What Can We Learn from Benefit Transfer Errors? Evidence from 20 Years of Research on Convergent Validity[*]

Sapna Kaul, Dept. of Agricultural and Applied Economics, Virginia Tech, sapnak@vt.edu[♦]

Kevin J. Boyle, Dept. of Agricultural and Applied Economics, Virginia Tech, kjboyle@vt.edu

Nicolai V. Kuminoff, Dept. of Economics, Arizona State University, kuminoff@asu.edu

Christopher F. Parmeter, Dept. of Economics, University of Miami, cparmeter@bus.miami.edu

Jaren C. Pope, Dept. of Economics Brigham Young University, jaren_pope@byu.edu

[*] We especially appreciate the helpful comments of V. Kerry Smith. We also thank seminar participants at 5th Annual Choice Modeling Workshop hosted by the University of South Australia, Arizona State University and conference participants at the 2011 AERE summer meeting. Courtney Mitchell provided excellent research assistance. This work was sponsored by EPA Grant No. RD-83345901–0.

[♦] Corresponding author: 208 Hutcheson Hall (0401), Blacksburg, VA 24061. Phone: (540) 231-6301.

1. Introduction

Benefit transfer is one of the most common methods for conducting benefit-cost analysis at the U.S. Environmental Protection Agency [27]. Since a 1992 issue of *Water Resources Research* (Vol. 28, 3) raised benefit transfers to an area of academic interest to environmental economists, at least 40 studies have investigated the empirical accuracy of this method using tests of convergent validity.[1] Two stylized facts have emerged. Function transfers tend to be more accurate than value transfers, and similarity between the study and policy cases tends to improve accuracy [4][13][14].[2] The apparent lack of consensus on other methodological features of the transfer process has made it difficult to define specific protocols for the conduct of benefit transfers and to develop a clear agenda for research.[3] Johnston and Rosenberger [13] observe, the "*complexity and relative disorganization of the (academic) literature may represent an obstacle to the use of updated (benefit-transfer) methods by practitioners.*"

The purpose of this paper is to investigate if specific modeling decisions can be identified that enhance or reduce the accuracy of benefit transfers, and this is done using a new statistical approach to meta-analysis. We begin by systematically reviewing all of the empirical studies on the convergent validity of benefit transfers conducted over the past 20 years. These studies tested a tremendous variety of methodological procedures, collectively reporting *more than one thou-*

[1] Convergent validity asks if two estimates of a specified value concept, using different data in the case of benefit transfers, provide statistically similar estimates. We are not aware of any criterion-validity studies of benefit transfer estimates that use a known or asserted true value to compare with a benefit-transfer estimate.
[2] Value transfers use point estimates from one or more previous studies to develop a benefit measure for a new policy. Function transfers take a preexisting estimate for the functional relationship between benefits and the characteristics of people and their choices and then calibrate the function to the population affected by the new policy. Readers seeking background on concepts, terminology, and methods in the benefit transfer literature are directed to the surveys prepared by Bergstrom and Taylor [2] , Boyle et al. [4], and Johnston and Rosenberger [13].
[3] The EPA's 2010 *Guidelines for Preparing Economic Analyses* [27] describe four steps for conducting a benefit transfer: (i) describe the policy case; (ii) select study cases; (iii) transfer values; and (iv) report the results. The guidelines are nearly devoid of recommendations for the actual conduct of benefit transfers.

sand benefit transfer errors. Most of these observations come from studies conducted in the United States and Western Europe. The applications cover a wide variety of amenities. Examples include access to forest, park, and lake recreation; hunting; changes in the quantity and quality of water in lakes, rivers, and coastal areas; air quality; exposure to ultraviolet radiation; freshwater fishing in streams and rivers; proximity to various types of open space; farmland amenities; and measures of the overall ecological health of watersheds, wetlands, and rivers.

It is standard practice in meta-analyses to use linear meta-regressions with robust standard errors to distill the collective findings on important questions in the field of environmental economics [19]. However, we have some concerns about the credibility of a *linear* meta-regression in the context of our analysis. The modeling decisions that comprise a benefit transfer are represented by binary variables that can interact in complex ways to influence the accuracy of transferred values [3]. Any specific variable can have a unique effect on the outcome being investigated (meta-equation regressand) when combined with sets of other regressors. For example, similarity between study and policy cases may be crucial to the accuracy of a value transfer, but not as important for a function transfer that can be calibrated to policy case conditions. A statistical approach that allows for these interaction effects has the potential to provide richer insights about the literature being investigated. The typical linear meta-regression imposes separability of each of the regressors. Capturing all potential interaction effects would require adding an intractable number of regressors to the meta-equation. Therefore, we propose a new nonparametric approach to meta-analysis that does not impose the linearity and separability assumptions. We contrast the insights from our new approach with a conventional parametric approach.

In our specific application, nonparametric meta-analysis avoids the need to impose a-priori restrictions on the functional relationship between benefit-transfer errors and the various modeling decisions that comprise the transfer process. Instead of generating a single point estimate for the average impact of a particular modeling decision, the nonparametric model generates a range of estimates. This allows us to recognize that there are multiple possible effects for each transfer procedure, with specific impacts depending on the empirical context defined by other methodological choices made in implementing a benefit transfer.

The cost of moving to a nonparametric framework is that it traditionally requires more from the data. While our meta-data represent the most comprehensive summary of the external validity of estimates for environmental values ever assembled, they do not allow us to identify ranges of effects for every benefit transfer procedure. Some procedures are not observed frequently enough in the meta-data to identify their full range of effects without adding parametric restrictions to the model. Therefore, we follow Charles Manski's [16] "bottom-up" approach to data analysis. First we use the nonparametric model to estimate the ranges of impacts for the benefit transfer procedures that can be identified from variation in the meta-data. Then we add conventional linearity and separability restrictions commonly used in meta-regressions and estimate this meta-equation using weighted least squares. Following Manski's logic, we recognize that the credibility of inference based on our results is decreasing in the strength of the parametric restrictions that we impose on the meta-regression.

Together, the parametric and nonparametric models suggest several important findings. First, transfer errors are not as large as one might expect. Out of 1071 observations, the mean absolute transfer error is 172% while the corresponding median is only 39%. Second, the nonparametric model generally confirms the stylized fact that function transfers outperform value transfers. This result holds regardless of the other modeling decisions that comprise the benefit transfer. A third result from the nonparametric analysis is that benefit transfers describing changes in environmental *quality* almost always have larger errors than transfers describing *quantity* changes. Moving to a linear specification for the meta-regression allows us to draw some additional conclusions. For example, we find that on average: (i) the accuracy of benefit transfers is improved by geographic proximity of the study and policy locations; (ii) the use of random utility models, travel cost models, contingent valuation methods, and choice modeling tends to be more accurate than meta-regression transfers; and (iii) drawing on information from multiple preexisting studies (as opposed to a single study) tends to reduced transfer errors.

The remainder of the paper is organized as follows. Section 2 develops a framework for characterizing the ways in which benefit transfer procedures can introduce errors into predictions of benefits for a new policy. Section 3 describes how we reviewed the convergent validity literature and developed the meta-data. Section 4 explains the mechanics of our nonparametric approach and presents regression results from parametric and nonparametric models. Section 5 summarizes and interprets the key findings. Finally, section 6 provides concluding remarks.

2. Conceptual Framework

A benefit transfer takes a value (or values or valuation equation) from a study case (or cases) to develop a customized benefit estimate for a new policy case.[4] The process begins by defining the relevant measure of benefits.[5] Consider a public policy that is expected to change the quality of an environmental amenity from q^0 to q^1 at policy case j. A Hicksian measure of willingness to pay (wtp) for this change is defined as:

[4] We adopt the phrases *study case* and *policy case* from the U.S. EPA's 2010 *Guidelines for Preparing Economic Analyses* [27]. This is a change from the conventional terminology, *study site* and *policy site*. We chose to adopt the new terminology to avoid confusion between the academic literature and practical applications. The word "case" also seems more accurate. Benefits are not always transferred to new geographic sites. Some transfers occur at the same physical location, using past values to assess current situations or predict future outcomes. Thus, benefits may be transferred to a new policy case at the same study site or to a new policy case at a different site. This is what the EPA Guidelines refer to as "describing the policy case" [27].

[5] This is what the EPA *Guidelines* refer to as "describing the policy case" [27].

$$V_{ij}\left(p_j, x_j, q_j^1, y_i - wtp_{ij};\ \alpha_i, d_i\right) = V_{ij}\left(p_j, x_j, q_j^0, y_i;\ \alpha_i, d_i\right), \tag{1}$$

where V_{ij} is the indirect utility for individual *i* at case *j* expressed as a function of market prices (p_j), other non-priced attributes (x_j), individual income (y_i), other demographic characteristics (d_i), and latent preferences (α_i).

Ideally, the analyst would estimate willingness to pay using the joint distributions of data describing the observable characteristics of individuals and their choices at policy case *j*, $G_j(y, d)$ and $F_j(p, x, q)$. If these data are unavailable, however, or if constraints on time and resources prohibit original estimation, then a benefit transfer is a second-best approximation to wtp_{ij} using preexisting information from a different study case, *k*. This approximation is a second-best alternative because of the benefit-transfer error (*BTE*) introduced by transferring information from the study case to the policy:

$$\begin{aligned} BTE\ &= \widehat{wtp}_{ij}^{T}\left[p_j, x_j, q_j^0, q_j^1, y_i, d_i, F_k(p, x, q), G_k(y, d) | \hat{\beta}_k, v\right] \\ &-\ \widehat{wtp}_{ij}\left[p_j, x_j, q_j^0, q_j^1, y_i, d_i, F_j(p, x, q), G_j(y, d) | \hat{\beta}_j, v\right], \end{aligned} \tag{2}$$

where $\hat{\beta}_k$ is a vector of parameters estimated from the study case data and *v* denotes the valuation methodology (e.g. travel cost, contingent valuation).[6] The benefit-transfer error is simply the difference between the approximation to wtp_{ij} based on available information from study case *k* and the counterfactual estimate for wtp_{ij} that would be obtained from information on policy case *j* if there were no constraints on time, resources, or data.

The *BTE* is a measure of convergent validity that compares two estimates of the same theoretical value defined in equation (1). Thus, *BTE*, as defined in (2), is a second-best metric because it does not reveal the difference between the transferred benefit measure and the true

[6] For example, $\hat{\beta}_k$ may define moments of statistical distributions used to describe sources of unobserved heterogeneity in the preferences of the study case population. In general, there may also be more than one study case.

willingness to pay. The true wtp_{ij} is always unknown. At best, the *BTE* can be viewed as an unbiased estimator of $(\widehat{wtp}_{ij}^T - wtp_{ij})$. The *BTE* defines the composite error in measuring $\widehat{wtp}_{ij}$ that is introduced by methodological features of the transfer process and by differences between the study and policy cases.

The general expression for the benefit transfer error in (2) illustrates four distinct ways in which an analyst's modeling choices can influence the magnitude of the *BTE*.[7] First, the size of the *BTE* may depend on how the change in the environmental amenity, Δq, is defined. Second, errors may stem from systematic differences between the observable characteristics of the study and policy case populations, $G_j \neq G_k$. Third, the *BTE* may vary with the valuation methodology, v. Finally, the error may depend on the transfer procedures embedded in $\widehat{wtp}_{ij}^T$. The individual modeling decisions reflected in the definitions for $\Delta q, v, G,$ and T may interact in ways that increase or decrease the *BTE*. Geographic similarity between the study and policy cases may produce a larger improvement in value transfer accuracy than in function transfer accuracy, for example. That is, a function transfer can be calibrated to policy-case conditions by assigning levels to covariates in the transfer equation; whereas, no similar calibration is possible for a value transfer so selection of the specific value to transfer that is "similar" becomes critically important.

The importance of benefit transfers for environmental policy has motivated a significant amount of research on measuring *BTEs* using research designs based on the concept of convergent validity.[8] These studies use estimates of willingness to pay for the study and policy cases.

---

[7] A fifth potential source of error that is arguably out of the practitioner's control is any systematic difference between the latent preferences of the study and policy case populations, $H_j(\alpha) \neq H_k(\alpha)$.

[8] The importance of benefit transfers for U.S. environmental policy is due, in part, to Presidential Executive Order 12866 (1993) [28], which requires federal agencies to assess "*costs and benefits*" of regulations that are based "*on the best reasonably obtainable scientific, technical, economic, and other information*." International interest is also growing. In 2005, EPA sponsored a forum entitled "Benefit Transfer and Valuation Databases: Are We Heading in the Right Direction", drawing presenters from Australia, Canada, France, Spain, Singapore, the U.K. and the U.S.

Then they compare the policy-case estimate to the transferred study-case estimate. The resulting transfer error is typically reported in percentage terms,

$$\text{\% Benefit Transfer Error} = \%BTE = \left[\left(\widehat{wtp}_{ij}^{T}/\widehat{wtp}_{ij}\right) - 1\right] \times 100. \qquad (3)$$

Studies that do not report $\%BTE$ almost always report sufficient information for readers to make this calculation on their own.

Each study in the convergent validity literature reports one or more transfer errors conditional on a specific set of modeling decisions. We have systematically reviewed these studies and assembled a database documenting transfer errors and transfer procedures. To extract the signals from the noise we use a meta-regression,

$$|\%BTE| = m(X) + \varepsilon, \qquad (4)$$

where X includes a vector of variables that we use to describe analysts' modeling decisions (variables representing $\Delta q, v, G,$ and T).

3. Data Description

3.1. Reviewing the Literature on Convergent Validity of Benefit Transfers

Through an exhaustive search, we identified 40 convergent-validity studies that were published or posted online between 1990 and 2009. Thirty eight of these studies were published in peer-reviewed journals or book chapters. We ultimately excluded the two studies that were not peer reviewed because they had insufficient documentation for some of the key variables ($\Delta q, v, G,$ and T).[9] Three peer reviewed studies were excluded for the same reason. A study by

[9] The inclusion/exclusion of gray literature studies is a point of debate in the literature. Bergstrom and Taylor [2] suggest that peer-reviewed studies are more likely to be error free, while Stanley [22] raises concerns over publication bias of peer-reviewed studies. While this debate is interesting, it is not relevant to our analysis. Even if the two convergent validity studies that were not peer reviewed had provided enough information to be included in our analysis, they would not have provided enough observations to recover the effects of publication bias.

Morrison et al. [17] was excluded in order to avoid duplication of the findings reported in Morrison et al. [18]. Studies by Engel [6] and Chattopadhyay [5] were excluded because benefit transfer errors were reported as ranges rather than point estimates. Finally, a study by Eshet et al. [7] was excluded because it would have been the only one to use the hedonic methodology, and a single study would not have allowed us to identify the effect of this valuation method on the *%BTE*.[10] These exclusions left us with a total of 31 studies with *%BTE* observations to analyze. Summary statistics describing the transfer procedures and results for each included study are reported in Appendix Table I, and excluded studies are summarized in Appendix Table II. The appendix also provides complete citations to all 40 studies.

A uniform coding protocol was implemented to ensure that the modeling choices made by the authors of each study, and the corresponding %*BTEs*, were recorded correctly and consistently. The data were double coded by two research assistants, who then met with us to resolve discrepancies. Then we cross checked the coding a second time.

Some studies report what we refer to as "flip" error calculations. The investigators would compute a transfer error with *j* as the policy case and *k* as the study case *a la* equation (3). Then they would flip the two cases, computing a second transfer error with *k* as the policy case and *j* as the study case. Flipping the study and policy cases changes the sign and absolute magnitude of the percentage error in (3). Unfortunately, it was not possible to infer the flip errors for most studies that did not make this calculation directly. Since we were unable to include flip errors for every study, we decided to use a single set of errors from studies that reported flips. Specifically, we used the first set of errors reported by the investigators.[11]

[10] Note, Chattopadhyay [5] is also a hedonic study.

[11] Chattopadhyay [5] proposes an alternative formula for the %*BTE* that would avoid the flip error problem: $|\%\ transfer\ error| = \left|\left[\left(\frac{study\ case\ value - policy\ case\ value}{study\ case\ \ vlaue + policy\ case\ value}\right)/2\right] \times 100\right|$.

Each of the 31 validity studies reported multiple transfer errors, from as few as 2 to as many as 178. The errors vary with the transfer procedures, selection of study cases and study case valuation methods, and amenity of interest. For example, Loomis et al. [15] tested the convergent validity of travel-cost estimates for the average consumer surplus associated with a single day of reservoir-based recreation in Sacramento, CA, Little Rock, AR, and Nashville, TN. After estimating travel-cost models for each of the three regions, they assessed the accuracy of benefit transfers by making all possible pair-wise comparisons of estimated and transferred consumer surplus for 28 reservoir sites. This yielded 112 distinct transfer errors, half of which were flips. Excluding the flip errors left us with 56 observations on $|\%BTE|$.

The final data set contains 1071 transfer errors. Of these observations, 55% describe applications in Europe, 37% are drawn from United States, and the remaining 8% are from Australia and the rest of the world. [12] The European data include observations from 12 Western European countries and the U.S. data include observations from all of the lower 48 states. The set of applications is also diverse. Eight studies considered the benefits of access to recreation sites (including forest recreation, reservoir based recreation, park recreation, and hunting); five evaluated prospective changes in the quality or quantity of water (including coastal areas, lakes, rivers, and groundwater); and three studies evaluated the benefits of reductions in sources of human health risk (air quality, water quality and ultra violet radiation). Other studies focused on opportunities for fresh water fishing (salmon, trout, big game, small game, flatfish, salmon, steelhead, walleye, pike, bass, and panfish), amenities associated with land preservation (farmland, forested

---

This formula produces the same measure of error, regardless of which value is defined as the study case. It would be convenient if future convergent-validity studies were to adopt this metric.

[12] We did not find evidence of systematic differences across these four regions in terms of the impact of modeling decisions on benefit transfer errors. Adding fixed effects for regions to the linear meta-regressions did not produce any statistically significant differences in coefficients on modeling procedures. Unfortunately, there was insufficient variation in the data to identify separate nonparametric models for each region.

land and coastal land), and the overall ecological health of watersheds, wetlands, and rivers.

3.2. The Distribution of Transfer Errors

Figure 1 illustrates the distribution of benefit transfer errors. In percentage terms, the errors range from 0% to 7,496%, with a mean of 172%. However, the mean reflects a few observations with extremely large errors. The median is only 39%, less than a quarter of the mean. The large difference between the mean and the median suggests a need to investigate outlying observations that could influence econometric inferences from the data.

We inspected the data for the presence of outliers using the inter-quartile range (IQR) criterion [21]. According to this criterion, an observation is classified as an outlier if $|\%BTE| < Q_1 - 1.5 \times IQR$ and/or $|\%BTE| > Q_3 + 1.5 \times IQR$, where $IQR = Q_3 - Q_1$, and Q_1 and Q_3 are the first and third quartiles of the $|\%BTE|$ distribution. This procedure detected 146 outliers.[13] Seventy-two percent of these were reported by just two studies, which also accounted for the largest percentage errors.[14]

Figure 2 graphs the distribution of transfer errors with outliers deleted. The mean and median $|\%BTE|$ are now reduced to 42% and 33%, respectively, and the maximum is reduced to 172%. One explanation for these large reductions is that some of the validity studies were conducted for situations where the policy case and study case values were conveniently available. These comparisons would not normally be considered good candidates for benefit transfers. Nevertheless, we estimate the meta-regression both with and without outliers.

[13] We also examined our models for leverage points and influential observations. There were some leverage points, but only a few were found to exert influence on OLS and WLS estimators. However, in general, it is difficult to identify influential and leverage data points when all of the explanatory variables are dichotomous. This observation, combined with the striking spread of the percentage transfer error, is what led us to adopt the IQR criterion.

[14] We considered dropping these two studies, but ultimately decided to keep them in the analysis because most of their observations are not defined as outliers by the IQR criterion (37% are outliers). To evaluate the sensitivity of our results to outlying observations we estimate the model with and without outliers.

3.3. Explanatory Variables

Table 1 defines the variables we use to explain the variation in benefit transfer errors, along with means and standard deviations for each variable. Of the 14 variables in the table, only $|\%BTE|$ is continuous. All 13 explanatory variables are binary descriptions of benefit-transfer applications and modeling decisions. They are grouped into the four categories that we defined earlier ($\Delta q, v, G,$ and T).

POLICYΔ indicates whether the analyst explicitly defines a baseline condition and a new policy condition. For example, in Johnston [12] Δq represents the difference between a current land use development plan and an alternative development plan. Therefore, POLICYΔ=1. Most studies do not consider specific policy changes, in which case POLICYΔ=0. The second variable in the Δq category, QUALITYΔ, equals one if and only if the transfer describes a change in *quality*, as opposed to a change in *quantity*. For example, changes in human health, river bank erosion, farming practices, air pollution, and water pollution are all defined as *quality* changes, whereas changes in fish catch rates, water supply, and access to recreation sites are all defined as *quantity* changes. Finally, USEVALUE indicates whether or not Δq affects the use value of a resource as opposed to a non-use or total value.

The second category of explanatory variables, *G*, assesses the similarity of the study and policy cases. POPULATION equals 1 if the study and policy case populations are essentially the same. STUDYAREA equals 1 if both cases occur in the same geographic region.

The third set of explanatory variables describes the valuation methodology. Most of the transfer errors are drawn from studies that used choice modeling (CM-31%) or contingent valuation (CV-29%). The remaining observations are based on reduced-form meta-analyses (META-

17%), travel-cost models (TC-12%), and random-utility models of site choice (RUM-11%).[15]

Finally, we use three variables to describe the nature of the transfer procedures. VALUETRANSFER indicates whether the procedure consisted of a value transfer or a function transfer; MULTIPLESTUDY equals 1 if the study case benefit measure is a composite of results from more than one study case; and MEAN equals 1 if the transfer error is reported as an average over two or more individual transfer errors.

4. Meta-Analysis of $|\%BTE|$

A nonparametric approach to meta-analysis avoids the need to restrict the functional relationship between benefit-transfer errors and features of the transfer process. However, it also presents a tradeoff. While nonparametric meta-analysis is robust to linear, additively separable functional forms assumed by most parametric meta-analyses, it typically requires more from the data.[16] To see this, consider the space of potential modeling decisions for benefit transfers. With 13 binary regressors representing modeling decisions, there are $2^{13} = 8{,}192$ "cells" in the data "grid" describing potential benefit transfer methods.[17] Since our data contain 1,071 observations on transfer errors, most of the cells in the grid are empty. Greater numbers of empty cells in the support of a particular variable make it tougher to determine the full range of impacts of that variable on the benefit transfer error.[18] Adding parametric restrictions to the meta-

[15] All of the meta-analysis studies use reduced-form regressions, as opposed to the "preference calibration" approach proposed by Smith, Van Houtven, and Pattanayak [23].

[16] In a nonparametric model *with continuous regressors* the complications would include a slower rate of convergence for estimators, selection of the smoothing parameters, and the curse of dimensionality. These complications are mitigated somewhat when all of the regressors are dichotomous. Details are provided in section 4.1.

[17] The methodological variables are mutually exclusive (RUM, TC, META, CV, CM), therefore, we have 13 binary regressors that fully characterize the data grid.

[18] In a parametric linear meta-regression with interaction effects for all possible combinations of modeling decisions, the impacts of modeling decisions on the transfer error cannot be identified for benefit transfer methods that correspond to empty cells on the data grid. Our nonparametric model "smoothes over" the empty cells by leveraging the information contained in nearby non-empty cells, providing a continuous approximation to the impacts of unobserved modeling decisions on transfer errors. This smoothing process is guided by our estimates of the optimal bandwidths.

regression reduces the number of cells that have to be filled in order to identify a variable's impact, while simultaneously increasing the potential for inconsistency due to functional form misspecification.

A second relevant tradeoff concerns finite sample bias. If some of the regressors in the nonparametric model are "irrelevant" in the sense that they have no systematic impact on the benefit-transfer errors, then the rate of convergence of nonparametric model will be slower than the correctly specified parametric model, increasing finite sample bias. While the correctly specified parametric model would be preferred in this situation, its functional form is unknown.

Given these tradeoffs, we proceed in two stages. First we estimate the nonparametric model. The results allow us to characterize ranges of effects for most variables. A few variables appear to be irrelevant, although this may simply reflect sparseness in the data grid. Therefore, we follow our nonparametric estimation with a conventional linear meta-regression, recognizing that the credibility of inferences based on our results is decreasing in the strength of the functional form assumptions we maintain [16].

4.1. A Nonparametric Model for Binary Regressors

Our nonparametric analysis is based on the kernel estimator for models with unordered discrete regressors described in Ouyang, Li and Racine [20]. The estimation procedure recognizes that the true form of $m(X)$ in equation (4) is unknown. Equation (4) is estimated using a variant of the Aitchison-Aitken kernel function, which has a simple form and can be easily generalized. To see the mechanics for this process, let λ_r denote a smoothing parameter (or bandwidth) associated with the r^{th} component of X. The kernel function for r is defined as

$$l(X_{ir}, X_{hr}, \lambda_r) = \begin{cases} 1 & if\ X_{ir} = X_{hr} \\ \lambda_r & if\ X_{ir} \neq X_{hr} \end{cases}, \tag{5}$$

using $i = 1, 2, \dots, N$ to index individual observations and X_{hr} to denote the value of the r^{th} component of X for the data point that neighbors X_i. The corresponding product kernel can be expressed as

$$L(X_{ir,} X_h, \lambda) = \prod_{r=1}^{R} l(X_{ir}, X_{hr,} \lambda_r) = \prod_{r=1}^{R} \lambda_r^{I(X_{ir} = X_{hr})}, \tag{6}$$

where $I(\cdot)$ is an indicator function that equals 1 iff the condition in parentheses is true. Finally, the estimator for the unknown function $m(X)$ is

$$\widehat{m}(X) = \frac{\sum_{i=1}^{N} |\%BTE|_i \, L(X_{ir}, X_h \ \lambda)}{\sum_{i=1}^{N} L(X_{ir}, X_h, \lambda)} \ , \tag{7}$$

which can be coded using the local constant Nadaraya-Watson estimator.

The estimated function $\widehat{m}(X)$ may be discontinuous or incur jumps because all of our regressors are binary. To define the effect of a binary variable, let $X_{\sim r}$ denote the subset of regressors in X after removing the r^{th} variable such that $X = [X_{\sim r}, X_r]$. The response of $\widehat{m}(X)$ to changing X_r from 0 to 1 can be written as:

$$\Delta_r = \widehat{m}(0, X_{\sim r}) - \widehat{m}(1, X_{\sim r}). \tag{8}$$

Δ_r can be estimated at every data point. Instead of having a single point estimate for each regressor as in a linear meta-regression, we have a vector of responses. We refer to these as "response effects".

In the language of nonparametric analysis, an explanatory variable is said to be "irrelevant" if $m(X)$ is constant with respect to that particular variable; i.e. if the response effects are zero everywhere. Hall, Li and Racine [9] formalize the distinction between relevant and irrelevant variables. To see their distinction, first partition the set of explanatory variables into two components, $X = [\ddot{X}, \dot{X}]$. The variables in $\dot{X}$ are irrelevant, if Y and $\ddot{X}$ are independent of $\dot{X}$. A weaker condition for irrelevance is that Y and $\dot{X}$ are independent, conditional on the set of relevant variables, $\ddot{X}$.[19] A variable may also be irrelevant if a large number of cells are empty in the support of that variable, which prevents us from learning the true range of effects. In either case, irrelevant variables can be identified by inspecting their bandwidths. The bandwidth for an irrelevant binary variable approaches the upper bound of "1" [20].[20] Therefore, if we find that $\lambda_r \approx 1$, we conclude that r is "irrelevant" in the sense that its response effects cannot be identified without adding parametric restrictions to the model. It is common in empirical work to use 0.8 or 0.9 as the cut-off for determining relevancy [9] .

The selection of smoothing parameters is one of the key empirical challenges with nonparametric estimation. We estimate bandwidths using the data-driven least-squares, cross-validation method (LSCV). LSCV employs the leave-one-out technique to find the optimal value of the smoothing parameters.[21] Ouyang, Li and Racine [20] demonstrate that the rate of convergence to the optimal value of λ is of the order $O_p(n^{-1})$ when all regressors are relevant and $O_p(n^{-0.5})$ when some regressors are irrelevant.[22] While nonparametric regression models with irrelevant variables are still econometrically consistent, their slower convergence rate implies an increase in finite sample bias relative to the correctly specified parametric model.

4.2. Nonparametric Results

[19] Hall, Li and Racine [9] proceed with the stronger definition and prove that their results still hold for the weaker condition.

[20] In this case, the model's predictions are essentially the same as if we had excluded the irrelevant variable. At the opposite extreme, a bandwidth of 0 would be equivalent to splitting the sample and estimating two separate models.

[21] Ouyang, Li and Racine [20] prove that smoothing parameters for irrelevant regressors approach 1 with positive probability. The exact probability depends on the function and the covariance matrix. For instance when the error is symmetric around zero and independent of the regressors, this probability is expected to be greater than 0.5.

[22] These convergence rates are faster than for nonparametric models with continuous regressors. If some regressors are continuous, then the rate of convergence for the smoothing parameter associated with binary variables is of the order $O_p(n^{-2/(4+q)})$, where q>1 is the number of regressors [9]][11]. Clearly, the rate of convergence to the true value (in probability) of the smoothing parameter is faster when all variables are binary. Moreover, if all regressors are discrete and relevant, the nonparametric model converges at the same rate as the correctly specified parametric model!

LSCV bandwidths were estimated separately for the full data set (N=1071) and for the data set excluding outliers (N=925).[23] The two sets of bandwidths differed substantially. Therefore, we adapted Hartarska et al.'s [10] procedure for dealing with outliers in nonparametric models. Following their logic, we attempted to mitigate the influence of outliers by estimating the optimal bandwidths without outliers and then using the estimated bandwidths to estimate $m(X)$ on the full data set (including outliers). Table 2 reports the resulting LSCV bandwidths, along with descriptive statics for the distribution of response effects (mean, median, 25th quartile and 75th quartile). Wild bootstrap standard errors are provided for each statistic.[24] The estimated bandwidths suggest that the data allow us to identify response effects for most variables.

Looking at the quartiles of the response effect distributions reveals four clear results. The first result is one of the stylized facts in the literature, the second result follows logically from what is known about the complexity of research design, and the last two results are new findings that have not been discussed previously. First, value transfers tend to be less accurate (have larger *%BTE*s) than function transfers. Second, transfer errors tend to be larger for studies that consider changes in environmental quality, rather than the quantity of a particular amenity. Third, travel-cost models tend to produce estimates that are more consistent (much lower *%BTE*s) between the study and policy sites than choice modeling. Finally, *%BTE*s tend to be larger for studies that add variables to the transfer function to control for differences between

---

[23]As a robustness check, bandwidths were also estimated using Akaike Information Criterion (AIC). The results were found to be very close to the bandwidths estimated via LSCV.

[24]After estimating equation (7), a bootstrap sample is generated as $|\%BTE|^* = \hat{m}(X) + \hat{u}\epsilon^*$, where $\epsilon^*$is a white noise term and $\hat{m}(X)$ and $\hat{u}$ are the nonparametric fitted value and residual terms, respectively. The white noise term is defined such that $\mathrm{E}(\epsilon^*) = 0$ and $\mathrm{E}(\epsilon^{*2}) = 1$. Values are randomly selected using a two point distribution given by: $\epsilon^* = \frac{1-\sqrt{5}}{2}$ with probability $p = (1+\sqrt{5})/2\sqrt{5}$ and $\epsilon^* = \frac{1+\sqrt{5}}{2}$ with probability $(1-p)$. The first bootstrap sample is then used to estimate the new response effect and residuals, which are then used to build the second bootstrap sample. This procedure is repeated 999 times. Thus, for each estimated response effect of the relevant variables, we will have 999 values. The standard error for an estimated response effect is calculated by taking the standard deviation of the corresponding 999 bootstrapped response effects [8].

baseline conditions and the new policy condition. All four of these results are remarkably robust; the interquartile ranges and means all indicate response effects greater than zero.

Other response effects are less robust. Convergent validity studies where the study and policy cases describe the same geographic area are associated with smaller transfer errors sometimes, but not always. The 25th quartile of the response effect distribution is negative, but the 50th and 75th quartiles are positive. The same is true for situations where the study case value was defined using data from multiple studies and in situations where the transfer error was reported as a mean over multiple individual errors.

Finally, as can be seen from the bandwidths, the data do not allow us to recover response effect distributions for USEVALUE, POPULATION, META, RUM, and CV. Estimated bandwidths for each of these variables are close to 1. Therefore, in the hope of recovering reasonable approximations to the means of their response effect distributions, we repeat the estimation after adding the conventional linearity and separability restrictions to the meta-regression ($\widehat{m}(X)$).

4.3. Parametric Results: Weighted Least Squares

Recall that the 31 convergent validity studies vary considerably in the number of values they report for the %*BTE* (from 2 to 178). Since ordinary least squares estimation assigns equal weight to each observation, studies that provide more observations have greater influence on the results from linear estimation. To mitigate this influence, we estimate equation (4) using weighted least squares (WLS). Each observation is weighted by the total number of observations contributed by the corresponding study. Thus, individual observations from studies that provide more observations receive less weight in the estimation.[25] The last two columns of Table 2 re-

[25] Not surprisingly, a consistent kernel test [11] soundly rejects the null hypothesis that the WLS model is correctly specified against the nonparametric alternative. However, such tests do not allow us to assess the magnitude of the potential bias from functional form misspecification. For example, one of the few previous studies to estimate a non-

port the WLS results, with and without outliers. Not surprisingly, dropping outliers decreases the absolute magnitudes of point estimates for the regression coefficients and improves model fit.[26] The 13 modeling decisions explain three quarters of the variation in the percentage transfer error when outliers are removed.

Imposing the linearity and separability restrictions on the model produces substantial changes in the estimated mean response effects. Nevertheless, with or without outliers, the WLS results are still consistent with all but one of the qualitative findings from the nonparametric regression. The difference is that the estimated coefficient on POLICYΔ is now close to zero and statistically insignificant. This difference arises because WLS assigns less weight to observations drawn from studies that report large numbers of transfer errors. The vast majority of the observations with POLICYΔ=1 are drawn from two studies that, together, report more transfer errors than any other study, plus the largest values for the transfer errors. Due to the discreteness in the other explanatory variables, the nonparametric response effects for POLICYΔ operate similarly to fixed effects for subsets of observations reported in those two studies. When the observations from the two studies are down-weighted by the WLS procedure, their influence on the average response effect diminishes, resulting in point estimates close to zero.

While we certainly must be more cautious in drawing strong conclusions from the WLS model, the additional parametric structure produces results that seem intuitively plausible. It suggests that geographic similarity between the study and policy cases tends to reduce transfer errors. It also suggests that, all else constant, using data from multiple study cases tends to decrease transfer errors. Both decreases are close to 10%. However, the most striking changes are

---

linear meta-regression, Smith and Osborne [26], soundly rejected linear models against a Box-Cox model, but found that the choice of model did not affect their qualitative results.

[26] We also estimated the model using OLS. The results were much more sensitive to outliers. This is partly due to the fact that the study with the largest error (7496%) also had the largest number of observations.

in the coefficients for the (formerly irrelevant) variables describing study case valuation methods. The results indicate that RUM, travel cost, and contingent valuation methods all tend to produce smaller transfer errors than reduced form meta-analyses.

5. Summary and Interpretation

The nonparametric response effect quartiles and WLS point estimates in Table 2 are informative, but they do not provide a complete picture of the results. In this section we use a novel approach to graphical analysis—45 degree plots of whisker figures—to summarize and interpret our findings. We focus on the variables that have significant point estimates in the WLS models and a discernable pattern in the distribution of response effects. This includes eight of the twelve explanatory variables. We are unable to draw clear inferences on POLICYΔ, USEVALUE, POPULATION, and MEAN based on our results.

The basic idea for the 45 degree plots is simple. Scaling both axes of a 2-dimensional diagram to represent the same range of $|\%BTE|$ response effects makes it possible to visualize the entire distribution of response effects (and whisker plots of their confidence intervals) arrayed along the 45-degree line. Consider Figure 3. It depicts the WLS point estimates for the coefficient on the value transfer indicator variable, the full range of nonparametric response-effect estimates, and 95% confidence intervals for all estimates. The WLS point estimates, based on the full data set (47%) and the data set without outliers (11%), are denoted by the shaded square and circle, respectively. The corresponding horizontal bars define the upper and lower bounds of their 95% confidence intervals, using White's corrected standard errors.

The numbers in the other whisker figures indicate the percentages of the nonparametric response effects located at those points. Since all of the regressors are binary, $\hat{m}(X)$ is discontinuous. As a result, the response effects tend to be clustered at values that correspond to specific combinations of explanatory variables. For example, 35% of the nonparametric response effects are located at the largest effect. A single cluster may represent multiple observations from the same study or may represent observations from different studies that have identical values for the regressors. In situations where clusters of response effects are too dense for their individual percentages to be legible, we use brackets to indicate the cumulative percentage associated with the group of effects. In Figure 3, nearly half (47%) of the response effects are located near zero. Ninety-five percent confidence intervals for each cluster are constructed with wild bootstrap standard errors and are denoted by the upper and lower bounds of the whisker plots.

5.1. Function Transfer versus Value Transfer

Figure 3 comes about as close as possible to confirming the stylized fact that function transfers outperform value transfers. Eighty-five percent of the response effects are in the positive quadrant and so are the point estimates from both WLS models. Based on the 95% confidence intervals, we would only reject the null hypothesis that value transfers have larger errors than function transfers for 7% of the response effects!

While many of the response effects are similar to the WLS estimates (indicating that value transfers increase errors between 0% and 50%) the distribution has a fat right tail. For the cluster of points representing the top 35% of response effects, even the lower bound of the 95% confidence interval exceeds errors of 100%. The bottom line is that the cumulative findings from the convergent-validity literature suggest that benefit transfers can almost always be improved by choosing to perform a function transfer rather than a value transfer. The strength of this conclusion is underscored by the fact that it does not rely on global parametric assumptions on the meta-regression function.

5.2. Quantity Changes versus Quality Changes

An equally robust result is that benefit transfers are almost always more accurate when they describe changes in the quantity of environmental amenities, rather than changes in quality. This result also does not depend on parametric restrictions in the meta-regression or other types of modeling decisions made by benefit-transfer practitioners. Figure 4 illustrates that 89% of the nonparametric response effects are positive, along with both estimates from the WLS model, indicating larger transfer errors for valuing changes in quality.

This finding has not been widely recognized in the literature, but it makes sense. Quantity changes are usually easier to describe than quality changes. Anglers can easily understand a change in catch rates or a permanent closure of a fishing site, for example. In contrast, it may be difficult for them to assess a change in water quality at a fishing site, especially when the change is not visible and the effect of fishing may not be explicit. This forces the benefit-transfer analyst to make an assumption about the metrics that consumers use to judge environmental quality. A quality assumption that holds for the study case may be invalid at the policy case, increasing the transfer error.[27]

The benefit-transfer practitioner may or may not be able to control whether their assessment is framed as a quantity change. At the very least, the results in Figure 4 suggest that extra caution and additional sensitivity analyses are warranted if the transfer involves valuing a change in environmental quality. This is especially important when roughly half of the estimated transfer effects exceed 100%

5.3. Geographic Similarity

Figure 5 illustrates that the results on geographic similarity are mixed. The WLS point estimates and many of the nonparametric response effects are negative, consistent with the sty-

[27] A less optimistic perspective would be that there is no reason to expect the errors from violations of the usual "full information" assumption to be the same at the study and policy cases.

lized fact that geographic similarity between the study and policy cases should improve transfer accuracy. However, 54% of the nonparametric response-effects are clustered at a large positive value (nearly 150%). The counterintuitive cluster of extreme positive values reflects two features of the nonparametric analysis. First, the response effects in Figure 5 are only calculated for the observations with STUDYAREA=1.[28] This is a small share of the data (18%). Second, of this 18%, just over half of the observations come from a single study that happens to have an extreme response effect. In contrast, many of the negative response effects are from value transfers. Thus, our results reinforce the importance of geographic similarity between study and policy cases, and this stylized fact is particularly relevant when study case values are transferred directly to the policy case and a function is not available to adjust the transferred value between cases.

5.4. Valuation Methodology

The WLS results suggest that random utility models, travel cost models, and contingent valuation all perform better than reduced form meta-analysis. After removing outliers, choice modeling also produces lower transfer errors than meta-analysis. The estimated bandwidths for RUM, CV and META suggest they are nonparametrically irrelevant.[29] If the irrelevancy is due to sparseness in the data "grid" used to describe modeling decisions, then parametric restrictions have the potential to help identify the true response effects. The nonparametric irrelevancy simply underscores the caveat that the WLS results for these variables are contingent on the maintained assumption that they influence the transfer errors independently of the other model-

[28] Similar to dummy variables in a linear model, calculation of response effects requires there to be a base group with a value of 0 for each variable. We defined the base group to be the subset of observations that did not make the modeling decision represented by a 1 for the corresponding variable. If we were to switch the base group, we could calculate response effects for the remaining observations, allowing us to "fill in" the response effect distribution. However, the qualitative results would be unchanged.

[29] We do not report 45-degree plots for each pair wise comparison of modeling decisions. This is partly in the interest of brevity and partly because, in this case, the general pattern of results can be seen from Table 2.

ing decisions we consider.

5.5. Combining Data from Multiple Study Cases

Figure 6 presents the distribution of response effects for the subset of studies that used data from multiple study cases (MULTIPLESTUDY) to estimate the benefits for the policy case. Intuition would suggest that combining data from multiple study cases would reduce transfer errors as long as all of the selected study cases are appropriate for the transfer. Moreover, Smith, Van Houtven, and Pattanayak's [23] logic for "preference calibration" suggests there are gains from using multiple study cases to span the relevant portion of the policy case benefit function. Indeed, the WLS models suggest a 10% to 30% improvement in accuracy, on average, and 55% of the nonparametric response effects confirm this finding.

The nonparametric model illustrates that the WLS averages reflect considerable heterogeneity in the response effects. Upon closer inspection, this heterogeneity can be explained by two other features of the transfer process. If we focus on the negative response effects, 98% correspond to function transfers and 90% correspond to studies that evaluated quantity changes. In contrast, all of the positive response effects are for studies that conducted value transfers and/or considered quality changes. These insights indicate that averaging values from several studies reduces accuracy of value transfers; it is better to select a single value estimate where the study case matches the policy case.

6. Conclusions

What are the practical implications from the past 20 years of research on the convergent validity of benefit transfers? The evidence overwhelmingly supports the stylized fact that function transfers are more accurate than values transfers. The literature also suggests that benefit transfers are better able to predict the willingness to pay for quantity changes than for changes in environmental quality. The former is a choice made by the benefit-transfer practitioner, while the latter is defined by the policy question being addressed. That said, if the practitioner must perform a value transfer then it becomes especially important to ensure that the study-case value estimate matches the policy-case value definition. For example, taking a simple average over values from multiple preexisting studies is unlikely to reduce transfer errors because this is simply averaging the benefit transfer errors from multiple study-case values. However, we do find that using information from multiple study cases can improve the accuracy of function transfers. If the analyst is able to perform a function transfer, then the evidence suggests that the various structural frameworks (RUM, travel cost, contingent valuation, and choice modeling) are all likely to generate smaller transfer errors than reduced form meta-analysis. This result is likely due to the fact that limited reporting in original studies often leaves the analyst to primarily record study case valuation methods in the meta-equation regressors with limited information on the item valued and the population whose values are estimated.

It is important to distinguish between the way that meta-analysis is used for benefit transfers and the way that we have used the methodology in this study. We have developed a new nonparametric approach to meta-analysis and demonstrated that it can extract important signals from the data that remain hidden in conventional linear models. This advancement continues a long tradition of refining the econometrics of meta-analysis in order to distill key findings from research on important questions in environmental economics [1][25][26][29]. In contrast, when meta-analysis is used to conduct a benefit transfer the methodology must provide policy-relevant transfer estimates. Parametric estimation is desirable for providing point estimates of meta-equations parameters, but the robustness of these parameter estimates should be investigated using nonparametric estimation. As Smith and Pattanayak [24] explain, the benefit concept and the amenity being valued must be held constant in order to make consistent predictions for policy case benefits. This can only happen if study cases do a better job of defining the environmental change valued and the affected population, and there are multiple studies valuing the same environmental change, but for different increments of change. Thus, effective use of meta-analysis equations in benefit transfer requires improved data reporting in study cases and more refined specification and estimation of meta-equations.

We recommend that future studies evaluating the validity of benefit transfers be more thorough in their documentation and analyses. In our review of the literature, we observed several cases where transfer procedures were not clearly documented. Wide ranges of transfer errors were presented, but not explained. A documentation protocol is needed in order for future convergent-validity studies to enhance the credibility of benefit transfers. Each study should define the criteria they use to identify study and policy cases that are good candidates for benefit transfers, and then justify each comparison in the context of those criteria. Additionally, investigators need to go beyond simply reporting transfer errors to explain why some comparisons have small errors and others have large errors. Taking these steps would provide insights to refine the criteria for when appropriate data are available to conduct credible benefit transfers.

References

[1] H.S. Banzaf, V.K. Smith, Meta-analysis in model implementation: choice sets and the valuation of air quality improvements, Journal of Applied Econometrics 22 (6) (2007) 1013-1031.

[2] J.C. Bergstrom, L.O. Taylor, Using meta-analysis for benefits transfer: theory and practice, Ecological Economics 60(2) (2006) 351-360.

[3] K.J. Boyle, N.V. Kuminoff, C.F. Parmeter, J.C. Pope, Necessary Conditions for Valid Benefit Transfers, American Journal of Agricultural Economics 91 (2009) 1328-1334.

[4] K.J. Boyle, N.V. Kuminoff, C.F. Parmeter, J.C. Pope, The benefit-transfer challenges, Annual Review of Resource Economics 2 (1) (2010) 161-182.

[5] S. Chattopadhyay, A repeated sampling technique in assessing the validity of benefit transfer in valuing non-market goods, Land Economics 79 (4) (2003) 576-596.

[6] S. Engel, Benefit function transfers versus meta-analysis as policy making tools: a comparison, in: R. J. G. M. Florax, Peter Nijkamp, Kenneth George Willis (eds.), Comparative Environmental Economic Assessment, Edward Elgar, Cheltenham, UK, 2002.

[7] T. Eshet, M.G. Baron, M. Shechter, Exploring benefit transfer: disamenities of waste transfer stations, Environmental and Resources Economics 37(3) (2007) 521-547.

[8] E. Flachaire, Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap, Computational Statistics & Data Analysis 49 (2) (2005) 361 – 376.

[9] P. Hall, Q. Li, J. Racine, Nonparametric estimation of regression functions in the presence of irrelevant regressors, The Review of Economics and Statistics 89(4) (2007) 784–789.

[10] V. Hartarska, C. F. Parmeter, D. Nadolnyak, B. Zhu, Economies of scope for microfinance: differences across output measures, Pacific Economic Review 15 (4) (2010) 464–481.

[11] C. Hsiao, Q. Li, J. Racine, A consistent model specification test with mixed discrete and continuous data, Journal of Econometrics 140(2) (2007) 802–826.

[12] R.J. Johnston, Choice experiments, site similarity and benefits transfer, Environmental and Resources Economics 38(3) (2007) 331-351.

[13] R.J. Johnston, R.S. Rosenberger, Methods, trends and controversies in contemporary benefit transfer, Journal of Economic Surveys 24 (3) (2010) 479-510.

[14] S. Kirchhoff, B.G. Colby, J.T. LaFrance, Evaluating the performance of benefit transfer: an empirical inquiry, Journal of Environmental Economics and Management 33 (1) (1997) 75-93.

[15] J.B. Loomis, B. Roach, F. Ward, R. Ready, Testing transferability of recreation demand models across regions: a study of corps of engineer reservoirs, Water Resources Research 31 (3) (1995) 721–730.

[16] C.F. Manski, Identification for Prediction and Decision, Harvard University Press, Cambridge, 2007.

[17] M. Morrison, J. Bennett, R. Blamey, J. Louivere, Choice modeling, non-use values and benefit transfer, Economic Analysis and Policy 30 (1) (2000) 13-32.

[18] M. Morrison, and J. Bennett, R. Blamey, J. Louviere, Choice modeling and tests of benefit transfer, American Journal of Agricultural Economics 84 (1) (2002) 161-170.

[19] J.P. Nelson, P.E. Kennedy, The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment, Environmental and Resource Economics 42(3) (2009) 345-377.

[20] D. Ouyang, Qi Li, J.S. Racine, Nonparametric estimation of regression functions with discrete regressors, Econometric Theory 25(1) (2009) 1-42.

[21] N. C. Schwertman, M. A. Owens, R. Adnan, A simple more general boxplot method for identifying outliers, Computational Statistics & Data Analysis 47 (1) (2004) 165–174.

[22] T.D. Stanley, Beyond publication bias, Journal of Economic Surveys 19 (3) (2005) 309-345.

[23] V.K. Smith, G. Van Houtven, S. K. Pattanayak, Benefit transfer via preference calibration: 'Prudential algebra' for policy, Land Economics 78 (1) (2002)132-152.

[24] V.K. Smith and S. K. Pattanayak, Is Meta-Analysis a Noah's Ark for Non-Market Valuation, Environmental and Resource Economics 22 (2002) 271-296.

[25] V.K. Smith, Y. Kaoru, Signals or noise? Explaining the variation in recreation benefit estimates, American Journal of Agricultural Economics 72(2) (1990) 419-433.

[26] V.K. Smith and L. Osborne, Do Contingent Valuation Estimates Pass a Scope Test? A Meta Analysis. Journal of Environmental Economics and Management 31 (1996): 287-301.

[27] U.S. Environmental Protection Agency, Guidelines for Preparing Economic Analyses, EPA 240-R_10-001 (prepublication edition), 2010. http://yosemite.epa.gov/ee/epa/eed.nsf/pages/Guidelines.html/$file/Guidelines.pdf, accessed December 31, 2010.

[28] U.S. Executive Office of the President, Executive Order 12866: Regulatory Planning and Review, Federal Register 58(190): 51735- 51744, 1993.

[29] R. G. Walsh, D. M. Johnson, J. R. McKean, Issues in nonmarket valuation and policy application: a retrospective glance, Western Journal of Agricultural Economics 14(1) (1989) 178-188.

Table 1. Summary Statistics and Definitions for Key Variables (N=1071)

Variable	Definition	Mean (s.d.)
|%Benefit Transfer Error|	$\left\lvert\left[\left(\frac{\widehat{wtp}_{ij}^{T}}{\widehat{wtp}_{ij}}\right)-1\right]\times 100\right\rvert$	171.90 (555.20)
	Definition for the Change in the Amenity (Δq)	
POLICYΔ	1 if transfer includes variables that define baseline and policy (new) conditions; 0 otherwise	0.24 (0.43)
QUALITYΔ	1 if a change in quality; 0 if change in quantity	0.51 (0.50)
USEVALUE	1 if use value; 0 if non-use value	0.66 (0.47)
	Similarity between Study and Policy Cases (G)	
POPULATION	1 iff study and policy case populations are the same	0.09 (0.29)
STUDYAREA	1 iff study and policy cases describe the same geographic area	0.18 (0.38)
	Valuation Methodology (v)	
META	1 iff the valuation method is a meta-analysis	0.17 (0.38)
RUM	1 iff the valuation method is a random-utility model	0.11 (0.32)
TC	1 iff the valuation method is a travel-cost model	0.12 (0.32)
CV	1 iff the valuation method is contingent valuation	0.29 (0.45)
CM	1 iff the valuation method is choice modeling	0.31 (0.46)
	Transfer Procedures (T)	
VALUETRANSFER	1 if value transfer; 0 if function transfer	0.38 (0.48)
MULTIPLESTUDY	1 iff two or more study cases are used to estimate study-case value.	0.27 (0.45)
MEAN	1 iff transfer error is reported as a mean of two or more transfer errors	0.15 (0.36)

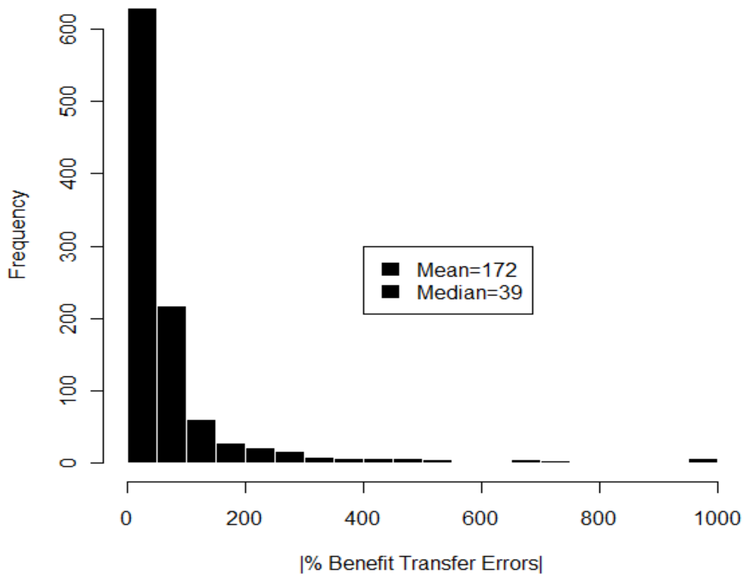
Table 2. Nonparametric and Parametric Meta-Regression Results

	Nonparametric Regression[a]					WLS[b]	
	BW	Response Effect Quartiles				All Data	Without Outliers
		Mean	25%	50%	75%		
POLICYΔ	0.11	258.53 (49.32)	206.47 (49.32)	322.25 (64.82)	322.25 (64.82)	11.13 (12.30)	-3.19 (10.05)
QUALITYΔ	0.00	233.25 (40.14)	6.22 (5.73)	109.18 (21.56)	579.04 (96.60)	47.72** (23.08)	24.50** (9.50)
USEVALUE	1.00	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	45.75** (20.46)	9.93 (7.13)
POPULATION	0.96	-0.14 (0.07)	-0.03 (0.07)	-0.03 (0.07)	-0.03 (0.07)	44.28 (28.26)	19.80* (12.01)
STUDYAREA	0.05	62.12 (6.00)	-44.59 (137.86)	147.89 (20.17)	147.89 (20.17)	-26.26** (11.69)	-11.45** (5.43)
RUM	1.00	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-132.05* (68.10)	-56.05*** (19.41)
TC	0.00	13.48 (7.90)	29.34 (7.90)	49.34 (11.93)	49.34 (11.93)	-150.22** (71.64)	-74.05*** (19.55)
CV	0.95	-0.40 (0.16)	-0.60 (0.08)	-0.03 (0.06)	-0.30 (0.08)	-146.24** (66.17)	-66.97*** (17.00)
CM	0.04	220.50 (38.82)	216.89 (38.35)	254.51 (44.28)	254.51 (44.28)	-78.59 (56.06)	-26.35* (14.36)
VALUETRANSFER	0.34	115.37 (31.79)	0.77 (0.19)	18.09 (6.25)	262.23 (49.08)	47.12*** (13.63)	11.21** (5.58)
MULTIPLESTUDY	0.01	-0.67 (4.47)	-14.23 (4.90)	5.48 (2.13)	5.48 (2.13)	-28.14** (12.05)	-13.28** (5.69)
MEAN	0.02	0.89 (5.07)	-11.05 (5.07)	20.88 (5.24)	20.88 (5.24)	-33.10*** (12.68)	-10.61 (6.45)
Intercept						114.12*** (41.40)	74.49*** (12.74)
$\overline{R}^2$						0.33	0.76
F - value						45.50	251.07
N	925	1071	1071	1071	1071	1071	925

[a] The variable META is excluded from this table as it is an irrelevant variable for nonparametric regression and an omitted category for WLS models. Bandwidths are estimated with outliers excluded. Bootstrapped standard errors for response effect estimates are given in parenthesis. The fact that the 50th and 75th quartiles in the distribution of response effects are identical for some explanatory variables reflects the discreteness in variable space.

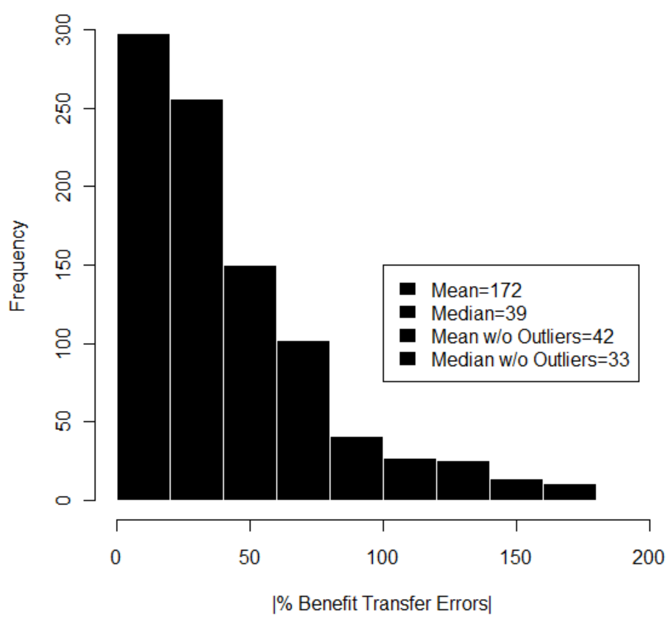
[b] Significance codes: '***' 0.01, '**' 0.05, '*' 0.1. White's corrected standard errors are given in parenthesis.

Figure 1. Distribution of Percentage Transfer Errors (n=1071)[a]



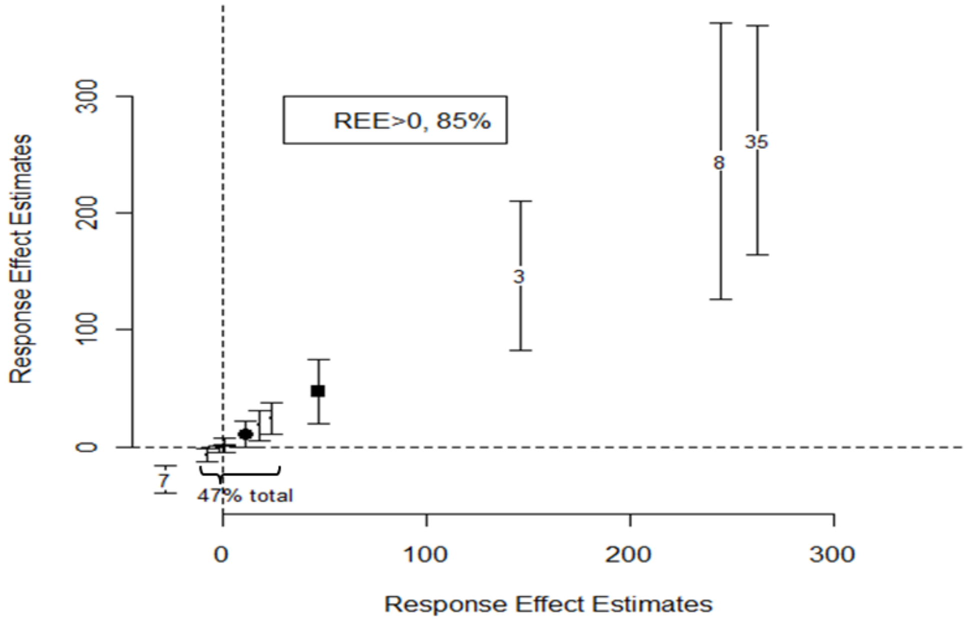
[a] The scaling of the horizontal axis excludes 3.83% of observations with errors exceeding 1000%.

Figure 2. Truncated Distribution of Transfer Errors, Excluding Outliers[a]



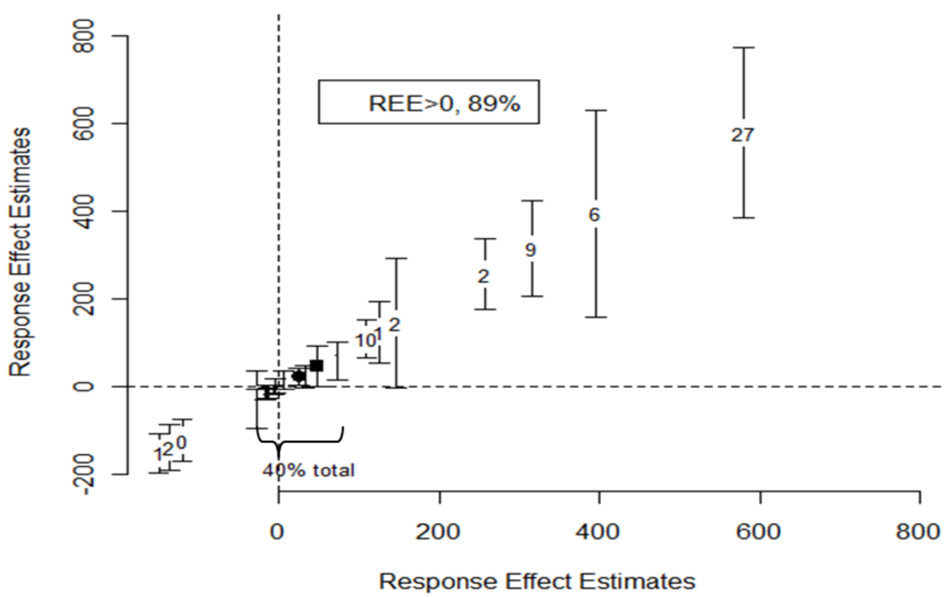
[a] This histogram is drawn for $|\%BTE|$ without outliers. Of 1071 observations 13.63% are outliers according to the interquartile range criterion.

Figure 3. Benefit Transfer Error Response Effects for VALUETRANSFER[a]



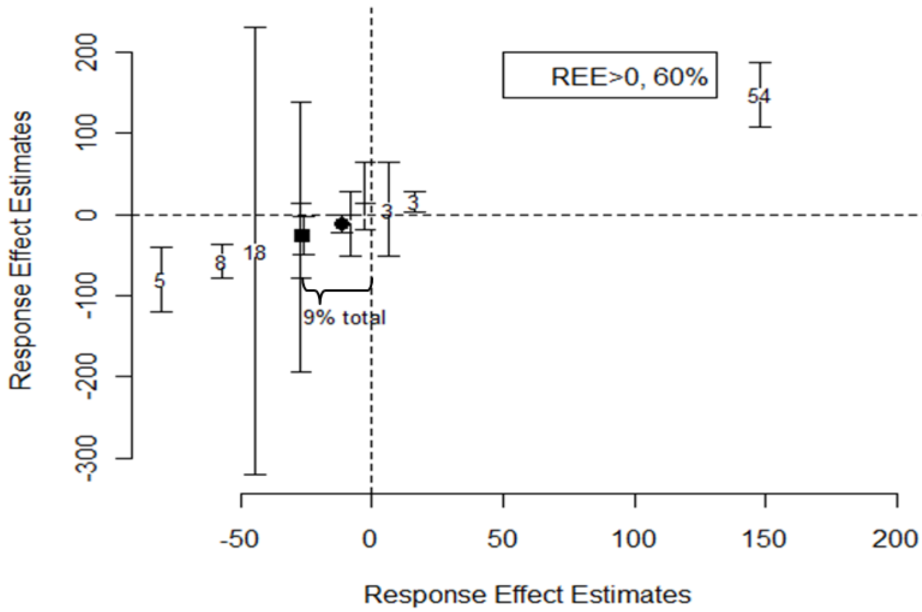
[a] The figure plots WLS point estimates for the VALUETRANSFER variable, nonparametric response effect estimates (REE), and 95% confidence intervals for each. The square and circle represent the WLS point estimates based on the data with and without outliers, respectively. The horizontal bars above and below denote 95% confidence intervals. The numbers indicate the share of the response effects at the point where the number is located. Clustering of REEs occurs because all of the independent variables are binary.

Figure 4. Benefit Transfer Error Response Effects for QUALITYΔ [a]



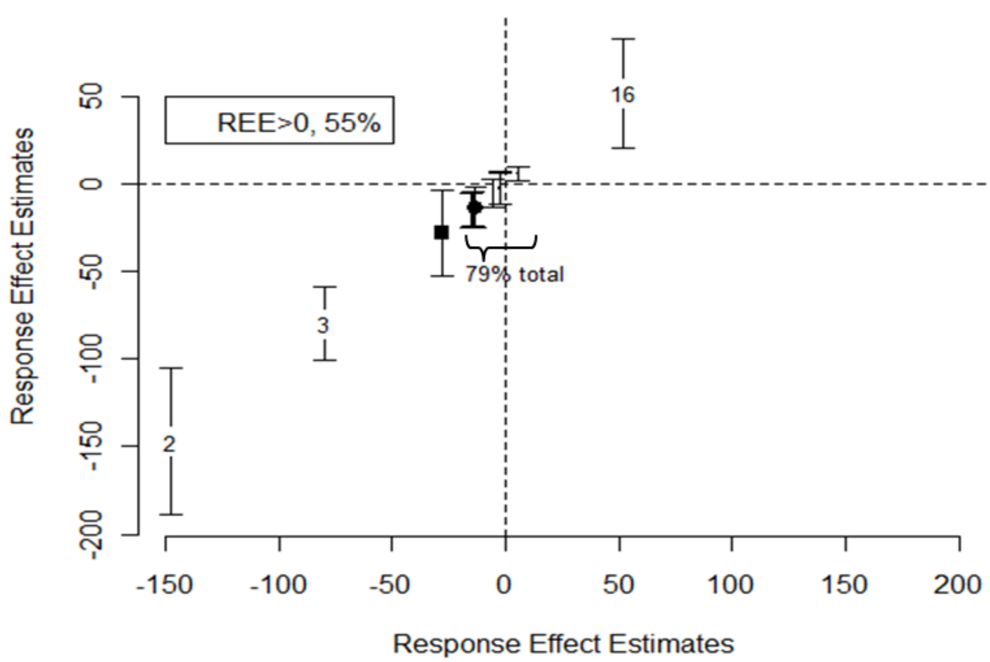
[a] The figure plots WLS point estimates for the QUALITYΔ variable, nonparametric response effect estimates (REE), and 95% confidence intervals for each. See the footnote to Figure 3 for additional explanation.

Figure 5. Benefit Transfer Error Response Effects for STUDYAREA [a]



[a] The figure plots WLS point estimates for the STUDYAREA variable, nonparametric response effect estimates (REE), and 95% confidence intervals for each. See the footnote to Figure 3 for additional explanation.

Figure 6. Benefit Transfer Error Response Effects for MULTIPLESTUDY [a]



[a] The figure plots WLS point estimates for the MULTIPLESTUDY variable, nonparametric response effect estimates (REE), and 95% confidence intervals for each. See the footnote to Figure 3 for additional explanation.

Appendix Table I. Characteristics of 31 Benefit Transfer Validity Studies

Authors (year)	Valuation Method	Resource	N	% transfer error mean (min, max)	|% transfer error| mean (min, max)
Barton (2002)	CV	Beach Quality	8	-20 (-23,-10)	20 (10,23)
Bergland, Magnussen and Navrud (1995)	CV	Water Supply	2	-21 (-24,-18)	21 (18,24)
Brouwer and Bateman (2005)	CV	Human Health	85	17 (-41,123)	34 (0.4,123)
Brouwer and Spaninks (1999)	CV	Farm Land	8	3 (-59,60)	42 (22,60)
Colombo and Hanley (2008)	CM	Farm Land	178	680 (2, 7496)	680 (2,7496)
Colombo, Calatrava-Requens, and Hanley (2007)	CM	Soil Conservation	54	110 (8,1148)	110 (8,1148)
Groothuis (2005)	CV/TC	Deer Hunting	120	-10 (-75,136)	30 (0.1,136)
Hanley, Wright, and Alvarez-Farizo (2006)	CM	Ecosystem Health	2	-72 (-78,-67)	72 (67,78)
Jiang, Swallow, and McGonagle (2005)	CM	Coastal Land	5	-68 (-85,-53)	68 (53,85)
Johnston and Duke (2009)	CM	Farm Land	4	-76 (-100,-29)	76 (29,100)
Johnston (2007)	CM	Mixed Resources	24	-12 (-101,58)	37 (7,101)
Kerr and Sharp (2006)	CM	Ecosystem Health	22	79 (-63,704)	120 (2,704)
Kristofersson and Navrud (2007)	CV	Fishing/ Ecosystem Health	21	125 (7,319)	125 (7,319)
Lindhjem and Navrud (2008)	META	Multiple Use Forestry	16	73 (25,266)	73 (25,266)
Loomis et al. (1995)	TC	Reservoir	56	106 (-50,475)	115 (0.5,475)

Loomis (1992)	TC	Sport Fishing	10	0.2 (-18,9)	6 (1, 18)
Matthews, Hutchinson, and Scarpa (2009)	CV	Forest Recreation	84	12 (-42,125)	27 (0.0,125)
Morrison and Bennett (2006)	CM	Ecosystem Health	28	-25 (-171,30)	35 (1,171)
Morrison et al. (2002)	CM	Wetlands	9	-32 (-66,-4)	32 (4,66)
Parsons and Kealy (1994)	RUM	Water Recreation	11	-4 (-66,75)	21 (1,75)
Piper and Martin (2001)	CV	Water Supply	8	35 (-9,149)	39 (3,149)
Ready and Navrud (2007)	CV	Human Health	2	37.95 (37.7,38.2)	37.95 (37.7,38.2)
Ready et al. (2004)	CV	Human Health	21	37 (20,83)	37 (20,83)
Rosenberger and Loomis (2000)	META	Mixed Re-sources	115	17 (-79,319)	49 (0.0,319)
Rozan (2004)	CV	Air Quality	4	-2 (-28,30)	25 (16,30)
Shrestha and Loomis (2003)	META	Mixed Re-sources	34	60 (-74,411)	84 (12,411)
Shrestha and Loomis (2001)	META	Outdoor Recreation	18	6 (-46,81)	28 (0.5,81)
Stapler and Johnston (2009)	META	Sport Fishing	4	228 (64,572)	228 (64,572)
Vandenberg, Poe and Powell (2001)	CV	Ground Water	8	29 (16,44)	29 (16,44)
Zanderson, Termansen and Jensen (2007a)	RUM	Forest Recreation	6	4 (-66,55)	30 (4,66)
Zanderson, Termansen and Jensen (2007b)	RUM	Forest Recreation	104	180 (-73,1111)	194 (1.3,1111)

Appendix Table II. Validity Studies Excluded from the Meta-Analysis [a]

Study	Reason For Excluding
Bowker, English and Bergstrom (1997)	Insufficient documentation of key variables
Chattopadhyay (2003)	No point estimate for transfer error
Downing and Ozuna (1996)	Insufficient documentation of key variables
Engel (2002)	No point estimate for transfer error
Eshet, Baron and Shechter (2007)	Only study to use hedonic methodology
Jeong and Haab (2004)	Insufficient documentation of key variables
Kirchhoff, Colby and LaFrance (1997)	Insufficient documentation of key variables
Leon et al. (2002)	Insufficient documentation of key variables
Morrison et al. (2000)	Redundant, given Morrison et al. (2002)

[a] Complete references are provided in the supplemental online appendix.

Reference List for Benefit Transfer Convergent Validity Studies

D.N. Barton, The transferability of benefit transfer: contingent valuation of water quality improvements in Costa Rica, *Ecological Economics* 42 (1-2) (2002) 147-164.

O. Bergland, K. Magnussen, S. Navrud, Benefit transfer: testing for accuracy and reliability, in: R.J.G.M. Florax, P. Nijkamp, K.G. Willis (Eds.), Comparative Environmental Economic Assessment, Edward Elgar, Cheltenham, UK, 2002.

J.M. Bowker, D.B.K. English, J.C. Bergstrom, Benefits transfer and count data travel cost models: an application and test of a varying parameter approach with guided whitewater rafting, Faculty Series (16703), Department of Agricultural and Applied Economics, University of Georgia, Athens, 1997.

R. Brouwer, and F.A. Spaninks, The validity of environmental benefits transfer: further empirical testing, *Environmental and Resource Economics* 14 (1) (1999) 95-117.

R. Brouwer, I. J. Bateman, Benefits transfer of willingness to pay estimates and functions for health-risk reductions: a cross-country study, *Journal of Health Economics* 24 (3) (2005) 591–611.

S. Chattopadhyay, A repeated sampling technique in assessing the validity of benefit transfer in valuing non-market goods, *Land Economics* 79 (4) (2003) 576-596.

S. Colombo, N. Hanley, How can we reduce the errors from benefits transfer? An investigation using the choice experiment method, *Land Economics* 84 (1) (2008) 128-147.

S. Colombo, J. Calatrava-Requens, N. Hanley, Testing choice experiment for benefit transfer with preference heterogeneity, *American Journal of Agricultural Economics* 89 (1) (2007) 135-151.

M. Downing, T. Ozuna, Testing the reliability of the benefit function transfer approach, *Journal of Environmental Economics and Management* 30 (3): (1996) 316-322.

S. Engel, Benefit function transfers versus meta-analysis as policy making tools: a comparison, in: R. J. G. M. Florax, Peter Nijkamp, Kenneth George Willis (eds.), Comparative Environmental Economic Assessment, Edward Elgar, Cheltenham, UK, 2002.

T. Eshet, M.G. Baron, M. Shechter, Exploring benefit transfer: disamenities of waste transfer stations, *Environmental and Resources Economics* 37(3) (2007) 521-547.

P.A Groothuis, Benefit transfer: a comparison of approaches, *Growth and Change* 36 (4) (2005) 551-564.

N. Hanley, R.E. Wright, B. Alvarez-Farizo, Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive, *Journal of Environmental Management* 78 (2) (2006) 183-193.

H. Jeong, T. Haab, The economic value of marine recreational fishing: applying benefit transfer to marine recreational fisheries statistics survey, Working Paper Series (28322), Department of Agricultural, Environmental and Development Economics, Ohio State University, Columbus, 2004.

Y. Jiang, S.K. Swallow, M.P. McGonagall, Context-sensitive benefit transfer using stated choice models: specification and convergent validity for policy analysis, *Environmental and Resource Economics* 31(4) (2005) 477–499.

R.J. Johnston, Choice experiments, site similarity and benefits transfer, *Environmental and Resources Economics* 38(3) (2007) 331-351.

R.J. Johnston, J. M. Duke, Willingness to pay for land preservation across states and jurisdictional scale: implications for benefit transfer, *Land Economics* 85 (2) (2009) 217-237.

G.N. Kerr, Basil M. H. Sharp, Transferring mitigation values for small stream, in: J. Rolfe, and J. Bennett (eds.), Choice Modeling and The Transfer of Environmental Values, Edward Elgar, Cheltenham, UK, 2006.

S. Kirchhoff, B.G. Colby, J.T. LaFrance, Evaluating the performance of benefit transfer: an empirical inquiry, *Journal of Environmental Economics and Management* 33 (1) (1997) 75-93.

D. Kristofersson, S. Navrud, Can use and non-use value be transferred across countries?, in: S. Navrud and R. Ready (eds.), Environmental Value Transfer: Issues and Methods, Springer, Dordrecht, The Netherlands, 2007.

C. J. Leon, F. J. Vazquez-Polo, N. Guerre, P. Riera, A Bayesian model for benefit transfer: application to national parks in Spain, *Applied Economics* 34 (6) (2002) 749-757.

H. Lindhjem, S. Navrud, How reliable are meta-analyses for international benefit transfers?, *Ecological Economics* 66 (2-3) (2008) 425-435.

J.B. Loomis, B. Roach, F. Ward, R. Ready, Testing transferability of recreation demand models across regions: a study of corps of engineer reservoirs, *Water Resources Research* 31 (3) (1995) 721–730.

J.B. Loomis, The evolution of a more rigorous approach to benefit transfer: benefit function transfer, *Water Resources Research* 28 (3) (1992) 701–705.

D.I. Matthews, W.G. Hutchinson, R. Scarpa, Testing the stability of the benefit transfer function for discrete choice contingent valuation data, *Journal of Forest Economics* 15 (1-2) (2009) 131-146.

M. Morrison, J. Bennett, Valuing New South Wales Rivers for use in benefit transfer, in: J. Rolfe, and J. Bennett (eds.), Choice Modeling and The Transfer of Environmental Values, Edward Elgar, Cheltenham, UK, 2006.

M. Morrison, and J. Bennett, R. Blamey, J. Louviere, Choice modeling and tests of benefit transfer, *American Journal of Agricultural Economics* 84 (1) (2002) 161-170.

M. Morrison, J. Bennett, R. Blamey, J. Louivere, Choice modeling, non-use values and benefit transfer, *Economic Analysis and Policy* 30 (1) (2000) 13-32.

G.R. Parsons, M.J. Kealy, Benefits transfer in a random utility model of recreation, *Water Resource Research* 30 (8) (1994) 2477-2484.

S. Piper, W.E. Martin, Evaluating the accuracy of the benefit transfer method: a rural water supply application in the USA, *Journal of Environmental Management* 63 (3) (2001) 223-235.

R. Ready, S. Navrud, Morbidity value transfer, in: S. Navrud and R. Ready (eds.), Environmental Value Transfer: Issues and Methods, Springer, Dordrecht, The Netherlands, 2007.

R. Ready, S. Navrud, B. Day, R. Dubourg, F. Machado, S. Mourato, F. Spanninks, M.X.V. Rodriquez, Benefit transfer in Europe: how reliable are transfers between countries?, *Environmental and Resource Economics* 29 (1) (2004) 67–82.

R.S. Rosenberger, J.B. Loomis, Using meta-analysis for benefit transfer: in-sample convergent validity tests of an outdoor recreation database, *Water Resources Research* 36 (4) (2000) 1097-1107.

A. Rozan, Benefit transfer: a comparison of wtp for air quality between France and Germany, *Environmental and Resource Economics* 29 (3) (2004) 295–306.

R. K. Shrestha, John B. Loomis, Meta-analytic benefit transfer of outdoor recreation economic values: testing out-of sample convergent validity, *Environmental and Resources Economics* 25 (1) (2003) 79-100.

R.K. Shrestha, J.B. Loomis, Testing a meta-analysis model for benefit transfer in international outdoor recreation, *Ecological Economics* 39 (1) (2001) 67-83.

R. W. Stapler, Robert J. Johnston, Meta-analysis, benefit transfer, and methodological covariates: implications for transfer error, *Environmental and Resources Economics* 42 (2) (2009) 227-246.

T.P. VandenBerg, G.L. Poe, J.R. Powell, Assessing the accuracy of benefits transfers: evidence from a multi-site contingent valuation study of groundwater quality, in: J.C. Bergstrom, K.J. Boyle, G.L. Poe (eds.), The Economic Value of Water Quality, Edward Elgar, Northampton, MA, 2001.

M. Zandersen, M. Termansen, F.S. Jensen, Evaluating approaches to predict recreation values of new forest sites, *Journal of Forest Economics* 13 (2-3) (2007a) 103-128.

M. Zandersen, M. Termansen, F.S. Jensen, Testing benefits transfer of forest recreation values over a twenty-year time horizon, *Land Economics* 83 (3) (2007b) 412-440.