D. KRISTÓFERSSON AND S. NAVRUD

CAN USE AND NON-USE VALUES BE TRANSFERRED ACROSS COUNTRIES?

1. INTRODUCTION

An alternative to conducting a new environmental valuation study is to use existing values as an approximation. This method has been termed benefit transfer (or more general; value transfer) because estimates are transferred from a site where a valuation study has been conducted to a site of policy interest. Benefit transfer has become popular in practice due to the high costs and time associated with conducting original studies, regardless of the numerous difficulties associated with obtaining valid estimates. Numerous studies have tested the validity and magnitude of errors in benefit transfer for example Loomis (1992), Loomis et al. (1995), Bergland et al. (2002), Downing and Ozuna (1996), Kirchhoff et al. (1997), Brouwer and Spaninks (1999) and Ready et al. (2004) to name a few. It is easily argued that benefit transfer can only produce valid estimates in the few cases where the environmental good and the population are virtually identical. Otherwise there is no reason to believe that the willingness to pay (WTP) is the same for non-identical goods and populations. This raises a question about the way in which the validity is tested. Usually, a null hypothesis of no difference between the original and the transferred estimate is tested. Valid benefit transfer is reported in the cases where the null hypothesis has not been rejected at the chosen level of significance, most often $\alpha = 0.05$. However, non-rejection of a null hypothesis is not proof of its truth as a rejection is proof of its untruth (Lehman 1983, Hoenig and Heisey 2001). It is, therefore, not possible to prove equality when such a null hypothesis is not rejected. When we are interested in the validity of the null hypothesis itself, as is the case for studies of benefit transfer validity, it is appropriate to test for equivalence and not for difference. Such equivalence tests have been developed and used for quite some time in pharmaceutical research (Hauck and Anderson 1984, Schuirmann 1987, Welling et al. 1992 and Berger and Hsu 1996) and in psychological research, for example Stegner et al. (1996), but have no widespread use in economics.

We will here test the equivalence of benefit transfer values and original values for use and non-use values of freshwater fish populations between three Nordic countries; Norway, Sweden and Iceland. Identical contingent valuation surveys were conducted at the same time in these three countries. Observed differences in willingness-to-pay (WTP) should therefore be due to such factors as demographic differences between the populations, non-quantifiable differences in underlying preferences and the institutional organization of recreational fishing in different countries.

S. Navrud and R. Ready (eds.), Environmental Value Transfer: Issues and Methods, 207–225.

To our knowledge, this is both the first application of equivalency analysis to environmental value transfers, and the first study to compare the transferability of non-use values versus use values.

2. BENEFIT TRANSFER

Assume that all respondents in all countries have identical underlying preferences. Using country m as the baseline country, WTP for individual i living in country m for an improvement in environmental quality from Q_0 to Q_1 is defined using indirect utility functions by:

$$(1) \qquad V(p_m, I_i, Q_0) = V(p_m, I_i - WTP, Q_1)$$

where $\boldsymbol{p}_m$ is a vector of prices for goods and services in country m and I_i is the individuals income or wealth. Let us suppose that individual j has the same preferences as individual i but lives in country n. He faces prices $\boldsymbol{p}_n = \boldsymbol{p}_m$. Because the indirect utility function is homogeneous of degree 0 in prices and income, known as the absence of money illusion, it will not influence the result. The WTP of individual k will be $WTP_i = WTP_k$.

Benefit transfer methods can be divided into two major types: i) unit value transfer and ii) value function transfer.

Unit value transfer methods estimate total benefits at the policy site by aggregating exiting standard values per unit. These values are derived from study site data. For example, the total benefits of fishing at the policy site may be estimated as the product of some standard value for a fishing day at the study site and the number of fishing days at the policy site. The obvious problem with this method is that individuals at the policy site may not value the good in question in the same way as individuals at the study site. There are two principal reasons for this. First, the characteristics of the population may differ in terms of income, education, religion, demographic composition and so forth. Second, even if the individual preferences are the same, the supply of the good in question may differ (Kirchhoff et al. 1997).

A more sophisticated approach would be to adjust the value before transferring it to the policy location. There are two different types of adjustments that can be made. First, the analyst may regard the unit value available from the study site to be biased, or estimated inaccurately. This might be based on an evaluation of the methodology used in the original study. Second, the value may have to be adjusted to better reflect the conditions at the policy site. Four potential differences should be addressed in this kind of adjustments:

- The quality/quantity of the environmental good affected
- What caused the environmental change
- The socioeconomic characteristics of the households affected
- The availability of substitutes

In *value function transfer* methods, estimator models derived from study site data are used with explanatory variables collected at the policy site to estimate both value per unit and total units. For example, an estimated recreational demand function from the study site, may be used on data from the policy site to estimate both the value of a fishing day and total value of fishing (Brookshire and Neill 1992). Value function transfer is viewed as the best approach to benefit transfer as it relies on a better theoretical basis than unit transfers. However, the benefit estimates derived from contingent valuation studies are often a complex function of the site and user characteristics and the spatial and temporal setting. Not to take these into account is to make very strong assumptions about preferences, that is: preferences are universally stable over populations and time (Loomis 1992).

While rigorous guidelines exist on how to carry out original valuation studies (Arrow et al 1993), no such protocol exists for benefit transfer as of yet. However, as the number of studies on the validity of benefit transfer increases rules of practice emerge, see for example Desvousges et al. (1992), Bergland et al. (2002), Brouwer (2000) and Ready et al. (2004).

3. THE SURVEY

Toivonen et al. (2000) performed identical Contingent Valuation (CV) surveys in all Nordic countries during the period September–December 1999 to estimate the total economic value of the Nordic freshwater fish stocks. We have used the data from Norway and Sweden to test benefit transfer between two countries under close to ideal conditions. These countries are very similar with regards to geographical, ecological, cultural, and institutional context for this environmental good. We have then added Iceland as the "odd ball out" among the Nordic countries, with a larger degree of privatization of recreational fishing and having less threatened freshwater fish stocks.

Random samples of the national population (between age 18 and 69) of 2,500, 5,000 and 7,500 persons were selected for the mail surveys in Iceland, Norway and Sweden, respectively. The response rates were 34.2, 44.6 and 46.7 %, respectively.

The questionnaire had four main parts. In the first part, general attitudes towards wildlife and fishing were assessed. The second part, which was for anglers only, was aimed at identifying their angling activity, preferences for different types of recreational fishing, and annual expenses for recreational fishing. The third part contained four CV scenarios; three use-value scenarios aimed at anglers and one non-use scenario for all respondents. Only two of the use-value scenarios are relevant for all three countries analyzed here. Table 1 provides an overview of the three CV scenarios used in the comparative study reported here. The last part contained questions about socio-economic data, such as, sex, age, education and income.

A multiple bounded approach based on a payment card was used to elicit respondents' WTP. The payment card consisted of ten offered amounts and five levels of certainty for each offered amount. The amounts were the same in all countries, and converted to national currencies using Purchase Power Parity (PPP)-adjusted exchange rates. However, there is one exception. The highest amounts were higher

Table 1. Characteristics of the scenarios used in the contingent valuation study

	Scenario 1: Use Value Local River fishing	Scenario 2: Use Value Local Lake fishing	Scenario 3: Non-use Value Nordic fish stocks
Description	A stream is opened for fishing after being closed for many years. It is near your home. It has high water quality. It has a restricted number of anglers	A lake is opened for fishing after being closed for many years. It is near your home. It has high water quality. It has a restricted number of anglers	The natural fish stocks of the Nordic countries are threatened in several ways: Low water quality Regulation of water flow Eutrophication Acid rain Parasites Disease
Fish stock	Salmon and trout. Above average chance of catch	Grayling, brown trout and artic char. Above average chance of catch	All
Methods allowed	Rod and line	Rod and line	–
You buy	12 months access	12 months access	Preservation of natural fish stocks
Payment vehicle	Annual rent per person paid into a local fund, where the anglers get representation at the board of the fund.	Annual rent per person paid into a local fund, where the anglers get representation at the board of the fund	Increase in annual, personal income tax

in Iceland to account for the generally higher price level of fishing licenses. This was done to avoid a “thick tail” of the distribution of WTP, which can influence the estimated mean WTP. The respondents were asked to assign a certainty level to each amount. The certainty levels were “I would certainly pay”, “I would almost certainly pay”, “I am unsure”, “I would almost certainly not pay” or “I would certainly not pay” the offered value. The problem with this approach is choosing the level of certainty that corresponds to the respondents “true” WTP. Welsh and Poe (1998) and Notaro and Signorello (1999) compared multiple bounded question format with open-ended and single bounded formats. Their levels of certainty correspond to those used in the study. Their results show that the first level of certainty produces lower mean WTP then an open-end format. In both cases the single bounded model produces mean WTP higher than the “not sure” level of certainty, in the multiple bounded format. The results of Ready et al. (2001) indicate that this difference is due to “yea-saying” resulting from uncertainty. They conclude that applying the level of “almost certain” to closed-ended questions reduces the estimated WTP to the same level as for the payment card response. Thus, in our study, we use the second level of certainty (“I would almost certainly pay”) in the estimation of values, as the best approximation of “true” WTP.

The data is well suited for studies on the validity of benefit transfer as the surveys are identical, conducted at the same time with the same scenarios, using the same sampling procedure and in similar populations. The study was conducted in accordance with existing guidelines and the dataset is adequate in size. This should result in data that is free of bias resulting from temporal differences or differences in the offered good. The size of the data set ensures that parameter estimates are stable and statistical tests can be used.

However, the data also have some weaknesses. Although the scenarios are identical in all countries, they need not represent the same *relative* changes in environmental quality. That will ultimately depend on the initial level in each country. This may lead to differences in the relative change in environmental quality perceived by respondents in different countries. Sample selection is another potential problem. With the relatively low response rates in these mail surveys of the general public, there could be an over-representation of respondents with an interest in recreational fishing. However, if this takes place in *all* countries, it should not have any impact on the benefit transfer *validity tests* between countries.

4. MODEL AND TEST PROCEDURES

4.1. Statistical Model

Ready et al. (2004) stress the importance of correcting WTP for differences in purchase power when attempting benefit transfer between countries. Purchasing power parity (PPP) is a theory that states that exchange rates between currencies are in equilibrium when their purchasing power is the same in each of the two countries. This means that the exchange rate between two countries should equal the ratio of the two countries' price level of a fixed basket of goods and services. When a country's domestic price level is increasing (i.e., a country experiences inflation), that country's exchange rate must depreciate in order to return to PPP (Burda and Wyplosz 1997).

All monetary values had to be adjusted for different exchange rates as well as purchase power. It was decided to use Norwegian kroner (NOK) as the monetary unit of the study. The exchange rate used for Icelandic and Swedish kroner was the mean exchange rate of the Norwegian central bank for the duration of the survey (September – December 1999). Comparative price levels based on purchase power parity (PPP) were obtained from the OECD. The estimates that were used are shown in Table 2.

Multiple-bounded responses do not provide point estimates, but identify the intervals within which the "true" WTP lies. An upper and lower bound on maximum WTP is obtained from the payment card responses. Respondents with zero WTP can be identified using an open-end question following the payment card. These individuals have to be accounted for when estimating the expected WTP. Some respondents did not state their WTP, and were deleted from the data set.

Reiser and Shechter (1999) and Kriström (1997) showed the importance of not excluding true zero bids from the statistical analysis. The solution they suggest is to estimate a spike model. The first part of such a model identifies those with positive and those with zero WTP. The second part estimates a value function for the first

Table 2. Mean exchange rate and purchase power indices used for conversion into Norwegian kroner (NOK); 1 euro = 8.70 NOK (exchange rate January 2004)

	Mean exchange rate	PPP based comparative price levels	Overall adjustment
	To NOK	Norway=1,00	To NOK
Iceland	0,10962	1,01	0,11072
Sweden	0,94833	1,10	1,04317

Source: OECD, Norwegian Central Bank.

group and assigns zeros to the second. By the method of Reiser and Shechter (1999) the likelihood function breaks up into two parts, in correspondence to the two parts of the model. The log of the first part is:

$$(2) \qquad \ln L = \sum_{i=1}^{n} [(1 - S_i)\, p + S_i\,(1 - p)]$$

where S_i is a dummy variable taking the value 1 if WTP<0, p is the percentage of zero bids and n is the number of observations.

For the second part we need to define the intervals within which each bid falls. The payment cards included ten offered values. Eleven intervals can be constructed, nine where upper and lower values are known, and two where only the upper or lower value is known. A population consisting of t observations can be indexed by the set T. T can then be portioned into eleven disjoint subsets depending on the interval into which the true WTP falls.

Observations belonging to subset T_1 are left censored, observations belonging to $T_2 - T_{10}$ are interval censored and observations belonging to T_{11} are right censored.

The log-likelihood function for the second part can be written as:

$$(3) \qquad \ln L = \sum_{i \in T_1} \ln \mathrm{F}(h_{1i}) + \sum_{k=2}^{10} \sum_{i \in T_k} \ln\,(\mathrm{F}(h_{ki}) - \mathrm{F}(l_{ki})) + \sum_{i \in T_{11}} \ln\,(1 - \mathrm{F}(l_{11i}))$$

where $F(.)$ is the cumulative density function of the distribution of WTP across the population and l, h are lower and upper levels of the intervals. This model can be estimated by some statistical packages, for example the PROC LIFEREG procedure in the SAS® statistical system. A Weibull probability density function was used (Greene 2000, Bergland et al. 2002, Allison 1995, Lindsey and Ryan 1998).

The maximum likelihood estimator for probability of zero WTP, p, is according to Reiser and Shechter (1999):

$$(4) \qquad \hat{p} = \frac{n - \sum_{i=1}^{n} S_i}{n}$$

Combining the results from the estimation of zero bidders and those with positive WTP we can estimate the mean WTP. Let $f(x)$ be the probability density function associated with the distribution of the WTP. The mean WTP is given by:

$$(5) \qquad [WTP] = (1 - \hat{p}) \int_0^{\infty} x\mathrm{f}(x)dx$$

The integration was done using numerical integration and ∞ approximated with a large number.

This type of spike model uses more of the available information, but has the advantage of being completely combinable with existing estimation procedures (Reiser and Shechter 1999).

A comparison was made of three scenarios in all the countries, two use-values, and one nonuse-value scenario. The criteria used in constructing the models for each scenario was that the included parameters for the explanatory variables were significantly different from zero (10% level of significance was loosely applied), that they were not seriously correlated with each other (with $r = 0{,}3$ as upper limit) and that the model was applicable as a pooled model for all countries. Some variables were included for the sake of completeness, for example if one level of a classification variable was included, then all the other levels were automatically included as well.

The expected value of WTP is calculated by numerical integration. This value is then adjusted for missing values, as previously described.

Goodness of fit is calculated in a likelihood ratio index;

$$(6) \qquad I_{LH} = -2(\ln L_0 - \ln L_1)$$

where $\ln L_0$ is the log likelihood for a model that only includes a constant and $\ln L_1$ is the log likelihood for a model with covariates. The index indicates how much the model improved with added information. The index can be tested for significance as it has a χ^2-distribution with degrees of freedom equal the number of covariates.

The calculated WTP function is highly nonlinear in the estimated parameters. This makes estimation of standard errors difficult. Therefore, bootstrap distributions of mean WTP are obtained from 1000 bootstrap iterations that resemble the original dataset with replacement (see e.g. Bergland et al. 1993). Each iteration includes a full model is estimation. These results are then used to estimate mean WTP for an original model and a value function transfer model. This is a preferable method when model residuals are not well defined. Cooper (1994) conducted a simulation study of the validity of four different approaches for calculating confidence intervals for closed-ended WTP models. He concluded that this method showed best overall performance when the true underlying distribution was Weibull.

Transfer error is used as a measure of the accuracy of the benefit transfer estimates the definition for transfer error used here is

$$(7) \qquad \text{transfer error} = \frac{|WTP_E - WTP_T|}{WTP_T}$$

where the subscript T stands for true value as estimated by an original study and the subscript E for estimated value as given by benefit transfer.

4.2. Equivalence Testing and Acceptable Transfer Errors

In equivalence testing one reverses the roles of the null hypothesis (H_0) and the alternative hypothesis (H_A). Equivalence is demonstrated by testing a set of these reversed hypotheses with a predetermined significance level. It is not sufficient to fail to show a difference one must be fairly certain that a large difference does not exist. Suppose that we are willing to conclude that a difference is negligible if its absolute value is no greater than a small positive value Δ. In contrast to the traditional casting of the null hypothesis, the null hypothesis becomes[1]

$$H_0 : D \leq \Delta \text{ or } D \geq -\Delta$$

$$H_A : \Delta < D < -\Delta$$

where D is the absolute transfer error. The structure of the statistical hypothesis is determined by the objective of the analysis. The null hypothesis states that D is not within the interval that has been determined as equivalent. The alternative hypothesis states that D is within the interval, which implies that the two parameters are equivalent. If we can reject the null hypothesis on the basis of a study result, then we conclude that H_A is true, i.e. the two are equivalent. If the null hypothesis is not rejected we do not conclude that H_0 is true. Rather, we say that H_A has not been shown to be true. This procedure is exactly the same as the classical methodology of testing, only with reversed null and alternative hypothesis.

Hauck and Anderson (1984) and Schuirmann (1987) showed that if a 1–2α confidence interval lies entirely between $-\Delta$ and Δ, we could reject the null hypothesis of non-equivalence in favour of equivalence at the α level. The equivalence test is at the α level because it involves two one tailed tests, which together describe the 1–2α level confidence interval. A simple version of this kind of test is the two one-sided test or TOST. In its simplest form it involves conducting two one-sided t-tests at the α level of significance.

$$t_1 = \frac{D - \Delta}{s_p\sqrt{1/n_1 + 1/n_2}} \geq t_{1-}$$

and

$$t_2 = \frac{\Delta - D}{s_p\sqrt{1/n_1 + 1/n_2}} \geq t_{1-} \qquad (8)$$

where t_{1-} is the t-value associated with the chosen significance level and degrees of freedom, s_p is an estimate of pooled standard deviation and n_1 and n_2 are the number of observations in the two samples used to achieve the estimates that are being compared. Equivalence tests have some appealing properties over traditional non-rejections. It becomes increasingly difficult to reject the null hypothesis with increasing variance. A well-preformed CV study is therefore more likely to show equivalence given that it is the true state of nature while the reverse is true for non-rejections of classical null hypothesis.

The TOST is only one of the ways in which equivalence can be tested. Other, more powerful parametric test exist as well as non-parametric tests, see for example Berger and Hsu (1996). The simplicity and widespread application of the TOST in, for example pharmaceutical research, and its basis in the well-known t-test makes it a good choice for the purpose of our study.

In order to assess the equivalence of two test groups we must first define what would be considered equivalent. In the pharmaceutical industry the agreed upon standard is that the population mean tested must be within 20% of the mean of the reference group ($\Delta = 0, 2\mu_{\text{ref}}$). Such a standard must be set for each application. This standard must be based on what is considered theoretically relevant. This should not impose a problem since the theoretical interpretation of statistical results is contingent upon that such a definition exists. Kristofersson and Navrud (2005) suggest that in the case of benefit transfer, it should be left to the users of the estimates to determine the acceptable level. Thus, the results of the equivalence test could be presented in a table where one shows the highest transfer error level that can be applied in order to show equality. The rationale is that the acceptable level varies dependent on the policy use of the benefit transfer estimate. Generally speaking; a lower level of accuracy is needed for benefit estimates in cost-benefit analyses (CBAs) of projects and policies than for environmental costing and green accounting. However, even for CBAs the level of accuracy needed would be high if benefits and costs are very close. The highest level of accuracy is needed in Natural Resource Damage Assessments (NRDA), which are used to derive the compensation the polluter should pay to compensate damages to the general public from the pollution incident. (see also Navrud and Pruckner 1997). The acceptable transfer error should depend on the costs of making the wrong decision if the decision is based on benefit transfer instead of a new valuation study. We suggest testing at two levels, 20% transfer error and 40% transfer error. The 20% level identifies the cases where benefit transfer could produce estimates that could be used in a similar way as original estimates. The second level of 40% identifies those cases where the benefit transfer estimate gives approximation that could be used in applications where the need for cost effectiveness outweighs the demand for accuracy, as is the case for some CBAs.

4.3. T-tests and Hypotheses

Hypothesis of benefit transfer will be tested using both the classical procedure of non-rejection of a null hypothesis and the two one-sided equivalence test. The pooled estimator for standard deviation used is:

$$(9) \qquad s_p^2 = \frac{(n_a - 1)\, s_a^2 + (n_b - 1)\, s_b^2}{n_a + n_b - 2}$$

The t-statistic used for comparison was the pooled version:

$$(10) \qquad t = \frac{|WTP_a - WTP_b|}{s_p\sqrt{1/n_a + 1/n_b}}$$

where *WTP* are willingness to pay estimates from the two sites, s_p is pooled standard deviation and n is sample size.

The following hypotheses were tested:

Hypothesis 1 (Transferability of unit values)

Under the null hypothesis estimated mean values can be transferred between sites and countries, and the WTP estimates from the study and policy sites are not significantly different and/or directly equivalent at $\Delta=\theta$ where $\theta \in \{0.2\mathrm{WTP_P}, 0.4\mathrm{WTP_P}\}$ and $\mathrm{WTP_P}$ is the estimate from the policy site.

A more sophisticated approach is to use the value function from the study site with sample information from the policy site to predict the mean willingness to pay at the policy site:

Hypothesis 2 (Transferability of value functions).

Under the null hypothesis of transferability, the predicted WTP at the policy site using the parameters from the study site is not significantly different and/or directly equivalent to the WTP at the policy site for $\Delta=\theta$.

5. RESULTS

Table 3 presents the descriptive statistics for each country, and for anglers, non-anglers and the combined sample, respectively. Table 3 shows that the descriptive statistics of the samples are quite similar across countries. The only large differences found are fishing expenses and lower and upper bounds of WTP between countries, which reflect the much higher value of fishing licenses in Iceland compared to Norway and Sweden.

It is worth noting that the mean cost falls within the higher and lower bounds of stated WTP (over and above costs) for fishing in Iceland and Norway, and for river fishing in Sweden. This shows that WTP is large compared to current fishing expenses.

Results from the model estimation, presented in tables 4, show that income has the expected positive sign in all models. Parameters for both personal and other household income have positive signs. Considerable information is contained in the covariates, as can be seen from the likelihood ratio indices.

Men seem to have significantly higher WTP for both fishing and conservation of fish stocks. This is clearest for the Swedish and Norwegian respondents. Men also dominate the group that says they are anglers (Table 3).

Age has a negative parameter where it is significant, suggesting that WTP decreases with age. Years of education do not seem to have any clear effect except in the non-use value scenario. Here, the respondents with higher education have

Table 3. Descriptive statistics for anglers, non-anglers and the combined sample from the surveys in Iceland, Norway and Sweden. Mean values of all variables, upper and lower bounds of willingness-to-pay (WTP) for the three scenarios (Table 1), and number of observations are reported. (The mean values for a dummy variables are percentages)

	Iceland			Norway			Sweden		
Variables	Non anglers	Anglers	All	Non anglers	Anglers	All	Non anglers	Anglers	All
Number of observations	571	268	839	1010	1148	2158	2166	1280	3446
Sex (male = 1, female = 0)	0,32	0,69	0,44	0,37	0,66	0,52	0,37	0,69	0,49
Age	41,7	40,9	41,4	43,4	42,1	42,8	44,5	42,9	43,9
Personal income (in NOK)	157 982	193 680	169 385	207 943	245 253	227 791	171 535	194 158	179 938
Other household income (in NOK)	196 511	193 904	195 678	156 991	161 467	159 372	168 922	157 062	164 516
Number of persons in household.	3,27	3,45	3,32	2,48	2,82	2,66	2,65	2,81	2,71
11–13 years of education [a]	0,31	0,31	0,31	0,30	0,32	0,31	0,38	0,41	0,39
> 14 years of education	0,50	0,48	0,49	0,39	0,41	0,40	0,34	0,31	0,33
Subsistence fisherman [b]	N.A. [d]	0,04	0,04	N.A.	0,06	0,06	N.A.	0,05	0,05
Generalist	N.A.	0,10	0,10	N.A.	0,15	0,15	N.A.	0,15	0,15
Occasional fisherman	N.A.	0,48	0,48	N.A.	0,53	0,53	N.A.	N.A.	N.A.
Semi-urban residential area [c]	0,27	0,27	0,27	0,30	0,29	0,30	0,27	0,27	0,27
Rural residential area	0,13	0,12	0,13	0,21	0,29	0,25	0,21	0,29	0,24
Fisherman	N.A.	N.A.	0,32	N.A.	N.A.	0,53	N.A.	N.A.	0,37
Total number of fishing days in the last 12 months.	N.A.	7,67	13,78	N.A.	13,78	13,78	N.A.	13,17	13,17
Fishing expenses for the last 12 months	N.A.	3 770	3 770	N.A.	1 307	1 307	N.A.	1 433	1 433

(Continued)

Table 3. *(Continued)*

Variables	Iceland Non anglers	Iceland Anglers	Iceland All	Norway Non anglers	Norway Anglers	Norway All	Sweden Non anglers	Sweden Anglers	Sweden All
No stated fishing costs	0,94	0,03	0,65	0,89	0,05	0,44	0,97	0,08	0,64
Likes river-fishing best	N.A.	0,56	0,56	N.A.	0,22	0,22	N.A.	0,26	0,26
Likes lake-fishing best	N.A.	0,32	0,32	N.A.	0,26	0,26	N.A.	0,36	0,36
Lower bound of WTP for scenario 1	N.A.	3 325	3 325	N.A.	1 014	1 014	N.A.	816	816
Upper bound of WTP for scenario 1	N.A.	5 697	5 697	N.A.	1 762	1 762	N.A.	1 483	1 483
Lower bound of WTP for scenario 2	N.A.	2 402	2 402	N.A.	957	957	N.A.	796	796
Upper bound of WTP for scenario 2	N.A.	4 444	4 444	N.A.	1 794	1 794	N.A.	1 419	1 419
Upper bound of WTP for scenario 3	1 656	1 808	1 709	684	946	826	631	859	722
Lower bound of WTP for scenario 3	2 765	3 444	2 999	1 139	1 604	1 391	1 084	1 475	1 233

[a] The “hidden” education level is the respondents with less then 10 years of education.
[b] The “hidden” category of fishermen was sports fishermen.
[c] The “hidden ” variable is respondents living in urban residential areas.
[d] N.A. indicates that the statistic is not available.

Table 4. Sign and significance of estimated parameters. Three signs indicate significance at the 0,1% level, two signs indicate 5% and one sign indicates significance at 10% level. No sign indicates a non-significant parameter while N.A. indicates that the variable was not included in the estimated model

	River fishing scenario			Lake fishing scenario			Non-use value scenario		
	Isl	Nor	Swe	Isl	Nor	Swe	Isl	Nor	Swe
Intercept	+++	+++	+++	+++	+++	+++	+++	+++	+++
Sex, man = 1		+++	++		++	+++		+++	+
Age	N.A.[c]	N.A.	N.A.	–		–		–	–
Fishing expenses last 12 months (in NOK)	+++	+++	+++	+++	+++	+++		+++	+++
No stated fishing costs			++		–			++	+++
Personal income (in NOK)	+	++		+			+++	+	++
Other income (in NOK)	+	++			++	+	+++	+++	+
11–13 years of education	N.A.	N.A.	N.A.	++	++				++
> 14 years of education	N.A.	N.A.	N.A.	+	+		++	+++	+++
Likes river-fishing best	++	+++		N.A.	N.A.	N.A.		++	++
Likes lake-fishing best	N.A.	N.A.	N.A.		+		N.A.	N.A.	N.A.
Fisherman	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.		++	+++
Subsistence fisherman	–	++		N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Generalist				N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Occasional angler		–	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Semi-urban residential area				N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Rural residential area	–			N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Log-likelihood of estimated model	−414	−1911	−2251	−416	−1879	−2171	−1145	−3554	−5554
$-2^*(\text{log-likelihood}_1 - \text{log-likelihood}_0)$ [a]	87	215	114	34	98	94	78	165	225
Zero WTP [b]	0,034	0,117	0,077	0,026	0,105	0,080	0,162	0,133	0,125
The number of observation with positive WTP	247	955	1117	237	952	1077	641	1735	2760

[a] −2 times the difference in log likelihood between a model with no covariates and this model. The resulting number is χ^2 distributed with degrees of freedom equal the number of covariates.

[b] An estimate for the perectage of true zero bidders obtained from the effected population. For the treatment of zero bidders see equations 4 and 5 and relevant discussion in text.

[c] N.A. indicates that the variable was not included in the relevant model.

significantly higher WTP in all countries. This could indicate a larger level of environmental awareness among the respondents with higher education.

Residential area produces some significant parameters. The signs indicate that respondents in rural Iceland and Norway have lower WTP than semi-urban and urban respondents.

Fishing expenses clearly influences WTP for both use and non-use value. This shows that anglers have a clear idea of the actual cost of fishing and their bids are strongly influenced by this.

A stated preference for one type of fishing increases WTP for that type of fishing, as expected. This effect is largest for those who prefer river fishing compared to those that prefer lake fishing. This might be caused by the fact that river fishing is more exclusive and often more expensive than lake fishing. The hidden variable is a stated preference for sea fishing.

Mean WTP per person per year with bootstrap standard errors are reported in Table 5. Table 5 shows that there is a strikingly large overall difference in WTP between Iceland and the other two countries. This reflects the large differences in costs associated with recreational fishing as reported in table 3. This seems also to affect the WTP for preserving the Nordic fish stocks (i.e. non-use value scenario). However, the relative difference in WTP between Iceland and the two other countries is smaller for the non-use values than the two use value scenarios. The difference in WTP is consistent between countries. It is always largest for Iceland, then Norway and smallest for Sweden.

The two separate hypothesis tested were equality and non-equivalence. Equality was tested by a classical t-test while non-equivalence was tested by the two one-sided procedure (TOST) at 20% and 40% transfer error. The tests were performed both for unit value transfer (hypothesis 1), Table 6, and value function transfer (hypothesis 2), Table 7.

Table 6 shows the results for all possible transfer combinations. The first column defines the country used for estimation while the rows define the country the value is transferred to. For example a transfer of values from Norway to Iceland for the river fishing scenario (row two, column one to three) would result in a

Table 5. Results of mean WTP per person per year with bootstrap standard errors and test for the normality of the bootstrap distribution

Scenario	Country	Mean WTP	Bootstrap standard error	Shapiro-Wilk statistic (W)[a]	Prob < W[b]
	Iceland	4 251	352	0,980	0,00
River	Norway	1 120	64	0,931	0,00
	Sweden	1 014	61	0,987	0,62
Lake	Iceland	3 119	245	0,984	0,15
	Norway	1 090	49	0,989	0,87
	Sweden	927	48	0,984	0,10
	Iceland	1 772	149	0,983	0,05
Non-use value	Norway	870	42	0,931	0,00
	Sweden	761	31	0,985	0,21

Notes:
[a] A test for normality applicable in small samples. The test is described in section 4.
[b] The probability of making a type I error by rejecting normality.

Table 6. Transfer error (Error) as given by equation 7, t-values from a classical test of equality and smallest level of significant equivalence (TOST) for unit value transfer

River fishing scenario

		To								
		Iceland			Norway			Sweden		
		Error [a]	t-value [b]	TOST [c]	Error	t-value	TOST	Error	t-value	TOST
From	Iceland	–	–	–	280%	16.21***	N.Eq.	319%	17.24***	N.Eq.
	Norway	−74%	16.21***	N.Eq.[d]	–	–	–	10%	1.20	40%***
	Sweden	−76%	17.24***	N.Eq.	−9%	1.20	40%***	–	–	–

Lake fishing scenario

		To								
		Iceland			Norway			Sweden		
		Error	t-value	TOST	Error	t-value	TOST	Error	t-value	TOST
From	Iceland	–	–	–	186%	14.25***	N.Eq.	236%	15.51***	N.Eq.
	Norway	−65%	14.25***	N.Eq.	–	–	–	18%	2.37*	40%**
	Sweden	−70%	15.51***	N.Eq.	−15%	2.37*	40%***	–	–	–

Non-use value scenario

		To								
		Iceland			Norway			Sweden		
		Error	t-value	TOST	Error	t-value	TOST	Error	t-value	TOST
From	Iceland	–	–	–	104%	8.50***	N.Eq.	133%	11.35***	N.Eq.
	Norway	−51%	8.50***	N.Eq.	–	–	–	14%	2.12*	40%***
	Sweden	−57%	11.35***	N.Eq.	−13%	2.12*	40%***	–	–	–

* = 5%, ** = 1% and *** =< 0.1% level of significance.

[a] Transfer error associated with the benefit transfer.

[b] The t-value for the hypothesis that the benefit transfer estimate and the original estimate are equal.

[c] The lowest level of significant equivalence. The levels tested are 10% 20% and 40% of the original study estimate of mean WTP.

[d] Not equivalent at any tested level.

transfer error of −51%. This transfer error is statistically different from zero and the hypothesis of non-equivalence cannot be rejected. The conclusion is therefore that such a transfer is not acceptable. The differences seen in Table 5 are also seen in the result in Table 6. Benefit transfers to and from Iceland result in very large transfer error that is highly significant. Both the t-test and the equivalence test clearly suggest that such a transfer is not acceptable. On the other hand, none of the t-tests of equality are rejected when transferring between Norway and Sweden. This, in addition to the small transfer errors, would have been interpreted as evidence of valid benefit transfer. The equivalence tests reveal a weakness of this argumentation. Non-equivalence is only significantly rejected at a 40% level of transfer error suggesting that the variance of the WTP makes it impossible to accurately predict a small transfer error. The estimates should only be used in cases where approximate figures are required, for example in a CBA where the estimated benefits and costs are far apart.

Table 7. Transfer error (Error), t-values from a classical test of equality, and smallest level of significant equivalence (TOST) for value function transfer

		To Iceland			To Norway			To Sweden		
River fishing scenario										
		Error [a]	t-value [b]	TOST [c]	Error	t-value	TOST	Error	t-value	TOST
From	Iceland	–	–	–	147%	9.09***	N.Eq.	210%	7.22***	N.Eq.
From	Norway	−35%	3.28**	N.Eq.[d]	–	–	–	33%	3.53***	N.Eq.
From	Sweden	-64%	9.4***	N.Eq.	-22%	3.27**	40%**	–	–	–
Lake fishing scenario										
		Error	t-value	TOST	Error	t-value	TOST	Error	t-value	TOST
From	Iceland	–	–	–	123%	9.61***	N.Eq.	165%	12.98***	N.Eq.
From	Norway	−33%	2.34*	N.Eq.	–	–	–	34%	4.10***	N.Eq.
From	Sweden	−61%	15.22***	N.Eq.	−16%	2.37*	40%***	–	–	–
Non-use value scenario- full sample										
		Error	t-value	TOST	Error	t-value	TOST	Error	t-value	TOST
From	Iceland	–	–	–	149%	9.40***	N.Eq.	137%	8.77***	N.Eq.
From	Norway	−43%	3.91***	N.Eq.	–	–	–	11%	1.60	40%***
From	Sweden	−52%	8.06***	N.Eq.	−8%	1.29	20%*	–	–	–
Non-use value scenario- non-anglers only										
		Error	t-value	TOST	Error	t-value	TOST	Error	t-value	TOST
From	Iceland	–	–	–	162%	6.48***	N.Eq.	148%	8.77***	N.Eq.
From	Norway	−55%	5.45***	N.Eq.	–	–	–	7%	0.82	40%***
From	Sweden	−60%	9.28***	–	−8%	0.91	40%***	–	–	–

* = 5%, ** = 1% and *** =< 0.1% level of significance.

[a] Transfer error associated with the benefit transfer.

[b] The t-value for the hypothesis that the benefit transfer estimate and the original estimate are equal.

[c] The lowest level of significant equivalence. The levels tested are 10% 20% and 40% of the original study estimate of mean WTP.

[d] Not equivalent at any tested level.

Table 7 shows the results of the tests of value function transfer (hypothesis 2). The results for value function transfer in Table 7 are slightly different than the ones for unit value transfer in Table 6. Transfer error between Iceland and Norway is considerably reduced for the use value scenarios. The estimated models explain a considerable amount of the differences in mean WTP. Transfer error is however not reduced to a level that results in valid benefit transfer. For Norway and Sweden, on the other hand, using the value function transfer approach increases transfer errors for the use value scenarios. This indicates that Icelanders and Norwegians have similar preferences when it comes to fishing, while Swedes and Norwegians do not. All the t-tests for the use value scenarios reject equality of the benefit transfer and original estimates of mean WTP at a 5% significance level. The results of the equivalence tests are not symmetric. Non-equivalence is rejected when transferring from Sweden to Norway at a 40% level of the original mean WTP. If the transfer is from Norway to Sweden the non-equivalence is not rejected at any tested level. Swedish values could potentially be used in Norway but nor vice versa.

The results for the non-use value scenario are quite different. Transfer error is reduced in all cases, except when transferring from Iceland to Norway. The largest reduction is achieved when only a sub sample of non-anglers is used. Then the transfer error between Norway and Sweden is reduced to 7–8%. This produces the smallest t-values, taken as an indication of validity of the null hypothesis. Similarly non-equivalence is generally rejected at 40% transfer error, indicating significant equivalence. Further, non-equivalence is rejected at 20% transfer error for a transfer from Sweden to Norway using the full sample estimate. This clearly indicates a fully acceptable benefit transfer. The results for the non-use value suggest that preferences for that environmental good are similar in Norway and Sweden.

6. CONCLUSIONS

Several tests of both unit value transfer and value function transfer have been conducted for cases where the study site and policy site are in a different country. The general conclusion is that the accuracy of benefit transfer relies heavily on the similarity of populations and described scenarios in respect to environmental conditions in each country. This corresponds to the results of previous studies. It is evident that the more similar the populations and environmental goods are the smaller the transfer errors. This similarity must also include the perceived price levels of the resource at hand.

The WTP estimates are consistent between countries for all tested scenarios. The Icelandic values are largest, followed by the Norwegian and then Swedish as the smallest. Added information generally reduces transfer error. This is most obvious when the errors are large. This underlines the necessity of well defined and correctly specified value functions.

The suggested method of using equivalence tests instead of non-rejection of a classical test of equality seems to result in more consistent outcomes. The rejection of non-equivalence only happens in cases where both transfer error is small and WTP estimates are stable while non-rejections of classical hypothesis may happen when transfer error is large and WTP estimates are unstable. This is clearly seen in tables 6 and 7. The equivalence tests clearly indicate that Swedish WTP estimates can be used in Norway in cases where the acceptable transfer error is large, especially for non-use values. Norwegian results can to a lesser extent be used in Sweden. No clear-cut outcome results from the classical test procedure because the results vary form scenario to scenario.

The transfer errors are consistently smaller for the non-use value scenario. The results for the non-use value scenario by non-anglers in Norway and Sweden produce the only statistically equivalent benefit transfer at the strictest level of 20% transfer error. Anglers being excluded, the estimated WTP should be a strict non-use value estimate. Recreational fishing is a fairly well defined good in comparison to the non-use value of preserving natural fish stocks. The purely hypothetical nature of the latter might make the preferences more general, and less influenced by specific factors regarding recreational fishing in each country. Brouwer (2000) suggests that non-use values reflect some kind of overall moral commitment to environmental causes. He hypothesizes that such values can be expected to stay more or less constant across social groups and environmental domain. Our results seem to support his hypothesis.

Daði Kristófersson is researcher, Agricultural University of Iceland, Borgarnes, Iceland. Ståle Navrud is associate professor, Department of Economics and Resource Management, Norwegian University of Life Sciences, Ås, Norway.

7. NOTE

[1] It is assumed that Δ is symmetrical. That implies that only the size of the error is important not its sign. It is equally simple to work with two Δ, one for the possessive error and another for the negative error, denoted here by $-\Delta$.

8. REFERENCES

Allison PD, 1995 Survival analysis using the SAS® system: a practical guide SAS Institute Inc, Cary, NC 292 pp ISBN: 1-55544-279-X

Arrow KJ, Solow R, Portney PR, Leamer EE, Radner R, ShumanH, 1993 Report of the NOAA panel on contingent valuation Federal register, 58:4601–4614

Berger RL, Hsu JC (1996) Bioequivalence trials, intersection-union tests and equivalence, Statistical Science, 11(4):283–302

Bergland O, Magnussen K, Navrud S (2002) Benefit transfer: testing for accuracy and reliability In: Florax, RJGM, P Nijkamp and K Willis (eds) 2002: Comparative Environmental Economic Assessment, Edward Elgar, Cheltenham, UK and Northampton, MA, USA 117–132

Bergland O, Romstad E, Kim S, McLeod D (1993) The use of bootstrapping in contingent valuation studies Unpublished working paper Department of Economics and Social Sciences, Agricultural University of Norway, Ås

Brookshire DS, Neill HR (1992) Benefit transfers—conceptual and empirical issues Water resources research, 28(3):651–655

Brouwer R, Spaninks A (1999) The validity of environmental benefit transfer: further empirical testing Environmental and resource economics, 14:95–117

Brouwer R (2000) Environmental value transfer: state of the art and future prospects Ecological Economics, 32(1):137–152

Burda M, Wyplosz C (1997) Macroeconomics: a European text – 2nd edn Oxford University Press, Oxford, p. 613, ISBN: 0-19-877468-0

Cooper JC (1994) A comparison of approaches to calculating confidence-intervals for benefit measures from dichotomous choice contingent valuation surveys Land economics,70(1):111–122

Desvousges WH, Naughton MC, Parsons GR (1992) Benefit Transfer: Conceptual Problems in Estimating Water Quality Benefits Using Existing Studies. Water Resources Research, 28(3): 675–683

Downing M, Ozuna T (1996) Testing the reliability of the benefit transfer approach Journal of Environmental Economics and management, 30:316–322

Greene WH (2000) Econometric Analysis Prentice Hall International New Jersey ISBN: 0-13-015679-5

Hauck WW, Anderson S (1984) A new statistical procedure for testing equivalence in two-group comparative bioavailability trials, Journal of Pharmacokinetics and Biopharmaceutics 12(1): 83–91

Hoenig JM, Heisey DM (2001) The abuse of power: The pervasive fallacy of power calculations for data analysis American Statistician 55(1): 19–24

Kirchhoff S, Colby BG, LaFrance JT (1997) Evaluating the performance of benefit transfer: An empirical inquiry Journal of environmental economics and management, 33(1):75–93

Kristofersson D, Navrud S (2005) Validity Tests of Benefit Transfer: Are we performing the wrong tests? Environmental and Resource Economics 30(3):279–286

Kriström B (1997) Spike models in contingent valuation American Journal of Agricultural Economics, 79:1013–1023

Lehman EL (1983) Theory of point estimation, Wiley: New York

Lindsey JC, Ryan LM (1998) Tutorial in biostatistics–Methods for interval-censored data Statistics in medicine, 17(2):219–238

Loomis JB (1992) The evolution of a more rigorous approach to benefit transfer—Benefit function transfer Water resources research, 28(3):701–705

Loomis J, 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

Navrud S, Pruckner GJ (1997) Environmental valuation - To use or not to use? A comparative study of the United States and Europe Environmental and Resource Economics 10:1–26

Notaro S, Signorello G (1999) Elicitation effects in contingent valuation: a comparison among multiple bounded, double bounded, single bounded and open-ended formats EAERA 9th annual conference, Oslo 25–27, June 1999

Ready RC, Navrud S, Dubourg R (2001) How Do Respondents with Uncertain Willingness To Pay Answer Contingent Valuation Questions? Land Economics, 77 (3):315–326

Ready R, Navrud S, Day B, Dubourg R, Machado F, Mourato S, Spanninks F,Rodriquez MXV (2004) Benefit Transfer in Europe How Reliable Are Transfers Between Countries? Environmental and Resource Economics 29:67–82

Reiser B, Shechter M (1999) Incorporating zero values in the economic valuation of environmental program benefits, Environmetrics, 10:87–101

Schuirmann DJ (1987) A comparison of the two one-sided procedure and the power approach for assessing equivalence of average bioavailability, Journal of Pharmacokinetics and Biopharmaceutics 15(6):657–680

Stegner, BL, A G Bostrom and T K Greenfield 1996 Equivalence testing for use in psychosocial and services research: An introduction with examples, Evaluation and Program Planning 19(3):193–98

Toivonen, Anna-Liisa; Appelblad, Håkan; Bengtsson, Bo; Geertz-Hansen, Peter; Gudbergsson, Gudni; Kristofersson, Dadi; Kyrkjebø, Hilde; Navrud, Ståle; Roth, Eva; Tuunainen, Pekka & Weissglas, Gösta (2000) The Economic Value of Recreational Fisheries in the Nordic Countries TEMA Nord Rapport 2000:604, 68 pp ISSN: 0908-6692; Nordisk Ministerråd; København

Welling, PG, FLS Tse and SV Dighe (1992) Pharmaceutical Bioequivalence, Marcel Dekker: New York