homes—something which might involve less focus in this cognitively fairly demanding exercise; compare information in three health care programmes before stating an ordinal ranking.

3. Discussion

Complete agreement cannot be expected between two different preference elicitation techniques. Even in the relatively simpler *explicit ranking* technique people make mistakes; they misread or they rank in the wrong direction. However, such inaccuracies in the answers stated on the underlying ER-yardstick were sought reduced by comparing three programmes only, and by having this as the first exercise in the interview, when respondents presumably had a fresh mind.

The most consistent result of these findings is the inconsistency of WTP and explicit ranking in all six countries. The convergent validity of WTP is low, particularly among those who did not state different WTP-values on the three programmes. When these results became evident from the first country surveys, the EuroWill group decided to try a different approach in Ireland. A so-called 'incremental approach' was developed (and used alongside the standard approach reported from in this paper), whereby WTP was first asked for the lowest ranked programme, followed by WTP being asked for the second ranked in terms of how much more they would be willing to pay to have that *rather than* their less preferred option, with the same incremental procedure being followed for their first ranked option (Shackley and Donaldson, 2002). However, this incremental version did not give a higher proportion of consistent respondents when comparing explicit ranking with the ranking

implied from WTP than the standard version did. One note of caution, though, is that respondents in Ireland seemed to perform better than in most other countries in terms of consistency in the standard version. Therefore, the scope for the incremental version to improve things was less in that country.

The chosen yardstick against which the convergent validity of WTP was tested can be criticised for its bias towards encouraging the respondents to make strong rankings, although equal ranks were allowed. However, this is exactly the type of choices that decision makers themselves are forced to make – and which, it is claimed, WTP assists them in doing. Rather, the question is whether preferences expressed on this ER-yardstick can be interpreted as reflecting the true underlying preferences – of relevance in this policy context. The case for this would be that explicit ranking is cognitively relatively simple: (i) it has a clear connotation to *real life* in health policy, i.e. ‘resources are limited, and one must prioritise between competing programmes’; (ii) the actual choice calls for *ordinal* rather than cardinal valuations; (iii) the valuation involves a *direct* comparison of alternative programmes, rather than *indirect* via one’s willingness to sacrifice own income. Three partial money trade-offs are cognitively more difficult, in that it changes people’s frame of reference to a choice between own income versus the existence of each of the described public programmes.

While ER may force respondents to make priorities, it would be fairly easy – in theory – to back up one’s preference ordering in the subsequent WTP-exercise by stating different values. However, the majority of respondents ‘weakened up’ their rankings, when they came to the WTP-exercise—*despite* the interviewer made them aware of any inconsistencies between the two approaches and also gave them the option of changing any such inconsistencies. In the Norwegian survey, those respondents whose explicit ranking did not match their implied ranking, and who thought that “surveys of this type should be used in the health service to help set priorities”, were subsequently asked: “which of the two methods would you like to be used?” The option “the one based on your ranking of the programmes” received 75% support, while the option “the one implied by your WTP-values” received 19% support. The remaining 6% did not know. Hence, the vast majority of respondents consider their answers to the explicit ranking exercise as the best reflection of their underlying priority preferences.

The high proportions of respondents stating the same positive WTP-values on three different programmes is a striking result, because there is *no* theoretical reason, a priori, why consumers should be prepared to sacrifice exactly the same level of income for three randomly selected goods. Thus, it seems like such respondents express a generally positive *attitude* to all programmes rather than well-behaved *preferences*.

It is worth noting that such difficulties in obtaining well-behaved consistent preferences are not restricted to WTP-surveys. Other preference elicitation methods used in economic evaluations, e.g. conjoint analyses and the various health status valuation techniques, appear to have problems with consistency (see, e.g. Lloyd, 2003; Bleichrodt and Johannesson, 1997). The question that remains is which methods should we use to elicit people’s preferences in priority setting contexts. The evidence presented in this paper poses serious challenges for WTP in this respect. Whilst convergent validity remains a key issue, an important, and parallel strand of research would be to establish more rigorously why these methods arrive at different results.

Appendix A. EuroWill partners and researchers

Professor Cam Donaldson, University of Newcastle upon Tyne, UK [EuroWill Coordinator]

Professor Jan Abel Olsen, University of Tromsø, Norway

Dr Phil Shackley, University of Newcastle upon Tyne, UK

Dr Kristian Kidholm, Odense University Hospital, Denmark

Professor Andrew Jones, University of York, UK

Professor Michael Jones-Lee, University of Newcastle upon Tyne, UK

Professor Graham Loomes, University of East Anglia, UK

Professor Jean-Paul Moatti, Inserm Research Unit 379, Marseilles, France

Dr Eamon O'Shea, National University of Ireland Galway, Republic of Ireland

Dr Joao Pereira, Universidade Nova de Lisboa, Portugal

Dr Christel Protiere, Inserm Research Unit 379, Marseilles, France

Professor Mandy Ryan, University of Aberdeen, UK

Dr David Scott, Fourth Hurdle Consulting, London, UK

Dr Jennifer Stewart, McMaster University, Ont., Canada

References

- Bleichrodt, H., Johannesson, M., 1997. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *Journal of Health Economics* 16, 155–175.
- Donaldson, C., 1999. Developing the method of 'willingness to pay' for assessment of community preferences for health care. Final Report to Biomed 2 Programme (PL950832) of the European Commission. Health Economics Research Unit, University of Aberdeen and Departments of Economics and Community Health Sciences, University of Calgary (available at website address: available by e-mailing the author: cam.donaldson@ncl.ac.uk).
- Donaldson, C., Shackley, P., 2003. Willingness to pay for health care. In: Scott, A., Maynard, A., Elliott, R. (Eds.), *Advances in Health Economics*. John Wiley, London.
- Lloyd, A.J., 2003. Threats to the estimation of benefit: are preference elicitation methods accurate? *Health Economics* 12, 393–402.
- Olsen, J.A., 1997. Aiding priority setting in health care: is there a role for the contingent valuation method? *Health Economics* 6, 603–612.
- Olsen, J.A., Donaldson, C., 1998. Helicopters, 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., Smith, R., 2001. Theory versus practice: a review of 'willingness-to-pay' in health and health care. *Health Economics* 10, 39–52.
- Olsen, J.A., Donaldson, C., Periera, J., 2004a. The insensitivity of 'willingness to pay' to the size of the good: new evidence for health care. *Journal of Economic Psychology* 25, 445–460.
- Olsen, J.A., Kidholm, K., Donaldson, C., Shackley, P., 2004b. Willingness to pay for public health care: a comparison of two approaches. *Health Policy* 70, 217–228.
- Protiere, C., Donaldson, C., Luchini, S., Moatti, J.P., Shackley, P., 2004. The impact of information on non-health attributes on willingness to pay for multiple health care programmes. *Social Science and Medicine* 58, 1257–1269.
- Ryan, M., Scott, D., Donaldson, C., 2004. Valuing health care using willingness to pay: a comparison of the payment card and dichotomous choice methods. *Journal of Health Economics* 23, 237–258.
- Shackley, P., Donaldson, C., 2002. Using willingness to pay to elicit community preferences for health care: is an incremental approach the way forward? *Journal of Health Economics* 21, 971–991.
- Smith, R.D., 2003. Construction of the contingent valuation market in health care: a critical assessment. *Health Economics* 12, 609–628.
- Stewart, J., O'Shea, E., Donaldson, C., Shackley, P., 2002. Do ordering effects matter in willingness to pay studies of health care? *Journal of Health Economics* 21, 585–599.