E C O N O M I C S B U L L E T I N

Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta−Analysis

Joseph Little
University of New Mexico

Robert Berrens
University of New Mexico

Abstract

Spurred by the need to account for non−market values in various policy applications, a lively and extended debate has surrounded the presence and magnitude of hypothetical bias in stated value studies (e.g., applications of the survey−based contingent valuation method). Using the rapidly accumulating set of comparison studies, List and Gallet (2001) conducted an initial meta−analysis of the experimental protocol that may be influencing the disparity between real and hypothetical values in stated value studies. We expand the original meta−analysis by using a significantly larger (29%) data set, including variables to account for referendum formats, certainty corrections, and cheap talk scripts.

The authors would like to thank Carsten Lange (California State University at Pomona) and an anonymous reviewer for helpful comments, and Craig Gallet (California State University at Sacramento) and John List (University of Maryland at College Park) for providing their data set. Address correspondence to Joseph Little, Department of Economics, University of New Mexico, Albuquerque, NM 87131. Phone: 505−277−6458, email: jlittle@unm.edu.
Citation: Little, Joseph and Robert Berrens, (2004) "Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta−Analysis." *Economics Bulletin,* Vol. 3, No. 6 pp. 1−13
Submitted: December 19, 2003. **Accepted:** March 1, 2004.
URL: http://www.economicsbulletin.com/2004/volume3/EB−03C90005A.pdf

1. Introduction

The use of stated value studies to estimate values for non-market goods remains common in benefit-cost analyses, natural resource damage assessments, and other policy applications. Yet, persistent concerns remain that stated value results (e.g., from contingent valuation studies) may contain an upward hypothetical bias. Hypothetical bias can be defined as the disparity between hypothetical statements and real values (or what an individual might actually pay for the provision of the good). Identifying the sources, direction, and magnitude of any hypothetical bias in stated value studies has been the focus of considerable research effort. To wit, there is a rapidly accumulating set of studies that directly compare hypothetical and real values, where the latter is assumed to represent true preferences. List and Gallet (2001) conducted a meta-analysis to identify the experimental protocol that may influence the disparity between actual and stated values. Their results provided an initial benchmark for researchers, and have been widely cited. The objective of this paper is to both update and expand the original meta-analysis.

A calibration factor, obtained by dividing sample estimates of hypothetical by real values, is the dependent variable in List and Gallet (2001). A large (small) calibration factor indicates a greater (lesser) disparity between real and hypothetical. Using the results from 29 studies (with 58 total observations), regressions were run on the median, maximum, and minimum calibration factors. They find that the disparity is significantly lower in studies estimating willingness to pay (versus willingness to accept), that used private (versus public) goods, and that used first price sealed bids (versus alternative elicitation mechanisms). Of note, the finding that studies obtaining values for public goods tend to exhibit a significantly larger disparity seems to contradict the arguments of Carson, Groves, and Machina (2000) and others. They argue that a referendum format for eliciting values for public goods provides the best possible case for reducing or eliminating hypothetical bias. Their reasoning is that referenda requiring a plurality have the potential of being incentive compatibility (Carson, Groves, and Machina 2000). While the original data set used by List and Gallet (2001) does not include any referendum-based studies, a number of recent comparisons have been conducted (with mixed results). Several additional approaches (certainty corrections and the use of cheap talk scripts) for reducing the disparity between real and hypothetical values have also received considerable recent attention (with mixed results). Given these important gaps in our understanding, it is opportune to re-visit the original meta-analysis of List and Gallet (2001).

The List and Gallet (2001) data set has been expanded to include 17 new observations (a 29% increase) from 12 studies. Calibration factors were calculated according to the List and Gallet (2001) criteria, and their general econometric approach is followed in examining this expanded data set. One difference is that since individual studies often produce multiple observations, we also explore weighting and clustering techniques in our estimation (e.g., to drop the assumption of independence between observations). For completeness, we also conduct a probit model, where the dependent variable is the absence or presence of a statistically significant finding of bias between hypothetical and real stated values. The probit model allows for the inclusion of previously excluded studies, where a test of hypothetical bias was conducted but a calibration factor could not be calculated (e.g., only acceptance rates to a single offered price may have been evaluated). Finally, both the extended calibration and probit models also include three new variables to account for the impact that referenda, cheap talk scripts, and certainty corrections, may have on any disparity between real and hypothetical valuation responses.

2. Background

The literature has continued to grow rapidly in the period since List and Gallet (2001) completed their meta-analysis, which included studies up to 2000. New studies were added to the original data set, and estimation approach, as long as they reported mean hypothetical and real values. They also needed to include a discussion of the experimental design.[1] To be included in the data set used for the probit analysis, a study had to report a test of significance relating to hypothetical bias. For example, Cummings, Elliot, and Harrison (1997) find a statistically significant disparity between hypothetical and real reported values, but were not included in List and Gallet (2001) because real and hypothetical willingness to pay were not estimated.[2] Besides adding new observations to the original data set, the discrete choice probit model allows for the inclusion of previously excluded studies (e.g. Cummings, Harrison, and Rutstrom 1995; Cummings, Elliot, Harrison, and Murphy, 1997; Haab, Huang, and Whitehead 1997).[3] These studies were excluded from List and Gallet (2001) because they did not estimate either willingness to pay (WTP) or willingness to accept (WTA). For example, this might be due to an insufficient variation in the bid (or payment) levels in a dichotomous choice study.

As noted, List and Gallet (2001) did not include any referendum-based studies in their analysis. This issue is important because the report of the National Oceanic and Atmospheric administration's blue-ribbon committee, known as the NOAA panel, (Arrow et al. 1993) and other prominent sources recommended the use of the referendum format. Recent work by Carson, Groves, and Machina (2000) echoes this sentiment. They argue that a public good referendum is potentially incentive compatible if a coercive payment is coupled with a binding plurality. For private goods, referendum formats are not incentive compatible because subjects are likely to vote "Yes" in order to expand their choice set (Carson, Groves, and Machina, 2000).

The question of incentive compatible public goods referenda remains an open one in that empirical test results have been mixed. Cummings, Elliot, and Harrison (1997), Taylor (1998), and, to an extent, Burton, Carson, Chilton and Hutchinson (2000b) find a significant disparity between hypothetical and real values elicited via referenda. Results from Vossler et al. (2000) were mixed. A statistically significant disparity (0.10 level) existed when the "Undecided" votes were excluded from the analysis. When "Undecided" were counted as a "No" the disparity was no longer significant. A third model, with the "undecided" votes excluded, presented a calibration factor but did not test for the disparity. Haab, Huang, and Whitehead (1999), Taylor et al. (2001), and Vossler and Kerkvliet (2002) failed to find a significant disparity between

---

[1] We follow the assumption that actual values are real economic commitments and unbiased.

[2] In the original data set, studies by Boyce et al. (1992), Brookshire and Coursey (1987), McClelland et al. (1993), Dickie et al. (1987), and Navrud (1992) were excluded from the discrete choice model because they did not test for the presence of a disparity.

[3] Table 1 found in List and Gallet (2001) presents the studies comprising their data set. It should be noted that the published version omits observations from studies by Irwin et al. (1992), Kealy et al. (1990), and Kealy et al. (1988). This was confirmed via personal correspondence with Dr. Gallet on October 16th, 2002. Additional studies not cited in the text include: Giraud et al., (2001); McMillan et al.,2002; Botehlo and Costa Pinto, (2002); Murphy et al., (2003); Champ and Bishop, (2001); Poe et al. (2002); Bhatia and Fox-Rushby, (2003); Paradiso and Trisorio, (2001); Aadland and Caplan, (2003); Brown et al. (2001); Bennet, (1987); Camacho-Cuena, (2003); Carlsson and Martinsson, (2000); Getzner, (2000).

hypothetical and real reported values. The two most notable explanations for the divergence of findings involve estimation techniques and the nature of the survey or experimental setting.

Recent efforts have attempted to use "certainty corrections" and "cheap talk" scripts to improve upon the credibility of hypothetical values. Certainty correction typically involves a follow up question asking a respondent just how certain they are, on some categorical or numerical scale, about their previous answer to a value elicitation question. The data can then be recoded so that only certain or relatively certain responses (e.g., Yes votes on a referendum) are used in the respective study. For example, Paradiso and Trisorio (2001), Blumenschein et al. (1998), Blumenschein et al. (2001), and Johannesson et al. (1999) each used a certainty correction mechanism in order to identify how hypothetical WTP responses changed when subjects were asked to consider the level of confidence they had in the value they had provided. In all these cases, certainty correction reduced the disparity between hypothetical and real stated values to a point where it was no longer significant.

While the findings have been mixed, cheap talk scripts have become more prevalent in recent years. List (2001) and Cummings and Taylor (1999) provide two of the most notable attempts to "talk down" hypothetical statements of value. Although Cummings and Taylor (1999) did have success in bringing down hypothetical value statements, List (2001) found that consumer experience may render such designs moot. Regardless of the findings it seems appropriate to account for such designs in the analysis.

3. Empirical Model and Results

Using calibration factors as the dependent variable, we follow the same reduced form model presented by List and Gallet (2001).[4] In order to maintain consistency the new studies were coded in the same fashion as those presented in the original analysis. The original data set included 58 reported calibrations with three alternative constructs (minimum, median, and maximum calibration factors) for a total of 174 observations. The expanded data set used here includes 75 calibration factors (a 29% increase) with the same three alternative constructs for a total of 225 observations. The baseline scenario for the expanded data set remains the same as that used in the original study.[5] For completeness, a probit analysis is used to model the probability of observing a statistically significant disparity between real and hypothetical values.[6] The probit data set contains 85 observations, with 12 referendum observations (all for public goods) included. Besides the 11 variables employed in List and Gallet (2001), three new

---

[4] In matrix notation the List and Gallet (2001) model is: $CF = X'\beta + u$

where x1=1 if lab, 0 otherwise, x2=1 if WTP, 0 otherwise, x3=1 if private good, 0 otherwise x4=1 if within group comparison, 0 otherwise, x5=1 if open ended, 0 otherwise, x6=1 if first price sealed bid, 0 otherwise, x7=1 if provision point mechanism, 0 otherwise, x8=1 if smith auction, 0 otherwise, x9=1 if random Nth price auction, 0 otherwise, x10=1 if BDM, 0 otherwise, x11=1 if dichotomous choice, 0 otherwise.

[5] The original List and Gallet (2001, 8) baseline consisted of a between group comparison, WTA study, public good, and Vickery auction.

[6] The maximum likelihood estimator is: $\max \beta = \prod_{i=1}^{T} [F(X'B)]^{y_i} [1 - F(X'\beta)]^{1-y_i}$

Where X is a vector of explanatory factors, β is the associated vector of coefficients, and y_i is a binary indicator variable for a statistically significant disparity (1=Yes, 0=No disparity).

variables accounting for certainty correction, cheap talk, and referendum protocols were included in both the extended calibration and probit models, with a one representing "yes" and zero representing "no".

A major concern in conducting meta-analyses pertains to the most appropriate way to build such models. Specifically, when making inferences from composite data researchers must clearly define the purpose of the analysis and strive for consistency in method (Smith, 2002). Although the question here is clear, the diverse set of studies included in the analysis presented a number of difficulties. Most importantly, the common presence of multiple observations from single studies implies that disproportionate weight may be assigned to such observations, or that individual observations may not be independent. Thus, we explore the use of weighting and clustering approaches in our econometric estimation. Most simply, in the weighting correction (available on STATA version 8.0), each observation is assigned a probability weight dependent upon the number of observations taken from a study. Hence, if four observations were taken from a study the calibration (disparity) factor was assigned a weight of one/fourth. As an alternative to probability weighting, the clustering correction, (available on STATA version 8.0) drops the assumption of independence between observations contained within a particular study, while maintaining this assumption across studies (Rogers, 1993).[7] This assumption is appropriate provided that the observed calibration factors are a function of the study design from which they came. When clustering observations, group error terms are summed and then used in the calculation of sample variance (Rogers, 1993). This preserves the information available across studies while, at the same time, eliminating potential bias that may be attributed to a single study with a large number of calculable calibration factors.

Another concern, raised by an anonymous reviewer, pertains to the inclusion of results from induced value experiments. Specifically there are several recent induced value referendum studies that make comparisons between hypothetical and actual data (Taylor et al., 2001; Burton et al., 2000b). These studies clearly place themselves within the debate over the validity of stated preferences. However, it may be argued that empirically they represent significant departures from the rest of the studies comparing hypothetical and real valuation responses. Thus, we present results only from models excluding observations from these studies.[8]

Results from the models examining the expanded calibration factor data set and descriptive statistics of the data are presented in table 1; in all cases use of the clustering

[7] It has been suggested that these types of meta-analyses should also control for potential individual author effects. As such, a variety of model specifications were run using dummy variables to account for observations from studies by List, Bishop, and Johanneson (arbitrarily examining the three authors with the most observations in the data set, although not necessarily with the same sets of co-authors) While we find no evidence of significant author effects, the probability weighting and clustering techniques would appear to be preferred approaches.

[8] Each of the models (calibration factor and disparity) was also run including the observations from these induced value referendum studies. For the models using the calibration factor as the dependent variable only observations from Burton et al. (2000b) were included since Taylor et al. (2001) did not report real and hypothetical WTP estimates. In the probit model, for the disparity, observations from both Burton et al. (2000) and Taylor et al. (2001) were included. In no case did the inclusion of the observations from the induced value studies significantly alter the findings (e.g., the sign and significance of the referendum variable).

correction (alone), provided the best fit.[9] The first column of the table provides the means and standard erros for each of the explanatory variables. The mean calibration factors for each construct are 2.93 (minimum), 3.13 (median), and 3.34 (maximum). Interpretation of the coefficients remains the same as that of List and Gallet (2001). Independent variables with negative (positive) coefficients signify calibration factors that are lower (higher) than those obtained from a comparable alternative protocol (List and Gallet, 2001). An F-test indicates that each of the overall models is significant (0.05 level), and R^2 values range from 0.35 to 0.42.

For the calibration factor model using the maximum construct the coefficient estimate for the first price variable is negative and significant (0.01 level). This finding is consistent with that of the original List and Gallet (2001) analysis. List and Gallet (2001) find that studies valuing private goods tended to have a smaller gap between real and hypothetical stated values.[10] Our findings provide no evidence that the use of private goods will significantly reduce the disparity between real and hypothetical values. Coefficient estimates on the referendum and certainty correction variables are negative and significant (0.05 level). Indicating that calibration factors obtained from public goods referendum studies are lower than those obtained from non-referendum public goods studies. The same hold true for studies employing certainty correction.

For the calibration factor model using the median calibration values the coefficient estimate on the first price auction variable is negative and significant (0.01 level). Again, this finding is consistent with that of List and Gallet (2001). The coefficient estimate on the random price auction variable is negative and significant (0.10 level). The coefficient estimates on the referendum and certainty correction variables are both negative and significant (0.05 level). Again, in contrast to List and Gallet (2001), the estimated coefficient on the private good variable is not significant at any reasonable level of confidence.

When the minimum calibration factor values are used, the coefficient on the first price auction variable is negative and significant (0.01 level). The coefficient estimates on the random price auction and referendum variables are also negative and significant (0.05 level). Lastly, the coefficient estimates on the dichotomous choice, certainty correction, and cheap talk variables are all negative and significant (0.10 levels). In comparison, List and Gallet (2001) found that first price sealed bids were less prone to hypothetical bias (as we do). However, our findings also differ from the original List and Gallet (2001) results in that the private goods and WTP variables are no longer significant at any reasonable level of confidence.

In summary, when compared to the original List and Gallet (2001) results our findings produce some notable similarities and differences. In each model (maximum, median, and minimum) the coefficient estimate on the first price sealed bid variable was negative and significant. This is consistent with the List and Gallet (2001) analysis. For the three new variables we included, results indicate that the use of public goods referenda and certainty corrections will reduce the disparity between hypothetical and real values. Although cheap talk scripts may also be used to serve this purpose, they were shown to significantly reduce the disparity in only one case (minimum calibration values).

While List and Gallet (2001) found that the use of private goods significantly reduced the disparity between hypothetical and real values (and presented this as one of their key findings), we find no evidence for this in our data. To be clear, we have increased the size of the data set by nearly one-third, use a clustering technique in estimation (which drops the assumption of

[9] Due to space limitations, data tables are not presented here, but are available upon request.

[10] For convenience we are only presenting differences relating to individual tests of significance.

independence between observations), and include three new variables to account for certainty corrections, referenda and cheap talk. However, the absence of any significant effect on the private goods variable was stable across a wide variety of specifications, and whether we used the clustering technique or the simple probability weighting alternative, or neither.

For completeness, a simple probit probability model examining the absence or presence of a significant disparity was estimated. Rather than using a calibration factor, the dependent variable is the presence or absence of a significant disparity between real and hypothetical stated values. Once again, both a clustering correction and probability weights were explored. Table 2 presents the descriptive statistics and probit estimates, using the clustering correction alone, which provided the best fit. The mean of the dependent variable is 0.671, which means that 57 of 85 observations indicated the presence of a statistically significant disparity. The coefficient estimate on the certainty correction variable is negative and significant (0.05 level). All other explanatory variables were found to be insignificant. Based on the results of a Wald test, the probit model is significant overall, but in terms of goodness of fit has a modest McFadden's R^2 of 0.14.

In terms of using a certainty correction to ameliorate the disparity between hypothetical and real values, the probit model results suggest that such methods will significantly reduce the probability of finding a statistically significant disparity. The marginal effect of this variable, when evaluated at the means of the other variables, is shown to be quite large at –0.48, or lowering the probability of observing a disparity by 48 percent.

4. Discussion and Conclusions

Benefit-cost analyses are required for major federal regulatory actions, and are increasing used at the state and regional level. The use of survey-based, contingent valuation studies to provide stated-value estimates for changes in the provision of non-market goods remains common. It has also been the focus of a lively ongoing debate over the degree of "hypothetical bias" in stated values. Using the rapid accumulation of comparison studies, List and Gallet (2001) took a necessary first step in identifying which experimental protocols influence the disparity between real and hypothetical values in stated value studies. As noted by List and Gallet (2001), the original data set did not contain enough observations for some of the variables. Further, the chosen dependent variable (calibration factor) used in their analysis excludes many studies from consideration. This paper builds upon the efforts of List and Gallet (2001) by expanding the original data set, replicating their general approach using the calibration factor, and adding a probit model, which examines the simple absence or presence of a statistically significant disparity. We also add three new variables to the analysis to account for referendum formats, certainty corrections, and cheap talk scripts.

Our results support the List and Gallet (2001) finding that the use of first price sealed bid auctions will reduce the disparity between hypothetical and real values. We find no evidence to support the previous finding that private goods (relative to public goods) will reduce the disparity between hypothetical and real values. Estimates from the calibration factor models indicate that referendum formats and certainty corrections will reduce the degree of observed disparity. Estimates from the probit model show that use of certainty corrections will reduce the probability of observing a statistically significant disparity between real and hypothetical reported values (and the marginal effect of doing so can be large). It should be noted that the various certainty corrections present in the literature are typically *ad hoc* corrections on Yes responses in simple

dichotomous choice or referendum formats to only allow for highly certain or relatively certain Yes responses. As such, by design they reduce the expected WTP in a hypothetical setting and thus the degree of upward bias. There is as yet no accepted theoretical model for such corrections, or even consensus in the literature on how much correction is necessary.

In closing, as with any meta-analysis, our purpose was to provide a more accurate picture of an accumulating body of statistical evidence than might be inferred from any single study. We believe that some important new results emerge for consideration. However, it is also clear that there is still much we don't know about the magnitude, extent and determinants of hypothetical bias. Like the original Gallet and List (2001) findings, we close with the caveat that collective understanding is likely to continue to evolve as additional comparison studies are completed.

References

Aadland, D., and Caplan, A. J. (2003) "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias." *American Journal of Agricultural Economics* **85**, 492-501 .

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R., and Schuman, H. (1993) "Report of the NOAA Panel on Contingent Valuation." *Federal Register* **58**, 4601-14.

Bennet, J.W. 1987. "Strategic Behavior: Some Experimental Evidence." *Journal of Public Economics* **32**, 355-68.

Bhatia, M.R. and Fox-Rushby, J.A. (2003) "Validity of Willingness to Pay: Hypothetical versus Actual Payment." *Applied Economics Letters* **10**, 737-40

Blumenschein, K., Johannesson, M., Blomquist, G.C., Liljas, B., and O'Conor, R. (1998) "Hypothetical Versus Real Payments in Vickery Auctions." *Economics Letters* **56**, 177-80.

Blumenschein, K., Johannesson, M., Blomquist, G.C., Liljas, B., and O'Conor, R. (2001) "Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation." *Southern Economic Journal* **65**,169-77.

Blumenschein, K., Johannesson, M., Yokoyama, K.K., and Freeman, P.R. (2001) Hypothetical Versus Real Willingness to Pay in the Health Care Sector: Results from a Field Experiment. *Journal of Health Economics* **20**, 441-57.

Botelho, A. and Costa Pinto, L. (2002) "Hypothetical, Real, and Predicted Real Willingness to Pay in Open-Ended Surveys: Experimental Results." *Applied Economics Letters* **9**, 993-996.

Brown, T.C., Ajzen, I., and Hrubes, D. (2001) "Further Tests of Entreaties to Avoid Hypothetical Bias in Referendum Contingent Valuation." *Journal of Environmental Economics and Management **46***, 353-61.

Burton, A.C., Carson, K.S., Chilton, S.M., and Hutchinson, W.G. (2002a) "Testing Estimation Methods on Experimental Referendum Data." *Manuscript* United States Air Force Academy.

Burton, A.C., Carson, K.S., Chilton, S.M., and Hutchinson, W.G. (2002b) "Divergent Behavior in Hypothetical Mechanisms: An Experimental Inquiry." *Manuscript* United States Air Force Academy.

Camacho-Cuena, E., Garcia-Gallego, A., Georgantzis, N., Sabater-Grande, G. "An Experimental Validation of Hypothetical WTP for a Recyclable Product." *Environmental and Resource Economics **00***, 1-23.

Carlsson, F., and Martinsson, P, (2000) "Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? Application to the Valuation of the Environment." *Journal of Environmental Economics and Management* **41**, 179-92.

Carson, R. T., Groves, T., Machina, M. T. (2000) "Incentive and Informational Properties of Preference Questions." *Manuscript,* University of California, San Diego.

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

Cummings, R. G., Harrison, G. W., and Rutstrom, E. (1995) "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible." *American Economic Review* **85**, 260-66.

Cummings, R. G., Elliott, S., Harrison, G. W., and Murphy, J. (1997) "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy* **105**, 609-21.

Cummings, R. G., and Taylor, L.O. (1999) "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *The American Economic Revie **89***, 649-664.

Getzner, M. (2000) "Hypothetical and Real Economic Commitments, and Social Status, in Valuing a Species Protection Program." *Journal of Environmental Planning and Management* **43**, 541-59.

Giraud, K. L., Loomis, J. B., and Cooper, J. C. (2001) "A Comparison of Willingness to Pay Estimation Techniques from Referendum Questions: Application to Endangered Fish." *Environmental and Resource Economics* **20**, 331-346.

Haab, T.C., Huang, J.C., Whitehead, J. C. (1999) "Are Hypothetical Referenda Incentive Compatible? A Comment." *Journal of Political Economy* **107**, 186-96.

Johannesson, M., Liljas, B., and O'Conor, R. M. (1997) "Hypothetical Versus Real Willingness to Pay: Some Experimental Results." *Applied Economics Letters* **4**, 149-51.

Johannesson, M., Blomquiest, G. C., Blumenschein, K., Johansson, P. O., Bengt, L., and O'Conor, R. (1999) "Calibrating Hypothetical Willingness to Pay Responses." *Journal of Risk and Uncertainty* **8**, 21-32.

List, J. A. and Gallet, C. A. (2001) "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis." *Environmental and Resource Economics* **20**, 241-54.

List, J. A. and Shogren J. F. (2002) "Calibration of Willingness to Accept." *Journal of Environmental Economics and Management* **43**, 219-33.

Macmillan, D.C., Smart, T.S., and Thorburn, A.P. (2002) "The Importance of Realism to Experiments Comparing Cash and CV Charitable Donations: The Case of the Isle of Eigg Trust." *Manuscript* University of Aberdeen, MacRobert.

Murphy, J. J., Stevens, T., and Weatherhead, D. (2003) "An Empirical Study of Hypothetical Bias in Voluntary Contribution Contingent Valuation: Does Cheap Talk Matter?" *Land Economics,* forthcoming.

Paradiso, M., and Trisorio, A. (2001) "The Effect of Knowledge on the Disparity Between Hypothetical and Real Willingness to Pay." *Applied Economics* **33**, 1359-64.

Poe, G.L., Clark, J.E., Rondeau, D., and Schulze, W. D. (2002) "Provision Point Mechanisms and Field Validity Tests of Contingent Valuation." *Environmental and Resource Economics* ***23***, 105-131.

Rogers, W.H. (1993) "Regression Standard Errors in Clustered Samples." *Stata Technical Bulletin* ***13***, 19-23.

Smith, V.K. and Pattanayak, S.K. (2002) "Is Meta-Analysis a Noah's Ark for Non-Market Valuation? *Environmental and Resource Economics* ***1-2,*** 271-296.

Taylor, L.O. (1998) "Incentive Compatible Referenda and the Valuation of Environmental Goods." *Agricultural and Resource Economics Review* **27**, 132-39.

Taylor, L.O., McKee, M., Laury, S.K., and Cummings, R.G. (2001) "Induced-Value Tests of the Referendum Voting Mechanism." *Economic Letters* ***71***, 61-65.

Vossler, C.A., Kerkvliet, J. (2002) "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***, 631-649.

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

TABLE 1

Models Using Additional Calibration Factors

Variable	Variable Means [Standard Errors]	Maximum (t-statistics)	Median (t-statistics)	Minimum (t-statistics)
Laboratory	0.60	0.0650	0.0847	0.0915
	[0.4932]	(0.28)	(0.35)	(0.34)
WTP	0.867	-0.7014	-0.7296	-0.8287
	[0.3422]	(-1.16)	(-1.22)	(-1.41)
Private Good	0.6933	-0.5449	-0.5563	-0.6196
	[0.4642]	(-1.39)	(-1.40)	(-1.46)
Within Group	0.3333	0.1210	0.0967	0.1159
	[0.4746]	(0.42)	(0.36)	(0.45)
Elicitation Method				
Open Ended	0.2133	0.0841	-0.0388	-0.1650
	[0.4124]	(0.20)	(-0.10)	(-0.51)
First Price Bid	0.04	-1.5721***	-1.6314***	-1.6714***
	[0.1155]	(-2.87)	(-2.99)	(-3.19)
Provision Point	0.067	.6597	0.2846	0.2651
	[0.2511]	(1.33)	(0.48)	(0.45)
Smith Auction	0.053	0.5648	0.5082	0.4544
	[0.2262]	(1.01)	(0.92)	(0.85)
Random Price Auction	0.04	-1.2345	-1.4030*	-1.6430**
	[0.1973]	(-1.62)	(-1.89)	(-2.20)
BDM	0.04	-0.1666	-0.1978	-.1387
	[0.1973]	(-0.57)	(-0.66)	(-0.44)
Dichotomous Choice	0.3067	-0.2722	-0.3672	-0.4558*
	[0.4642]	(-0.99)	(-1.30)	(-1.70)
Referendum	0.067	-0.8891**	-0.9680**	-0.9098**
	[0.2511]	(-2.33)	(-2.44)	(-2.32)
Certainty Correction	0.0534	-0.2711*	-0.6100**	-0.3326*
	[0.1621]	(-1.93)	(-2.38)	(-1.78)
Cheap Talk	0.0267	0.3093	0.6434	0.4150*
	[0.1622]	(0.90)	(1.62)	(1.93)
Constant	---	1.7604**	1.8310**	1.9344***
		(2.40)	(3.22)	(2.71)
Sample Size		75	75	75
R-squared		0.3520	0.3857	0.4203
F-Statistic		2.33**	2.69**	3.11**

Notes: * Significant at 0.10
** Significant at 0.05
*** Significant at 0.01

TABLE 2
Results from the Probability of Disparity Model

Variable	Variable Means [standard errors]	Model 4 Coefficients (t-statistics)	Marginal Effects (z-statistics)
Laboratory	0.5852	-0.2840	-0.1001
	[0.5186]	(-0.81)	(-0.81)
WTP	0.8706	-0.4002	-0.014
	[0.3376]	(-0.05)	(-0.06)
Private Good	0.5412	0.5448	0.1922
	[0.5013]	(1.32)	(1.33)
Within Group	0.4353	-0.2549	-0.0903
	[0.4987]	(-0.77)	(-0.78)
Elicitation Method			
Open Ended	0.2353	0.2312	0.0788
	[0.4267]	(0.50)	(0.51)
First Price Bid	0.0353	-1.2582	-0.4692
	(0.1856]	(-1.08)	(-1.30)
Provision Point	0.0582	0.0183	0.0064
	(0.2367]	(0.02)	(0.02)
Smith Auction	0.047	0.8026	0.2193
	[0.2130]	(0.86)	(1.23)
Random Price Auction		-0.0747	-0.0268
		(-0.08)	(-0.07)
BDM	dropped	dropped	dropped
Dichotomous Choice	0.4353	0.4196	0.1450
	[0.4987]	(1.17)	(1.17)
Certainty Correction	0.1176	-1.2805**	-0.4780**
	[0.3241]	(-2.41)	(-2.76)
Referendum	0.1412	0.1349	0.0463
	[0.3503]	(0.21)	(-0.21)
Cheap Talk	0.0471	0.0224	0.0079
	[0.2130]	(0.04)	(0.04)
Constant	---	0.4181	
		(0.49)	
Sample Size		85	
McFadden's R^2		.1386	
Wald Statistic		20.70*	

Notes: * Significant at 0.10.
** Significant at 0.05.
*** Significant at 0.01.
dropped= Predicted failure perfectly and the observations were dropped.