Environ Resource Econ (2007) 38:331–351
DOI 10.1007/s10640-006-9073-4

Choice experiments, site similarity and benefits transfer

Robert J. Johnston

Received: 18 April 2006 / Accepted: 20 December 2006 / Published online: 13 January 2007

Abstract Choice experiments are designed to account for variations in environmental resources and site characteristics, as well as potential implications of these variations for willingness to pay. As a result, choice experiment results may be well suited for benefits transfer. It is unclear, however, whether the flexibility of choice experiments renders the similarity of study and transfer sites less critical for transfer validity. Drawing from identical choice experiments conducted in different Rhode Island communities, this model assesses the extent to which error in function-based benefits transfer is related to the similarity of communities across a variety of observable dimensions. Results suggest that site similarity, at least across some dimensions, influences the validity of choice experiment benefits transfers. However, the use of some measures of similarity as indicators of transfer error may provide misleading results.

Keywords Benefit transfer · Choice experiment · Willingness to pay · Land use · Choice modeling · Equivalence test

1 Introduction

Generalization error, also referred to as transfer error, may be defined as the error that occurs when benefit estimates from a study site (or combination of sites) are used or adapted to forecast benefits at a policy or transfer site; it is the difference between the transferred and actual, generally unknown, value (Rosenberger and Stanley 2006). The likelihood and magnitude of such errors are critical to both the validity and accuracy of benefits transfer. As a general consensus, generalization errors are assumed to be smaller in cases where transfer and study sites are more

R. J. Johnston (✉)
Connecticut Sea Grant Office,
University of Connecticut at Avery Point,
1080 Shennecossett Rd, Groton, CT 06340-6048, USA
e-mail: robert.johnston@uconn.edu

similar (Boyle and Bergstrom 1992; Desvousges et al. 1992; Kask and Shogren 1994; Rosenberger and Loomis 2001). Indeed, site similarity is often considered a fundamental prerequisite of benefits transfer (Jiang et al. 2005). We denote this consensus the "similarity hypothesis." Transfers conducted over dissimilar sites or contexts—even if addressing willingness to pay (WTP) for otherwise similar resources—are often treated with skepticism (e.g., Bergstrom and De Civita 1999; Rosenberger and Loomis 2003). This skepticism persists despite the ability of some transfer methods (e.g., meta-analysis; function-based transfer) to adjust for attributes of the valuation context, and for some valuation models (e.g., choice experiments) to adjust for multi-attribute distinctions among resources and/or valuation contexts (Rosenberger and Stanley 2006; Johnston et al. 2005).

Site similarity may be discussed in terms of affected populations, the valuation context (including the availability of substitutes and complements to studied attributes), and/or geographical proximity. A significant literature demonstrates the importance of associated measures of site similarity for transfer error (e.g., Barton 2002; Loomis 1992; VandenBerg et al. 2001; Piper and Martin 2001; Rosenberger and Loomis 2001). Transfers reported in this work, however, rely almost exclusively on methodologies unable to adjust for differences between transfer and policy sites, including contingent valuation and travel cost methods (Morrison and Bennett 2004). In contrast, only a relatively small number of peer-reviewed studies (e.g., Hanley et al. 2006a,b; Jiang et al. 2005; Morrison et al. 2002; Morrison and Bennett 2004; Van Bueren and Bennett 2004) have assessed the performance of benefits transfer using choice experiments—a methodology designed to account for variations in environmental resources and site characteristics, as well as potential implications of these variations for willingness to pay (Morrison and Bergland 2006).

Although stochastic elements inherent in empirical welfare estimation always lead to the potential for measurement error to influence transfer validity (Rosenberger and Stanley 2006), the ability of choice experiments to account for inter-site variations may render the resulting benefit functions highly suitable for transfer.[1] It is unclear, however, whether this flexibility renders the similarity of study and transfer sites less critical for transfer validity. While a variety of authors in the choice experiment benefits transfer literature note the potential impact of policy context on transfer errors (e.g., Hanley et al. 2006a,b; Morrison et al. 2002; Morrison and Bennett 2004), studies assessing transfer validity across sites provide limited and sometimes conflicting insights regarding the role of site similarity.[2] Moreover, no published work formally assesses the potential influence of specific metrics of site similarity on choice experiment transfer error and validity. For example, Morrison and Bennett (2004, p. 610) indicate that implicit prices for specific attributes of river health improvement, estimated using choice experiments, should only be transferred "between similar rivers," where similarity refers to the likeness of rivers in terms of such elements as general location (northern versus southern) and watershed type (inland versus

[1] To the extent that choice experiments allow welfare estimates to differ systematically according to population attributes (e.g., Morrison et al. 2002), such methods may also adjust for differences among populations at study and transfer sites. However, in this regard, choice experiments are less distinct from other valuation methodologies which may also incorporate elements allowing welfare estimates to vary across population groups or individuals.

[2] A number of other studies contrast values held by different population groups for resource changes at identical sites (Morrison and Bergland 2006). Such analyses, however, provide little insight for the validity of transfers across different sites.

coastal). However, this conclusion is based on the rejection of equivalent implicit prices across the majority of watersheds in the sample, not a systematic assessment of the impact of similarity in specific watershed characteristics.[3] In contrast, Hanley et al. (2006b) "firmly reject" the equivalence of implicit prices for river improvements, even for rivers that are ex ante judged to be very similar. Other assessments of choice experiment benefits transfer (e.g., Jiang et al. 2005) draw no systematic conclusions regarding the role of site similarity.

The difficulty in reaching general conclusions regarding the role of site similarity for choice experiment benefit transfer is often compounded by a lack of formality in definitions of similarity, and by variation in the vectors of elements over which similarity may be measured. Hanley et al. (2006b), for example, discuss a variety of attributes over which rivers and watersheds might be considered similar or dissimilar. Existing studies are also commonly limited by the lack of a "control case" in which at least two of the sites in question are nearly identical in all regards—leading some, for example, to speculate as to whether the unanticipated rejection of equivalent implicit prices might be due to unexpected differences between sites judged ex ante to be similar (e.g., Hanley et al. 2006b).

This paper seeks to provide a more systematic perspective on transfer error and site similarity, at least for a single case study. Drawing from identical choice experiments conducted in distinct Rhode Island communities, the model assesses the extent to which transfer error in function-based benefits transfer is related to the similarity of communities across three quantifiable dimensions. Of particular emphasis is similarity with regard to attributes that might reflect the relative availability of substitutes or complements in different communities, here denoted policy context similarity (Bergstrom and De Civita 1999). The model also distinguishes between context similarity and geographical proximity.[4] Results suggest that context similarity, at least across some indicators, influences the validity of choice experiment benefits transfer. However, reliance on other measures of likeness as indicators of the potential performance of benefits transfer, including geographical proximity alone, may provide misleading results.

2 Choice experiments and benefits transfer

The suitability of choice experiments for benefits transfer is discussed by Morrison et al. (2002), Morrison and Bennett (2004) and Jiang et al. (2005) among others; these

[3] The rejection of equivalent implicit prices, and Morrison and Bennett's (2004) conclusion regarding site similarity, applies primarily to their in-watershed samples. A subsequent pooled model suggests that implicit prices may vary according to such attributes as "northern" versus "southern" watersheds, and "inland" versus "coastal". These results suggest that such elements as geographical proximity and context similarity may influence the validity of transfer, although the incorporated attributes (e.g., inland versus coastal) appear to confound potential effects of context and proximity, and more specific tests addressing the role of site similarity were not conducted.

[4] Although past studies (and an anonymous reviewer) have also stressed the potentially important role of similarity in population attributes and size (e.g., Barton 2002; Morrison et al. 2002; Morrison and Bennett 2004), such issues are not a primary focus of this analysis. An anonymous reviewer also emphasizes that similarity may be assessed in terms of the specific attributes whose values are being estimated. However, the literature generally assesses transfer error though the application of otherwise identical choice experiments in different regions and/or populations, thereby holding the definition of choice experiment attributes constant.

discussions are only summarized here. Choice experiments ask respondents to evaluate alternative goods or programs (often including a status quo option) that may differ across a variety of attributes, and choose the option that offers the greatest satisfaction or utility (Adamowicz et al. 1998).[5] The framework forces respondents to acknowledge and react to tradeoffs among attributes, including money cost. As a result, unlike contingent valuation—which typically estimates values for a single or very small number of policy or good configurations—choice experiments generate an empirical estimate of a valuation or utility function. This function typically allows analysts to estimate utility theoretic values for a wide range of policy or environmental good outcomes, and assess how these values change when policy configurations are altered. This property of choice experiments renders them highly suitable for benefits transfer, at least in theory. Simply put, the ability of choice experiments to explicitly adjust for differences in the attributes of environmental goods or policies provides an increased capacity to adjust for differences between study and policy sites—thereby improving the potential accuracy of benefits transfer (Morrison et al. 2002; Jiang et al. 2005).

As a practical matter, however, choice experiments cannot account for all possible attributes that distinguish study and policy contexts. Assessments of choice experiment benefits transfer provide promising, but not universally positive results (Morrison and Bergland 2006).[6] Statistically significant transfer errors remain common, particularly for compensating surplus (Morrison et al. 2002; Jiang et al. 2005; Hanley et al. 2006a,b). Moreover, the literature provides little information regarding general conditions under which choice experiment transfers are likely to incorporate substantial errors. This raises the question as to whether choice experiment transfer validity is improved in cases in which study and transfer sites are more similar.

3 Methods and conceptual approach

The data are drawn from the *Rhode Island Rural Land Use Survey*, a choice experiment survey designed to assess rural residents' tradeoffs among attributes of community-level development and conservation.[7] Respondents from four Rhode Island rural communities (Burrillville, Exeter, West Greenwich, and Coventry) were asked to consider alternative, multiattribute development options for hypothetical tracts of forested land located in their local town. Attributes of choice options characterized land use features and amenities identified as important by focus groups and interviews with growth management practitioners.

The four sampled communities provide exemplars of contexts with varying degrees of likeness. Table 1 illustrates demographic, geographical and development characteristics of the four communities. Two of the communities (Exeter and West Greenwich)

[5] Those interested in discussions of contemporary issues in stated preference modeling are referred to a recent special issue of *Environmental and Resource Economics* (Adamowicz and DeShazo 2006).

[6] Considering the performance of transfers across sites (the case addressed here), Morrison and Bergland (2006) find a relatively high level of transfer validity across studies in the literature. As a generalization, most studies (e.g., Morrison et al. 2002; Morrison and Bennett 2004; Van Bueren and Bennett 2004) find statistical equivalence in at least half of implicit prices tested (for an exception, see Hanley et al. 2006b). Compensating surplus measures, where tested, are less likely to be found equivalent (Morrison and Bergland 2006).

[7] Additional details of the choice experiment are provided by Johnston et al. (2003a,b).

Table 1 Demographic and land use indicators for sampled communities

	Burrillville	Coventry	Exeter	West Greenwich
Population	15,796	33,668	6,045	5,085
Population density (persons/sq. mile)	284	566	105	100
Housing units	5,821	13,059	2,196	1,809
Housing density (units / sq. mile)	104.77	219.33	38.05	35.73
Agricultural and Forest land density (acres / sq. mile)	465.39	391.89	490.59	482.09
Mean family income	58,979	60,315	74,157	71,332
Distances to sampled communities[a]	Coventry: 13.25 W. Greenwich: 17.50 Exeter: 21.25	Burrillville: 13.25 W. Greenwich: 0 Exeter: 3.75	Burrillville: 21.25 W. Greenwich: 0 Coventry: 3.75	Burrillville: 17.50 Coventry: 0 Exeter: 0

[a] Linear distances between closest geographic points, rounded to the nearest quarter mile
Source: US Census data reported by the RI Economic Development Corporation (2006) and RIGIS (2000) land use data

are neighboring communities that share a long border and are highly similar over a range of attributes relevant to land development and conservation choices. Indeed, many consider these communities to be virtually identical, distinguished only by arbitrary lines on a map. Population and housing densities are nearly identical across the two communities, and both have similar numbers of housing units and mean family incomes (Table 1). These communities provide a control case of sites that are nearly identical in almost all attributes related to land use.

A third community (Coventry) borders West Greenwich to the north, and hence is geographically proximate to the first two communities. However, the population and development characteristics of this community differ markedly from both Exeter and West Greenwich (Table 1). The population is density in Coventry, for example, is nearly six times greater than that in West Greenwich—a difference that may be highly relevant for residents' land use preferences. The density of open space and agricultural land (in acres per square mile) is also lower than that in all other communities. The fourth community (Burrillville) is located in the far northwestern corner of Rhode Island, and is hence geographically separated from the first three communities. However, with regard to population and development attributes, it provides a middle-ground between the rural communities of Exeter and West Greenwich and the more developed town of Coventry.

The selection of communities allows for an assessment of two distinct types of similarity that often remain undistinguished in the literature. The first is *geographical proximity* of the study and policy site across geographical space. The second is *context similarity*, as characterized above. In the present case, the choice experiment survey addressed development and conservation issues. Hence, as an indicator of context similarity one might consider the likeness of communities with respect to potential substitutes and complements to goods offered by land use policies (e.g., indicators of development and open space). The four communities provide examples of both: communities that are geographically similar yet divergent in land use indicators, and communities that are geographically distant yet more similar in terms of land use.

Methods follow established statistical approaches (cf. Morrison et al. 2002; Morrison and Bennett 2004; Jiang et al. 2005; Hanley et al. 2006a,b). Hypothesis tests address differences in estimated preference functions (i.e., model coefficients), scale parameters (Swait and Louviere 1993), implicit prices of policy attributes, and compensating surplus estimates for a sample of development and conservation policies. Finally, we quantify correlations among quantified indicators of site similarity and the magnitude of error in implicit prices.

To conduct these analyses, the data are systematically split such that individual choice models may be estimated for each of the four communities, together with a set of pooled models that imposes identical preference structures across community pairs. Given four distinct communities, this provides six different pair-wise contrasts, allowing tests of transfer validity across each pair. A contrast of these six models provides information necessary to assess the role of site similarity on the validity of function-based benefits transfer between communities, as well as to identify those aspects of similarity that appear most relevant.

For example, if geographical proximity is an important indicator of potential transfer error, one would expect to see substantial and statistically significant transfer errors involving the community of Burrillville, a community spatially distant from the other three contiguous communities. If, in contrast, context similarity is a more important indicator of transfer validity, one might expect to see more substantial transfer errors between highly developed Coventry and its much less developed neighbors—Exeter and West Greenwich—with greater similarity between Coventry and Burrillville. Clear differences between values estimated in Coventry and Burrillville, and between these two communities and Exeter/West Greenwich, would suggest that both geographical and context similarity are relevant. Finally, significant differences between results for Exeter and West Greenwich would suggest that transfers are likely to involve substantial error even between sites that are nearly identical.

3.1 The random utility model

Survey responses are analyzed using a standard random utility model. To model a respondent's choice, we define a utility function that includes attributes of a development or conservation plan and the net cost of the plan to the respondent (Hanemann 1984; McConnell 1990):

$$U(\cdot) = U(\mathbf{X}_c, Y - F_c) = v(\mathbf{X}_c, Y - F_c) + \varepsilon_c \tag{1}$$

where

$\mathbf{X}_c$ = a vector of variables describing attributes of development or conservation plan c;

Y = disposable income of the respondent.

F_c = the change in mandatory taxes paid by the respondent under plan c;

$v(\cdot)$ = a function representing the empirically measurable component of utility;

ε_c = econometric error.

If one compares Plan A ($c = A$) to Plan B ($c = B$), the difference in utility (dU) may be modeled as

$$\begin{aligned} dU &= U(\mathbf{X}_A, Y - F_A) - U(\mathbf{X}_B, Y - F_B) = [v(\mathbf{X}_A, Y - F_A) - v(\mathbf{X}_B, Y - F_B)] \\ &\quad -[\varepsilon_B - \varepsilon_A] \\ &= dv - \theta \end{aligned} \tag{2}$$

The model assumes a respondent assesses the difference between utility under the two plans and indicates the sign of dU by either choosing Plan A ($dU > 0$) or Plan B ($dU < 0$). If θ is assumed to have a logistic distribution then the familiar logit model applies, in which the probability of selecting a given option is a logistic function of the utility difference dv (Maddala 1983).

Although the literature offers no firm guidance regarding the choice of specific functional forms for dv, in practice linear forms are often used. Hence,

$$dv = v(\mathbf{X}_\mathrm{A}, F_A) - v(\mathbf{X}_\mathrm{B}, F_B) = \boldsymbol{\beta}_\mathrm{x}(\mathbf{X}_\mathrm{A} - \mathbf{X}_\mathrm{B}) + \beta_f(F_B - F_A), \quad (3)$$

where $\boldsymbol{\beta}_\mathrm{x}$ is a conforming vector of coefficients associated with the vector of attribute differences $(\mathbf{X}_\mathrm{A} - \mathbf{X}_\mathrm{B})$, and β_f is a scalar coefficient associated with the tax difference $(F_B - F_A)$. The parameter vector $\boldsymbol{\beta}_\mathrm{x}$ may be interpreted as the marginal utility of development or conservation attributes, while β_f represents the marginal utility of income.

Six models are estimated—one for each possible pair of communities. Each model pools data from the relevant community pair, but allows systematic variations in slope and intercept coefficients. Formally, this approach redefines dv in (3) to provide a separate utility estimate for respondents in each community. We define a binary variable D_j to equal one for respondents from community j, and zero for those from the second community $i \neq j$. We then estimate a simple extension of (3) allowing for systematically varying slopes,

$$dv = \boldsymbol{\beta}_\mathrm{x}(\mathbf{X}_\mathrm{A} - \mathbf{X}_\mathrm{B}) + \beta_f(F_B - F_A) + \boldsymbol{\beta}_\mathrm{xj}D_j(\mathbf{X}_\mathrm{A} - \mathbf{X}_\mathrm{B}) + \beta_{fj}D_j(F_B - F_A) \quad (4)$$

where $\boldsymbol{\beta}_\mathrm{x}$ and β_f represent marginal utility parameters for respondents from community i, and the conforming sums $(\boldsymbol{\beta}_\mathrm{x} + \boldsymbol{\beta}_\mathrm{xj})$ and $(\beta_f + \beta_{fj})$ represent marginal utilities for respondents from community j. When testing for the equivalence of model parameters, we use established methods (e.g., Allison 1999; Swait and Louviere 1993) to account for the potential effect of the scale parameter (or heteroskedasticity in the residual variance) on coefficient estimates. Given convergence difficulties associated with mixed logit models for some pooled community datasets, results are based on fixed parameters logit estimation, following prior examples in the literature (e.g., Jiang et al. 2005; Morrison and Bennett 2004; Morrison et al. 2002).[8]

4 The survey

Survey development required approximately eighteen months and involved background research, interviews with policy makers and stakeholders, and a large number

[8] Prior attempts at estimating random parameters specifications for these data failed to converge for pooled models including the community of Exeter. Various attempts at model specification and likelihood maximization failed to result in convergence. Given these convergence difficulties, various alternative, preliminary models were used to verify the appropriateness of the iid (independent and identically distributed errors) assumption of the fixed parameters logit model. These models provide no evidence of statistically significant iid violation, supporting the use of the simpler fixed parameters logit approach. For example, preliminary random effects logit models failed to reject the null hypotheses of zero panel-level variance components for individual respondents, with p-values ranging from 0.18 to 0.43 across different models. Random effects models involving observations from Exeter failed to estimate panel-level variance components, mirroring convergence difficulties in random parameters specifications. Similarly, preliminary heteroskedastic probit models failed to identify any statistically significant relationship between error variances and respondent characteristics including age, income and gender.

of focus groups. Individual and group pretests ensured that survey language and format could be easily understood by respondents, and that respondents shared consistent interpretations of survey scenarios (cf. Johnston et al. 1995). Each choice experiment scenario presented respondents with a choice between two development options for land parcels in their community—a current development plan (CDP) and an alternate development plan (ADP)—where each plan could differ across a set of land use (development and conservation) attributes. Attributes distinguishing management plans were chosen based on focus groups and interviews, and characterized such features as protected open space, residential development, unprotected undeveloped land, scenic views, wildlife habitat, and household cost (taxes). Table 2 characterizes attributes distinguishing management plans. Additional details of choice question design are provided by Johnston et al. (2002).

Prior to presenting choice questions, the survey provided background information on community land use and tradeoffs implicit in development choices. Instructions and choice questions were then presented. Each respondent considered three pairs of current and alternate plans for the same 400 acre undeveloped site (one binary CDP versus ADP choice for each question). Respondents were instructed to consider each pair independent of previous choices, and to assume that all choices applied to the same parcel. Respondents were told that "if you do not vote for either plan, development will automatically occur as shown by the current development plan," thereby specifying the status quo that would occur if no choice were made. This framework was chosen to mimic actual community considerations of development proposals, wherein a landowner possesses the property rights necessary to permit development. However, officials may seek to influence development configuration, delaying permits unless changes are made. As a result, officials may exert some control over development form, but cannot reject all development options (Johnston et al. 2003b).

A fractional factorial design was used to construct survey questions with an orthogonal array of attribute levels, with identical experimental designs across the four communities.[9] Attributes were free to vary over their full range for both the current and alternate plans, with no imposed ordering of attribute levels between the two plans. This resulted in 128 unique contingent choice questions divided among 43 different survey booklets. Surveys were mailed to 4000 randomly selected residents of the four Rhode Island towns during March–May 2000 (1000 surveys per town), following the total survey design method (Dillman 2000). Of 3702 deliverable surveys, 2157 were returned, providing 6062 (94% of the potential 6471) complete and usable responses to choice questions. The number of completed surveys per town ranged from 505 in Coventry to 580 in West Greenwich, with response rates varying from 53 to 61% across communities.

5 Assessments of convergent validity and transfer error

This section highlights methods used to evaluate transfer errors between communities. The focus on benefits transfer implies a comparison of results across communities (convergent validity) rather than detailed individual results for each community. As a basis for initial comparison, however, Table 3 presents individual results for each of

[9] The statistical design was conducted by Don Anderson of STATdesign, Inc.

Table 2 Model variables: definitions and summary statistics

Variable name	Description	Units and measurement[a]	Mean (Std. Dev.)
adj_open	The difference between acres of open space adjacent to developments and roads in the CDP and ADP.	Acres in CDP minus acres in ADP. (Range: −200–200)	−3.41967 (95.091)
iso_open	The difference between acres of open space not adjacent to developments and roads in the CDP and ADP.	Acres in CDP minus acres in ADP. (Range: −200–200)	2.62028 (53.724)
size_dif	The difference between acres of residential development in the CDP and ADP.	Acres in CDP minus acres in ADP. (Range: −200–200)	−1.77646 (90.806)
dense_dif	The difference in housing density in developments illustrated the CDP and ADP.	Houses/acre in CDP minus houses/acre in ADP. (Range: −2–2)	−0.00666 (0.9759)
lg_mammal	Difference between habitat quality for large mammals in CDP and that in ADP.	Difference in wildlife habitat quality scale (1 = worst; 5 = best).	0.00370 (1.2193)
sm_mammal	Difference between habitat quality for small mammals in CDP and that in ADP.	Difference in wildlife habitat quality scale (1 = worst; 5 = best).	−0.01628 (1.2194)
com_bird	Difference between habitat quality for common birds in CDP and that in ADP.	Difference in wildlife habitat quality scale (1 = worst; 5 = best).	0.05107 (1.7511)
uncom_bird	Difference between habitat quality for uncommon birds in CDP and that in ADP.	Difference in wildlife habitat quality scale (1 = worst; 5 = best).	0.00370 (1.7038)
wet_sp	Difference between habitat quality for wetland species in CDP and that in ADP.	Difference in wildlife habitat quality scale (1 = worst; 5 = best).	−0.04663 (1.7359)
tax_dif	Difference in additional annual taxes and fees between CDP and ADP (resulting from the management plan).	Dollars in CDP minus dollars in ADP. (Range: −$325–$325)	−1.22132 (154.33)
lowvis	Difference between dummy variables indicating the presence of development either highly screened or not visible from the main road; in the CDP and ADP. Survey versions included eight different photographs characterizing different development visibility levels; four of these photographs are characterized as low visibility development.	Difference between dummy variables for CDP and ADP.	−0.00740 (0.6928)

[a] CDP = Current Development Plan; ADP = Alternate Development Plan

Table 3 Choice model (logit) results: independent community models

Variable	Parameter estimates (std. error)			
	Burrillville	Coventry	Exeter	West Greenwich
intercept	−0.0272 (0.0627)	0.0192 (0.0662)	−0.1600(0.0641)**	−0.1266 (0.0599)**
dense_dif	−0.7253 (0.0705)***	−0.6806 (0.0747)***	−0.7404 (0.0722)***	−0.6260 (0.0658)***
size_dif	−0.0094 (0.0007)***	−0.0059 (0.0008)***	−0.0099 (0.0008)***	−0.0089 (0.0007)***
iso_open	0.0052 (0.0015)***	0.0031 (0.0014)**	0.0034 (0.0015)**	0.0046 (0.0013)***
adj_open	0.0028 (0.0008)***	0.0044 (0.0009)***	0.0051 (0.0009)***	0.0055 (0.0008)***
lowvis	0.0830 (0.0886)	0.2047 (0.920)**	0.1749 (0.0877)**	0.2075 (0.0834)**
lg_mammal	0.1337 (0.0524)**	0.0821 (0.0546)	0.1223 (0.0517)**	0.0827 (0.0492)*
sm_mammal	−0.0573 (0.0511)	0.0329 (0.0532)	−0.0664 (0.0511)	0.0263 (0.0483)
com_bird	0.0893 (0.0363)**	0.0681 (0.0379)*	0.1348 (0.0369)***	0.1082 (0.0350)***
uncom_bird	−0.0023 (0.0360)	0.0474 (0.0377)	0.0134 (0.0359)	0.0552 (0.0344)
wet_sp	0.0449 (0.0372)	0.0081 (0.0382)	0.0849 (0.0379)**	0.0476 (0.0353)
tax_dif	−0.0044 (0.0004)***	−0.0061 (0.0005)***	−0.0051 (0.0004)***	−0.0052 (0.0004)***
Likelihood Ratio χ^2	438.10 ($p < 0.0001$)	393.20 ($p < 0.0001$)	502.92 ($p < 0.0001$)	505.37 ($p < 0.0001$)
Obs (N)	1431	1297	1453	1593

* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$

the four communities, based on the random utility model outlined above. All models are statistically significant at $p < 0.01$. In all cases, the majority of variables are statistically significant, with the sign of significant variables matching prior intuition. Here, we emphasize only the general similarity of results across communities, subject to more rigorous subsequent testing. Primary results are not based on these independent models, however, but rather on models that pool community data pair-wise to test hypotheses relevant to benefits transfer. The six pooled models used for convergent validity testing follow (4) above, and are suppressed for the sake of conciseness.

Based on the pooled models for each community pair, a variety of tests relevant to the validity of benefits transfer are conducted. Following Jiang et al. (2005) and Morrison et al. (2002), we first test for differences in estimated utility parameters across community pairs. This is followed by tests of implicit prices for selected attributes, using both traditional hypothesis tests and equivalence tests (Kristofersson and Navrud 2005). Finally, we test the convergent validity of compensating surplus measures for a set of illustrative policy changes. This implies six sets of hypothesis tests associated with unique community pairs, for each of the four categories noted above (i.e., model parameters, implicit prices, implicit price equivalence tests, and compensating surpluses).

As shown by Swait and Louviere (1993), the confounding effect of the scale parameter (or residual variance) on coefficient estimates requires that tests of the equivalence of utility function parameters across community pairs be decomposed into two parts. Specifying the vector of coefficient estimates (associated with the vector of model variables) as $\boldsymbol{\beta}$ and the scale parameter as μ, the first test is of hypothesis H_{1A}, that $\boldsymbol{\beta_i} = \boldsymbol{\beta_j} = \boldsymbol{\beta}$ for communities $i \neq j$, allowing μ to vary across communities.[10] If we reject H_{1A}, then we may conclude that the parameters of the utility function differ across communities i and j. If we fail to reject H_{1A}, we continue to test hypothesis

[10] That is, $\boldsymbol{\beta} = [\boldsymbol{\beta_x}\ \beta_f]$ from (4).

Table 4 Hypothesis test results: equivalence of estimated coefficients allowing residual variance to vary across community samples ($\boldsymbol{\beta}_i = \boldsymbol{\beta}_j = \boldsymbol{\beta}$)

	Burrillville	Coventry	Exeter
Coventry	$\chi^2 = 27.11$ $p < 0.01$	–	–
Exeter	$\chi^2 = 6.42$ $p = 0.70$	$\chi^2 = 22.99$ $p < 0.01$	–
West Greenwich	$\chi^2 = 12.77$ $p = 0.17$	$\chi^2 = 14.05$ $p = 0.12$	$\chi^2 = 5.91$ $p = 0.75$

Table 5 Hypothesis test results: equivalence of residual variance across community samples ($\mu_i = \mu_j = \mu$)

	Burrillville	Coventry	Exeter
Coventry	$\chi^2 = 0.12$ $p = 0.73$	–	–
Exeter	$\chi^2 = 1.24$ $p = 0.27$	$\chi^2 = 1.89$ $p = 0.17$	–
West Greenwich	$\chi^2 = 0.16$ $p = 0.69$	$\chi^2 = 0.37$ $p = 0.54$	$\chi^2 = 0.72$ $p = 0.39$

H_{1B}, that $\mu_i = \mu_j = \mu$. Rejection of H_{1B} also implies rejection of the equivalence of utility parameters.

Models required for the testing of H_{1A} and H_{1B} are estimated following Allison (1999). Table 4 summarizes results for hypothesis H_{1A}. Results show only two instances in which we reject the null hypothesis (the Coventry–Exeter and Burrillville–Coventry pairs); in four out of six instances we fail to reject the null hypothesis that $\boldsymbol{\beta_i} = \boldsymbol{\beta_j} = \boldsymbol{\beta}$, if μ is free to vary across communities. Table 5 summarizes results for hypothesis H_{1B}. Wald χ^2 tests (Allison 1999) universally fail to reject the null hypothesis of equal residual variances across community pairs. Combining results for H_{1A} and H_{1B}, we fail to reject parameter equality for four out of six community pairs. This is an encouraging result for benefits transfer, and stands in contrast to results such as those of Jiang et al. (2005) and Hanley et al. (2006b), which show broad rejection of parameter equality across groups.

The second set of tests involves the equivalence of implicit prices, or WTP for marginal changes in individual attributes. It is well known that the implicit price for the kth attribute, assuming a linear approximation for utility, is given by $-\beta_k/\beta_{tax\_dif}$, where β_k is the parameter on the kth attribute, and β_{tax_dif} is the parameter on the household cost of the program (i.e., change in mandatory taxes). As above, results are drawn from models pooled pair-wise that allow parameter estimates to vary systematically across communities.

To test the convergent validity of implicit prices, each implicit price difference is parametrically bootstrapped directly from model results, following Krinsky and Robb (1986), with p-values estimated from the empirical distribution following the percentile method (Efron and Tibshirani 1993). This approach avoids the potential for misleading statistical inferences related to inappropriate normality assumptions or the use of non-overlapping confidence intervals (Poe et al. 2005). For illustration, we test the equivalence of implicit prices for four attributes that are highly significant in all four community models (Table 3) and are also of primary focus in land use policy: open space isolated from developments (*iso_open*), open space adjacent

Table 6 Hypothesis test results for implicit price differences: open space and development attributes[a,b]

	Burrillville	Coventry	Exeter
Coventry			
iso_open	0.677 (0.453)	–	–
adj_open	−0.073 (0.253)	–	–
size_dif	−1.150 (0.274)***	–	–
dense_dif	−53.05 (23.290)**		
Exeter			
iso_open	0.514 (0.469)	−0.163 (0.388)	–
adj_open	−0.353 (0.266)	−0.281 (0.240)	–
size_dif	−0.177 (0.314)	0.974 (0.251)***	–
dense_dif	−19.48 (26.556)	33.58 (21.210)*	
West Greenwich			
iso_open	0.299 (0.466)	−0.379 (0.364)	−0.216 (0.403)
adj_open	−0.423 (0.267)	−0.350 (0.224)	−0.070 (0.246)
size_dif	−0.400 (0.298)	0.750 (0.223)***	−0.223 (0.265)
dense_dif	−43.33 (23.893)*	9.72 (19.653)	−23.86 (22.201)

[a] Implicit price differences are denominated in dollars and are calculated as the implicit price calculated for the community named in the row subtracted from the implicit price for the community named in the column, for each attribute. For example, results suggest that the implicit price of *iso_open* in Burrillville is $0.677 greater than that in Coventry. In all cases, implicit prices for *iso_open* and *adj_open* are positive, and for *size_dif* are negative. As the implicit prices of *size_dif* and *dense_dif* are negative, results must be interpreted accordingly

[b] Numbers in parentheses are standard errors, calculated using the parametric bootstrap of Krinsky and Robb (1986) with 10,000 draws. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

to developments (*adj_open*), development size (*size_dif*), and development housing density (*dense_dif*) (Table 2).[11,12] Hypothesis tests are conducted for each community pair. Results are shown in Table 6, along with point estimates of implicit price differences.

In addition to traditional hypothesis tests presented in Table 6, we also present an alternative set of hypothesis tests based on the method of equivalence testing (Kristofersson and Navrud 2005; Muthke and Holm-Mueller 2004; Hanley et al. 2006a). Equivalence testing reverses traditional null and alternative hypotheses and the burden of proof. More specifically, implicit prices are assumed *different* unless hypothesis tests can demonstrate, with a chosen probability level, that the difference is smaller than a specified tolerance limit Δ within which values are considered equivalent (Kristofersson and Navrud 2005). For benefit transfer, Δ represents the maximum difference (transfer error) permitted while nonetheless maintaining the equivalence of predicted and "true" WTP estimated at the policy site (Muthke and Holm-Mueller 2004). For example, Muthke and Holm-Mueller (2004) test the equivalence of benefit

[11] In all cases, marginal WTP estimates for *iso_open* and *adj_open* are positive, and marginal WTP estimates for *size_dif* and *dense_dif* are negative. This is the expected result associated with positive preferences for preserved open space and negative preferences for developed acres.

[12] Hypothesis tests are not conducted for attributes that show uneven statistical significance across communities, to avoid difficulties noted in Jiang et al. (2005), specifically, a failure to reject the null hypothesis of implicit price equality due primarily to large standard errors associated with such attributes. Attributes with uneven statistical significance across communities include indicators of wildlife habitat quality for various species types (*lg_mammal*, *sm_mammal*, *com_bird*, *uncom_bird*, *wet_sp*), and a dummy variable indicating the presence of low visibility developments (*lowvis*).

Table 7 Equivalence test results and percentage differences in implicit prices: open space and development attributes[a]

	Burrillville (%)	Coventry (%)	Exeter (%)
Coventry			
iso_open	57.61	–	–
adj_open	−11.50	–	–
size_dif	54.48	–	–
dense_dif	32.40		
Mean Absolute Value Difference	39.00		
Exeter			
iso_open	43.73	−32.73	–
adj_open	−55.60	−39.55	–
size_dif	8.36*	−101.33	–
dense_dif	11.89*	−30.33	
Mean Absolute Value Difference	29.90	50.99	
West Greenwich			
iso_open	25.41	−75.96	−32.57
adj_open	−66.55	−49.37	−7.04*
size_dif	18.94*	−78.09	11.55*
dense_dif	26.46	−8.78*	16.54*
Mean Absolute Value Difference	34.34	53.05	16.93
Absolute value of percentage transfer errors: means by community[b]			
Burrillville	34.41		
Coventry	47.68		
Exeter	32.61		
West Greenwich	34.77		

[a] Percentage differences are calculated as the difference in implicit price for a given attribute between community pairs, divided by the baseline implicit price for the community in the column. For example, the implicit price of *iso_open* in Burrillville is $0.677 greater than that in Coventry (table 6), a 57.61% error compared to baseline WTP of $1.18 per acre in Burrillville. Equivalence tests are based on the two one-sided t-test (TOST) method, with implicit price differences and estimated standard errors (required to calculate test statistics) from table 6. Mean absolute value differences are not subject to equivalence testing

[b] Calculated as the mean of the absolute value of all percentage differences in implicit prices (*iso_open*, *adj_open*, *size_dif*, *dense_dif*), between the noted community and all other communities. This value is not subject to equivalence testing

* Equivalence test rejects the null hypothesis that implicit price difference exceeds 40% limit of tolerance, at $p < 0.10$

estimates for water quality improvements at tolerance limits of 20, 40, and 60% transfer error, at a significance level of $p < 0.05$. Here, we use the standard two one-sided *t*-test (TOST) method detailed by Schuirmann (1987) and Stegner et al. (1996) to test the null hypothesis that the difference in implicit prices across community pairs is greater than 40%; this is the median tolerance limit applied by Muthke and Holm-Mueller (2004). A significance level of $p < 0.10$ is chosen for the test.[13] Test results are shown in Table 7, along with associated percentage differences in implicit price estimates.

As an additional quantitative assessment of relationships between transfer error and community similarity, we also estimate the correlation between differences in implicit prices for land use attributes and differences in quantifiable indicators of

[13] For pharmaceutical research, the typical tolerance limit is 20%, with a p-value of 0.05 (Kristofersson and Navrud 2005).

Table 8 Correlations across community pairs: community indicators and differences in implicit prices

Attributes	Pearson correlation coefficients		
	Difference in Housing Density and Difference in Implicit Prices	Difference in Agricultural and Forest Density and Difference in Implicit Prices	Geographical Distance and Difference in Implicit Prices (absolute value)
iso_open (acres open space isolated from developments)	−0.8234	0.7078	0.5418
adj_open (acres open space adjacent to developments)	−0.3703	0.6350	0.3593
size_dif (acres in housing developments)	0.8144	−0.9741	−0.2384
dense_dif (housing density in developments)	0.7201	−0.9057	0.4053
Mean Absolute Value of Correlation Coefficients[a]	0.6821	0.8057	0.3862

[a] Calculated as the mean of the absolute value of Pearson correlation coefficients for *iso_open*, *adj_open*, *size_dif* and *dense_dif*, over each metric of community similarity

existing community land use. Housing density and agricultural and forest land density (see Table 1) are used as illustrative indicators of community land use. Pearson correlations are estimated between differences in these indicators and differences in implicit prices, across community pairs. As above, we consider the implicit prices for *iso_open*, *adj_open*, *size_dif* and *dense_dif* (the same attributes in Tables 6 and 7). Correlation coefficients are also estimated between the absolute values of these implicit price differences and geographical distances (Table 1) between each community pair, measured as linear distances (in miles) between the closest geographical points. The latter analysis provides a measure of correlation between similarity in implicit prices for land use attributes and a quantifiable indicator of geographical proximity. Results are shown in Table 8.

The final assessment of community similarity and transfer error tests the convergent validity of compensating surplus measures for selected land use policies. While compensating surplus measures are among those used most commonly for cost benefit analysis, and hence assessment of transfer error in these welfare measures can be among the most relevant for applied benefit transfer, such assessments are often complicated by the large number of policies that can be evaluated using results from most choice experiments (Morrison et al. 2002; Jiang et al. 2005). To address this potential complication, we follow Morrison et al. (2002) and select a relatively small number of illustrative policy alternatives for which to assess compensating surplus transfer error.

Compensating surplus is estimated for three potential policy alternatives, following the standard approach summarized by Boxall et al. (1996). In all cases, surplus is estimated relative to a baseline of no development or conservation activity. The first policy alternative for which compensating surplus is calculated includes the development of a new 40 acre, 20 house site (i.e., *size_dif* = 40, *dense_dif* = 0.5). The second alternative preserves 40 acres of open space isolated from developments (i.e., *iso_open* = 40), with no development. The third alternative combines the changes in the first two alternatives (i.e., *size_dif* = 40, *dense_dif* = 0.5, *iso_open* = 40). For each policy alternative, the surplus difference is calculated for each community pair. Using the same process applied to implicit prices above, compensating surplus differences are parametrically bootstrapped following Krinsky and Robb (1986), with p-values (for the null hypothesis of zero difference) estimated directly from the resulting empirical distribution using the percentile method (Efron and Tibshirani 1993).

6 Implications for site similarity and transfer error

The most immediate implication of model results is the relatively high degree of convergent validity of choice experiment results across communities, at least from a standard statistical perspective. Standard hypothesis test results are generally supportive of the use of choice experiment results from study communities to forecast marginal utilities, implicit prices, and compensating surplus measures in other Rhode Island rural communities. As expected, equivalence tests are less supportive of transfer validity, but still support transfer in some cases. The primary focus of this assessment, however, is on the importance of site similarity for transfer validity. Here, the analysis supports the similarity hypothesis; transfers are more appropriate between communities that are more similar. The analysis also suggests that similarity in terms of the policy context—here the similarity of land use attributes—is more critical than geographical proximity.[14]

6.1 Utility parameters

With respect to the equivalence of utility parameters, we fail to reject the null hypothesis of equal parameter estimates in four of the six community pairs. The two instances in which we reject the null hypothesis (i.e., find evidence of differences in the utility function) both involve the community of Coventry—a community distinguished by a much greater population and housing density than any of the other communities sampled (Table 1). A third hypothesis test involving this community (Coventry–West Greenwich) narrowly misses rejection of the null hypothesis of parameter equality ($p = 0.12$; Table 5).

In contrast, parameter estimates for Exeter and West Greenwich—neighboring communities that are nearly identical from a land use perspective—show a high degree of correspondence, with no evidence of significant differences. Geographical distance alone, however, has less apparent impact, with utility parameters for the least spatially proximate community (Burrillville) differing only from those of Coventry. Moreover, of the three neighboring communities (Exeter, West Greenwich, and Coventry), parameter estimates for Coventry are more likely to differ from those of other communities—regardless of proximity.

6.2 Implicit prices

Hypothesis test results (Table 6) show a high degree of statistical correspondence in implicit prices across community pairs. Of 24 individual WTP differences, only six are shown to be statistically significant at $p < 0.10$ or better. Hence, results here are generally supportive of the convergent validity of implicit prices across sites. This corresponds to similar findings of Morrison et al. (2002), Van Bueren and Bennett (2004), and Jiang et al. (2005) in the choice experiment benefit transfer literature, but diverges from findings of Hanley et al. (2006b).

[14] Barton (2002) finds similar results for the transferability of contingent valuation WTP estimates.

With regard to the impact of community similarity, results for implicit prices mirror those for utility parameters shown above. Five of the six statistically significant differences (out of 24 tested) involve marginal WTP to avoid additional acres or density of housing developments in the more heavily-developed community of Coventry (i.e., the implicit prices of *size_dif* and *dense_dif*). Results suggest that Coventry residents are willing to pay less to prevent increases in housing acres and density.[15] Statistical equivalence of implicit prices, however, cannot always be shown to be related to context similarity; in some cases statistically valid transfers can be conducted across sites that differ widely. We often fail to reject the null hypothesis of equivalent implicit prices, even for more divergent communities in the sample (a positive finding for benefit transfer). For example, marginal WTP for open space preservation (*iso_open*, *adj_open*) cannot be shown to differ across communities, at least from a statistical perspective. Such findings suggest that reliance on a small number of indicators of site context similarity (e.g., housing density, agricultural and forest land density, geographical distance) may not always provide appropriate guidance regarding the statistical equivalence of implicit prices.

As expected based on past findings (e.g., Hanley et al. 2006a; Kristofersson and Navrud 2005) equivalence test results differ from those of traditional hypothesis tests, with most tests (18 out of 24) failing to reject the null hypothesis that implicit prices differ across communities (Table 7). That is, equivalence tests are less supportive of transfer validity. These differences notwithstanding, test results suggest the general, but not universal importance of site similarity for valid transfer. We reject the null hypothesis of different implicit prices between Exeter and West Greenwich for three out of four land use attributes (*adj_open*, *size_dif*, *dense_dif*)—a result that supports widespread equivalence of implicit prices across these two similar communities. However, we also reject the null hypothesis for housing density (*dense_dif*) between the much less similar communities of Coventry and West Greenwich, and for housing density and development size (*size_dif*) between communities of Burrillville and Exeter. These latter results suggest that implicit prices can sometimes be equivalent, despite differences in policy context.

Point estimate percentages of implicit price differences again suggest the importance of site similarity for transfer error. Percentage differences in implicit prices for identical attributes, between different communities vary from 7.04 to 101.33% in absolute value, with an average value of 37.37%. Coventry has the largest average percentage transfer error across all implicit prices, with an average absolute value error of 47.68%, compared to errors ranging from 32.61 to 34.77% in other communities (Table 7). Geographical proximity alone plays a less clear role, however, with average (absolute value) implicit price transfer errors involving the more distant community of Burrillville (34.41%) similar to those involving both Exeter (32.61%) and West Greenwich (34.77%). Further supporting the similarity hypothesis, the two most similar communities in terms of both context and proximity—Exeter and West Greenwich—show the highest degree of similarity in implicit prices. The absolute value of implicit price differences between these two communities averages only 16.93%, smaller than any other community pair in the sample (Table 7). In contrast,

[15] This result is intuitive if viewed from the perspective of the substantial amount of preexisting housing stock. Compared to other communities in the sample, marginal increases in housing acreage or density in Coventry are more trivial relative to the existing baseline—and hence residents are willing to pay less to avoid such changes.

the analogous percentage difference between the also neighboring but more distinct communities of Coventry and West Greenwich is 53.05%.[16]

Pearson correlation coefficients (Table 8) still further support the intuition that context similarity is related to transfer error; results suggest a high degree of correlation between differences in indicators of community land use and differences in implicit prices associated with land use attributes. Such patterns suggest that communities that are more similar in terms of land use are likely to have implicit prices (for land use attributes) that comport more closely. Correlation coefficients between housing density differences and differences in estimated implicit prices range in absolute value from 0.37 to 0.82, with a mean absolute value of 0.68. Particularly strong correlations hold for implicit prices associated with housing acres (0.81), housing density (0.72) and open space acres isolated from open space (−0.82).[17] For example, the correlation coefficient associated with *iso_open* (−0.82) indicates that greater positive differences in community housing density are associated with greater negative differences in the implicit price of open space isolated from developments.

Still larger correlation coefficients are found between differences in agricultural and forest land density (acres/square mile) and differences in implicit prices (Table 8), with the mean absolute value of correlation coefficients greater than 0.80. Strong negative correlations are found between open land density and implicit prices for housing acres (−0.97) and housing density (−0.91), with somewhat smaller but still strong positive correlations with implicit prices for isolated (0.71) and adjacent (0.64) open space. These results indicate that residents of communities that are more similar in terms of the density of open land (agricultural and forest) also tend to have more similar implicit prices for land use attributes.[18] In contrast, Pearson correlations show less clear evidence that geographical proximity influences transfer error, with three out of the four estimated coefficients smaller than 0.5 in absolute value, and a mean absolute value just over 0.38. This supports prior findings in assessments of non-choice experiment benefit transfer that geographical proximity alone is insufficient to guarantee validity (VandenBerg et al. 2001; Piper and Martin 2001).

Overall, results provide relatively strong evidence of the importance of site similarity in the transferability of implicit prices. Indicators of policy context similarity, however, appear to be more relevant than geographical proximity alone. While similarity in agricultural and forest land density is on average most closely related to transfer error for the four attributes of interest (Table 8), similarity in housing density also shows strong correlations with transfer errors.

---

16 While the magnitude of an "acceptable" transfer error is typically determined by decision-makers in the context of specific policy contexts (Shrestha and Loomis 2003), average errors found in the present case are fairly modest compared to past findings in the benefit transfer literature (Rosenberger and Stanley 2006). Moreover, a reviewer emphasizes that one would expect some point-estimate divergence in implicit price estimates simply due to the inherent randomness in primary econometric estimates.

17 Marginal WTP estimates for *size_dif* and *dense_dif* are negative; respondents are willing to pay to *prevent* increases in housing acres and density. Results must be interpreted accordingly.

18 Results shown in table 8 also indicate that residents of communities with greater quantities of agricultural and forest land tend to have higher marginal WTP for additional open space acres. These findings are consistent with the idea of Tiebout sorting among Rhode Island communities based on open space amenities (Tiebout 1956), and counter the standard assumption that residents of communities with greater scarcity of certain attributes (e.g., open space acres) should be willing to pay greater amounts to preserve these attributes.

Table 9 Compensating surplus differences: open space preservation and development policies[a,b]

	Burrillville	Coventry	Exeter	West Greenwich
Coventry				
size_dif = 40 *dense_dif* = 0.5	−81.81 (27.26)***	–	–	–
iso_open = 40	17.82 (23.71)	–	–	–
size_dif = 40 *dense_dif* = 0.5 *iso_open* = 40	−54.71 (28.40)*	–	–	–
Exeter				
size_dif = 40 *dense_dif* = 0.5	8.22 (30.71)	90.03 (25.17)***	–	–
iso_open = 40	45.60 (24.99)*	27.78 (21.23)	–	–
size_dif = 40 *dense_dif* = 0.5 *iso_open* = 40	28.80 (31.21)	83.51 (26.27)***	–	–
West Greenwich				
size_dif = 40 *dense_dif*= 0.5	−19.45 (28.75)	62.35 (22.78)***	−27.68 (26.84)	–
iso_open= 40	30.15 (24.24)	12.35 (20.31)	−15.44 (22.03)	–
size_dif= 40 *dense_dif* = 0.5 *iso_open* = 40	−7.50 (29.47)	47.21 (24.21)**	−36.30 (27.75)	–

[a] Compensating surplus differences are denominated in dollars and are calculated as the compensating surplus for the community named in the row subtracted from the compensating surplus for the community named in the column. Alternative specific constants (the intercepts) are included in compensating surplus calculations

[b] Numbers in parentheses are standard errors. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

6.3 Compensating surplus

Hypothesis test results (Table 9) show a high degree of statistical correspondence in surplus measures across communities. Of 24 illustrated differences in compensating surplus, only seven are shown to be statistically significant at $p < 0.10$ or better. Of these, six involve pairs with the more heavily-developed community of Coventry, and incorporate some type of housing development (i.e., *size_dif* > 0, *dense_dif* > 0). For the most similar communities (Exeter and West Greenwich), no statistically significant differences may be found. Geographical proximity alone, however, has no clear impact on the significance of compensating surplus transfer errors between communities, with four of seven statistically significant differences occurring between Coventry and the closely proximate communities of Exeter and West Greenwich. All of these results parallel findings for implicit prices reported above. In summary, compensating surplus estimates are generally transferable across communities, and statistically significant differences are more likely between communities with more divergent policy contexts.

Interestingly, parallelism in results for implicit prices and compensating surplus is not the norm in the choice experiment benefit transfer literature. For example, both Morrison et al. (2002) and Jiang et al. (2005) find a high degree of equivalence in implicit prices across sites, but find a fairly low degree of correspondence in compensating surplus. Here, both implicit prices and compensating surplus show similar degrees of transfer validity (Tables 6 and 9) and reveal equivalent findings for the importance of site similarity.

7 Conclusions

Model results suggest that standard guidance regarding the importance of site similarity for transfer error is largely justified. Reliance on choice experiments for benefits transfer does not invalidate the similarity hypothesis. Transfer errors between communities are smaller and less likely to be statistically significant in cases where environmental characteristics—here proxied by housing density and agricultural and forest land density—are more similar. Residents of communities that are more similar across land use attributes are more likely to have similar WTP for land use policies, and hence WTP measures that may be transferred with greater confidence. For example, results across the neighboring and nearly identical communities of Exeter and West Greenwich differed to the smallest degree, with traditional hypothesis tests showing no sign of significant differences in utility parameters, implicit prices, or compensating surplus estimates, and equivalence tests rejecting the null hypothesis of divergent implicit prices for three out of the four attributes tested.

Results also suggest, however, that relationships between context similarity and transfer validity may be more complex than is often assumed (cf. Hanley et al. 2006b). Here, the importance of context similarity for transfer error appears to vary across attributes. For example, similarity appears critical for the transferability of implicit prices and compensating surplus measures involving development attributes (*size_dif*, *dense_dif*), but appears somewhat less important for open space attributes (*iso_open*, *adj_open*). In addition, while context similarity influences the validity of function-based transfers, it is also possible to conduct some statistically valid transfers across sites that are relatively dissimilar. Finally, geographical proximity does not appear sufficient to justify benefits transfer when other attributes of the valuation context are not comparable; for communities in close proximity but highly dissimilar in terms of land use attributes, the traditional hypothesis tests are more likely to reject the convergent validity of implicit prices.

The apparent complexity of relationships between metrics of site similarity and transfer validity may be related to factors not addressed in the present analysis. For example, while simple demographic and community structure indicators may not always have clear influences on welfare estimates (Kline and Wichelns 1998), latent factors such as attitudes (Jiang et al. 2005), shared experiences (VandenBerg 2001; Rosenberger and Stanley 2006), and population attributes (Barton 2002; Morrison et al. 2002; Morrison and Bennett 2004) have been shown to influence welfare differences across sites. Nonetheless, even in works that adjust for population, demographic, and/or attitudinal characteristics, a substantial proportion of transfer error sometimes remains (particularly in compensating surplus measures), in some cases even across relatively similar sites (e.g., Jiang et al. 2005; Morrison et al. 2002). Although not formally assessed here, potential interactions between site and population similarity in choice experiment benefit transfer, and the potential benefits of adjusting for measurable differences in associated metrics, represent a potentially significant area for future work.

Results, of course, must be viewed within the context of the small sample of original studies (4) from which they are drawn. It should also be emphasized that these communities—all similar size communities in a single northeastern US state—are somewhat more similar overall than typical study and policy sites between which benefits transfer is conducted. Hence, it is perhaps not surprising that the convergent validity of welfare measures was largely supported by traditional hypothesis tests.[19] Also, there are a number of quantifiable vectors across which sites may differ, only a small number of which are addressed here. These caveats notwithstanding, model results suggest that common intuition regarding the role of site similarity in benefit transfer is for the most part appropriate, and holds for function-based, choice experiment benefit transfers as it does for simpler, fixed-value transfers.

Acknowledgements This research was funded by the USDA Fund for Rural America #97-36200-5219. Opinions belong solely to the author and do not imply endorsement by the funding agency. The author thanks Stephen K. Swallow and Dana Marie Bauer for assistance in the original research effort. Any remaining errors are those of the author.

References

Adamowicz W, Boxall P, Williams M, Louviere J (1998) Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. Am J Agr Econ 80(1):64–75

Adamowicz W, DeShazo JR (2006) Frontiers in stated preferences methods: an introduction. Environ Resour Econ 34(1):1–6

Allison PD (1999) Comparing logit and probit coefficients across groups. Sociolo Method Res 28(2):186–208

Barton DN (2002) The transferability of benefit transfer: contingent valuation of water quality improvements in Costa Rica. Ecol Econ 42(1–2):147–164

Bergstrom JC, De Civita P (1999) Status of benefits transfer in the United States and Canada: a review. Can J Agr Econ 47(1):79–87

Boxall PC, Adamowicz WL, Swait J, Williams M, Louviere J (1996) A comparison of stated preference methods for environmental valuation. Ecol Econ 18(3):243–253

Boyle KJ, Bergstrom JC (1992) Benefit transfer studies: myths, pragmatism, and idealism. Water Resour Res 28(3):657–663

Desvousges WH, Naughton MC, Parsons GR (1992) Benefit transfer: conceptual problems in estimating water quality benefits using existing studies. Water Resour Res 28(3):675–683

Dillman DA (2000) Mail and internet surveys: the tailored design method. John Wiley and Sons, New York, NY

Efron B, Tibshirani R (1993) An Introduction to the Bootstrap. Chapman and Hall, New York

Greene WH (2003) Econometric Analysis, 5th ed. Prentice Hall, Upper Saddle River, NJ

Hanemann WM (1984) Welfare evaluations in contingent valuation experiments with discrete responses. Am J Agr Econ 66(3):332–341

Hanley N, Colombo S, Tinch D, Black A, Aftab A (2006a) Estimating the benefits of water quality improvements under the water framework directive: are benefits transferable? Eur Rev Agr Econ 33(3):391–413

Hanley N, Wright RE, Alvarez-Farizo B (2006b) Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive. J Environ Manage 78(2):183–193

Jiang Y, Swallow SK, McGonagle M (2005) Context-sensitive benefit transfer using stated choice models: specification and convergent validity for policy analysis. Environ Resour Econ 31(4):477–499

Johnston RJ, Swallow SK, Bauer DM (2002) Spatial factors and stated preference values for public goods: considerations for rural land development. Land Econ 78(4):481–500

Johnston RJ, Besedin EY, Iovanna R, Miller C, Wardwell R, Ranson M (2005) Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: a meta-analysis. Can J Agr Econ 53(2–3):221–248

Johnston RJ, Swallow SK, Marie Bauer D, Anderson CM (2003a) Preferences for residential development attributes and support for the policy process: implications for management and conservation of rural landscapes. Agr Resour Econ Rev 32(1):65–82

[19] We note, however, that this general result does not hold for equivalence tests, in which the burden of proof is reversed.

Johnston RJ, Swallow SK, Tyrrell TJ, Bauer DM (2003b) Rural amenity values and length of residency. Am J Agr Econ 85(4):1009–1024
Johnston RJ, Weaver TF, Smith LA, Swallow SK (1995) Contingent valuation focus groups: insights from ethnographic interview techniques. Agr Resour Econ Rev 24(1):56–69
Kask SB, Shogren JF (1994) Benefit transfer protocol for long-term health risk valuation: a case of surface water contamination. Water Resourc Res 30(10):2813–2823
Kline J, Wichelns D (1998) Measuring heterogeneous preferences for preserving farmland and open space. Ecol Econ 26(2):211–224
Krinsky I, Robb AL (1986) On approximating the statistical properties of elasticities. Rev Econ Stat 68(4):715–719
Kristofersson D, Navrud S (2005) Validity tests of benefit transfer—are we performing the wrong tests? Environ Res Econ 30(3):279–286
Loomis JB (1992) The evolution of a more rigorous approach to benefit transfer: benefit function transfer. Water Resourc Res 28(3):701–705
Maddala GS (1983) Limited dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge
McConnell KE (1990) Models for referendum data: the structure of discrete choice models for contingent valuation. J Environ Econ Manage 19(1):19–34
Morrison M, Bennett J (2004) Valuing new south wales rivers for use in benefit transfer. The Aust J Agr Resour Econ 48(4):591–611
Morrison M, Bennett J, Blamey R, Louviere J (2002) Choice modeling and tests of benefit transfer. Am J Agr Econ 84(1):161–170
Morrison M, Bergland O (2006) Prospects for the use of choice modelling for benefit transfer. Ecol Econ 60(2):420–428
Muthke T, Holm-Mueller K (2004) National and international benefit transfer testing with a rigorous test procedure. Envir Resour Econ 29(3):323–336
Piper S, Martin WE (2001) Evaluating the accuracy of the benefit transfer method: a rural water supply application in the USA. J Environ Manage 63(3):223–235
Poe GL, Giraud KL, Loomis JB (2005) Computational methods for measuring the difference in empirical distributions. Am J Agr Econ 87(2):353–365
Rhode Island Geographic Information System (RIGIS) (2000) 1995 Land use by Rhode Island community. Providence, RI: Rhode Island geographic information system and the Rhode Island board of governors for higher education
Rhode Island Economic Development Corporation (2006) State and community profiles. http://www.riedc.com/riedc/ri_databank/31/. Providence, RI: Rhode Island Economic Development Corporation
Rosenberger RS, Loomis JB (2001) Benefit transfer of outdoor recreation use values: a technical document supporting the forest service strategic plan (2000 Revisions). General technical report RMRS-GTR-72. Fort Collins, CO: U.S. Department of agriculture, forest service, rocky mountain research station. 59p. (www.fs.fed.us/rm/pubs/rmrs_gtr72.html)
Rosenberger RS, Loomis JB (2003) Benefit transfer. In: Champ PA, Boyle KJ, Brown TC (eds) A Primer on non-market valuation. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 445–482
Rosenberger RS, Stanley TD (2006) Measurement, generalization and publication: sources of error in benefit transfers and their management. Ecol Econ 60(2):372–378
Schuirmann DJ (1987) A comparison of the two one-sided procedure and the power approach for assessing equivalence of average bioavailability. J Pharmacokinet Biop 15(6):657–680
Shrestha RK, Loomis JB (2003) Meta-analytic benefit transfer of outdoor recreation economic values: testing out-of-sample convergent validity. Environ Resour Econ 25:79–100
Stegner BL, Bostrom AG, Greenfield TK (1996) Equivalence testing for use in psychosocial and services research: an introduction with examples. Eval Program Plann 19(3):193–198
Swait J, Louviere JJ (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. J Market Res 30(3):305–314
Tiebout C (1956) A pure theory of local expenditures. The J Polit Econ 64(5):416–424
Van Bueren M, Bennett J (2004) Towards the development of a transferable set of value estimates for environmental attributes. Aust J Agr Resour Econ 48(1):1–32
VandenBerg TP, Poe GL, Powell JR (2001) Assessing the accuracy of benefits transfers: evidence from a multi-site contingent valuation study of groundwater quality. In: Bergstrom JC, Boyle KJ, Poe GL (eds) The economic value of water quality. Edward Elgar, Northampton, MA