

available at [www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/ecolecon](http://www.elsevier.com/locate/ecolecon)

# Do stated preference methods stand the test of time? A test of the stability of contingent values and models for health risks when facing an extreme event

Roy Brouwer \*

Institute for Environmental Studies (IVM), Vrije Universiteit, De Boelelaan 1087, 1081 HV, Amsterdam, The Netherlands

## ARTICLE INFO

### Article history:

Received 9 November 2005  
 Received in revised form  
 28 March 2006  
 Accepted 4 April 2006  
 Available online 7 July 2006

### Keywords:

Benefits transfer  
 Temporal reliability  
 Contingent valuation  
 Extreme event  
 Health risks

## ABSTRACT

This paper focuses on the temporal stability of dichotomous choice (DC) contingent values and models and the implication of time related aspects for benefits transfer as a viable valuation method. From a practical point of view, time plays an important role in benefits transfer. Results are used from studies often carried out years ago and the obvious question is to what extent the estimated values and models stand the test of time. In the case of value function transfer, this implies testing the robustness of the estimated coefficient estimates over time. The novelty of the temporal stability work presented here is that two identical DC contingent valuation (CV) studies of the health risks associated with bathing water quality were carried out before and during extreme weather conditions causing the closure of bathing locations for public health reasons. Differences in contingent values and models before and during the extreme event are tested. The extreme event causes the seasonal good in question to become temporarily scarcer and marginal WTP is therefore expected to be higher during the extreme event than before, but not that underlying preferences have changed. Hence, basically two things are tested: the effect of time on stated preferences for a seasonal good and the effect of the extreme event on these preferences. WTP values before and during the event appear to be robust, also when accounting for substitution effects. The results before and during the extreme event remain transferable when accounting for theoretically expected factors in a simple multivariate transfer model. However, when introducing additional and contextual 'ad hoc' factors in the model, the estimated models and their random components become significantly different and hence not transferable, suggesting that unobserved preferences may not have been randomly distributed across the two surveys after all.

© 2006 Elsevier B.V. All rights reserved.

## 1. Temporal stability of contingent WTP values and models

The practice of benefits transfer generally implies that existing Willingness to Pay (WTP) values or models of WTP for specific environmental changes estimated with the help of nonmarket valuation methods are transferred across time, space, people and sometimes even natural resources, habitats and ecosystems. Values and value models from one constructed market

are used in another. Early guidelines (e.g. Boyle and Bergstrom, 1992; Loomis, 1992; Pearce et al., 1994) suggest that value model (function) transfer is more robust than simple average value transfer in view of the fact that effectively more information is transferred. In an overview of existing testing procedures, based on both contingent valuation (CV) and travel cost (TC) studies, I was unable to find conclusive empirical evidence for this claim or hypothesis (Brouwer, 2000). Most studies testing the reliability of benefits transfer focus on the spatial characteristics of

\* Tel.: +31 20 5985608; fax: +31 20 5989553.  
 E-mail address: [roy.brouwer@ivm.vu.nl](mailto:roy.brouwer@ivm.vu.nl).

ecosystem habitats and their beneficiaries (e.g. Loomis et al., 1995; Kirchoff et al., 1997; Bateman et al., 2000). Increasingly, meta-analysis techniques are used to also account for methodological differences between studies (e.g. Smith and Osborne, 1996; Woodward and Wui, 2000). New examples and methodological developments are found in this special issue.

In this paper, I will focus on the temporal stability of contingent values and models and the implication of time related aspects for benefits transfer as a viable valuation method. From a practical point of view, time plays an important role in benefits transfer. Results are used from studies carried out often years ago and the obvious question is to what extent the estimated values and models stand the test of time. When propagating the use of value functions in benefits transfer, the question is how robust the estimated coefficient estimates are in time. The National Oceanic and Atmospheric Administration (NOAA) Panel on Contingent Valuation (Arrow et al., 1993) raises some concern about the temporal stability of CV estimates. To date, test–retest studies, considering time periods ranging from 2 weeks to 2 years, have supported the replicability of findings and stability of values (McConnell et al., 1998). Yet there is no reason to assume that preferences and values for non-market goods should remain constant over extended periods of time, as shown in recent work (Brouwer and Bateman, 2005a).

The work presented in this paper is based upon two identical dichotomous choice (DC) CV surveys carried out at two different points in time, examining public preferences and values for bathing water quality improvements and associated health risks in the Netherlands. The novelty of the temporal stability work presented here is that the studies were carried out before and during an extreme event, i.e. the hot and dry summer in Europe of 2003. The Exxon Valdez study (Carson et al., 1997) also examined the stability and reliability of stated preference results in the face of an extreme event (large oil spill), but in this latter study the extreme event had already occurred when the CV studies were carried out at different points in time. Here, I present stated preference outcomes before and during an extreme event and test for possible differences.

During the bathing season in 2003, extremely high surface water temperatures were recorded and river flows were at a historical low, resulting in algal blooms and botulism, the death of fish, shellfish and birds, and the closure of bathing locations along the coast and inland for public health reasons (LCW, 2004). The Government issued a ‘code red’ for surface water intake in order to safeguard national electricity production. The extreme weather conditions and the effects on economic activities (shipping, irrigation, electricity production), wildlife (birds and fish) and people (recreational bathing in open waters, water use in and around people’s homes) were in the news every day for months, hence public awareness of the extremity of the event was very high.

Six months before this extreme event a large-scale CV mail survey was carried out, targeting 5000 randomly selected Dutch households, as part of a cost–benefit analysis (CBA) of the revision of the European Bathing Water Directive (76/160/EEC) in which more stringent bathing water quality standards were proposed (for more details about the CBA, see Brouwer and Bronda, 2005). During the extremely hot summer months, the same CV mail survey was carried out once again, again targeting 5000 randomly selected Dutch households, in order to test the stability and

#### Box 1 Valuation scenarios and WTP questions used in the CV survey

On average, currently one in every ten bathers (10%) runs the risk of getting ill from bathing in coastal and inland waters if current bathing water quality standards are not met (ear, eye or throat infections and/or stomach upset such as diarrhea). This risk occurs for instance during hot summer days in combination with heavy rainfall. If the sewer system is unable to process the excess rainwater, it is discharged directly in open water with a part of the sewer’s wastewater. Also excessive algae growth in standing waters as a result of hot weather conditions can cause failure of meeting bathing water quality standards.

Plans exist to strengthen the existing bathing water quality standards in order to reduce the public health risks associated with bathing in coastal and inland waters. Although these health risks can never be reduced completely, more stringent bathing water quality standards imply that in the future on average one in every twenty bathers (5%) instead of one in every ten (10%) runs the risk of getting ill during the bathing season under the mentioned circumstances.

The implementation of more stringent bathing water quality standards means that additional measures have to be taken to reduce the current health risks by half. These additional measures cost money. Examples of additional measures include:

- Increasing the sewer storage capacity to reduce storm water outfall when there is heavy rainfall
- Extra wastewater treatment (disinfection) to reduce the discharge of bacteria and other germs in open waters
- Installation of wastewater tanks on boats and wastewater collection points at marinas.

Q. How familiar were you before reading this questionnaire with the information provided above? [not familiar at all, not familiar, somewhat familiar, familiar, very familiar, DK].

Q. How difficult is it for you to translate the information above to your own situation and assess your own health risks when bathing in open waters? [not difficult at all, not difficult, somewhat difficult, difficult, very difficult, DK].

Q. If everybody contributes, would you be willing to pay in principle for more stringent bathing water quality standards? Note: this money will be used exclusively for the funding of the extra costs of the necessary measures to reduce the public health risks of bathing in open waters by half [yes, no, DK].

Q. Are you in that case willing to pay every year €... extra in terms of additional income taxation? [yes, no, DK].

Q. Can you explain why you are willing to pay this amount of money?

Q. Can you explain why you are not willing to pay?

temporal reliability of stated WTP values and estimated models for the same good (reduced health risks associated with bathing water quality improvements) under extreme conditions, i.e. when respondents face an extreme natural event.

Note that I have no a priori expectations for public demand for the proposed bathing water quality improvement to shift as a result of the extreme hot summer period and the deteriorated state of bathing water quality during the bathing season. The seasonal good in question has become temporarily scarcer as a result of the extreme event, and hence I expect marginal WTP to be higher during the extreme event than before, but not that underlying preferences have changed (i.e. change along the same demand curve for the good in question, not an upward shift of the demand curve for example). So, I basically test two things: the effect of time on stated preferences for a seasonal good and the effect of the extreme natural event on these preferences. For this I use two different samples, which makes the work presented in this paper different from previous test–retest exercises, but not less relevant for the practice of benefits transfer (temporal stability in-season and off-season).

The remainder of this paper is organized as follows. Section 2 presents details of the sampling procedure, the valuation scenarios and the sample populations. Section 3 briefly discusses the test procedures underlying the testing of the temporal reliability of stated WTP values and models. Section 4 presents the test results and Section 5 concludes.

## 2. Sampling procedure and sample population characteristics

A thoroughly developed and pre-tested DC CV questionnaire consisting of 45 questions was sent to 5000 randomly selected Dutch households by mail at the beginning of December 2002 and once again to another random sample of 5000 Dutch households by the end of August 2003.<sup>1</sup> Hereafter I will refer to these two surveys as ‘survey 1’ and ‘survey 2’ respectively. Besides eliciting standard demographic and socio-economic characteristics, respondents are asked for their bathing behavior, experiences, attitudes, awareness, knowledge, opinions and perceptions of bathing water quality and associated health risks, distinguishing between coastal and inland bathing water quality and using conventional measurement scales from socio-psychology (e.g. Summers, 1970; Nunnally, 1978; Slovic, 1987). At the end of the questionnaire, respondents are asked for their opinion about the questionnaire self and in particular the WTP questions. Response rates (31.3 and 29.6% respectively) are in the range usually found for CV mail surveys in the Netherlands (e.g. Hoevenagel, 1994; Brouwer and Spaninks, 1999; Brouwer, 2006).

Based on previous work, two WTP questions are asked. First, respondents are asked in a DC format whether they are willing to pay in principle for bathing water quality improvements without mentioning the specific money amount. This question

<sup>1</sup> Development of the questionnaire took several months, involving bathing water quality and epidemiological experts and professional survey and marketing experts. The questionnaire was pre-tested through 100 face-to-face street interviews on a random ‘next-to-pass’ basis in the cities of Amsterdam and Groningen. The questionnaire follows Dillman’s (1978) ‘total design method’.

**Table 1 – Summary of respondent characteristics in the two surveys and in the national population from which the respondents were randomly drawn**

Sample characteristic	Dec 2002 survey	Aug 2003 survey	National average <sup>a</sup>
Percentage women	40.2	43.1	50.5
Average age	53	52	39
Average household income (€/month) <sup>b</sup>	1935	1935	1900
Average household size	2.72	2.65	2.28
Percentage households with children	40.3	41.3	36.0
Average number of children	2	1	2
Percentage higher educated <sup>c</sup>	34,2	36,4	23,4

<sup>a</sup> Statistics Netherlands (year 2002–2003).  
<sup>b</sup> €1=US\$1.10 (01-12-2002–31-08-2003).  
<sup>c</sup> Percentage that finished high school or university with a diploma.

is used to distinguish protest bidders from legitimate zero bidders. Secondly, those respondents who said yes to the first WTP question are then asked for their WTP for a specific bid amount. Eleven bids were used, ranging from €1 to €200, based on the pretest results. Respondents who are not willing to pay the specified bid, but who are willing to pay another money amount are given the opportunity to state their maximum WTP in an open-ended follow-up question. The valuation scenarios are presented in Box 1.

Corresponding with results found in previous CV work in the Netherlands, ‘the polluter should pay’ is the most important reason for people to say ‘no’ to the first WTP question, resulting in a protest rate in both surveys of 10.3 and 11.2% respectively. Compared to non-protest respondents, protest respondents in both surveys appear to be slightly, but significantly younger men and women, with slightly lower education and income levels, with less interest in environmental issues (and hence usually not a member of any environmental or wildlife protection organization), a majority of whom do not want to pay for a cleaner environment in general.

Sample population characteristics are presented in Table 1. Although small significant differences are found between survey 1 and survey 2 regarding respondent age and the number of children in their households, the surveys are the same in terms of other demographic and socio-economic characteristics such as geographical spread (place of residence), household income and education.<sup>2</sup> Both samples represent a cross-section of the total Dutch household population in terms of geographical spread and household income. However, compared to the national population, women are slightly underrepresented in both surveys and the respondents are slightly older and higher educated.

Also no significant differences can be detected between the two samples when looking at bathing behavior and experiences (e.g. how often people swim in open water and where, whether people ever became ill after swimming in open water

<sup>2</sup> Mann–Whitney and Kolmogorov–Smirnov tests were used to test equality of central tendency and distributions of the variables discussed in this section. Test results are available from the author.

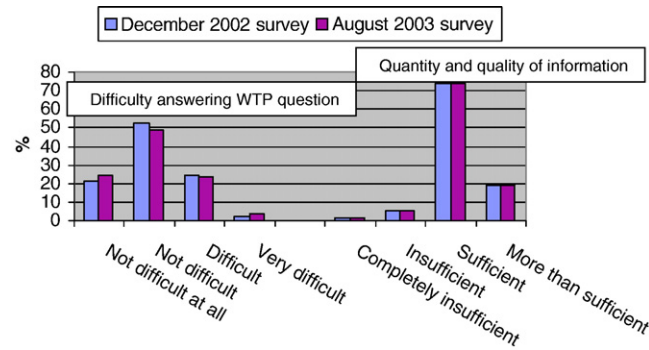


Fig. 1 – Respondent opinion regarding the WTP question and background information provided in the two surveys.

and, as a result, whether they went to see a doctor), respondent perception of bathing water quality (coastal and fresh inland waters) and the health risks associated with bathing in open waters. Respondents attach equal weight in both surveys to the importance of being informed about bathing water quality and the risks of swimming in open water, they are equally knowledgeable about existing bathing water quality standards (European blue flag system) and are equally upset when they are unable to swim in coastal or inland waters if poor water quality does not allow this. No difference furthermore exists regarding the availability of alternative outdoor and indoor bathing locations (substitutes) between the two surveys. Hence, no indication is found that preferences and underlying experiences, perceptions and motivations have changed as a result of the time elapsed and the extreme event.

Finally, no significant differences are found between the two surveys when examining respondent opinion about the questionnaire (e.g. general impression, how long to complete) and the WTP questions in the questionnaire (difficulty answering the WTP question and the quantity and quality of the provided information) (see Fig. 1).

### 3. Test procedures

Temporal reliability of DC CV models is tested by examining the statistical equality of unadjusted average WTP values (hypothesis 1) and the DC WTP functions (hypothesis 2). Comprehensive statistical testing procedures were originally proposed by Bergland et al. (1995). Testing the first hypothesis is relatively straightforward (e.g. Greene, 1993). The second hypothesis is tested in different ways in the literature. The Likelihood Ratio (LR) and Chow tests are most widely applied (e.g. Loomis et al., 1995; Kirchoff et al., 1997; McConnell et al., 1998). A similar test of the transferability of the estimated DC WTP functions is obtained by pooling the data from the two surveys as per Downing and Ozuna (1996) or Carson et al. (1997) and simply including a dummy variable representing whether the survey was carried out before or during the extreme event. If the survey dummy has a significant impact on respondent WTP, this implies that the study results are not transferable. The pooled regression results are the same as the outcomes of the LR test.

However, testing the equality of coefficient estimates may not be guaranteed when pooling data sets (Bergland et al., 1995). Showing that two samples originate from the same underlying

distribution does not guarantee equality of either their coefficient estimates nor their explained variance. All it shows is that the estimated models based on the individual samples equal the pooled model, not that the individual models themselves are the same. It has been shown that equality of coefficients is a necessary if not, on its own, sufficient condition for function transfer (Brouwer and Spaninks, 1999). Therefore, the Wald test is also used in this case study in order to compare the estimated coefficients of the two surveys directly. As in previous work, I will show that this produces different transfer results.

Mean and median WTP measures for DC responses are inferred from the underlying statistical distribution of the probability that respondents say yes or no to the different bid amounts (Hanemann and Kanninen, 1999). Different mean WTP values can be calculated depending on the statistical specification of the valuation function and the truncation strategies used. The calculation procedures for mean and median values based on binary WTP response data were first outlined by Hanemann (1984). Various statistical models exist to estimate underlying distributions. These include parametric approaches such as logit or probit models and non-parametric (distribution-free) approaches such as those developed by Kriström (1990) or the Turnbull model (Haab and McConnell, 1997). These models often produce significantly different results, with the Turnbull estimator yielding a lower bound on WTP. In this study, both parametric (logistic regression) and non-parametric (Turnbull

Table 2 – Mean WTP values (€/household/year) in the two surveys obtained from the logistic, Turnbull and open-ended WTP model

	Linear-logistic DC model		Turnbull DC model		Open ended WTP model	
	Survey 1	Survey 2	Survey 1	Survey 2	Survey 1	Survey 2
Mean WTP	70.5	71.4	32.5	30.4	32.9	29.0
Standard error	7.9	9.4	1.7	1.9	2.0	2.1
95% conf interval	55.0–86.0	53.0–89.8	29.1–35.9	26.7–34.1	28.8–37.0	24.8–33.3
Min–max values	–∞+∞	–∞+∞	0–200	0–200	3–100	2–100
n	1045	902	1045	902	139	110

€1 ≈ US\$1.10 (01-12-2002–31-08-2003).

estimation) approaches are used to test the temporal reliability of WTP values. The standard errors of the Turnbull models are furthermore estimated using non-parametric bootstrapping (e.g. Efron and Tibshirani, 1993).

Turning now to the testing of model transferability, an iterative approach is applied as introduced in previous work (Brouwer and Bateman, 2005a,b) in order to see how much control is needed to produce transferable models of WTP. These models are generated by progressively blending theoretically expected determinants of WTP with additional ad hoc variables. This approach involves a gradual expansion in the number of explanatory variables added to the DC model of WTP, starting from models solely utilizing variables that economic theory suggests should influence WTP (initially a two-variable model relating WTP to the DC bid amount and household income). The resulting model is then extended through the progressive addition of various other ad hoc control variables (such as use of the good in question, respondent demographic characteristics and attitude and opinion-related variables). At each addition of a variable, the transferability is assessed by applying the model to both survey data and undertaking various tests as described before. Theoretically, this progressive expansion approach should allow the identification of some optimal level of control for transferability.

#### 4. Temporal stability of estimated WTP values

No significant difference can be found between the mean WTP values before and during the extreme event, using either the linear-logistic or the Turnbull model at the conventional five percent significance level using the Normal (Z) test (Table 2).<sup>3</sup> As expected, mean WTP is significantly higher based on the linear-logistic than the Turnbull model. The latter is known to provide a lower bound for mean WTP (Carson et al., 1997). The difference is considerable. The linear-logistic mean value is twice as high as the Turnbull mean estimate. The accuracy of the linear-logistic mean value estimate as measured through the variation coefficient is lower than that of the Turnbull estimate.

As a sideline, another interesting finding is that the same results are found for the sub-sample of respondents who stated a maximum WTP after they said no to the DC WTP question for a specific bid (Table 2). No significant difference can be found between survey 1 and survey 2 using either Student's t-test or the non-parametric Mann–Whitney test.<sup>4</sup> Corresponding with previous findings (e.g. McFadden, 1994; Bateman et al., 1995), these open-ended WTP values are significantly lower than the DC estimates. Hence, corresponding with previous test–retest studies, the findings here show that time and the extreme event have had no impact on average stated WTP, irrespective of the applied estimation procedure.

Another interesting finding is that also no significant difference exists between both surveys when accounting for available substitution opportunities. Those respondents who said that they would go to an alternative indoor or outdoor bathing location if they would be unable to swim in open water

<sup>3</sup> All calculated z statistics are smaller than 1.96.  $|z|=0.11$  when comparing the difference between mean WTP values based on the linear-logistic model and  $|z|=1.2$  for the Turnbull model.

<sup>4</sup>  $t=1.292$ ;  $p=0.197$ ;  $MW Z=-1.542$ ;  $p=0.123$ .

**Table 3 – Mean Turnbull WTP values (€/household/year) from the two surveys when accounting for substitution opportunities**

	Survey 1	Survey 2
No substitution available	29.3 (2.9)	30.1 (3.2)
Substitution available	34.4 (2.7)	34.3 (3.0)
Standard error between brackets. €1 ≈ US\$1.10 (01-12-2002–31-08-2003).		

as a result of poor water quality are willing to pay, on average, slightly, but significantly more than those respondents who indicated that they do not have an alternative available (Table 3).<sup>5</sup> However, no significant differences can be found for these two groups between surveys, i.e. those respondents who have substitution possibilities are willing to pay, on average, the same before and during the extreme event. The same applies for those respondents who have no substitution possibilities.

#### 5. Temporal stability of estimated multivariate dichotomous choice models

The results in the previous section are confirmed when estimating a multivariate DC WTP model and using the ‘pooled model’ approach (see Section 3) by simply including a dummy variable for the two surveys carried out at different points in time before and during the extreme event (Table 4). The dummy variable is not statistically significant in any of the specified models to which the Wald test was applied.<sup>6</sup> All explanatory variables are highly significant in the statistically best-fit model presented in Table 4 except the dummy variable for the two surveys.<sup>7</sup> Hence, no significant differences exist before and during the extreme event and also the estimated multivariate model is transferable (see Downing and Ozuna, 1996).

<sup>5</sup> These results have to be interpreted with the necessary care. Proper testing of the effect of substitution possibilities on stated WTP for reduced health risks involves tests of the welfare changes associated with substitution. However, more detailed information about these welfare changes is not available from the questionnaire survey, e.g. about the opportunity costs of having to visit another outdoor or indoor bathing location. The substitution effects are furthermore not significant when accounting for other explanatory factors in the multivariate DC CV model.

<sup>6</sup> Only the results for the statistically best-fit model are presented here. The results for the other transfer model specifications are available from the author.

<sup>7</sup> Other variables for which no significant relationships were found include: respondent age, household size, number of children in household, place of residence (including clustering of places of residence in for example coastal and non-coastal provinces), respondent perception of bathing water quality, respondent awareness, knowledge and perception of health risks associated with bathing, the importance attached to information about bathing water quality and health risks, the extent to which respondents are upset if they are unable to swim as a result of poor water quality and the availability of alternative (substitute) bathing locations, outdoor and indoor. Respondent membership of an environmental organization has a significant impact on stated WTP, but is significantly correlated with household income and respondent interest in the environment in general and was therefore excluded from the model.

**Table 4 – Statistically best-fit pooled multivariate regression DC results**

Explanatory variables	Value range	Marginal effect
Constant	–	–0.174 (0.378)
Survey	0–1	–0.070 (0.125)
Bid level (€)	1–200	–0.020 (0.001)***
Household income (income intervals)	1–11	0.114 (0.022)***
Respondent sex (1=man)	0–1	0.407 (0.128)***
Swimmer in open water (1=yes)	0–1	0.413 (0.128)***
Interest in environment (Likert scale)	0–4	0.512 (0.104)***
Difficulty answering WTP question (0=not difficult at all ... 4=very difficult)	0–4	–0.816 (0.062)***
–2 Log Likelihood	1608.9	
Chi-square	749.4 (7 df, $p < 0.01$ )	
Percentage correct predicted	78.6	
Nagelkerke R-square	0.47	
n	1743	

\*\*\*  $p < 0.001$

Most coefficient estimates in Table 4 have the expected signs. The lower the bid amount or the higher household income, the higher the probability of a positive WTP (i.e. saying yes to the DC WTP question, *ceteris paribus*). Positive relationships are found for men, recreational swimmers and respondents who are more interested in the environment. Interestingly, the more difficulty someone experiences to answer the WTP question, the lower the probability of saying yes to a specific bid amount.

When using the Wald test instead of the pooled model approach, the results are very different and the estimated best-fit multivariate model presented in Table 4 is not transferable, either from survey 1 to survey 2 or the other way around (Table 5). Table 5 presents the Wald test results when progressively including more control variables in the transfer models. Only the theoretically expected model, including bid and household

income level, appears to be transferable, corresponding with previous findings (Brouwer and Bateman, 2005a). When introducing additional ‘ad hoc variables’, such as whether or not someone is an outdoor swimmer (user of the good in question), man or woman (demographic variable), how interested someone is in the environment in general (attitude variable) and the difficulty experienced in answering the WTP question (method related variable), these models appear to be non-transferable. Hence, transferability does not increase the more contextual control is included in the estimated transfer models, corresponding with recent findings by Ready et al. (2004).

In Table 5 also the LR test results are included. In principle, these should give the same results as the pooled approach. All the estimated models are transferable, except the model, which includes the variables bid, income and swimmer.

When examining the residual variance in the estimated transfer models, there appears to exist a systematic difference in the random component of the estimated transfer models (referred to as ‘unobserved preferences’ by McConnell et al. (1998)). The estimated models based on data from survey 2 result in a systematically higher residual variance (Fig. 2). The residual variance decreases only slightly with a few percent points (i.e. explained variance increases in Fig. 2) when including more explanatory variables in the transfer models besides the theoretically expected ones (bid and income). The inclusion of the method related variable (difficulty answering the WTP question) at the end, resulting in the statistically best-fit model, has a more substantial effect on the explanatory power of the transfer model. This systematic difference in the random component of the transfer models is expected to be an important reason for nontransferability.

## 6. Discussion and conclusions

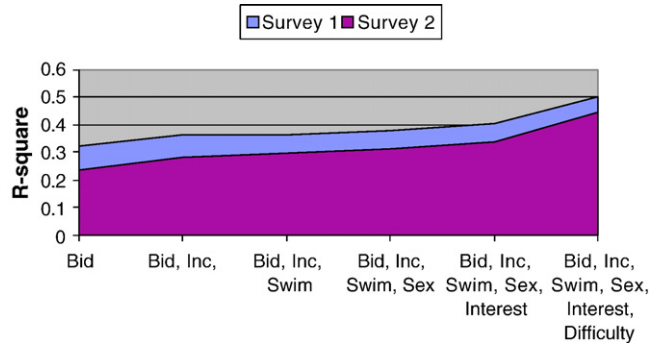
In the study presented in this paper, the transferability was examined of readily understandable use values of the environmental benefits associated with improved bathing water quality before and during an extreme event, i.e. extremely hot weather conditions resulting in a serious deterioration of bathing water quality. The valuation scenario (the ‘with’ situation) presented in

**Table 5 – Wald and Likelihood Ratio (LR) test results for different transfer model specifications<sup>1</sup>**

Estimated multivariate model <sup>2</sup>	Transfer		
	Survey 1→Survey 2 (Waldtest) <sup>3</sup>	Survey 2→Survey 1 (Waldtest) <sup>4</sup>	Survey 1↔Survey 2 (LRtest)
Bid	7.09>5.99	6.88>5.99	3.50<5.99
Bid, Inc	6.75<7.82	6.60<7.82	3.34<7.82
Bid, Inc, Swim	22.51>9.49	21.30>9.49	10.94>9.49
Bid, Income, Swim, Sex	21.57>11.07	20.81>11.07	10.60<11.07
Bid, Income, Swim, Sex, Inter	24.47>12.59	22.51>12.59	11.72<12.59
Bid, Inc, Swim, Sex, Inter, Diffic	29.02>14.07	26.69>14.07	13.87<14.07

Explanatory notes:

- If the Wald or LR test statistic < critical (Chi-square) value at 5%, then the model is transferable. Shaded rows indicate that the model is transferable.
- Bid: Bid level; Inc: Household income level; Swim: Swimmer in open water; Sex: Respondent sex; Inter: Interest in environment; Diffic: Difficulty answering WTP question.
- Survey 2 is unrestricted model.
- Survey 1 is unrestricted model.



**Fig. 2 – Development of the explained variance in the estimated transfer models based on data from survey 1 and survey 2 (measured through Nagelkerke R-square).**

the two surveys and the baseline scenario (the ‘without’ situation) are identical and well defined. Possible scope effects should therefore not play any role of significance in the valuation results. Public preferences (demand) for the proposed bathing water quality improvement are not expected to fundamentally change over the considered time period of nine months as a result of the extreme hot summer period and the deteriorated state of bathing water quality during the bathing season. Given the good’s seasonal character and the fact that the extreme event in effect made the good scarcer, average WTP is expected to be higher in-season (during the bathing period) than off-season (before the bathing period) whilst accounting for possible substitution effects. Notwithstanding the extreme circumstances and increased scarcity conditions, the average stated preference values from the two identical dichotomous choice CV studies appear to be robust and stand the test of time. Similar evidence, albeit under less extreme circumstances and for different environmental goods, is found in previous test–retest studies carried out over a time period shorter than 2 years.

The results before and during the extreme event remain transferable (from off-season to in-season and the other way around) when accounting for theoretically expected factors in a simple multivariate transfer model. When introducing additional ‘ad hoc’ factors in the multivariate transfer models, i.e. contextual factors which are not directly put forward by economic theory, but which appear to be significant determinants of stated WTP in the statistical analysis, the estimated models become significantly different and hence not transferable. I found similar results in previous transfer tests in other countries looking at other types of health risks and recreational amenities.

On the other hand, the estimated models are transferable when using the less stringent ‘pooled’ approach used in the literature (equivalent to the widely applied LR test). However, using this approach does not prove that the individual models are transferable, e.g. that the estimated off-season model is transferable to the extreme in-season situation, only that the estimated models based on the individual samples equal the pooled model. In practice, it is this former direct comparison of values and value models and their stability over time that is of primary interest in benefits transfer applications.

The finding that theoretically expected multivariate CV models are transferable and that the inclusion of ‘ad hoc’ variables yields non-transferable models seems to provide

important support for the practical use of CV results for the purpose of benefits transfer. If true, there is no need to look for additional contextual factors, including factors, which have been shown to influence behavioral intentions as measured through CV and for which usually no secondary data are available (contrary to income for example). Under certain circumstances even average value transfer instead of value function transfer may produce acceptable transfer errors.

However, the question in that case is how acceptable the use of a theoretically expected transfer model is in practice, which transfers reliably forwards and backwards in time and is able to cope with an extreme event, but remains somewhat of a black box when considering the systematic difference in the unobserved preferences between the transfer models. The random component of the transfer models appeared to be systematically higher when using data from survey 2 compared to the same model estimated with the help of data from survey 1, suggesting that there may exist significant differences in unobserved preferences throughout the entire testing procedure. The contribution of the occurrence of the extreme event on this observed systematic difference needs to be further examined. For instance, in the underlying study, the extreme event may have changed people’s perception of the probability of such an event happening and hence the risk involved and consequently their willingness to pay to avoid this risk. Given the ‘split sample’ nature of the surveys, I was unable to test this explicitly for the same respondents.

### Acknowledgements

I am grateful to Stavros Georgiou for his useful comments on a first draft of this paper and two anonymous reviewers. As always, I remain solely responsible for the content of this paper.

### REFERENCES

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R., Schuman, H., 1993. Report of the NOAA panel on contingent valuation. Federal Register 58 (10), 4601–4614 (January 15).  
 Bateman, I.J., Langford, I.H., Turner, R.K., Willis, K.G., Garrod, G.D., 1995. Elicitation and truncation effects in contingent valuation studies. *Ecological Economics* 12 (2), 161–179.

- Bateman, I.J., Langford, I.H., Nishikawa, N., Lake, I., 2000. The Axford debate revisited: a case study illustrating different approaches to the aggregation of benefits data. *Journal of Environmental Planning and Management* 43 (2), 291–302.
- Bergland, O., Magnussen, K., Navrud, S., 1995. Benefit transfer: testing for accuracy and reliability. Discussion Paper, #D-03/1995, Department of Economics and Social Sciences, Agricultural University of Norway.
- Boyle, K.J., Bergstrom, J.C., 1992. Benefit transfer studies: myths, pragmatism, and idealism. *Water Resources Research* 28 (3), 675–683.
- Brouwer, R., 2000. Environmental value transfer: state of the art and future prospects. *Ecological Economics* 32, 137–152.
- Brouwer, R., 2006. Valuing water quality changes in the Netherlands using stated preference techniques. In: Pearce, D.W. (Ed.), *Valuing the Environment in Developed Countries*. Edward Elgar, Cheltenham, UK.
- Brouwer, R., Bateman, I.J., 2005a. The temporal stability and transferability of models of willingness to pay for flood control and wetland conservation. *Water Resources Research* 41 (3), W03017.
- Brouwer, R., Bateman, I.J., 2005b. Benefits transfer of willingness to pay estimates and functions for health-risk reductions: a cross-country study. *Journal of Health Economics* 24, 591–611.
- Brouwer, R., Bronda, R., 2005. The costs and benefits of a revised European bathing water directive in the Netherlands. In: Brouwer, R., Pearce, D.W. (Eds.), *Cost-Benefit Analysis and Water Resources Management*. Edward Elgar Publishing, Cheltenham, UK.
- Brouwer, R., Spaninks, F.A., 1999. The validity of environmental benefits transfer: further empirical testing. *Environmental and Resource Economics* 14 (1), 95–117.
- Carson, R.T., Hanemann, W.M., Kopp, R.J., Krosnick, J.A., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K., Conaway, M., Martin, K., 1997. Temporal reliability of estimates from contingent valuation. *Land Economics* 73 (2), 151–163.
- Dillman, D.A., 1978. *Mail and telephone surveys: the total design method*. Wiley and Sons, New York.
- Downing, M., Ozuna, T., 1996. Testing the reliability of the benefit function transfer approach. *Journal of Environmental Economics and Management* 30, 316–322.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Greene, W.H., 1993. *Econometric Analysis*, Second edition. Macmillan Publishing Company, New York.
- Haab, T.C., McConnell, K.E., 1997. Referendum models and negative willingness to pay: alternative solutions. *Journal of Environmental Economics and Management* 32, 251–270.
- Hanemann, W.M., 1984. Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics* 66, 332–341.
- Hanemann, W.M., Kanninen, B., 1999. The statistical analysis of discrete-response CV data. In: Bateman, I.J., Willis, K.G. (Eds.), *Valuing environmental preferences, Theory and practice of the contingent valuation method in the US, EU, and Developing Countries*. Oxford University Press.
- Hoevenagel, R., 1994. The contingent valuation method: scope and validity. PhD Thesis, Faculty of Economics and Econometrics, Free University, Amsterdam.
- Kirchhoff, S., Colby, B.G., LaFrance, J.T., 1997. Evaluating the performance of benefit transfer: an empirical inquiry. *Journal of Environmental Economics and Management* 33, 75–93.
- Kriström, B., 1990. A non-parametric approach to the estimation of welfare measures in discrete response valuation studies. *Land Economics* 66, 135–139.
- LCW, 2004. Evaluatie van de werkzaamheden van de Landelijke Coördinatiecommissie Waterverdeling. RIZA report 2004.043x, Lelystad, The Netherlands.
- Loomis, J.B., 1992. The evolution of a more rigorous approach to benefit transfer: benefit function transfer. *Water Resources Research* 28 (3), 701–705.
- Loomis, J.B., Roach, B., Ward, F., Ready, R., 1995. Testing transferability of recreation demand models across regions: a study of corps of engineer reservoirs. *Water Resources Research* 31 (3), 721–730.
- McConnell, K.E., Strand, I.E., Valdes, S., 1998. Testing temporal reliability and carryover effect: the role of correlated responses in test-retest reliability studies. *Environmental and Resource Economics* 12, 357–374.
- McFadden, D., 1994. Contingent valuation and social choice. *American Journal of Agricultural Economics* 76, 689–708.
- Nunnally, J.C., 1978. *Psychometric Theory*, Second edition. McGraw-Hill, New York.
- Pearce, D.W., Whittington, D., Georgiou, S., 1994. *Project and policy appraisal: Integrating economics and environment*. OECD, Paris.
- Ready, R., Navrud, S., Day, B., Dubourg, R., Machado, F., Mourato, S., Spaninks, F., Vázquez Rodríguez, M.X., 2004. Benefit transfer in Europe: how reliable are transfers between countries? *Environmental and Resource Economics* 29, 67–82.
- Slovic, P., 1987. Perception of risk. *Science* 236, 280–285.
- Smith, V.K., Osborne, L., 1996. Do contingent valuation estimates pass a 'scope' test? A meta-analysis. *Journal of Environmental Economics and Management* 31, 287–301.
- Summers, G.F., 1970. *Attitude Measurement*. Rand-McNally, Chicago.
- Woodward, R.T., Wui, Y.-S., 2000. The economic value of wetland services: a meta-analysis. *Ecological Economics* 37, 257–270.