Validating Hypothetical Surveys Using Binding Public Referenda:
Implications for Stated Preference Valuation

Robert J. Johnston
&
Deepak P. Joglekar

Department of Agricultural and Resource Economics
University of Connecticut

Corresponding Author:

Dr. Robert J. Johnston
Connecticut Sea Grant Office
University of Connecticut at Avery Point
1080 Shennecossett Rd.
Groton, CT 06340-6048
(860) 405-9278 (office) / (860) 405-9109 (fax) / robert.johnston@uconn.edu

Short Abstract: This study presents a criterion validity test in which stated choices are compared to subsequent binding referendum votes. The study is distinguished by identical hypothetical and actual choice contexts, and results that show no evidence of hypothetical bias. Results suggest a number of possibilities for amelioration of hypothetical bias.

Tracking Number: #132887

Keywords: Criterion Validity, Hypothetical Bias, Contingent Valuation, Stated Preference.

Abstract

This study presents a criterion validity test in which stated choice responses are compared to aggregated votes in a subsequent binding referendum. The assessment is distinguished by an identical good and information content in hypothetical and actual contexts; a familiar good provided by a trusted institution; a genuine contingent choice survey implemented before the referendum was scheduled or announced; and hypothetical and actual responses representing a large proportion of a well-defined population. The comparison is designed to be simple and unambiguous—no response re-coding is required, no cheap-talk mechanisms are used, and a single choice per respondent parallels the binding referendum. The study is also distinguished by results that show no statistical evidence of hypothetical bias. Estimates of hypothetical and actual WTP differ by less than 10%, and hypothesis tests cannot reject the equivalence of hypothetical and actual choices. Results suggest a number of possibilities for amelioration of hypothetical bias.

Introduction

Prior assessments have demonstrated often substantial hypothetical bias in stated preference

willingness to pay (WTP) estimates, calling into question the appropriateness of such estimates

for welfare evaluation. Reviews provided by Murphy and Stevens (2004), List and Gallett

(2001), and Little and Berrens (2004) illustrate the ubiquity of research addressing the presence

of—and potential calibrations for—hypothetical bias in stated preference valuation. While not

all research finds clear evidence of hypothetical bias (e.g., Champ and Bishop 2001; Smith and

Mansfield 1998; Vossler and Kerkvliet 2003; Johannesson 1997), and some finds that

hypothetical bias may be ameliorated using cheap-talk, certainty adjustments, or other

mechanisms (Champ et al. 1997; Champ et al. 2004; Cummings and Taylor 1999; Loomis et al.

1996), the preponderance of evidence concludes that there are often significant differences

between stated (hypothetical) and actual market behavior (e.g., Murphy and Stevens 2004; List

and Gallett 2001; Little and Berrens 2004).

      While hypotheses regarding hypothetical bias may be tested using assessments of either

convergent[1] or criterion validity, criterion validity tests are generally preferred in cases where

appropriate criterion values exist. Formally, an assessment of criterion validity is "concerned

with whether the measure … is related to other measures which may be regarded as criteria"

(Mitchell and Carson 1989, p. 190). In stated preference research , criterion validity tests

typically require comparisons between contingent (hypothetical) markets and markets that

require some type of real payment (Carson et al. 1986). However, notwithstanding the common

and perhaps mistaken[2] presumption that markets requiring real payments necessarily reveal

criterion values (i.e., actual WTP or Hicksian surplus), markets used in purported criterion

---

[1] Convergent validity considers whether a measure (e.g., a stated preference WTP estimate) is correlated with other
measures of the same theoretical construct (Mitchell and Carson 1989).
[2] For example, see Murphy and Stevens (2004) and Harrison (2002).

1

validity tests most often incorporate simulated or experimental elements as a means to permit comparisons with hypothetical markets (Murphy and Stevens 2004).

While simulated markets may in some instances allow the elicitation of unbiased WTP estimates, they are often conducted in institutional settings unfamiliar to respondents. Such settings may generate novel or unusual behavior, confusion, distrust, or other behaviors that may lead to misstatements of true WTP. For such reasons Bateman et al. (2002) and others argue that comparisons involving WTP estimates elicited through simulated or experimental markets rarely qualify as true tests of criterion validity. The most common alternative—the use of individual or group donation mechanisms to validate hypothetical WTP (e.g., Champ et al. 1997; Champ and Bishop 2001; Seip and Strand 1992; Foster et al. 1997; Spencer et al. 1998)—may suffer from confounding effects of free-riding and the associated lack of incentive compatibility (Bateman et al. 2002; Champ and Bishop 2001). As a result, it not clear that such assessments provide appropriate criterion values to which hypothetical WTP may be compared.

Compared to assessments of hypothetical bias using simulated markets or donation mechanisms, a relatively small number of studies have attempted to validate hypothetical survey responses using criterion values inferred from binding referenda or public votes (Bateman et al. 2002). Recent studies include Schläpfer et al.'s (2004) comparison of contingent valuation (CVM) responses and binding referendum responses for similar but not identical landscape protection programs. Others, including Champ and Brown (1997), Carson et al. (1986) and Vossler and Kerkvliet (2003) compare measures of intended voting behavior to actual referendum results.[3] However, despite a variety of past comparisons of hypothetical and actual

---

[3] Schläpfer et al. (2004, p. 4-5) characterize a genuine CVM survey as one conducted "before the actual referendum proposition becomes the subject of public debate." They distinguish such surveys from those assessing "intended voting behavior," in which stated preference surveys are conducted after an upcoming referendum has been scheduled and announced.

voting behavior, the authors are aware of no published comparisons of responses in *genuine* stated preference survey contexts (cf. Schläpfer et al. 2004, p. 4) and exactly parallel binding referendum votes for identical environmental programs.

This study compares hypothetical, discrete choice CVM responses to aggregated votes in a subsequent, official and binding public referendum, where identical goods are considered in both the hypothetical and actual choice contexts, and respondents are drawn from an identical, well-defined population. The good in question is the provision of a public water supply to the Village of North Scituate, Rhode Island. Comparison of behavior in these parallel hypothetical and real choice contexts provides a criterion validity test that does not incorporate the unfamiliar elements and/or incentive compatibility concerns that may be associated with comparisons of hypothetical and real WTP using simulated markets or donation mechanisms.

The study presented here represents an unambiguous and uncluttered assessment of criterion validity, characterized by clearly consequential, familiar, salient, and exactly parallel hypothetical and actual choice contexts. More specifically, the comparison of hypothetical choices and binding votes presented here is distinguished from prior referendum-based criterion validity tests by: (1) an identical quasi-public good and information content in both hypothetical and actual contexts; (2) a highly familiar and salient good provided by a trusted local institution[4]; (3) a genuine contingent choice survey implemented before the referendum was scheduled or announced; (4) hypothetical and actual responses representing a large proportion of the same, well-defined population.

The research presented here is also distinguished by its simplicity. No re-coding or

---

[4] The familiarity and salience of the good in question renders the potential for confounding effects related to methodological misspecification (Mitchell and Carson 1989), scope insensitivity, distrust in the institution or payment vehicle (Johnston et al. 1999), or other features related to a misunderstanding of the good in question less likely.

transformation of survey responses is required, no cheap-talk or similar mechanisms are applied, and a straightforward "one vote per survey" format eliminates the need to adjust for correlation among responses by individual respondents or to allow for sequence effects. Finally, this study is distinguished by results that show no statistical evidence of hypothetical bias. While actual and intended voting behavior have often been shown to comport quite closely (e.g., Vossler and Kerkvliet 2003), the authors are aware of no prior work that shows similar correspondence between binding votes and hypothetical choices in genuine stated preference contexts. These findings contradict the preponderance of evidence from simulated markets, which shows clear evidence of hypothetical biases across a range of stated preference approaches (Murphy and Stevens 2004). Contrary findings from the present study provide evidence that hypothetical bias is not universal, and hint at potential avenues for the amelioration of such biases in future work.

Comparing Hypothetical Choices and Actual Votes

Numerous authors, including most recently Murphy and Stevens (2004) and Little and Berrens (2004), have outlined conceptual issues underlying the presence of hypothetical biases in stated preference responses and associated WTP estimates. These discussions reflect issues well-known in the literature, and are not reprised here. Rather, we emphasize that which may be learned from a direct comparison of hypothetical choice behavior in a genuine contingent choice survey context and a directly analogous, officially sanctioned and binding vote.

For purposes of the present analysis, we define hypothetical bias as one of two interrelated conditions: (1) a case in which sample mean or median WTP estimated using a stated preference survey instrument may be shown to differ systematically from that estimated using actual markets or behavior, and/or (2) a case in which the percentage or likelihood of 'yes' (or approval) votes associated with a specific and identical public or quasi-public good, at a

4

given and identical household cost, differs systematically between a hypothetical choice and actual voting context. We further constrain our definition of hypothetical bias with an additional condition that the actual behavior occur in a familiar and trusted environment or institution, to avoid potential biases in WTP that may result from the eliciting of behavior in unfamiliar or simulated market contexts.

Conditions for Appropriate Comparison of Hypothetical Choices and Referendum Votes

Champ and Bishop (2001, p. 385) note that past attempts to validate stated preference responses using referendum behavior have "have been met with serious limitations," related to a typical lack of parallelism between hypothetical and referendum choice contexts. More recently, Schläpfer et al. (2004) note three conditions relevant to criterion validity tests involving hypothetical choices and actual votes that are often violated.[5] The first involves the *types of referenda* that are appropriate for tests of criterion validity. Whereas many referenda involve programs whose costs to the household are unspecified or otherwise unclear, direct comparison to stated preference formats requires a voting choice that may be placed in the "same utility maximization framework as the CV responses" (Schläpfer et al. 2004, p. 4). Respondents in the real voting context must consider a choice involving the provision of a known good, with the expected increase in household costs clearly specified.

The second condition relates to the distinction between *genuine* stated preference survey contexts, and surveys (or pre-election polls) in which respondents are asked about planned or intended votes. As stated by Schläpfer et al. (2004, p. 4), "it appears to be impossible to put respondents in the mindset of a hypothetical choice on the basis of controlled survey information when they are, at the same time, making a real voting choice that is influenced by information

---

[5] Schläpfer et al. (2004) also mention a fourth issue, relevant only where within-subject comparisons are made. Since the present assessment does not make within-subject comparisons, this issue is not relevant here.

from a variety of sources." For example, Vossler and Kerkvliet (2003, p. 633) state that "it is likely that many of [their] respondents were aware of the upcoming vote and many may have studied the issue and decided how to vote." Based on such arguments, Schläpfer et al. (2004, p. 4-5) recommend that assessments of criterion validity involve stated preference surveys conducted "*before* the actual referendum proposition becomes the subject of public debate."

The third condition involves common differences between elicitation and information formats in stated preference surveys and binding referenda. For example, where tax consequences and public good outcomes are rarely precisely specified in public referenda, they are typically specified in detail by stated preference surveys (Schläpfer et al. 2004). In contrast, the most appropriate comparison of hypothetical versus real choices involves close correspondence between the elicitation and information formats of the two choice contexts, inasmuch as is possible given that one choice is hypothetical.

A fourth condition, and one not emphasized by Schläpfer et al. (2004), is that tests of criterion validity should ideally involve identical goods in both hypothetical and actual contexts. Past validity tests involving referenda or donation mechanisms have often incorporated non-trivial differences between goods considered in the hypothetical and actual contexts. For example, Schläpfer et al. (2004) report on a study in which the good considered in the hypothetical good was a "regional landscape conservation program," while the good considered in the referendum was an "increase of cantonal fund for nature and heritage protection." While such distinctions might be considered trivial, it is also possible that they might have significant implications for willingness to pay and findings of hypothetical bias.

The North Scituate Water Supply Study

Although tests of criterion validity need not satisfy all of the conditions characterized above,

significant departures from these conditions may render conclusions regarding hypothetical bias of questionable validity. Based on these conditions, the North Scituate Water Supply Study provides a nearly ideal situation to assess criterion validity using actual voting behavior. The study is characterized by: (1) an actual voting choice that may be placed in the same utility maximization context as stated preference responses, (2) a stated preference survey implemented before the public referendum was approved or scheduled, (3) a referendum that precisely specified household tax and quasi-public good outcomes with an elicitation and information content similar to that of the hypothetical survey, and (4) hypothetical and actual choice contexts addressing an identical and salient quasi-public goods.

The study was conducted to assist the Town of Scituate Water Study Committee in assessing public support for provision of public water supply to the Village of North Scituate.[6] The project would have installed and operated infrastructure necessary to provide public water to approximately 500 homes within the Village of North Scituate. Based on State of Rhode Island regulatory requirements, implementation of the proposed water project would have required a majority vote of affected property owners in an officially sanctioned referendum. The Water Supply Committee funded the contingent choice survey as a practical means to gauge public willingness to pay for public water, prior to incurring the significant cost required to sanction, promote, schedule, and implement the actual referendum (Johnston 2001).

While the survey cover letter and instrument noted the possibility of a public vote as a "next step" in the process, this was the first indication that any official referendum might be forthcoming. As the survey was designed as a means to assess public preferences—before the official vote was approved or scheduled—it provides a nearly ideal context in which to assess the

---

[6] The Village of North Scituate is a geographically well-defined, small village of approximately 500 homes, located within the larger Town of Scituate.

validity of hypothetical survey responses in a genuine stated preference context. The Appendix provides the exact text incorporated in the survey cover letter, in which the rationale for the survey was described to respondents.

The Survey Instrument

The contingent choice survey was designed in collaboration with the Water Supply Committee, and was iteratively pretested and revised based on meetings with the Committee and water supply experts, focus groups with local residents, and individual pretests including verbal protocol analysis (Schkade and Payne 1994). The Committee provided engineering plans for a proposed water project and estimated costs based on consultations with water systems experts. These plans and estimates were used as the basis for both the contingent choice survey and the subsequent binding referendum.

Prior to presenting the contingent choice question, the survey (and an accompanying fact sheet) provided detailed information concerning public water and the proposed water project. As described to both survey respondents and referendum voters, the water supply project is best characterized as a quasi-public good. The Committee initially envisioned a project in which individual homeowners could "opt-out" of the system—making it more akin to a private good. At the time of the study and subsequently scheduled referendum, however, it was considered likely that connection to the system would be mandatory (Long 2001). Hence, to reduce the potential for incentive non-compatibility (Carson et al. 1999), the survey presented the choice as involving mandatory payment and water service. In addition, the project would have resulted in other changes characteristic of a public good, including the opening of new areas to potential development (areas unsuitable for private wells), and the nuisance associated with the installation of water supply lines.

The hypothetical choice question asked respondents to vote for or against the proposed

water project, given a specified quarterly household water bill that would cover all installation,

operation, and maintenance costs. Because costs were not known with 100% certainty, the

Committee estimated a realistic range of costs, based on data received from neighboring water

districts and engineering cost estimates (Maguire Group Inc. 2000). In order to forecast voting

behaviors at different cost levels within this range, five different survey versions were produced,

each with a different projected cost.   The five quarterly cost levels were $125, $175, $250, $325,

and $425. Respondents were informed that "because costs are not known with 100% certainty,

the Water Study Committee has estimated a realistic range of costs. Some receiving this survey

will see costs on the high end of this range. Some will see costs on the low end of this range."

This statement was included in order to provide a straightforward explanation for the varying

cost levels incorporated in versions of the survey received by different households.

After detailing the water supply project and its costs, the dichotomous choice question

was worded as follows:

> For this final question, please assume that your household's average **quarterly bill
> (paid every three months, or four times per year)** would be **$___.** Considering the
> costs and benefits of the proposed public water supply, how would you vote?
>
> ❑ YES—I would vote in favor of the public water supply in my area. I
>   understand that my household would pay approximately **$___ per
>   quarter** in new district water bills.
>
> ❑ NO—I would vote against the public water supply in my area. I
>   understand that my household would continue to pay the costs of
>   operating and maintaining our private well.

The Data and Empirical Model

The North Scituate Water Supply Survey was implemented as a mail survey, following the

tailored survey design method of Dillman (2000).   As only quarterly cost varied across survey

versions the experimental design was straightforward, with a roughly equal number of surveys

incorporating each of the five quarterly cost levels. Other attributes of the project remained

constant over all survey versions. In addition to the contingent choice question, survey responses

provided information concerning respondents' satisfaction with the water from their private

wells, the ability of their wells to provide an adequate quantity of water year-round, the

household's use of water filters and/or bottled water, and other elements of potential relevance.

Surveys were distributed to all homeowners in the region who would be affected by the

proposed water supply project. In total, 435 surveys were mailed on July 2, 2001.   Of these, 341

surveys were returned, for a final response rate of 78%. Three-hundred surveys were returned

based on a preview letter, an initial survey mailing and a reminder letter. An additional 41 were

returned subsequent to a final reminder postcard sent on July 28. This distinction is noted as it

appears that there is a systematic difference between responses received before and after the

reminder postcard. Of all returned surveys, 314 provide sufficient information for inclusion in

the empirical model of stated preference (hypothetical) responses.

The North Scituate Water Supply Referendum

Subsequent to the implementation of the survey (and indeed informed by survey results), a

binding referendum to authorize the water supply project was sanctioned and scheduled by the

Town of Scituate. The referendum concerned "whether to establish an independent water

district…[with an]…approximate cost per year to the people receiving the service [of] $1000."[7]

The referendum was held during a Special Financial Town Meeting on November 13, 2001

---

[7]   It was emphasized that as the project proceeded, certain additional elements of the proposed water supply project would likely be placed "before the voters of the district again for final approval" (Long 2001). The specific elements that would require additional voter approval were left unspecified. However, it was made clear that prior to the completion of the project (a multiple year task), subsequent votes would likely be necessary to address specific issues. Regarding the issue of mandatory versus voluntary participation in the public water system, once established, it was emphasized that mandatory connection with the system "was a possibility."

(Long 2001). The water supply project was identical to that considered in the hypothetical

survey instrument implemented approximately four months prior. Description of the water

supply project in the public referendum was drawn from the same information found in the

hypothetical survey instrument, and specified an anticipated household cost of $1000 per year, or

$250 per quarter—the median value presented in hypothetical survey versions. Those invited to

the Town Meeting were the same homeowners to whom the survey instrument had been mailed.

Aggregate data for the November 13, 2001 referendum (actual votes) were obtained from

official minutes of the Special Financial Town Meeting (Long 2001). (All those eligible to vote

in the Town Meeting had already received the prior survey describing the water supply project.)

After the provision of a matching verbal description of the project, and a limited question-and-

answer session, the vote was initiated at 8:45pm. The final ballot results, at an anticipated

household cost of $250 per quarter, were 122 in favor and 145 opposed, for a 45.69% approval.

The Empirical Model

As each respondent considered only a single contingent choice question, the data may be

analyzed using a standard random utility model (Hanemann 1984; McConnell 1990), with all

observations considered independently and identically distributed (*iid*). The model presumes

that respondents choose whether to vote for mandatory public water supply based on the

difference in utility resulting from public water supply and that resulting from retaining private

(well) supply. This difference may be represented by

$$dU \quad = \quad U_p(W_p,\ Y\text{-}C_p,\ D) \text{ - } U_w(W_w,\ Y\text{-}C_w,\ D) \tag{1}$$

where *dU* represents the utility difference associated with public water supply, $U_p(\cdot)$ represents

utility realized from the provision of public water, and $U_w(\cdot)$ represents utility realized from the

retention of private water. Utility from public water is assumed to be determined by the attributes of the public water supply (vector $\mathbf{W_p}$, assumed fixed across all households), the household's demographic attributes (vector $\mathbf{D}$), and household income (Y) minus the additional cost of public water (C_p). Utility from private water supply is determined by the attributes of the household's well-water supply (vector $\mathbf{W_w}$, which may vary across households), the household's demographic attributes (vector $\mathbf{D}$), and household income (Y) minus the current cost of operating a private well (C_p). Based on interviews with the Water Supply Committee and focus groups, we assume that this cost (C_p) is fixed and approximately constant across households.

The random utility model presumes that utility $U(\cdot)$ may be divided into empirically measurable (observable) and stochastic (unobservable) components, such that

$$
\begin{aligned}
dU &= v_p(\mathbf{W_p},\ Y\text{-}C_p,\ \mathbf{D}) - v_w(\mathbf{W_w},\ Y\text{-}C_w,\ \mathbf{D}) - [\varepsilon_w - \varepsilon_p] \\
&= dv - \theta.
\end{aligned}
\tag{2}
$$

Here, $v_p(\cdot)$ and $v_w(\cdot)$ represent the empirically measurable components of utility associated with public and well water supply, respectively. The terms ε_p and ε_w represent the associated stochastic, or unobservable components. Assuming the standard linear approximation for $v_p(\cdot)$ and $v_w(\cdot)$ common in random utility models, and that $\mathbf{W_p}$ and C_w are fixed and constant across all observations, dv may be represented by a reduced-form econometric specification

$$
dv = \beta_0 + \beta_1(\mathbf{W_w}) + \beta_2(C_p) + \beta_3(\mathbf{D}).
\tag{3}
$$

Given (1)-(3), if one assumes that θ follows a logistic distribution, one may model the probability of choosing public water using the familiar logit model, with the standard likelihood function provided by Maddala (1983). As each survey incorporates only a single choice question, each response is treated as independently and identically distributed (*iid*) for purposes

of estimation. Following the model of (1)-(3) above, variables included in the statistical model characterize such features as the cost of public water supply, the attributes of each household's existing private well-water supply, and other household attributes. Table 1 characterizes independent variables included in one or more variants of the statistical model.

Results from the model are compared to aggregated results from the binding referendum. To provide the most conservative test of criterion validity—here the equivalence between hypothetical choices and binding votes—the results of the binding vote are treated as non-stochastic (i.e., a fixed point of 0.457). While one might also treat referendum results as the result of a stochastic process, the resulting variance would increase the ease with which one might reject the null hypothesis of zero hypothetical bias.

Model Results

Table 2 illustrates the raw choice results associated with each of the five quarterly cost levels. As shown by the table, the data are well-behaved, with the proportion of 'yes' responses decreasing with increases in cost. For example, 48.44% of survey respondents shown a cost of $250/quarter voted in favor of the water supply project. This closely approximates the 45.69% of respondents in the binding referendum who also voted in favor of the project, at the same quarterly cost. While not representing a full, formal assessment of hypothetical bias, the similarity of these "raw voting" outcomes suggests a high degree of similarity between hypothetical and real behavior.

Results of the logit model are presented in table 3. Model one is the final, unrestricted specification. The model is statistically significant at p<0.0001 (likelihood ratio $\chi^2 = 88.898$, df = 6), and correctly predicts 77.2% of observed choices. All independent variables in the final

model are statistically significant at p<0.10, with most significant at p<0.01. Model two represents a restricted specification from which two variables (*late*; *late×cost*) have been excluded. As indicated by variable descriptions in table 1, these variables allow estimated parameters for the equation intercept and program cost to vary systematically according to whether the survey was received after a final reminder postcard was mailed (July 28, 2001).

A likelihood ratio test of restrictions in model two rejects the null hypothesis of zero joint influence (χ^2 = 12.322, df = 2, p=0.002). Hence, we conclude that there are systematic differences between the preferences of those who returned surveys after July 28 and those who returned surveys between July 2 and July 28. Of particular relevance here, results suggest that late respondents were much less influenced by the household cost of the project (*cost*), as indicated by the near-zero sum of (*cost + late×cost*). While the explanation for this finding is unknown, it suggests that the timing of survey responses may influence the potential for hypothetical bias, as in this case it influences the responsiveness of choices to program cost. Given statistical significance of the restrictions implicit in model two (see above), model one is selected as the final model for assessments of criterion validity.

Assessments of Criterion Validity and Hypothetical Bias

As referendum results are only available as an aggregate sum, comparisons of individual responses are not possible. Hence, results are based on aggregate responses to in both hypothetical and binding choice contexts, preventing a detailed assessment of nonresponse such as that of Vossler and Kerkvliet (2003). Nonetheless, while the specific identities and demographic attributes of the referendum respondents are unknown, both the survey and referendum capture a large percentage of the households in the affected region (78% and 61% of the 435 households, respectively). The fact that hypothetical and actual responses represent a

large proportion of this well-defined population should at least partially ameliorate concerns

raised by Carson et al. (1999) and others regarding potentially significant differences between

the "random sample of the public" responding to a stated preference survey and "the sample of

the public that actually votes." Moreover, given that the sole purpose of the November 13, 2001

special town meeting was to vote on the water supply referendum, there is no evidence of

divergence between incentives for survey response and incentives for voting.

Unlike assessments of hypothetical biases in simulated or experimental markets—in

which complete demand or preference equations may be estimated—here a complete preference

function may only be estimated for hypothetical responses. As noted above, referendum results

are characterized only by a single point: 45.69% approval of the proposed water supply project at

an estimated household cost of $250 per quarter. This limits the extent of comparisons that may

be made between hypothetical and binding responses. Nonetheless, even within these

limitations, one may assess differences between responses in the two choice contexts.

The most direct comparison involves the probability of supporting the proposed water

supply project (or, alternatively, the percentage of representative respondents who would support

the project). The proportion of votes supporting the water supply project is known in the

referendum—0.4569 at $250/quarter. An analogous proportion may be estimated based on

stated preference model results from table 3, characterizing hypothetical choices. Given a

$250/quarter cost, and mean values for other model variables (table 1), the proportion of support

('yes' votes) predicted by the logit model is 0.4978. Based on model results, a Wald test (Greene

2003) fails to reject the equality of the hypothetical (estimated) and actual proportions of

supporting votes ($\chi^2=0.04$, df = 1, prob = 0.8474). An analogous result holds if one conducts a

comparable hypothesis test using standard errors for the predicted (hypothetical) probabilities,

15

following the resampling approach of Krinsky and Robb (1986).[8] That is, a statistical comparison of hypothetical choices and binding votes provides no evidence of hypothetical bias in the probability of supporting the provision of identical quasi-public goods, at identical household costs. These results are summarized in table 4.

An even stronger refutation of hypothetical bias emerges if one only considers surveys received on-time (i.e., the variable *late* is set equal to 0). Given a $250/quarter cost, and mean values for other model variables (table 1), the proportion of support predicted by a logit model excluding late surveys is 0.4796. Based on model results, a Wald test (χ^2=0.02, df = 1, prob = 0.9153) fails to reject the equality of the hypothetical (estimated) and actual proportions of supporting votes.

One may also assess criterion validity in terms of implied WTP. Traditionally, one assesses criterion validity based on standard estimates of either mean or median WTP. Here, however, standard WTP estimates may not be derived from binding referendum results. Rather, as noted above, one only observes a single point estimate. This point estimate indicates that 45.69% of respondents vote 'yes' at a $250/quarter cost. The appropriate WTP comparison, then, is the estimated *hypothetical* cost level at which survey respondents would provide an identical level of support (i.e., 45.69%). Given model results, this may be calculated as

$$WTP_{0.457} = \frac{\hat{\beta}_0 + (\hat{\beta}_{bottled} \times bottled) + (\hat{\beta}_{supply} \times supply) + (\hat{\beta}_{h\_size} \times h\_size) + (\hat{\beta}_{late} \times late) + 0.1727}{-\hat{\beta}_{\cos t} - \hat{\beta}_{cost \times late} \times late} \quad (4)$$

where independent variables (*bottled, supply, h_size, late*) are set to mean sample values,

---

[8] The latter test is conducted for comparison. We randomly draw 10,000 sets of coefficient estimates from the estimated distribution of parameters obtained from maximum likelihood estimation. Probability estimates (i.e., probability of a 'yes' vote) are calculated for each draw, resulting in an empirical distribution of probability from which the estimated standard error is calculated. Based on this estimated standard error (0.0419), a t-test fails to reject the equality of the hypothetical (estimated) and actual probabilities (p=0.25).

corresponding parameter estimates are given by $\hat{\beta}_i$ ($\hat{\beta}_0$ is the equation intercept), and the constant 0.1727 represents the value of *dv* in (3) that leads to 45.69% support based on the logistic probability function. It is easily shown that (4) follows directly from the specification of *dv* in (3) and the logistic probability function of the logit model.

Based on (4), 45.69% support for the water supply project is projected at a quarterly cost of $268.01. This represents a 7.2% increase over the $250/quarter cost at which actual 45.69% approval was found—a small difference in point-estimate WTP associated with this level of support. A Wald test (χ^2=0.04, df = 1, prob = 0.8505) fails to reject the equality of these hypothetical (estimated) and actual WTP estimates. These results are also illustrated in table 4. Mirroring results found above, assessment of WTP finds no evidence of significant hypothetical bias in a direct comparison of identical hypothetical and binding choices. This stands in clear contrast to prior work showing that hypothetical values are typically 1.5 to 3 times greater than actual market values (e.g., Loomis et al. 1996; Murphy and Stevens 2004; List and Gallet 2001).

Also as above, hypothetical and binding results comport even more closely if one omits late survey responses. When *late*=0, 45.69% support for the water supply project is projected at a quarterly cost of $258.59. Again, Wald test results (χ^2=0.01, df = 1, prob = 0.9162) fail to reject the equality of hypothetical (estimated) and actual estimates. Analogous hypothesis tests using resampling methods (Krinsky and Robb 1986) generate qualitatively identical results.

Implications and Discussion

Results found here mirror those of Champ and Brown (1997), Vossler and Kerkvliet (2003), and others who contrast intended voting behavior (or CVM votes in contexts where an upcoming referendum was anticipated) with binding referendum votes, and find close correspondence between hypothetical and binding responses. However, to our knowledge this is the first

assessment in which hypothetical responses in a genuine CVM context have been compared to

binding referendum results for identical programs, and in which there is no evidence of

significant hypothetical bias. Moreover, the combination of parallel hypothetical and real choice

contexts, a salient and familiar good, and an identical sampled population (both hypothetical and

real samples were drawn from the same 435 households) suggests that model results provide a

valid assessment of criterion validity. This result begs the question as to why hypothetical bias is

not present here, despite overwhelming evidence of hypothetical bias in simulated markets

(Murphy and Stevens 2004).

Although the present data do not allow unambiguous determination of the reason(s) why

hypothetical bias is not significant here, there are a variety of potential explanations that

correspond to prior findings in the literature. These might include: 1] the salience and familiarity

of the good in question and equivalence of information content in hypothetical and binding

choice contexts, 2] the explicit linkage between the CVM survey and an official government

process, including notification that an official vote might be forthcoming.

Salience and Familiarity of the Good and Equivalence of Information Content

According to Arrow et al. (1993, p. 4605), "[i]f CV surveys are to elicit useful information

about willingness to pay, respondents must understand exactly what it is they are being asked to

value." The survey and focus group literature suggests that first hand experience with natural

resources may influence values and survey responses (Cameron and Englin 1997; Johnston et al.

1995). The literature also argues that familiarity with goods and/or behavior tends to create

closer correspondence between behavioral intentions and actual behavior (Mitchell and Carson

1989, p. 186). As stated by Blamey et al. (2001, p. 128), "familiarity may be beneficial in terms

of the validity and reliability of results." Here, the survey presents a quasi-public good that is

18

highly familiar and salient to many respondents. Moreover, many respondents will have had direct experience with the attributes of this good—public water supply—in other homes or businesses. This stands in contrast to goods often valued using stated preference methods, which may be unfamiliar and with which respondents may have little direct experience.

It has also been established that information provision can influence WTP estimates derived from stated preference survey instruments (e.g., Bergstrom and Stoll 1989; Bergstrom et al. 1989; Hoehn and Randall 2002). Moreover, it is common for information provision in stated preference survey instruments to differ substantially from that in public referenda (Champ and Bishop 2001; Schläpfer et al. 2004). Here, the information content underlying hypothetical and binding choices is identical. Indeed, all households in the affected neighborhood were provided with extensive and accurate information regarding the proposed water supply project as part of the survey. Information provided during the public meeting was also drawn from the informational materials developed for the survey. Hence, the level and type of information underlying the hypothetical and binding choices in this case are highly similar—if not identical.[9] Based on empirical results found here and prior results from the literature, it is possible— perhaps even likely—that the familiarity and salience of the good in question, combined with extensive and equivalent information provided in both choice contexts, may have contributed to a closer correspondence between hypothetical and actual responses.

Explicit Linkage Between the CVM Survey and an Official Government Process

The contingent choice survey described here is relatively unusual in that it was implemented as part of an official government consideration of a specific and known public project. Also, while

---

[9] It is of course possible that some additional information or opinions may have been available as a result of media reports that surfaced subsequent to the survey and before the binding referendum. However, much of the information provided in local media reports were drawn from the same planning documents upon which the survey information materials were based.

the decision to hold the public referendum had not been made at the time of survey implementation, the possibility of a forthcoming officially sanctioned referendum was mentioned as part of the survey materials (see Appendix). Hence, while the survey is appropriately characterized as a genuine CVM instrument conducted prior to the announcement and scheduling of a subsequent referendum, the clear association with an official government process and the possibility of a subsequent referendum may have placed respondents in a mindset closer to that which would apply during a binding vote.

As noted by Schläpfer et al.(2004, p. 4), "[i]t appears to be impossible to put respondents in the mindset of a hypothetical choice … when they are, around the same time, making a real voting choice." While this statement is presented as a critique of validity tests based on surveys conducted *after* a referendum had been announced and scheduled, it might also be interpreted as a potential means to reduce hypothetical biases in genuine CVM survey contexts. That is, the simple mention that a CVM survey has been commissioned by a government agency, and that the scheduling of a binding referendum might be a subsequent step in the process, may be sufficient to place respondents in a mindset equivalent to that which would occur during an analogous binding vote. While survey results cannot confirm such a supposition, it is consistent with the both results found here and the close correspondence between binding votes and responses in CVM surveys conducted after public referenda had been announced and promoted (e.g., Vossler and Kerkvliet 2003).

This reasoning is also consistent with Carson et al.'s (1999) arguments concerning the role of consequentiality in stated preference surveys and implications for incentive compatibility. Carson et al. (1999) define consequentiality as a case in which "survey results are seen by the agent as potentially influencing agency action" and in which the agent treats the survey as "an

opportunity to influence those actions." From this definition, it follows that the level of

consequentiality depends on the perceived probability of survey responses influencing agency

actions, where this perceived probability may vary across survey contexts. Here, survey

materials present a clear description of the role of survey results in the policymaking process.

Moreover, the "agency" in consideration was a local government body comprised of known and

largely trusted individuals. Hence, respondents may have viewed the survey as being more

highly consequential than typical stated preference instruments. Based on the arguments of

Carson et al. (1999), this may have led to a greater incentive for respondents to answer truthfully

concerning their willingness to support the water supply project.

Conclusion

This paper compares hypothetical, discrete choice CVM responses to aggregated votes in a

subsequent, official and binding public referendum, where identical goods are considered in both

the hypothetical and actual choice contexts. In contrast to prior stated preference validity tests

conducted using simulated markets, here we find no evidence of statistically significant

hypothetical bias. Even on a point-estimate basis, implied hypothetical and actual WTP differ by

less than 10%—a result of some significance given that may past assessments of CVM have

shown mean hypothetical values "are about 2.5 to 3 times greater than actual values." (Murphy

and Stevens 2004). While reasons for the lack of hypothetical bias cannot be unambiguously

established in the present case—attributes of the study provide a number of compelling

possibilities for the potential amelioration of hypothetical biases.

Among the possibilities suggested by the present study are that reductions in hypothetical

biases may be related to the salience and familiarity of the good in question and equivalence of

information content in hypothetical and binding choice contexts, and an explicit linkage between

the survey and an official government process, including notification that an official vote might be forthcoming. Additional research is required, however, to establish which, if either of these factors might provide a practical means to reduce or eliminate hypothetical biases in broader CVM research. In contrast, subsequent research might also find that the present results are due primarily to unique elements of the present study, and that these results hold little promise for improvements to CVM methods in general.

Empirical results shown here may offer more in terms of compelling questions and possibilities than unambiguous proof that hypothetical biases may be overcome in broader stated preference research. While the relatively simple, unambiguous results provided by the present research might be considered compelling evidence that hypothetical bias can be avoided in some cases, the lack of data for individual respondents also leaves open the possibility—even if remote—that results reflect other influences such as differing rates of nonresponse. Such possibilities notwithstanding, model results at a minimum suggest that hypothetical bias is not universal, and that researchers should neither abandon the search for means to ameliorate hypothetical biases, nor should accept such biases as a foregone conclusion.

Appendix: Text of Selected Survey Materials

The following appendix provides the text incorporated in the survey cover letter, in which the rationale for the survey was described to respondents.

Text of the Cover Letter

"I am writing to ask for your input on an important decision that affects your household—the potential creation of a water district to provide public water to the Village of North Scituate and points North along Route 116. The Town of Scituate Water Study Committee has prepared this survey to assess public support or opposition to this project. You have received this survey because your home is in the region under consideration for public water.

The proposed water district will not proceed without the majority support of property owners who live within the boundaries of the proposed district. This survey is the first step in determining whether such support exists. Depending on survey results, the next step would be a public information forum, followed by an official vote of affected homeowners in November, 2001. Only homeowners within the proposed water district would be allowed to vote for or against the project.

Included in this package are a one-page survey and a brief fact sheet providing some important information regarding the proposed water district. It is important that we hear from all residents who may be affected by the proposed water supply project. Your answers to this survey are strictly confidential and anonymous. If you have any questions or concerns regarding the survey, please don't hesitate to contact The Town of Scituate Water Supply Board at _____. We hope that you will take the time to complete this important survey."

References

Bateman, I.J., R.T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Leee, G. Loomes, S. Mourato, E. Ozdemiroglu, D.W. Pierce, R. Sugden, and J. Swanson. 2002. *Economic Valuation with Stated Preference Surveys: A Manual.* Northampton, MA: Edward Elgar.

Bergstrom J.C., and J.R. Stoll. 1989. Application of Experimental Economics Concepts and Precepts to CVM Field Survey Procedures. *Western Journal of Agricultural Economics* 14 (1): 98-109.

Bergstrom, J.D., J.R. Stoll, and A. Randall. 1989. Information Effects in Contingent Markets. *American Journal of Agricultural Economics* 71(3): 321-323.

Blackburn, M., G. W. Harrison, and E. E. Rutstrom. 1994. Statistical Bias Functions and Informative Hypothetical Surveys. *American Journal of Agricultural Economics* 76(5):1084-1088.

Blamey, R., J. Bennett, J.J. Louviere, and M. Morrison. 2001. Green Product Choice, in J. Bennett and R. Blamey, eds. *The Choice Modelling Approach to Environmental Valuation*. Northampton, MA. Edward Elgar.

Cameron, T.A. and J. Englin. 1997. Respondent Experience and Contingent Valuation of Environmental Goods, *Journal of Environmental Economics and Management* 33(3), 296-313.

Carson, R. T., T. Groves and M. J. Machina (1999), Incentive and Informational Properties of Preference Questions, Plenary Address at European Association of Resource and Environmental Economists. Oslo, Norway.

Carson, R.T., W.M. Hanemann, and R.C. Mitchell. 1986. The Use of Simulated Political Markets to Value Public Goods. Working Paper, Economics Department, University of California, San Diego.

Champ, P.A. and R. C. Bishop. 2001. Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias. *Environmental and Resource Economics* 19: 383-402.

Champ, P. A., R.C. Bishop, T.C. Brown, and D.W. McCollum. 1997. Using Donation Mechanisms to Value Nonuse Benefits from Public Goods. *Journal of Environmental Economics and Management* 33:151-162.

Champ, P.A. and T.C. Brown. 1997. A Comparison of Contingent and Actual Voting Behavior. Proceedings from the W-133 Benefits and Cost Transfer in Natural Resource Planning, 10th Interim Report, Rocky Mountain Research Station, USDA Forest Service.

Champ, P.A., R. Moore, and R. C. Bishop. 2004. Hypothetical Bias: The Mitigating Effects of Certainty Questions and Cheap Talk. Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, August 2004.

Cummings, R. G., and L. O. Taylor. 1999. Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. *American Economic Review* 89(3):649-665.

Cummings, R. G., S. Elliott, G. W. Harrison, and J. Murphy. 1997. Are Hypothetical Referenda Incentive Compatible? *Journal of Political Economy* 105(3):609-621.

Dillman. D.A. 2000. *Mail and Internet Surveys: The Tailored Design Method.* New York, NY: John Wiley and Sons.

Foster, V., I. J. Bateman, and D. Harley. 1997. Real and Hypothetical Willingness to Pay for Environmental Preservation: A Non-experimental Comparison. *Journal of Agricultural Economics* 48(2):123-138.

Fox, J. A., J. F. Shogren, D. J. Hayes, and J. B. Kliebenstein. 1998. CVM-X: Calibrating Contingent Values with Experimental Auction Markets. *American Journal of Agricultural Economics* 80(3):455-465.

Greene, W.H. 2003. *Econometric Analysis, 5th ed*. Upper Saddle River, NJ: Prentice Hall.

Haab, T. C., J. Huang, and J. C. Whitehead. 1999. Are Hypothetical Referenda Incentive Compatible? A Comment. *Journal of Political Economy* 107(1):186-196.

Hanemann, W.M. 1984. Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. *American Journal of Agricultural Economics* 66(3):332-41.

Harrison, G.W. 2002. Experimental Economics and Contingent Valuation. Working Paper, Department of Economics, University of Central Florida.

Hoehn, J.P. and A. Randall. 2002. The Effect of Resource Quality Information on Resource Injury Perceptions and Contingent Values. *Resource and Energy Economics* 24 (ER1-2), 13-31.

Johannesson, M. 1997. Some Further Experimental Results on Hypothetical Versus Real

Willingness to Pay. *Applied Economics Letters* 4: 535-536.

Johnston, R. J. 2001. Final Report of the North Scituate Water District Survey. The Scituate
Water Study Committee, Scituate, Rhode Island.

Johnston, R.J., T.F. Weaver, L.A. Smith, and S.K. Swallow. 1995. Contingent Valuation Focus
Groups: Insights from Ethnographic Interview Techniques. *Agricultural and Resource
Economics Review* 24(1): 56-69.

Krinsky, I. and R. Robb. 1986. On Approximating the Statistical Properties of Elasticities.
Review of Economics and Statistics 68(2): 715-719.

Little, J., and R. Berrens. 2004. Explaining Disparities between Actual and Hypothetical Stated
Values: Further Investigation Using Meta-analysis. *Economics Bulletin* 3(6):1-13.

List, J.A. and C. Gallet. 1998. What Experimental Protocols Influence Disparities Between
Actual and Hypothetical Stated Values? *Environmental and Resource Economics* 20:
241-254.

Loomis, J., T. Brown, B. Lucero, and G. Peterson. 1996. Improving Validity Experiments of
Contingent Valuation Methods: Results of Efforts to Reduce the Disparity of
Hypothetical and Actual Willingness to Pay. *Land Economics* 72(4):450-461.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*.
Econometric Society Monographs No. 3, Cambridge University Press, Cambridge.

Maguire Group Inc. 2000. Report on Water Supply for Scituate, Rhode Island, 1975 Revised to
2000. Final Report June 30. 225 Chapman Street, Providence, RI 02905.

McConnell, K.E. 1990. Models for Referendum Data: The Structure of Discrete Choice Models
for Contingent Valuation. *Journal of Environmental Economics and Management* 18(1):
19-34.

Mitchell, R. C., and R. T. Carson. 1989. *Using Surveys to Value Public Goods : The Contingent
Valuation Method.* Resources for the Future. Distributed worldwide by the Johns
Hopkins University Press.

Murphy, J.J. and T.H. Stevens. 2004. Contingent Valuation, Hypothetical Bias, and
Experimental Economics. *Agricultural and Resource Economics Review* 33(2): 182-192.

Newell, L.W., and S.K. Swallow. Are Stated Preferences Invariant to the Prospect of Real-
Money Choice? Paper presented at AAEA annual meeting, Long Beach, CA, 28-31 July
2002.

Schkade, D.A. and J.W. Payne. 1994. How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation. *Journal of Environmental Economics and Management* 26(1): 88-109.

Schläpfer, F., A. Roschewitz, and N. Hanley. 2004. Validation of Stated Preferences for Public Goods: A Comparison of Contingent Valuation Survey Response and Voting Behavior. *Ecological Economics* 51(1): 1-16.

Seip, K. and Strand, J. 1992. Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payments. *Environmental and Resource Economics* 2: 91-106.

Smith, V. K., and C. Mansfield. 1998. Buying Time: Real and Hypothetical Offers. *Journal of Environmental Economics and Management* 36(3): 209-224.

Spencer, M. A., S. K. Swallow, and C. J. Miller. 1998. Valuing Water Quality Monitoring: A Contingent Valuation Experiment Involving Hypothetical and Real Payments. Agricultural and Resource Economics Review 27(1):28-42.

Long, M.M. 2001. Minutes of Special Financial Town Meeting. November 13, 2001. Office of the Town Clerk, Town of Scituate, Rhode Island.

Vossler, C.A. and J. Kerkvliet. 2003. A Criterion Validity Test of the Contingent Valuation Method: Comparing Hypothetical and Actual Voting Behavior for a Public Referendum. *Journal of Environmental Economics and Management* 45(3): 631-649.

Vossler, C.A., J. Kerkvliet, S. Polasky, and O. Gainutdinova. 2003. Externally Validating Contingent Valuation: An Open-Space Survey and Referendum in Corvallis, Oregon. *Journal of Economics Behavior and Organization* 51: 261-277.

Table 1. Model Variables and Summary Statistics

Variable	Definition	Mean	Std. Dev.
Bottled	Binary variable indicting that the household uses bottled water for drinking (1 = yes, 0 = no).	0.1404	0.3479
Supply	Binary variable identifying households whose well provides an adequate quantity of water during all months of the year (1 = yes, 0 = no).	0.8788	0.3269
H_Size	Number of persons living in the household.	2.9266	2.2594
Late	Binary variable identifying surveys received "late," or after July 28.	0.1257	0.3320
Cost	Estimated quarterly household cost of the proposed water supply project, including all installation, maintenance, and operation costs.	248.0994	101.9605
Cost×Late	Quadratic interaction between *Cost* and *Late*.	32.7485	95.4149

Table 2.   Raw Results: Proportion of 'Yes' Votes at Different Quarterly Cost Levels

	Proportion of 'Yes' Votes	
	Hypothetical Survey Responses	Binding Referendum Responses
Quarterly Cost Level		
$125	0.6528	--
$175	0.5867	--
$250	0.4843	0.4569
$325	0.4085	--
$425	0.3261	--

Table 3. Maximum Likelihood Logit Results

	Unrestricted Model			Restricted Model		
Variable	Coefficient	Std. Err.	Prob>|z|	Coefficient	Std. Err.	Prob>|z|
Intercept	5.0533	0.8717	<0.001	4.2656	0.8009	<0.001
Bottled	0.9017	0.4140	0.029	0.9467	0.4073	0.020
Supply	-3.5329	0.6746	<0.001	-3.1893	0.6477	<0.001
H_Size	0.1701	0.0909	0.061	0.1472	0.0870	0.091
Late	-2.4447	0.9446	0.010	--	--	--
Late×Cost	0.0121	0.0036	0.001	--	--	--
Cost	-0.0106	0.0017	<0.001	-0.0080	0.0015	<0.001
-2LnL χ^2: model (df = 6)	88.8981		<0.001	76.5759		<0.001
-2LnL χ^2: for restrictions (df = 2)				12.3222		0.002
Pseudo R^2	0.2109			0.1817		
N (obs.)	304			304		
N per completed survey	1			1		

Table 4. Assessments of Hypothetical Bias: Empirical Results

	Full Model (Includes Late Surveys)	Full Model (*Late*=0)
Probability of 'Yes' Vote at $250/Quarter Cost		
Contingent Choice (hypothetical)	0.4978	0.4796
Referendum (actual)	0.4569	0.4569
Point Estimate Difference	0.0409	0.0227
Wald χ^2 for H$_0$: [Hypothetical = Actual Probability]	0.04	0.02
Prob $> \chi^2$ (df=1)	0.8474	0.9153
WTP Associated with 45.69% 'Yes' Votes		
Contingent Choice (hypothetical)	$268.01	$258.59
Referendum (actual)	$250.00	$250.00
Point Estimate Difference	$18.01	$8.59
Wald χ^2 for H$_0$: [Hypothetical = Actual WTP]	0.04	0.01
Prob $> \chi^2$ (df=1)	0.8505	0.9162