Systematic Variation in Willingness to Pay for Aquatic Resource Improvements and Implications for Benefit Transfer: A Meta-Analysis

Robert J. Johnston,[1] Elena Y. Besedin,[2] Richard Iovanna,[3] Christopher J. Miller,[4] Ryan F. Wardwell[2] and Matthew H. Ranson[2]

[1]*Department of Agricultural and Resource Economics, University of Connecticut (corresponding author: Robert J. Johnston, Connecticut Sea Grant Office, University of Connecticut at Avery Point, 1080 Shennecossett Road, Groton, CT 06340-6048; phone: (860) 405-9278; fax: (860) 405-9109; e-mail: robert.johnston@uconn.edu).*
[2]*Environment, Trade, and Agriculture Division, Abt Associates, Inc. (phone: (617) 349-2770; e-mail: elena_besedin@abtassoc.com, ryan.wardwell.2006@marshall.usc.edu, matthew_ranson@abtassoc.com).*
[3]*National Center for Environmental Economics, U.S. Environmental Protection Agency (phone: (202) 564-9887; e-mail: iovanna.rich@epamail.epa.gov)..*
[4]*Forest Service, United States Department of Agriculture (phone: (801) 517-1034; e-mail: chrismiller@fs.fed.us).*

Researchers are increasingly considering benefit transfer approaches that allow welfare measures to be adjusted for characteristics of the policy context. The validity and reliability of such adjustments, however, depends on the presence of systematic variation in underlying WTP. This paper describes a meta-analysis conducted to identify systematic components of WTP for aquatic resource improvements. Model results reveal systematic patterns in WTP unapparent from stated preference models considered in isolation, and suggest that observable attributes account for a substantial proportion of the variance in WTP estimates across studies. The analysis also exposes challenges faced in development, estimation, and interpretation of meta-models for benefit transfer and welfare guidance. These challenges remain salient even in cases where the statistical performance of meta-models is satisfactory.

Les chercheurs examinent de plus en plus les méthodes de transfert des valeurs qui permettent le rajustement des mesures du bien-être selon les caractéristiques du contexte politique. Toutefois, la validité et la fiabilité de ce genre de rajustements dépendent de la présence de la variation systématique de la volonté de payer (VDP) sous-jacente. Le présent article décrit une méta-analyse réalisée dans le but d'identifier les composantes systématiques de la VDP pour l'amélioration des ressources aquatiques. Les résultats du modèle ont révélé des profils systématiques de la VDP non apparents dans les modèles de préférence déclarée examinés isolément et ils ont semblé indiquer que les caractéristiques observables ont représenté une partie substantielle de la variance des estimés de la VDP dans les études. Cette analyse a également mis en lumière les difficultés à surmonter quant à l'élaboration, à l'estimation et à l'interprétation des méta-modèles pour le transfert des valeurs et l'orientation du bien-être. Ces difficultés demeurent même dans les cas où la performance statistique des méta-modèles est satisfaisante.

INTRODUCTION

Despite the mixed performance of benefit transfer in past assessments (Smith et al 2002), welfare measures estimated using such methods are increasingly incorporated as central components of benefit–cost analyses (Bergstrom and De Civita 1999). Given the generally unreliable performance of unadjusted single-site transfers, however, researchers are increasingly considering approaches that allow welfare measures to be adjusted for characteristics of the policy context under consideration. For example, U.S. EPA (2000, 87) notes that analysts often "adjust [WTP] point estimates based on judged differences between the study and policy cases."

The validity of willingness to pay (WTP) adjustments and their appropriateness for benefit transfer depends on the presence of systematic, identifiable variation in underlying WTP. If WTP cannot be shown to vary systematically according to attributes distinguishing study and policy sites, the justification for such adjustments—and for benefit transfer in general—becomes more tenuous. The validity of benefit transfer may also be called into question if a large proportion of WTP variation is associated with otherwise unexplained study-level effects, rather than identifiable differences in resource, context, and study design attributes. Nonetheless, transfers are often conducted without assessment of whether welfare measures display systematic variation associated with observable resource, context, and study design attributes, or whether (in contrast) WTP variation is due largely to unobservable, stochastic, or study-specific elements.

Meta-analysis[1] has been drawing particular attention as a potential means to assess systematic variation in WTP (Brouwer 2002; Johnston et al 2003). Such methods have been applied extensively in fields such as epidemiology and education, where applications typically involve studies conducted under controlled conditions with standardized experimental designs (Glass et al 1981; Bateman and Jones 2003). However, because of heterogeneity of research methods in economics and a lack of standard data reporting, meta-analysis is still used sparingly in resource and environmental economics (Button 2002, 83–5).

Within a benefit transfer context, meta-analysis may be applied to identify systematic influences of study, economic, and resource attributes on WTP. Such information may allow researchers to more appropriately adjust WTP estimates, providing an improved mechanism for benefit transfer (Rosenberger and Loomis 2003). Based on this potential, U.S. EPA (2000) guidelines characterize meta-analysis as "the most rigorous benefit transfer exercise" (p. 87). Another advantage of meta-analysis is that it "may . . . provide insights into phenomena for which no current studies exist" (Button 2002, 78). Nonetheless, review of the literature reveals some controversy with regard to the use of meta-models for applied welfare analysis. For example, while many authors (e.g., Poe et al 2001) advise caution in direct policy applications of meta-analysis, others recognize its increasing policy use (Bergstrom and De Civita 1999).

This paper describes a meta-analysis conducted to identify systematic patterns in WTP for aquatic resource improvements. The analysis was initially conducted to explore benefit transfer methods for estimating WTP for fish and related resources affected by U.S. EPA regulations. The broader purpose of the analysis, however, is to assess whether variation in WTP for aquatic resources may be explained sufficiently by systematic variation in policy, context, and other observable attributes to justify attempts at benefit transfer, or whether WTP variation is dominated by unexplained or study-level factors.

A secondary goal is to assess the potential sensitivity of WTP to specification issues that are not determined by characteristics of the policy context, nor may be decided based on theory alone—issues for which researcher judgment is critical in a benefit transfer context.

Results of the analysis are promising with regard to the ability of meta-analysis to synthesize information regarding WTP for aquatic resource improvements and reveal systematic relationships unapparent from individual stated preference studies. Results further suggest that WTP variation for aquatic resource improvements is largely systematic. However, results also expose challenges in the design and estimation of meta-models for welfare and policy analysis.

DATA AND CONCEPTUAL APPROACH

The applicability of meta-analysis to any particular research question is dependent on the quality and comparability of the available data. While a certain degree of heterogeneity in observations is desirable, extensive divergence in study approaches and attributes may cast doubt on the appropriateness and validity of meta-models (Bateman and Jones 2003). Here, the goal of the analysis is to estimate the relative influence of resource, context, and study characteristics on per household WTP for water quality improvements that affect aquatic species.

Given this emphasis, the data are drawn from nonmarket valuation studies that estimate total WTP for water quality changes that affect aquatic life habitats and/or recreational fishing and other recreational uses. From a universe of greater than 300 identified surface water valuation studies addressing such resource types, 34 were found to be suitable for inclusion in the metadata. Specific criteria for inclusion were: (1) a requirement that the study estimate total (use and nonuse) per household WTP, (2) a requirement that the water quality change being valued affect aquatic life or habitat in a water body that provides recreational fishing uses or other recreational activities, (3) a requirement that the study was conducted in the United States, (4) a requirement that the study apply research methods generally accepted by journal literature, and (5) a requirement that the study provide sufficient information regarding resource, context, and study attributes to warrant inclusion in the metadata.

The resulting metadata comprise 81 observations from 34 unique studies conducted between 1973 and 2001. The studies include 18 journal articles, 10 research reports or academic papers,[2] four Ph.D. dissertations, one book, and one Master's thesis. The number of observations exceeds the number of studies because many studies provide more than one estimate of WTP. Multiple WTP estimates from a single study are available due to in-study variations in such factors as the extent of amenity change, elicitation methods applied, water-body type and number, recreational activities affected, and species affected. Because of the requirement that each study estimate total (use and nonuse) WTP, the data are limited to studies relying on stated preference methods; these include open-ended contingent valuation, choice-based survey methods, and combined revealed/stated preference techniques. Table 1 summarizes principal study characteristics for those studies included in the metadata.

Based on theory and findings from the literature, we expect that various attributes may be associated with systematic variations in WTP for water quality improvements

Table 1. Characteristics of surface water valuation studies included in meta-analysis

Citation for study	Number of observations in metadata	State	Water-body type	Species affected	Methodology	Adjusted raw WTP values[a]
Aiken (1985)	1	CO	All freshwater	Game fish	Contingent valuation (CVM)—multiple methods[b]	$167.98
Anderson and Edwards (1986)	1	RI	Salt pond/marshes marshes	Unspecified	Contingent valuation (CVM)—open ended	$157.14
Azevedo et al (2001)	5	IA	Lake	Game fish	CVM—discrete choice	$17.76–118.68
Bockstael et al (1989)	2	MD	Estuary	Unspecified	CVM—discrete choice	$65.80–209.51
Cameron and Huppert (1989)	1	CA	River/stream	Game fish	CVM—discrete choice	$43.07
Carson et al (1994)	2	CA	Estuary	Game fish; multiple categories	CVM—discrete choice	$35.83–67.47
Clonts and Malone (1990)	3	AL	River/stream	Unspecified	CVM—iterative bidding	$68.10–110.85
Croke et al (1987)	9	IL	River/stream	All recreational fish; none	CVM—iterative bidding	$53.31–81.46
Cronin (1982)	4	DC	River/stream	All recreational fish	CVM—open ended	$61.85–212.73
Desvousges et al (1983)	2	PA	River/stream	Unspecified	CVM—discrete choice	$111.41–220.24
De Zoysa (1995)	2	OH	Lake; river and lake	Multiple categories	CVM—discrete choice	$35.88–61.02
Farber and Griner (2000)	3	PA	River/stream	All recreational fish	CVM—discrete choice	$44.22–105.58
Hayes et al (1992)	2	RI	Estuary	Shellfish; none	CVM—discrete choice	$339.72–351.47
Herriges and Shogren (1996)	2	IA	Lake	All recreational fish	CVM—discrete choice	$53.66–180.35
Huang et al (1997)	2	NC	Estuary	Multiple categories	CVM—discrete choice; revealed and stated preference	$221.75–228.07
Kaoru (1993)	1	MA	Salt pond/marshes	Shellfish	CVM—open ended	$190.10
Lant and Roberts (1990)	3	IA/IL	River/stream	Game fish; all recreational fish	CVM—discrete choice	$107.86–134.18

(Continued)

Table 1. Continued

Citation for study	Number of observations in metadata	State	Water-body type	Species affected	Methodology	Adjusted raw WTP values[a]
Loomis (1996)	1	WA	River/stream	Game fish	CVM—discrete choice	$80.93
Lyke (1993)	2	WI	Lake	Game fish	CVM—discrete choice	$51.96–84.99
Magat et al (2000)	2	CO/NC	All freshwater	All aquatic species	CVM—iterative bidding	$114.49–376.61
Matthews et al (1999)	2	MN	River/stream	All aquatic species	CVM—discrete choice	$15.77–22.01
Mitchell and Carson (1981)	1	National	All freshwater	All aquatic species	CVM—discrete choice	$242.34
Olsen et al (1991)	3	Pacific NW	River/stream	Game fish	CVM—open ended	$34.48–107.59
Roberts and Leitch (1997)	1	MN/SD	Lake	Multiple categories	CVM—discrete choice	$7.26
Rowe et al (1985)	1	CO	River/stream	Game fish	CVM—open ended	$117.04
Sanders et al (1990)	4	CO	River/stream	Unspecified	CVM—open ended	$70.44–171.59
Schulze et al (1995)	2	MT	River and lake	Multiple categories	CVM—discrete choice	$15.08–21.16
Stumborg et al (2001)	2	WI	Lake	Multiple categories	CVM—discrete choice	$57.90–88.38
Sutherland and Walsh (1985)	1	MT	River and lake	Unspecified	CVM—open ended	$126.98
Welle (1986)	6	MN	All freshwater	Multiple categories; game fish	Multiple methods	$95.30–207.32
Wey (1990)	2	RI	Salt pond/marshes	Shellfish	Multiple methods	$55.61–200.50
Whitehead and Groothuis (1992)	3	NC	River/stream	All recreational fish	CVM—open ended	$27.74–46.23
Whitehead et al (1995)	2	NC	Estuary	Multiple categories	CVM—iterative bidding	$68.08–97.91
Whittington et al (1994)	1	TX	Estuary	All aquatic species	CVM—discrete choice	$169.32

[a] As noted in the text, reported WTP values apply to different levels of water quality change. All WTP estimates are converted to 2002 dollars and rounded to the nearest cent, and hence may not match exactly the raw WTP estimates reported in source studies. Where multiple WTP estimates are available from a given study, the range of values is presented.

[b] The author averaged WTP estimates derived from both open-ended and iterative bidding methods to obtain a single reported WTP estimate.

(Poe et al 2001; Johnston et al 2003). For ease of exposition, these attributes are categorized into those characterizing (1) study and methodology, (2) surveyed populations, (3) geographic region and scale, (4) water-body type, and (5) resource condition and change. *Study and methodology* attributes characterize such features as the year in which a study was conducted, payment vehicle and elicitation format, WTP estimation methods and conventions, and survey response rates. *Surveyed populations* attributes characterize such features as the average income of respondents and the representation of users and nonusers within the survey sample. *Geographic region and scale* attributes characterize such features as the number of water bodies affected by the policy and the geographic region in which the study was conducted. *Water body type* attributes characterize hydrological characteristics of the affected water body (e.g., river, lake, salt pond, estuary). Finally, *resource condition and change* attributes characterize baseline conditions, resource uses supported, and the extent of water quality change. Table 2 summarizes the set of independent variables included in the meta-analysis.

Although the interpretation and calculation of most independent variables requires little explanation, there are some variables for which additional detail is warranted. These include variables characterizing surface water quality and its measurement. To allow the partial slope associated with water quality changes to vary systematically as a function of the primary affected species group(s), we include water quality in the model as a set of interactions with binary variables characterizing the primary species affected by water quality change, as noted in the original studies. These interaction variables distinguish the effects of water quality change for fish (*WQ_fish*), shellfish (*WQ_shell*), multiple species (*WQ_many*), and nonspecified species (*WQ_non*) (Table 2).

Further explanation is also warranted for methods used to reconcile water quality measures across different studies. Many (26) observations in the metadata characterize quality changes using variants of the Resources for the Future (RFF) water quality ladder (Mitchell and Carson 1989, 342).[3] This scale is linked to specific pollutant levels which, in turn, are linked to presence of aquatic species and suitability for particular recreational uses. Other observations in the metadata, however, rely on ordinal rankings—often paired with verbal descriptions—to measure water quality. To reconcile measurements of water quality change (a prerequisite for this meta-analysis), we map all water quality measures to the RFF water quality ladder.

In most cases, the descriptions of water quality (present in the studies that did not apply the water quality ladder) rendered mapping of water quality measures to the RFF ladder straightforward. For example, studies often defined baseline and subsequent water quality in terms of suitability for recreational activities (e.g., boating, fishing, swimming) or corresponding qualitative water quality measures (e.g., poor, fair, good)—features corresponding to the RFF ladder. For studies in which such information was not provided, we used descriptive information available from studies (e.g., amount/indication of the presence of specific pollutants and historical decline of the quality of the resource) to approximate the baseline level of water quality and the magnitude of the change. However, to account for potential systematic biases involved in mapping those studies that are not based on the RFF water quality ladder, we define the binary variable *wq_ladder*. This variable identifies those studies in which RFF water quality ladder measurements were an original component of the survey instrument.

Table 2. Meta-analysis variables and descriptive statistics

Variable	Description	Units and measurement	Mean (SD)
ln_WTP	Natural log of willingness-to-pay for specified resource improvements. WTP for all studies was converted to 2002 dollars using the U.S. Bureau of Labor Statistics non-seasonally adjusted average CPI for all urban consumers.	Natural log of dollars (Range: 1.98 to 5.93)	4.43 (0.77)
year_indx	Year in which the study was conducted, converted to an index by subtracting 1970.	Year Index (Range: 3 to 31)	18.79 (6.57)
discrete_ch	Binary variable indicating that WTP was estimated using a discrete choice survey instrument.	Binary (Range: 0 or 1)	0.35 (0.37)
voluntary	Binary variable indicating that WTP was estimated using a payment vehicle described as voluntary.	Binary (Range: 0 or 1)	0.07 (0.26)
interview	Binary variable indicating that the survey was conducted through in-person interviews.	Binary (Range: 0 or 1)	0.19 (0.39)
mail	Binary variable indicating that the survey was conducted through the mail.	Binary (Range: 0 or 1)	0.56 (0.50)
lump_sum	Binary variable indicating that payments were to occur on something other than a long-term annual basis (e.g., a single lump sum payment).	Binary (Range: 0 or 1)	0.21 (0.41)
nonparam	Binary variable indicating that WTP was estimated using nonparametric methods.	Binary (Range: 0 or 1)	0.46 (0.50)
wq_change	Change in mean water quality, specified on the RFF water quality ladder. Defined as the difference between baseline and post-improvement quality. Where the original study (survey) did not use the RFF water quality ladder, we mapped water quality descriptions to analogous levels on the RFF ladder to derive water quality change (see text). Note that this variable was only included in the model as part of an interaction term (*WQ_fish, WQ_shell, WQ_many, WQ_non*).	Water quality ladder units (Range: 0.5 to 5.75)	2.42 (1.07)

Table 2. Continued

Variable	Description	Units and measurement	Mean (SD)
lnwq_change	The natural log of *wq_change* (see above).	Range: −0.69 to 1.75	0.77 (0.52)
wq_ladder	Binary variable indicating that the original survey reported resource changes using a standard Resources for the Future water quality ladder.	Binary (Range: 0 or 1)	0.32 (0.47)
protest_bids	Binary variable indicating that protest bids were excluded when estimating WTP.	Binary (Range: 0 or 1)	0.46 (0.50)
outlier_bids	Binary variable indicating that outlier bids were excluded when estimating WTP.	Binary (Range: 0 or 1)	0.22 (0.42)
median_WTP	Binary variable indicating that the study reported median, not mean, WTP.	Binary (Range: 0 or 1)	0.06 (0.24)
hi_response	Binary variable indicating that the survey response rate exceeds 74% (i.e., 75% or above).	Binary (Range: 0 or 1)	0.31 (0.47)
income	Mean income of survey respondents, either as reported by the original survey or calculated based on US Census averages for the original surveyed region.	Dollars (Range: 30396 to 137693)	47,034.10 (12,788.72)
nonusers	Binary variable indicating that the survey is implemented over a population of nonusers (default category for this dummy is a survey of any population that includes users).	Binary (Range: 0 or 1)	0.19 (0.39)
single_river	Binary variable indicating that resource change explicitly takes place over a single river (default is a change in an estuary).	Binary (Range: 0 or 1)	0.24 (0.43)
single_lake	Binary variable indicating that resource change explicitly takes place over a single lake.	Binary (Range: 0 or 1)	0.12 (0.33)
multiple_river	Binary variable indicating that resource change explicitly takes place over multiple rivers.	Binary (Range: 0 or 1)	0.09 (0.28)
salt_pond	Binary variable indicating that resource change explicitly takes place over multiple salt ponds.	Binary (Range: 0 or 1)	0.05 (0.22)

Table 2. Continued

Variable	Description	Units and measurement	Mean (SD)
num_riv_pond	Number of rivers or salt ponds affected by policy when *multiple_river* or *salt_pond*=1 (Only studies addressing rivers and salt ponds specified multiple water bodies.). Specified as the sum of the multiplicative interactions between *multiple_river* and the number of water bodies and that of *salt_pond* and the number of water bodies.	Number of specified rivers or ponds (Range: 0 to 15)	1.40 (3.56)
regional_fresh	Binary variable indicating that resource change explicitly takes place in a fresh waterbody .	Binary (Range: 0 or 1)	0.16 (0.37)
southeast	Binary variable indicating that survey was conducted in the USDA Southeast region (default is Northeast region).	Binary (Range: 0 or 1)	0.12 (0.33)
pacif_mount	Binary variable indicating that survey was conducted in the USDA Pacific/Mountain region.	Binary (Range: 0 or 1)	0.18 (0.40)
plains	Binary variable indicating that survey was conducted in the USDA Northern or Southern Plains region.	Binary (Range: 0 or 1)	0.02 (0.15)
mult_reg	Binary variable indicating that survey included respondents from more than one of the regions.	Binary (Range: 0 or 1)	0.04 (0.19)
WQ_fish	Interaction variable: *wq_change* multiplied by a binary variable identifying studies in which water quality improvements are stated to benefit only fin fish. Default is zero (i.e. water quality change did not affect fish).	Water quality ladder units (Range: 0.5 to 5.75)	1.15 (1.53)
WQ_shell	Interaction variable: *wq_change* multiplied by a binary variable identifying studies in which water quality improvements are stated to benefit only shellfish. Default is zero (i.e. water quality change did not affect shellfish).	Water quality ladder units (Range: 0.5 to 4.00)	0.12 (0.64)
WQ_many	Interaction variable: *wq_change* multiplied by a binary variable identifying studies in which water quality improvements are stated to benefit multiple species types. Default is zero (i.e. water quality change did not affect multiple species).	Water quality ladder units (Range: 0.5 to 4.00)	0.63 (1.20)

Table 2. Continued

Variable	Description	Units and measurement	Mean (SD)
WQ_non	Interaction variable: *wq_change* multiplied by a binary variable identifying studies in which species benefitting from water quality improvements remain unspecified. Default is zero (i.e. water quality change did not affect unspecified species).	Water quality ladder units (Range: 0.5 to 2.5)	0.52 (0.93)
lnWQ_fish	Interaction variable: *lnwq_change* multiplied by a binary variable identifying studies in which water quality improvements are stated to benefit only fin fish. Default is zero (i.e. water quality change did not affect fish).	Range: −0.69 to 1.75	0.37 (0.54)
lnWQ_shell	Interaction variable: *lnwq_change* multiplied by a binary variable identifying studies in which water quality improvements are stated to benefit only shellfish. Default is zero (i.e. water quality change did not affect shellfish).	Range: 0 to 1.39	0.03 (0.22)
lnWQ_many	Interaction variable: *lnwq_change* multiplied by a binary variable identifying studies in which water quality improvements are stated to benefit multiple species types. Default is zero (i.e. water quality change did not affect multiple species).	Range: −0.69 to 1.39	0.19 (0.46)
lnWQ_non	Interaction variable: *lnwq_change* multiplied by a binary variable identifying studies in which species benefitting from water quality improvements remain unspecified. Default is zero (i.e. water quality change did not affect unspecified species).	Range: 0 to 0.92	0.18 (0.33)
nonfish_uses	Binary variable identifying studies in which changes in uses other than fishing are specifically noted in the survey.	Binary (Range: 0 or 1)	0.73 (0.45)
fishplus	Binary variable identifying studies in which a fish population or harvest change of 50% or greater is reported in the survey.	Binary (Range: 0 or 1)	0.12 (0.33)
baseline	Baseline water quality, specified on the RFF water quality ladder.	Water quality ladder units (Range: 0 to 7)	4.60 (2.47)

THE EMPIRICAL MODEL

Past meta-analyses have incorporated a range of statistical methods, with none universally accepted as superior (e.g., Poole and Greenland 1999; Poe et al 2001; Bateman and Jones 2003; Johnston et al 2003). Indeed, the literature provides mixed guidance on several specification and estimation issues, leaving researchers to make sometimes *ad hoc* judgments regarding the most appropriate specification of meta-models. Despite the variation in statistical approaches to meta-analysis, the literature has reached consensus on many fundamental issues. For example, there is general consensus that meta-models must somehow address (or at a minimum, test for) potential correlation among observations provided by like authors or studies and the related potential for heteroskedasticity (Rosenberger and Loomis 2000b; Bateman and Jones 2003).

Here, we follow recent work of Bateman and Jones (2003) and apply a multilevel model to the metadata, to address potential correlation among observations gathered from single studies. Following Poe et al (2001) and Smith and Osborne (1996, 293), we also apply Huber–White robust variance estimation; this "approach treats each study as the equivalent of a sample cluster with the potential for heteroskedasticity... across clusters."

Given this empirical framework, we illustrate three models. In all cases, the dependent variable is the natural log of estimated household WTP for water quality improvements in aquatic habitat. For model one, all right-hand-side variables are linear, resulting in a semi-log functional form common in meta-analysis (e.g., Smith and Osborne 1996; Johnston et al 2003). While linear forms are also common (Rosenberger and Loomis 2000a, 2000b; Poe et al 2001; Bateman and Jones 2003), the semi-log form was chosen on the basis of its statistical performance, ability to capture curvature in the valuation function, and because it allows independent variables to influence WTP in a multiplicative rather than additive manner.

For comparison, two alternative specifications are illustrated. Model two is a trans-log model, identical to the semi-log specification save for the inclusion of water quality measures as natural logarithms. This form—common in the hedonic modeling literature—shares many advantages of the semi-log functional form, but also incorporates the desirable quality that WTP is constrained to zero when quality change is also equal to zero. For both models 1 and 2, weighting of observations is avoided following Bateman and Jones (2003). Model 3 is identical to the semi-log specification, save that observations are weighted following the approach of Poe et al (2001). Weights are defined such that each study is given identical weight in the analysis (i.e., weights on multiple observations within each study sum to 1). Although weighting methods prevent studies providing multiple observations from unduly influencing model estimation, they also imply that such studies are no more informative, overall, than others (Bateman and Jones 2003). Consideration of three model specifications allows at least a preliminary assessment of robustness in statistical results.

MODEL RESULTS

Regression results reveal numerous statistically significant and intuitive patterns that influence WTP for water quality improvements in aquatic habitats (Table 3). In general, the statistical fit of the three estimated equations is good; model results suggest a considerable systematic component of WTP variation. Likelihood ratio tests (Table 3) show that model

Table 3. Results for multilevel models: WTP for aquatic habitat improvements

Variable	Model 1 Semi-log unweighted Parameter estimate (std. error)	Model 2 Trans-log unweighted Parameter estimate (std. error)	Model 3 Semi-log weighted Parameter estimate (std. error)
intercept	6.0043***	6.0782***	6.0232***
	(0.6078)	(0.6813)	(0.4633)
year_indx	−0.1058***	−0.1220***	−0.1201***
	(0.0185)	(0.0152)	(0.0201)
discrete_ch	0.3713	0.7057**	0.4020
	(0.3306)	(0.2726)	(0.2800)
voluntary	−1.6422***	−1.5980***	−1.7320***
	(0.2255)	(0.2410)	(0.1461)
interview	1.3030***	1.3401***	1.2615***
	(0.1700)	(0.1880)	(0.1449)
mail	0.5627***	0.6353***	0.6809***
	(0.1753)	(0.1944)	(0.1906)
lump_sum	0.6180***	0.4826***	0.6878***
	(0.1710)	(0.1606)	(0.1224)
nonparam	−0.4650**	−0.2593*	−0.4057 **
	(0.1756)	(0.1365)	(0.1612)
wq_ladder	−0.3617*	−0.2148	−0.2333*
	(0.1795)	(0.1984)	(0.1321)
protest_bids	0.9390***	1.0556***	0.9464***
	(0.1325)	(0.1255)	(0.1092)
outlier_bids	−0.8814***	−0.8335***	−0.8729***
	(0.1103)	(0.1165)	(0.1041)
median_WTP	0.2193	0.1641	0.1339
	(0.1625)	(0.1609)	(0.1922)

(Continued)

Table 3. Continued

Variable	Model 1 Semi-log unweighted Parameter estimate (std. error)	Model 2 Trans-log unweighted Parameter estimate (std. error)	Model 3 Semi-log weighted Parameter estimate (std. error)
hi_response	−0.8020***	−0.8654***	−0.8246***
	(0.1190)	(0.1280)	(0.0698)
income	3.83×10^{-07}	5.04×10^{-06}	4.59×10^{-06}
	(4.88×10^{-06})	(4.63×10^{-06})	(4.84×10^{-06})
nonusers	−0.5019***	−0.5169***	−0.6215***
	(0.1176)	(0.1245)	(0.1149)
single_river	−0.3236*	−0.3250	−0.3738**
	(0.1791)	(0.2157)	(0.1703)
single_lake	0.2950	0.5420**	0.4062
	(0.2621)	(0.2523)	(0.2648)
multiple_river	−1.6155***	−1.3804***	−1.7595***
	(0.2951)	(0.3036)	(0.2085)
salt_pond	0.7613**	0.5510	0.5252
	(0.3366)	(0.3452)	(0.3231)
num_rivers_ponds	0.0791***	0.0789***	0.0821***
	(0.0094)	(0.0115)	(0.0145)
regional_fresh	−0.0069	0.0901	0.0143
	(0.1642)	(0.1967)	(0.1490)
southeast	1.1396***	1.3434***	1.2807***
	(0.2174)	(0.2379)	(0.1974)
pacif_mount	−0.3080**	−0.3143*	−0.3168***
	(0.1298)	(0.1610)	(0.1047)
plains	−0.7958**	−0.8544***	−0.9292***
	(0.2831)	(0.3058)	(0.2641)
mult_reg	0.6074**	0.5682*	0.7514***
	(0.2490)	(0.3040)	(0.2331)

(Continued)

Table 3. Continued

Variable	Model 1 Semi-log unweighted Parameter estimate (std. error)	Model 2 Trans-log unweighted Parameter estimate (std. error)	Model 3 Semi-log weighted Parameter estimate (std. error)
WQ_fish (*lnWQ_fish* for translog)	0.2095**	0.2274*	0.1726*
	(0.0809)	(0.1210)	(0.0998)
WQ_shell (*lnWQ_shell* for translog)	0.2610**	0.4567**	0.2127*
	(0.0984)	(0.2152)	(0.1109)
WQ_many (*lnWQ_many* for translog)	0.2400**	0.3093	0.2199*
	(0.0977)	(0.1893)	(0.1150)
WQ_non (*lnWQ_non* for translog)	0.4808**	0.6827*	0.4765**
	(0.1947)	(0.3396)	(0.1854)
nonfish_uses	−0.1541	−0.1375	−0.2072*
	(0.1225)	(0.1405)	(0.1111)
fishplus	0.7964***	0.8104***	0.9222***
	(0.1719)	(0.1845)	(0.1649)
baseline	−0.1240***	−0.1290***	−0.1168***
	(0.0407)	(0.0441)	(0.0289)
−2 log-likelihood			
Full model	65.8	70.7	63.2
Intercept and random effects only	167.6	167.6	176.6
−2 log likelihood χ^2	101.8***	96.9***	113.4***
Covariance factors			
Study level (σ_u^2)	7.71×10^{-18}	0.0	1.18×10^{-19}
Residual (σ_e^2)	0.1320	0.1402	0.0421
$\boldsymbol{R^2}$ (see note)	0.77	0.76	0.85
Observations (*N*)	81	81	81

Note: Because σ_u^2 approximates (or is equal to) zero in all cases, unadjusted R^2 estimates here are identical to those obtained from OLS. However, in the general case, R^2 obtained from multilevel or random-effects models is not equivalent to standard OLS R^2, and should not be interpreted equivalently (Statacorp 2001, 439). * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

variables are jointly significant at $p < 0.01$ in all cases. In all models, the majority of independent variables are statistically significant at $p < 0.10$, with most statistically significant at $p < 0.01$. Signs of significant parameter estimates generally correspond with intuition, where prior expectations exist. Considering these factors, the statistical performance of all models compare favorably to prior meta-analyses in the valuation literature.

While all models provide evidence of systematic WTP variation associated with resource, context, and study attributes, random effects associated with systematic study-level variance (σ_u^2) are not statistically significant in any of the estimated models. Indeed, σ_u^2 approximates (or is equal to) zero in all cases. This finding is similar to those of Bateman and Jones (2003) and Johnston et al (2003), and suggests that once one accounts for variation in observable resource, context, and study attributes, no additional systematic variation in WTP may be ascribed to study-level effects. This is a significant finding, as it suggests that systematic variation in WTP is not driven by unobservable attributes unique to particular studies or sets of study authors.

Contrasting Model Specifications

Despite differences in the three presented model specifications, statistical results are similar. In most all cases, coefficient magnitudes and standard errors vary to only a small degree. Measures of equation fit are similar, and all models are significant at $p < 0.01$. Moreover, additional preliminary models—suppressed from Table 3 for the sake of brevity—reveal that the signs and magnitudes of statistically significant parameter estimates are generally robust with regard to modest changes in model variables.[4] Such results mirror those of Johnston et al (2003), whose meta-analysis of use and nonuse WTP for water quality improvements finds a high degree of robustness to changes in model specification. This suggests that meta-analyses may provide useful insights into systematic elements of WTP variation, notwithstanding controversy in the literature regarding appropriate statistical methods.

For purposes of initial discussion, we emphasize results of the semi-log model (model one). The R^2 for this model is 0.77, indicating that approximately 3/4 of the variation in WTP is systematically correlated with model variables. Despite emphasizing results of a single model, we emphasize that—with a few notable exceptions to be discussed later—policy implications of the three model specifications are nearly identical.

Systematic Components of WTP: Resource Attributes

The variables *WQ_fish*, *WQ_shell*, *WQ_many*, and *WQ_non* indicate the effects of water quality improvements associated with gains in fish, shellfish, multiple species, and unspecified habitat, respectively (Table 2). All signs are as expected. The associated coefficients are positive and statistically significant ($p < 0.02$ or better), indicating that higher WTP is associated with larger gains in water quality, as measured on the RFF ladder (Table 3). This is a noteworthy result, as it indicates that WTP—compared systematically across studies—is sensitive to the scope of water quality improvements (cf. Smith and Osborne 1996; Johnston et al 2003).

Results also suggest that WTP for water quality improvement declines as baseline water quality increases. The variable *baseline* represents the baseline water quality from which water quality change would occur. The associated parameter estimate is significant ($p < 0.01$) and of the expected negative sign, revealing diminishing returns to scale for water quality improvements. This finding suggests that WTP across studies is not only systematically sensitive to scope at a broad level (i.e., larger water quality improvements generate larger WTP), but at a more subtle, if no less important, level associated with diminishing marginal returns to scale.

Finally, the variable *fishplus* identifies those studies for which the associated survey identified particularly large gains in fish populations or harvest rates ($>50\%$). The positive and statistically significant result ($p < 0.01$) indicates that particularly large gains in fish populations or harvests are associated with statistically significant increases in total WTP.

Systematic Components of WTP: Geographical and Water Body Type Attributes

Ten binary variables characterize geographic region and scale and water body type; eight are statistically significant at $p < 0.10$. The default category from which these variables allow systematic variations in WTP is an estuarine water body in the Northeast United States. Compared to this baseline, lower WTP is associated with rivers (*single_river*, *multiple_river*), while higher WTP is associated with water quality gains in salt ponds (*salt_pond*). *Single_lake* and *regional_fresh* both have positive values, but neither is statistically significant.

Results further suggest that WTP is sensitive to the number of water bodies under consideration. Of the water body categories distinguished above, both rivers and salt ponds include variation in numbers of affected water bodies explicitly described by the survey. This variation is captured by the variable *num_riv_pond* (Table 2). The associated parameter estimate is statistically significant ($p < 0.01$) and indicates that WTP increases with the number of water bodies considered (Table 3). This result, combined with the statistical significance of the water quality change variables noted above, suggests that WTP values in the metadata are strongly sensitive to scope—both in terms of the number of water bodies and the magnitude of quality change. Such multidimensional scope sensitivity extends findings such as those of Smith and Osborne (1996), which address sensitivity to scope in more limited dimensions.

Finally, the regional indicator variables *southeast*, *pacif_mount*, *plains*, and *mult_reg* are statistically significant at $p < 0.05$ (most at $p < 0.01$), suggesting that there are significant differences among WTP estimates from surveys in different geographical regions of the United States. While such effects may be related to systematic differences in preferences or resource characteristics across regions, they may also be related to otherwise unexplained characteristics of authors, methodology, or other factors that may be correlated with geographical region.[5]

Systematic Components of WTP: Population Attributes

WTP studies often differ with regard to the presence and type of demographic and other variables that characterize sampled populations. Given the wide disparity in the treatment of such factors, meta-analyses in the valuation literature typically include relatively few variables that characterize sampled populations (e.g., Smith and Osborne 1996; Poe et al 2001). Here, only two variables, *nonusers* and *income*, are used to characterize surveyed populations. The variable *nonusers* is of particular relevance. The negative and significant ($p < 0.01$) parameter estimate indicates that surveys of nonusers only—where nonusers by definition have only nonuse values for the resource improvements in question (Freeman

2003, 142)—generate lower WTP values than surveys that include users, who may have both use and nonuse values. Caution must be taken in using such estimates to provide guidance regarding general population nonuse values, however, as nonuser values may underestimate nonuse values of the general population, if nonuse values of users exceed those of nonusers (Whitehead and Blomquist 1991).

Systematic Components of WTP: Study Attributes

A variety of study and methodology effects may be shown to influence WTP for water quality improvements. While not surprising, this does indicate that methodological approach influences WTP, as indicated by prior meta-analyses (e.g., Smith and Osborne 1996; Rosenberger and Loomis 2000a; Brouwer 2002; Johnston et al 2003). Analysis shown later in this paper demonstrates the potential implications of such sensitivity for benefit transfer and WTP estimation. Of 12 variables characterizing study and methodological effects, 10 are statistically significant at $p < 0.10$. Among these is the year in which a study was conducted (*year_indx*), with later studies associated with lower WTP. This is an expected result, as the focus of stated preference survey design over time has often been on the reduction of survey biases that would otherwise result in an overstatement of WTP (Arrow et al 1993).

Model results reveal that voluntary (*voluntary*) payment vehicles (i.e., surveys that describe hypothetical payments as voluntary) are associated with reduced WTP estimates. This result counters common intuition that voluntary payment vehicles may be associated with overstatements of true WTP. The reason for this finding is unknown, but may indicate an unwillingness among respondents to proffer large voluntary payments, given the fear that others will free-ride. Reduced WTP estimates are also associated with studies applying nonparametric methods (*nonparam*). Survey response format (e.g., *discrete_ch*) does not have a statistically significant effect in the model.

Smaller WTP estimates are associated with studies that eliminate or trim outlier bids when estimating WTP (*outlier_bids*; $p < 0.01$). Conversely, increased WTP estimates are associated with studies that seek to eliminate protest bids (*protest_bids*; $p < 0.01$), suggesting a preponderance of zero protest bids. Especially when eliciting values that relate to ecological resources, such as fish species, such bids may be provided by respondents that have preferences structures at variance with consumer choice axioms; they may be essentially unwilling to equate an ecological change with *any* dollar amount (Spash 2000).

Studies with high response rates (*hi_response*; $p < 0.01$) are associated with lower WTP estimates, an expected result associated with limiting avidity bias. In addition, lower WTP is associated with the use of the RFF water quality ladder in the original survey (*wq_ladder*; $p < 0.10$). As is the case with a variety of study design variables, there is no necessary expectation with respect to the direction of this effect. Survey format variables also have an effect on WTP, as might be expected. *Interview* and *mail* both have positive and statistically significant coefficients ($p < 0.01$), compared to the default of telephone surveys.

WTP values for the majority of studies included in the analysis consist of annual payments over an indefinite duration. However, a small number of studies estimate WTP for payments over a short horizon—typically 3–5 years. The variable *lump_sum* identifies studies in which payments were to occur on something other than an indefinite annual basis (Table 2). The positive and statistically significant parameter for *lump_sum* indicates sensitivity to the payment schedule (Stevens et al 1997). Studies that ask respondents to report an annual payment (as opposed to a shorter *lump_sum* payment) have lower nominal WTP estimates ($p < 0.01$).

IMPLICATIONS FOR WTP ESTIMATION AND BENEFIT TRANSFER

Findings summarized above (Table 3) suggest a wide range of robust, systematic, and intuitive patterns influencing WTP for water quality improvements to aquatic habitats. Results suggest that while WTP is sensitive to survey and elicitation methods, it is also systematically influenced by scope in various dimensions, the type of habitat under consideration, the type of population sample (i.e., user vs. nonuser), and other attributes of the resource(s) and region(s) in question. Results further suggest that WTP is not systematically influenced by otherwise unobservable attributes of study authors. Based on such results, one might argue that such meta-analyses can provide useful guidance regarding the general magnitudes of welfare effects within a benefit transfer context—at least with regard to potential WTP adjustments associated with policy, resource, or context effects.

The statistical performance of this particular meta-analysis notwithstanding, however, there are a variety of issues that must be addressed if one seeks to use such results for benefit transfer or applied welfare guidance. Many of these issues may not be appropriately resolved based solely on empirical considerations, and involve such features as implications of functional form, the assignment of levels for study design attributes,[6] and methods used to reconcile environmental quality measures. Such issues remain relevant, even in instances where WTP variation is largely systematic and robust to changes in model specification.

Sensitivity of WTP to Study Methodology and Functional Form

As an example, consider that the literature provides little practical guidance with regard to the choice of functional forms for meta-models used in welfare analysis. Econometric functional forms are most appropriately interpreted as approximations of actual functional relationships. Nonetheless, there may be constraints or patterns imposed by specific functional forms that researchers may find desirable or undesirable in certain contexts.

For example, while many meta-analyses of WTP apply linear or semi-log functional forms, appropriately specified double log or trans-log models have the desirable quality that WTP may be constrained to zero when quality change is also equal to zero. Semi-log and linear specifications do not impose this exogenous—but theoretically attractive—restriction. In addition, constraints imposed on the second-derivatives of estimated WTP by semi-log or linear functions may be undesirable under certain circumstances. Such issues may be of particular relevance in cases where WTP is highly sensitive to functional form, or in which investigators are faced with a choice of one form that may offer superior empirical performance while an alternative form provides desirable theoretical properties.

The literature also provides little guidance with regard to the specification of variables characterizing study methodology, including those characterizing such factors as survey implementation, question formats, payment vehicles, and analytic methods. Here, WTP is sensitive to a wide array of such variables (Table 3). While this does not negatively affect the statistical properties of meta-models—and in fact may be expected—it does lead to questions regarding the most appropriate treatment of these variables for benefit transfers.

To illustrate potential implications of issues such as functional form and variable-level assignment in the present case, we use model results (Table 3) to estimate nonuser WTP associated with increasing levels of *WQ_fish*. Nonuser WTP estimates are calculated for both the semi-log and trans-log models. Given the choice of functional forms, it is expected that variable level assignments for resource and policy context attributes will be largely determined by characteristics of the policy and context for which WTP estimates are desired. However, these characteristics will not, in general, provide insight into the appropriate assignment of levels for variables characterizing study methodology.

For purposes of illustration, levels for policy and context variables are fixed at levels consistent with what might be expected from a regulation promulgated under the U.S. Clean Water Act. To illustrate the potential significance of level assignments for study methodology variables within this context, we calculate nonuser WTP given two different sets of level assignments for these variables. For simplicity, we show potential WTP variation associated with changes in only one set of methodology variables—those characterizing survey administration method (e.g., mail, phone, or in-person). Other methodological variables are set with the goal of providing conservative WTP estimates, subject to consistency with methodological guidance in the literature.[7]

Table 4 shows the four scenarios under which nonuser WTP is illustrated. These include (1) semi-log specification, mail survey; (2) trans-log specification, mail survey; (3) semi-log specification, telephone survey; and (4) trans-log specification, telephone survey. For each scenario, Table 4 illustrates estimated mean nonuser WTP for three different levels of *WQ_fish* (water quality improvements that benefit fish habitat), representing improvements of 0.5, 1.0, and 2.0 units. To further clarify WTP differences, Figure 1 illustrates estimated nonuser WTP for each scenario, as a continuous function of *WQ_fish*. As baseline water quality for WTP illustration is set at 7 on the 10-point RFF ladder, the maximum possible gain in *WQ_fish* is 3.[8]

In general, illustrated patterns in WTP (Figure 1) show little sensitivity to functional form; over most of the data range WTP forecasts are similar. While the choice between semi- and trans-log forms may have little practical consequence for policies involving moderate water quality change (between 0.5 and 2.5), implications are quite evident at the extremes of the data—particularly for very small changes in water quality. While more striking WTP differences occur in a data range for which there are no in-sample observations (the smallest water quality change present in the metadata is 0.5 units), they nonetheless exemplify the need to carefully consider choices affecting the development and application of meta-analysis for benefit transfer.

Central to such choices here is a tradeoff between congruence to accepted theory and model fit. To wit, the trans-log model offers desirable theoretical properties, as noted above. These include the properties that WTP approaches zero as quality change approaches zero, and the negative second derivative of the WTP function with regard to water quality change. In contrast, the semi-log model offers a somewhat improved fit to the data. The meta-analysis literature offers little to assist researchers in choosing among such contrasting specifications.

While choices among functional form influence WTP within certain policy contexts (i.e., small water quality changes), specification of levels for study methodology variables

Table 4. Specification of attribute levels and nonuser WTP forecasts

Variable	Specification 1 semi-log; telephone survey	Specification 2 trans-log; telephone survey	Specification 3 semi-log; mail survey	Specification 4 trans-log; mail survey
intercept	1	1	1	1
year_indx	31	31	31	31
discrete_ch	1	1	1	1
voluntary	0	0	0	0
interview	0	0	0	0
mail	**0**	**0**	**1**	**1**
lump_sum	0	0	0	0
nonparam	0	0	0	0
wq_ladder	0	0	0	0
protest_bids	1	1	1	1
outlier_bids	1	1	1	1
median_WTP	0	0	0	0
hi_response	1	1	1	1
income	53,840	53,840	53,840	53,840
nonusers	1	1	1	1
single_river	0	0	0	0
single_lake	0	0	0	0
multiple_river	0	0	0	0
salt_pond	0	0	0	0
num_rivers_ponds	0	0	0	0
regional_fresh	0	0	0	0
southeast	0	0	0	0
pacif_mount	0	0	0	0
plains	0	0	0	0
mult_reg	0	0	0	0
WQ_fish	0–3	0–3	0–3	0–3
nonfish_uses	0	0	0	0
fishplus	0	0	0	0
baseline	7	7	7	7
Nonuser WTP forecasts (2002 dollars)				
WTP for WQ_fish = 0.5	3.24	3.07	5.70	5.80
WTP for WQ_fish = 1.0	3.60	3.60	6.32	6.79
WTP for WQ_fish = 2.0	4.44	4.21	7.80	7.95

have significant implications for WTP over the entire range of quality change. Figure 1 illustrates substantial shifts in estimated nonuser WTP associated with changes in the method of survey administration, with mail surveys associated with as much as a 76% increase in predicted WTP, compared to the default of a telephone survey. Unlike sensitivity associated with functional form, WTP variation associated with the survey administration method applies over the full range of policy outcomes, with often-substantial implications for WTP.[9]

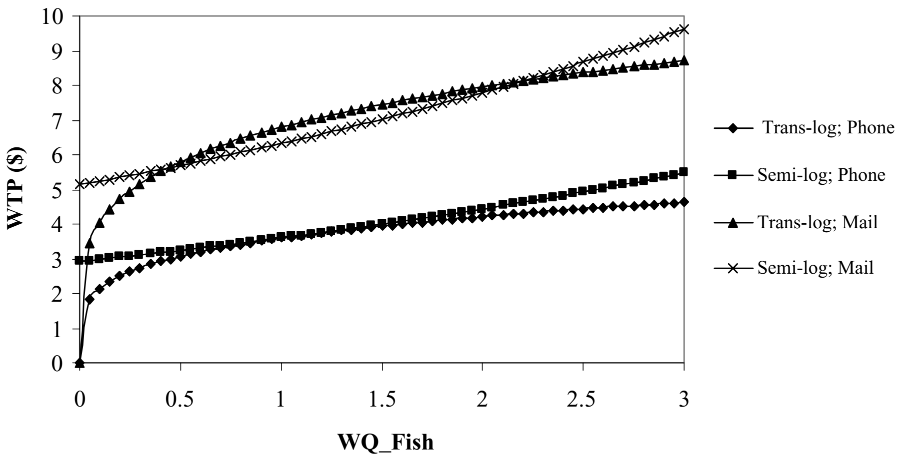

Figure 1. Estimated willingness to pay for improvements in water quality for fish habitat (*WQ_Fish*): four specifications

Researchers may address such sensitivity in a variety of ways. Where possible, one might choose variable levels based on guidance from prior work regarding the appropriateness of particular methodologies within stated preference research (Arrow et al 1993). Where such guidance is lacking, variables might be specified at mean values. A potential advantage of the mean-value approach includes reduced sensitivity to researcher judgment, as variable level assignments are determined by the data. Implications of this approach are illustrated in Table 5, which reports nonuse WTP estimates for a policy scenario analogous to that shown in Table 4, except that all methodological variables are set at mean values (Table 2).

Results of Table 5 parallel those of Table 4, showing similarity in WTP estimates across the two functional forms (semi-log vs. trans-log), but also revealing WTP estimates that vary according to changes in assigned values for methodological variables. Here, WTP forecasts associated with mean values for methodological variables exceeds estimates shown in Table 4, with the increase ranging from less than 1% (compared to the trans-log, mail survey scenario) to greater than 100% (compared to the semi-log, telephone survey scenario). Hence, while the use of mean values for methodological variables represents an alternative, perhaps compelling strategy for variable-level assignments, it does not ameliorate the sensitivity of WTP to such variables.

An alternative approach to the sensitivity of WTP to methodological variables would be to omit such variables from the statistical model(s). That is, variables characterizing study methodology—assuming negligible correlation to other model variables—might be dropped from the model, their influence instead subsumed under random effect terms specified at higher levels in a multilevel model. Statistical tests (e.g., Hausman; likelihood ratio) would be essential in such cases. Here, likelihood ratio ($\chi^2 = 51.17$; df = 11; $p = 0.0001$) and Hausman tests ($\chi^2 = 36.73$; df = 19; $p = 0.009$) performed on preliminary models indicate that such omissions are both statistically significant, and lead to

Table 5. Alternative specification of attribute levels: nonuser WTP forecasts

Variable	Specification 5 semi-log; mean methodology values	Specification 6 trans-log; mean methodology values
intercept	1	1
year_indx	31	31
discrete_ch	**0.35**	**0.35**
voluntary	**0.07**	**0.07**
interview	**0.19**	**0.19**
mail	**0.56**	**0.56**
lump_sum	**0.21**	**0.21**
nonparam	**0.46**	**0.46**
wq_ladder	**0.32**	**0.32**
protest_bids	**0.46**	**0.46**
outlier_bids	**0.22**	**0.22**
median_WTP	**0.06**	**0.06**
hi_response	**0.31**	**0.31**
income	53,840	53,840
nonusers	1	1
single_river	0	0
single_lake	0	0
multiple_river	0	0
salt_pond	0	0
num_rivers_ponds	0	0
regional_fresh	0	0
southeast	0	0
pacif_mount	0	0
plains	0	0
mult_reg	0	0
WQ_fish	0–3	0–3
nonfish_uses	0	0
fishplus	0	0
baseline	7	7
Nonuser WTP forecasts (2002 dollars)		
WTP for WQ_fish = 0.5	6.89	5.83
WTP for WQ_fish = 1.0	7.65	6.82
WTP for WQ_fish = 2.0	9.44	7.99

systematic changes in remaining model parameters, respectively.[10] Hence, in the present case, the omission of methodological variables appears unjustified from a statistical perspective.

The appropriateness and policy implications of such solutions may vary across data sets and policy contexts. Such variation notwithstanding, the potential sensitivity of WTP to variables characterizing study methodology remains a challenge to those seeking to apply meta-analysis for welfare estimation. The literature provides little guidance in the

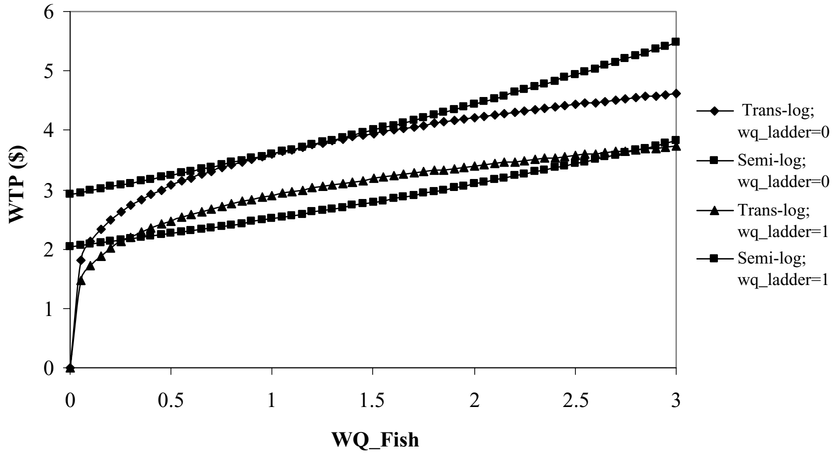

Figure 2. Estimated willingness to pay for improvements in water quality for fish habitat (*WQ_Fish*): *wq_ladder* = 0 versus *wq_ladder* = 1

resolution of such issues, which may affect meta-analytic benefit transfers even in cases where strong and intuitive systematic elements of WTP are demonstrated.

Reconciliation of Quality Measures: Implications for WTP

Similar sensitivity of WTP may be shown to choices made regarding the reconciliation of resource quality measures across studies. Here, differences in methods used to represent water quality change in original studies are captured by *wq_ladder*, which indicates that the original study reports resource changes using a variant of the RFF water quality ladder (Mitchell and Carson 1989, 342). To illustrate potential WTP differences associated with this variable, Figure 2 illustrates nonuser WTP for the cases in which *wq_ladder* = 0 and *wq_ladder* = 1. As above, estimated WTP is illustrated for both the semi-log and trans-log functional forms, as a function of *WQ_fish*. The illustration assumes a telephone survey instrument, with all other variables specified as shown in Table 4.

As shown in Figure 2, nonuser WTP varies substantially according to the specification of *wq_ladder*. At *WQ_fish* = 1, for example, WTP resulting from the semi-log model increases by approximately 43% when *wq_ladder* = 0, compared to *wq_ladder* = 1 (\$3.60 vs. \$2.51). The analogous increase in the trans-log model is approximately 23%. As one might expect, WTP divergence is greater for larger values of *WQ_fish* (Figure 2).

The sensitivity of WTP to variables such as *wq_ladder*—which account for mechanisms used to reconcile quality measures—has obvious implications for benefit transfer. However, the appropriate response to such sensitivity is unclear, and may depend on the suspected rationale for the statistical significance of such variables. For example, if the significance of *wq_ladder* were related to a true reduction in WTP associated with the use of the RFF water quality ladder in survey instruments, then an appropriate action might be to specify the variable at its mean value, unless one has *a priori* reason to believe that lower WTP estimates associated with ladder-using studies are more appropriate. In contrast, if one suspects that the significance of this variable is due to systematic biases involved in mapping water quality measurements from those studies that are not based on the RFF water quality ladder (see discussion of this mapping above), then a more appropriate action might be to set *wq_ladder* = 1, to offset this suspected bias.

In the present case, the rationale for the statistical significance is ambiguous, leading to uncertainty regarding its appropriate treatment in a benefit transfer context. Moreover, as above, the literature provides little guidance regarding appropriate treatment of such variables. This lack of attention notwithstanding, such issues may be encountered in any meta-analysis or benefit transfer in which study methodology (or context) affects the measurement or representation of environmental quality, and in which researcher judgment is required to reconcile quality measurements from different studies (Smith et al 2002). As shown above, ambiguity in the treatment of such effects may have substantial implications for the outcome of applied welfare analysis or benefit transfer.

CONCLUSION

This paper presents a meta-analysis conducted to estimate systematic components of WTP for aquatic resource improvements. Model results are promising with regard to the ability of meta-analysis to identify systematic components of WTP and reveal patterns that may be unapparent from stated preference models considered in isolation. We find intuitive and statistically significant relationships between a range of independent variables and WTP, including findings that indicate strong sensitivity to scope in various dimensions. WTP is shown to be sensitive to such factors as geographical region, sample characteristics, water body type, habitat type, and a variety of study design attributes.

While illustrating that meta-analysis can successfully explain a substantial proportion of the variance in WTP estimates, model results also expose challenges faced in the estimation and interpretation of meta-models for policy analysis. These challenges involve methodological choices faced by researchers, and remain salient even in cases where the statistical performance of meta-models may be exemplary. Researchers commonly face choices involving such factors as functional form, the use of weights in statistical models, and metrics used to reconcile resource quality across studies (Engel 2002; Smith et al 2002). In addition, application of meta-models to policy analysis typically requires professional judgment regarding selection of independent variable values, particularly for variables characterizing study methodology.

Currently, the meta-analysis literature provides minimal guidance regarding such issues, leaving researchers to make often *ad hoc* decisions. However, as meta-analysis and similar methods become more commonly used as central components of benefit–cost analyses, the need for research and guidance on such issues will almost certainly increase. Even given strong systematic variation in WTP, the ability of researchers and policymakers to agree on standard guidance for policy applications of meta-analysis and benefit transfer may have significant implications for the future role of such methods in applied welfare analysis.

NOTES

[1] Glass (1976) characterizes meta-analysis as "the statistical analysis of a large collection of results for individual studies for the purposes of integrating the findings. It provides a rigorous alternative

to the casual, narrative discussion of research studies which is commonly used to make some sense of the rapidly expanding research literature" (p. 3; cited in Poe et al 2001, 138).

[2]In some cases, peer-reviewed journal articles failed to provide sufficient information on study attributes, necessitating a review of more detailed technical reports (from which the journal articles were derived). In such cases, the original reports are referenced as the primary data source.

[3]Additional details on the RFF ladder are provided by McClelland (1974) and Vaughan (1986).

[4]As might be expected, results are somewhat less robust to omission of large and statistically significant groups of variables. However, such large-scale omissions are rarely advisable, given the potential for omitted variables bias. Statistical tests associated with the potential omission of various variable groups are discussed in subsequent sections.

[5]Preliminary tests of models omitting regional variables reveal few subsequent changes in either the sign or statistical significance of parameter estimates for remaining model variables, but some substantial effects on the magnitudes of parameter estimates.

[6]In order to use model results to derive WTP estimates, one must choose variable levels for each independent variable, including those characterizing study methodology.

[7]For example, in correspondence with typical guidance (e.g., Arrow et al 1993) we assume a nonvoluntary payment mechanism ($voluntary = 0$), a discrete choice instrument ($discrete\_ch = 1$), high response rate ($hi\_response = 1$), and elimination of protest and outlier bids ($protest\_bids = 1$; $outlier\_bids = 1$).

[8]This level of baseline water quality (i.e., $baseline = 7.0$) represents the highest level present in the metadata.

[9]There is no way to ascertain which of these values (i.e., those from mail versus telephone surveys) more closely approximates true underlying WTP, as criterion values are unavailable. Interestingly, however, model results indicate that in-person interview surveys—those recommended by the NOAA Blue Ribbon Panel (Arrow et al 1993)—generate higher WTP estimates than either mail or telephone surveys (Table 3).

[10]The Hausman test compares an unrestricted semi-log, unweighted model to an analogous model from which the 11 methodological variables (noted in Table 5) have been omitted. Full results of the restricted model are suppressed for brevity. The unrestricted model is identical to that shown in Table 3, save that Huber–White adjustment is not applied to the covariance matrix.

ACKNOWLEDGMENT

This research was partially funded under U.S. EPA contract no. 68-C-99-239. Opinions belong solely to the authors and do not necessarily reflect the views or policies of U.S. EPA or imply endorsement by the funding agency.

REFERENCES

Aiken, R. A. 1985. Public benefits of environmental protection in Colorado. Masters thesis, Colorado State University.

Anderson, G. D. and S. F. Edwards. 1986. Protecting Rhode Island's coastal salt ponds: An economic assessment of downzoning. *Coastal Zone Management* 14 (1/2): 67–91.

Arrow, K., R. Solow, E. Leamer, P. Portney, R. Rander and H. Schuman. 1993. Report of the NOAA panel on contingent valuation. *Federal Register* 58 (10): 4602–14.

Azevedo, C., J. A. Herriges and C. L. Kling. 2001. Valuing preservation and improvements of water quality in Clear Lake. Center for Agricultural and Rural Development (CARD), Iowa State University, Staff Report 01-SR 94.

Bateman, I. J. and A. P. Jones. 2003. Contrasting conventional with multi-level modeling approaches to meta-analysis: Expectation consistency in U.K. woodland recreation values. *Land Economics* 79 (2): 235–58.
Bergstrom, J. C. and P. De Civita. 1999. Status of benefits transfer in the United States and Canada: A review. *Canadian Journal of Agricultural Economics* 47 (1): 79–87.
Bockstael, N. E., K. E. McConnell and I. E. Strand. 1989. Measuring the benefits of improvements in water quality: The Chesapeake Bay. *Marine Resource Economics* 6 (1): 1–18.
Brouwer, R. 2002. Environmental value transfer: State of the art and future prospects. In *Comparative Environmental Economic Assessment*, edited by P. Nijkamp, R. Florax and P. Willis. Cheltenham, UK: Edward Elgar Publishers.
Button, K. 2002. An evaluation of the potential of meta-analysis. In *Comparative Environmental Economic Assessment*, edited by P. Nijkamp, R. Florax and P. Willis. Cheltenham, UK: Edward Elgar Publishers.
Cameron, T. A. and D. D. Huppert. 1989. OLS versus ML estimation of non-market resource values with payment card interval data. *Journal of Environmental Economics and Management* 17 (3): 230–46.
Carson, R. T., W. M. Hanemann, R. J. Kopp, J. A. Krosnick, R. C. Mitchell, S. Presser, P. A. Ruud and V. K. Smith. 1994. Prospective interim lost use value due to DDT and PCB contamination in the southern California bight. Vol. 2. Report to the National Oceanic and Atmospheric Administration, Produced by Natural Resources Damage Assessment, Inc., LA Jolla, CA.
Clonts, H. A. and J. W. Malone. 1990. Preservation attitudes and consumer surplus in free flowing rivers. In *Social Science and Natural Resource Recreation Management*, edited by J. Vining. Boulder, CO: Westview Press.
Croke, K., R. G. Fabian and G. Brenniman. 1987. Estimating the value of improved water quality in an urban river system. *Journal of Environmental Systems* 16 (1): 13–24.
Cronin, F. J. 1982. *Valuing Nonmarket Goods Through Contingent Markets*. Richland, WA: Pacific Northwest Laboratory, PNL-4255.
Desvousges, W. H., V. K. Smith and M. P. McGivney. 1983. *A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements*. Washington DC: U.S. Environmental Protection Agency, Economic Analysis Division.
De Zoysa, A. D. N. 1995. A benefit evaluation of programs to enhance groundwater quality, surface water quality and wetland habitat in northwest Ohio. Dissertation, Ohio State University.
Engel, S. 2002. Benefit function transfer versus meta-analysis as policy making tools: A comparison. In *Comparative Environmental Economic Assessment*, edited by P. Nijkamp, R. Florax and P. Willis. Cheltenham, UK: Edward Elgar Publishers.
Farber, S. and B. Griner. 2000. Valuing watershed quality improvements using conjoint analysis. *Ecological Economics* 34 (1): 63–76.
Freeman, A. M., III. **2003.** *The Measurement of Environmental and Resource Values: Theory and Methods*. Washington, DC: Resources for the Future.
Glass, G. V. 1976. Primary, secondary, and meta-analysis of research. *Educational Researcher* 5 (10): 3–8.
Glass, G. V., B. Shaw and M. L. Smith. 1981. *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage Publications.
Hayes, K. M., T. J. Tyrrell and G. Anderson. 1992. Estimating the benefits of water quality improvements in the upper Narragansett Bay. *Marine Resource Economics* 7: 75–85.
Herriges, J. A. and J. F. Shogren. 1996. Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management* 30 (1): 112–31.
Huang, J. C., T. C. Haab and J. C. Whitehead. 1997. Willingness to pay for quality improvements: Should revealed and stated preference data be combined? *Journal of Environmental Economics and Management* 34 (3): 240–55.

Johnston, R. J., E. Y. Besedin and R. F. Wardwell. 2003. Modeling relationships between use and nonuse values for surface water quality: A meta-analysis. *Water Resources Research* 39 (12): 1363–72.

Kaoru, Y. 1993. Differentiating use and nonuse values for coastal pond water quality improvements. *Environmental and Resource Economics* 3: 487–94.

Lant, C. L. and R. S. Roberts. 1990. Greenbelts in the cornbelt: Riparian wetlands, intrinsic values, and market failure. *Environment and Planning A* 22 (10): 1375–88.

Loomis, J. B. 1996. How large is the extent of the market for public goods: Evidence from a nation wide contingent valuation survey. *Applied Economics* 28 (7): 779–82.

Lyke, A. J. 1993. Discrete choice models to value changes in environmental quality: A Great Lakes case study. Dissertation, University of Wisconsin-Madison.

Magat, W. A., J. Huber, K. W. Viscusi and J. Bell. 2000. An iterative choice approach to valuing clean lakes, rivers, and streams. *Journal of Risk and Uncertainty* 21 (1): 7–43.

Matthews, L. G., F. R. Homans and K. W. Easter. 1999. Reducing phosphorous pollution in the Minnesota river: How much is it worth? Staff Paper, Department of Applied Economics, University of Minnesota.

McClelland, N. I. 1974. Water quality index application in the Kansas river basin. Report to U.S. EPA Region VII, Kansas City, MO.

Mitchell, R. C. and R. T. Carson. 1981. An Experiment in Determining Willingness to Pay for National Water Quality Improvements. Preliminary Draft of a Report to the U.S. Environmental Protection Agency. Washington DC: Resources for the Future.

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

Olsen, D., J. Richards and R. D. Scott. 1991. Existence and sport values for doubling the size of Columbia river basin salmon and steelhead runs. *Rivers* 2 (1): 44–56.

Poe, G. L., K. J. Boyle and J. C. Bergstrom. 2001. A preliminary meta analysis of contingent values for ground water quality revisited. In *The Economic Value of Water Quality*, edited by J. C. Bergstrom, K. J. Boyle and G. L. Poe. Northampton, MA: Edward Elgar Publishers.

Poole, C. and S. Greenland. 1999. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology* 150 (5): 469–74.

Roberts, L. A. and J. A. Leitch. 1997. Economic valuation of some wetland outputs of Mud Lake. Agricultural Economics Report No. 381, Department of Agricultural Economics, North Dakota Agricultural Experiment Station, North Dakota State University.

Rosenberger, R. S. and J. B. Loomis. 2000a. Panel stratification in meta-analysis of economic studies: An investigation of its effects in the recreation valuation literature. *Journal of Agricultural and Applied Economics* 32 (3): 459–70.

Rosenberger, R. S. and J. B. Loomis. 2000b. Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database. *Water Resources Research* 36 (4): 1097–107.

Rosenberger, R. S. and J. B. Loomis. 2003. Benefit transfer. In *A Primer on Non-Market Valuation*, edited by P. A. Champ, K. J. Boyle and T. C. Brown. Dordrecht, The Netherlands: Kluwer Academic Publishers (pp. 445–82).

Rowe, R. D., W. D. Schulze, B. Hurd and D. Orr. 1985. *Economic Assessment of Damage Related to the Eagle Mine Facility*. Boulder, CO: Energy and Resource Consultants.

Sanders, L. B., R. G. Walsh and J. B. Loomis. 1990. Toward empirical estimation of the total value of protecting rivers. *Water Resources Research* 26 (7): 1345–57.

Schulze, W. D., R. D. Rowe, W. S. Breffle, R. R. Boyce and G. H. McClelland. 1995. *Contingent Valuation of Natural Resource Damages Due to Injuries to the Upper Clark Fork River Basin*. Natural Resource Damage Litigation Program. Boulder, CO: RCG/Hagler Bailly.

Smith, V. K., G. Van Houtven and S. K. Pattanayak. 2002. Benefit transfer via preference calibration: "Prudential algebra" for policy. *Land Economics* 78 (1): 132–52.

Smith, V. K. and L. Osborne. 1996. Do contingent valuation estimates pass the scope test? A meta analysis. *Journal of Environmental Economics and Management* 31 (3): 287–301.

Spash, C. L. 2000. Ecosystems, contingent valuation and ethics: The case of wetland re-creation. *Ecological Economics* 34 (2): 195–215.

StataCorp. 2001. *Stata Reference Manual, Release 7*. College Station, TX: Stata Press.

Stevens, T., N. De Coteau and C. Willis. 1997. Sensitivity of contingent valuation to alternative payment schedules. *Land Economics* 73 (1): 140–8.

Stumborg, B. E., K. A. Baerenklau and R. C. Bishop. 2001. Nonpoint source pollution and present values: A contingent valuation of lake mendota. *Review of Agricultural Economics* 23 (1): 120–32.

Sutherland, R. J. and R. G. Walsh. 1985. Effect of distance on the preservation value of water quality. *Land Economics* 61 (3): 282–91.

U.S. EPA. 2000. *Guidelines for Preparing Economic Analyses (EPA 240-R-00-003)*. Washington, DC: U.S. EPA, Office of the Administrator.

Vaughan, W. J. 1986. The RFF water quality ladder, Appendix B. In *The Use of Contingent Valuation Data for Benefit/Cost Analysis in Water Pollution Control, Final Report*, edited by R. C. Mitchell and R. T. Carson. Washington: Resources for the Future.

Welle, P. G. 1986. Potential economic impacts of acid deposition: A contingent valuation study of Minnesota. Dissertation, University of Wisconsin–Madison.

Wey, K. A. 1990. Social welfare analysis of congestion and water quality of Great Salt Pond, Block Island, Rhode Island. Dissertation, University of Rhode Island.

Whitehead, J. C. and P. A. Groothuis. 1992. Economic benefits of improved water quality: A case study of North Carolina's Tar-Pamlico river. *Rivers* 3: 170–78.

Whitehead, J. C. and G. C. Blomquist. 1991. A link between behavior, information, and existence value. *Leisure Sciences* 13 (2): 97–109.

Whitehead, J. C., G. C. Blomquist, T. J. Hoban and W. B. Clifford. 1995. Assessing the validity and reliability of contingent values: A comparison of on-site users, off-site users, and non-users. *Journal of Environmental Economics and Management* 29 (2): 238–51.

Whittington, D., G. Cassidy, D. Amaral, E. McClelland, H. Wang and C. Poulos. 1994. The economic value of improving the environmental quality of Galveston Bay. Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, GBNEP-38, 6/94.