# Meta-Analysis of the Probability of Disparity between Actual and Hypothetical Valuation Responses: Extension and Preliminary New Results[1]

Joseph Little[2], Craig D. Broadbent[3] and Robert P. Berrens[4]

## Introduction and Background

How we think about, interact with, and value the open spaces and public places that characterize large swaths of the western United States and elsewhere help define us as a society. Likewise, consideration of those open spaces and public places also produces ongoing challenges for public lands management. Information about public preferences over access and preservation versus development issues on public lands, and the myriad ecosystem services they provide, can be important inputs to cost-benefit analyses, natural resource damage assessments and resource planning processes (Loomis, 2002; PCAST, 2011). As part of the battery of non-market valuation techniques developed by economists and others over the last 60 years to value changes in environmental goods and services, survey-based stated preference approaches, such as the contingent valuation (CV) method, can be highly flexible tools for collecting preference information. The available literature on CV and related approaches is extensive (Carson, 2012; Li and Ho, 2008), across both theoretical and applied domains (e.g., experimental design, survey collection and econometric issues).

For researchers and policy analysts, there are no shortages of interesting and important topics and applications for stated preference studies involving public lands management. These range from valuing access to a wide variety of outdoor recreation opportunities (e.g., see Loomis et al., 2008), including acceptance of recreational fees (e.g., Aadland et al., 2012), wilderness area preservation (see summary in Loomis, 2002), and applications of community forestry, and wildfire risk mitigation in the Wildland Urban Interface (WUI) (e.g., Loomis et al., 2011; Talberth et al., 2006;). Data from stated preference approaches, such as CV or Contingent Behavior (CB) studies, can also be combined with revealed preference information from trip-taking behavior (e.g., Grijalva et al. [2003] on rock-climbing trips), actual transactions (e.g., Little et al. [2006] examine elk hunting raffles on the Valles Caldera National Preserve), or with experimental laboratory results to help verify observed patterns (e.g, Talberth et al. [2006] examine wildfire risk mitigation behavior in the WUI).

Despite important and useful attempts to establish reference operating conditions (Cummings et al., 1986) and “blue-ribbon panel” guidelines (Arrow et al., 1993), there perhaps remains no single Method (with a capital “M”) to follow for applied CV studies. However, the applied researcher can access useful primers, reference volumes and manuals (e.g., Bateman et al., 2002; Boyle, 2003; Champ et al., 2003), participate in an open discourse community as it sorts through a kind of ever-evolving “local, provisional methodology” (LPM) (Randall, 1993), and strive to generally apply high quality social science methods (with a small “m”). Such an approach is consistent with arguments to avoid focusing on the results from single studies

---

[1] Address correspondence to Joseph Little, Department of Economics, 213C Bunnell Building, P.O. Box 756080, Fairbanks, AK 99775-6080. Email: jmlittle2@alaska.edu
[2] Assistant Professor, Department of Economics, University of Alaska-Fairbanks
[3] Assistant Professor, Department of Economics, Illinois Wesleyan University
[4] Professor, Department of Economics, University of New Mexico

(Randall, 1998) and instead trying to draw inferences from patterns across studies. Meta-analyses provide us with an important tool for doing that, and often are a good place to start for the applied researcher (e.g., see Loomis et al., 2008; and Richardson and Loomis, 2009).

In implementing CV and related approaches, the applied researcher often must focus on the proposed policy change or issue of interest, but should also be aware of persistent methodological concerns and emerging perspectives. Wherever possible, the applied researcher is encouraged to think carefully about experimental design (e.g., Rose et al., 2011), whether or not the respondent is likely to view the survey as consequential to a real outcome or decision (Carson and Groves, 2007), possible effects of alternative survey modes (e.g., Berrens et al., 2003; Champ, 2003; Loomis et al., 2011), possible follow-ups to the chosen valuation question (e.g., Champ et al., 2009) and building specific hypotheses and validity tests into their research.

Of particular note, concerns over upward, hypothetical bias in CV results are persistent and should be of interest to any applied researcher. Hypothetical bias is interpreted as the tendency for valuation responses in a survey setting to be different (typically larger) than in some actual setting involving real economic commitments (e.g., Arrow et al., 1993; Cummings et al., 1995; List et al., 2004). Loomis (2011) offers a useful recent review, including summarizing some early meta-analysis results, as well as adding additional insights. Further, viewing CV studies from a “consequentiality” perspective (Carson and Groves, 2007 and 2011) offers an important emerging perspective. Poe and Vossler (2011) provide an assessment of this perspective, which rejects simply bifurcating data into purely hypothetical versus actual, or viewing all stated preference surveys under the term “hypothetical.” Rather, hypothetical surveys can be differentiated between consequential and inconsequential. Consequential survey questions are seen by the respondent (or agent) as potentially influencing a policy outcome (i.e., there is a reason to take them seriously), and have incentive properties that can be theoretically evaluated. Assuming a survey has been designed so that the respondent will view it as potentially consequential to them (and see Herriges et al., 2010; Nepal et al., 2009), then it is possible to make theoretical predictions on truthful revelation of preferences based on the incentive structure of the elicitation format, type of good and all the information provided. This does not imply that all consequential surveys will be compatible with truthful revelation of preferences (Poe and Vossler, 2011). For example, following Carson and Groves (2007), it is expected that binding referenda for public goods will be incentive compatible, whereas the more common voluntary contribution dichotomous choice formats, or surveys involving possible introductions of private goods (e.g., Cummings et al., 1995) will not. To date, early tests of induced value public goods referenda appear to provide initial support for the consequentiality perspective or paradigm (Poe and Vossler, 2011; Taylor et al., 2001; and Vossler and McKee, 2006).

In the spirit of mapping performance characteristics across different study designs (Randall, 1998), this analysis attempts to improve our understanding of potential determinants of hypothetical bias in CV and other stated preference studies. Meta-analysis is used to investigate the determinants of the probability of observing a statistically significant disparity between hypothetical and actual valuation responses. Early results from induced value tests of the consequentiality perspective (Poe and Vossler, 2011) also help us re-think what we may be able to isolate in a meta-analysis of hypothetical versus actual comparisons. For example, understanding incentive-compatibility in any setting probably requires controlling for particular combinations of elicitation format and good type (public or private). Further, Carson and Groves, (2011) argue that inconsequential questions can easily be created in laboratory settings, but are much less likely to happen in field settings. Thus, it is important to control for general laboratory-type settings (with “homegrown values” brought to the experiment), versus explicit induced value experiments, or field settings, and possibly to further control for student versus non-student samples (Loomis, 2011).

## Data and Modeling Approach

There have been a number of prior studies that have used meta-analysis to investigate "calibration factors" of willingness to pay (WTP) or willingness to accept (WTA) values for the magnitude of hypothetical bias from various comparison studies (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005). This investigation extends a related analysis also presented in Little and Berrens (2004), which briefly analyzed the probability determinants of the presence of a significant hypothetical bias in comparison studies. Notably, since many studies only investigate dichotomous yes or no responses to a single payment amount for comparisons, and can not produce a calibration factor since WTP or WTA value estimates are not made, there are considerably more total comparisons that can be utilized. Here, we extend the analysis of Little and Berrens (2004) on the presence of hypothetical bias by increasing the original dataset from 85 observations from 53 studies, to 225 observations from 96 studies (220 usable observations).

Since List and Gallet (2001) published their first meta-analysis investigating hypothetical bias, the literature has continued to grow. In an effort to include as many new observations as possible, studies were added as long as they reported a test of significance related to hypothetical bias. Again, because we are using a probability of disparity model it was not necessary for the studies to report hypothetical and actual mean valuation amounts; rather, we only needed a test of significance to be reported in the study. Published studies were found by doing a search for 'hypothetical bias' using Econlit as the search vehicle. In addition, we also reviewed recent publications to find studies that may not have been indexed on Econlit. All studies were collected by the end of the 2011 calendar year. The data is available upon request, and the reference list of all studies used in this analysis can be found at: www.tech-teachers.net/craig.

The meta-analysis attempts to improve the understanding of observed disparities between actual and hypothetical stated values. Using 220 of the possible 225 observations, the dependent variable (DISPARITY) in a probit probability model is the presence (1) or absence (0) of a significant disparity between actual and stated values in a comparison study. Explanatory variables are chosen to allow preliminary investigation of arguments in recent theoretical (Carson and Groves, 2007, and 2011) and review (Loomis, 2011) studies; they include study and sample characteristics (laboratory, induced value, student/nonstudent samples, sample size, etc.), a set of dummy variables representing different pairings of elicitation format (open-ended, referenda, voluntary contribution dichotomous choice, etc.) the nature of the good (public or private), and use of attempted corrections methods (e.g., certainty corrections and cheap talk scripts).

The different pairings of elicitation formats and the nature of the good may be particularly important in beginning to explore hypotheses, such as incentive compatibility for truthful revelation of preferences in consequential CV designs (Carson and Groves, 2007). For example, in contrast to prior meta-analyses of hypothetical bias, which included a single binary variable to distinguish between private and public goods, our approach allows us to isolate studies, for example, that paired a referendum elicitation format with a public good (PUB-REF). We expect PUB-REF to have a negative impact on the probability of observing a disparity between actual and hypothetical responses relative to any of the cases that are not incentive compatible. The latter set would include: a public good with a generic dichotomous choice format (PUB-DC) which are essentially all voluntary contributions settings (but not always explicitly presented as such), or any private good case (where the respondent may be trying to influence the probability of provision) with either an open-ended format (PRIV-OE), or dichotomous choice format (PRIV-DC), or even the somewhat odd case of a referendum format (PRIV-REF).

Finally, any meta-analysis involves choices by the analyst between the lumping and splitting of observations in the creation of the explanatory variables. It would perhaps be ideal to follow prior meta-analyses exactly in the choice of explanatory variables for comparison purposes (Loomis, 2011). We do not follow that approach here for two reasons. First, as discussed above, we create pairings of elicitation format and good type in an attempt to be more sensitive to the emerging consequentiality paradigm (Carson and Groves, 2007 and 2011; Poe and Vossler, 2011). This allows us a much more nuanced look than early determinations about public and private goods, which were designated by a single dummy variable. For example, from meta-analysis results for calibration factors, List and Gallet (2001) found that private goods studies were subject to less hypothetical bias than public good studies. Second, prior meta-analyses (e.g., see List and Gallet, 2001) were dominated by the statistically significant effects of several dummy variables on elicitation formats (e.g., specific auction types or formats) that had only a small handful of observations (e.g., 3 or fewer) with that study attribute or characteristic. Examples include Smith auctions, Becker-DeGroot Marschak (BDM) mechanisms, and Random-$n^{th}$ price auctions. To avoid this problem, we somewhat arbitrarily only include variables with at least four percent of the total observations in our meta-data, with that attribute or study characteristic. Thus, in this preliminary analysis we lump all auction formats together. Certainly, not all auctions have the same demand revealing properties, and there is a need for further refinement as more comparison studies are completed. But, it does allow us to isolate or control the general auction group, and focus on particular comparisons of interest (e.g., PUB-REF against, PUB-DC or PRIV-DC). Following this rule (≥ 4%), we also lump a group of payment card and multiple category, and provision point mechanisms together in an "other" elicitation format grouping; however, given that only public goods (PUB-OTHER) matched with that lumping, there is no PRIV-OTHER variable (0 observations). Additionally, this preliminary rule also causes us to drop one of the possible pairings of elicitation format and good type – an auction format with a public good (PUB-AUCT), which only had 5 observations. This resulted in a set of 10 pairings we can use in our modeling.

## Results

Table 1 provides the variable definitions and their descriptive statistics out of 225 observations. Of note, the mean for the dependent variable, DISPARITY, is 0.6. That is, 60 percent of the comparisons in the literature have been stated by the authors to show a statistically significant disparity between hypothetical and actual valuation responses. Since the choice of statistical significance level (e.g., 0.05 or 0.10 level, etc.) varies across studies, in this preliminary analysis we let it vary by study, and define it here as whether or not (1 or 0, respectively) the author(s) state there is a disparity. Notably, of the different pairings of elicitation format and good type, the most common is PRIV-DC with 20 percent of the observations in the data.
Table 2 presents the results from two probit model specifications (both n= 220). Both estimated coefficients and marginal effects (showing the magnitude of the effect on the probability of observing a disparity) are presented. Model 1 is the baseline model specification, while Model 2 is an extended specification that includes several additional variables related to sample size, year of the study, and whether or not the study occurred after the "blue ribbon" panel report (Arrow et al., 1993), which greatly increased attention on hypothetical bias concerns and is captured in the dummy variable POST-NOAA. PRIV-DC provides the base case for the set of dummy variables on pairings of elicitation formats and good type, and is omitted from the specification. Note again that all observations (5) with PUB-AUCT are dropped from the analysis (and would get dropped from any regression due to collinearity since this variable would predict failure perfectly, with no variation [all 1's]).

**Table 1: Variable Definitions and Descriptive Statistics**

| Variable | Definition | Mean | Std. Err. |
|---|---|---|---|
| DISPARITY | =1 if disparity between hypothetical and real payments is stated by author, 0 if no disparity stated by author | 0.609 | 0.489 |
| LAB | =1 if experimental setting as stated by author is in a controlled laboratory, 0 otherwise | 0.516 | 0.499 |
| STUDENT | =1 if student population strictly used, 0 otherwise | 0.409 | 0.493 |
| INDUCED | =1 if an induced value experiment, 0 otherwise | 0.062 | 0.242 |
| WTP | =1 if WTP study, 0 if WTA study | 0.938 | 0.242 |
| W-GROUP | =1 if a within-group comparison, 0 if between group comparison | 0.253 | 0.436 |
| PRIV-DC | = 1 if a private good valued in a dichotomous choice format, 0 otherwise | 0.20 | 0.401 |
| PUB-DC | =1 if a public good valued in a dichotomous choice format, 0 otherwise | 0.138 | 0.345 |
| PRIV-REF | =1 if a private good valued in a referendum format, 0 otherwise | 0.044 | 0.207 |
| PUB-REF | =1 if a public good valued in a referendum format, 0 otherwise | 0.138 | 0.345 |
| PRIV-AUCT | =1 if a private good valued in any type of auction format, 0 otherwise | 0.169 | 0.375 |
| PUB-AUCT | =1 if a public good valued in any type of an auction format | 0.022 | 0.148 |
| PRIV-OE | =1 if a private good in an open ended format, 0 otherwise | 0.049 | 0.216 |
| PUB-OE | =1 if a public good in an open ended format, 0 otherwise | 0.062 | 0.242 |
| PUB-OTHER | =1 if a private good in any other elicitation format, 0 otherwise | 0.04 | 0.196 |
| PRIV-CE | =1 if a private good using a choice experiment format, 0 otherwise | 0.098 | 0.298 |
| PUB-CE | =1 if a public good using a choice experiment format, 0 otherwise | 0.071 | 0.258 |
| CERTAINTY | =1 if a certainty correction is utilized, 0 otherwise | 0.098 | 0.298 |
| CHEAP-TALK | =1 if a cheap talk script is utilized, 0 otherwise | 0.093 | 0.292 |
| POST-NOAA | =1 if study is after the NOAA Panel Report (Arrow et al. 1993), 0 otherwise | 0.844 | 0.363 |
| OBS | total number of observations in each study | 632.8 | 1036.2 |
| SAMPLE-SIZE | total number of participants in each study | 279.4 | 327.31 |
| YEAR | year study was published | 2002.0 | 7.63 |

**Table 2: Probability of Disparity Probit Models (n=220), Clustering Corrections**

| | Model 1: Baseline | | Model 2: Extended | |
|---|---|---|---|---|
| **Variable** | **Estimate** | **Marginal Effects** | **Estimate** | **Marginal Effects** |
| LAB | -0.162 | -0.062 | -0.351 | -0.135 |
| | (0.377)[a] | (0.145) | (0.389) | (0.148) |
| STUDENT | 0.472 | 0.179 | 0.627* | 0.235* |
| | (0.378) | (0.139) | (0.37) | (0.132) |
| INDUCED | -1.21** | -0.439** | -1.198** | -0.436** |
| | (0.553) | (0.152) | (0.592) | (0.164) |
| WTP | 0.534 | 0.211 | 0.50 | 0.198 |
| | (0.36) | (0.139) | (0.42) | (0.163) |
| W-GROUP | 0.293 | 0.111 | 0.309 | 0.116 |
| | (0.26) | (0.094) | (0.259) | (0.094) |
| PUB-DC | 0.168 | 0.637 | 0.142 | 0.054 |
| | (0.318) | (0.118) | (0.341) | (0.127) |
| PRIV-REF | -0.422 | -0.167 | -0.512 | -0.202 |
| | (0.458) | (0.18) | (0.506) | (0.196) |
| PUB-REF | -0.807** | -0.313** | -0.863** | -0.333** |
| | (0.362) | (0.132) | (0.395) | (0.141) |
| PRIV-AUCT | -0.34 | -0.134 | -0.309 | -0.122 |
| | (0.334) | (0.132) | (0.36) | (0.143) |
| PRIV-OE | 0.06 | 0.023 | 0.055 | 0.021 |
| | (0.484) | (0.184) | (0.51) | (0.194) |
| PUB-OE | -0.885* | -0.339** | -0.928* | -0.353** |
| | (0.499) | (0.169) | (0.486) | (0.161) |
| PUB-OTHER | -1.22** | -0.44** | -1.351** | -0.473** |
| | (0.457) | (0.122) | (0.474) | (0.113) |
| PRIV-CE | -0.539 | -0.212 | -0.119 | -0.047 |
| | (0.474) | (0.183) | (0.623) | (0.245) |
| PUB-CE | -0.591 | -0.232 | -0.469 | -0.185 |
| | (0.436) | (0.167) | (0.521) | (0.204) |
| CERTAINTY | -2.662** | -0.68** | -2.764** | -0.689** |
| | (0.562) | (0.053) | (0.644) | (0.053) |
| CHEAP-TALK | -1.183** | -0.435** | -1.266** | -0.46** |
| | (0.323) | (0.096) | (0.339) | (0.095) |
| POST-NOAA | | | 0.55 | 0.216 |
| | | | (0.524) | (0.203) |
| OBS | | | -0.0002 | -0.00007 |
| | | | (0.0002) | (0.00007) |
| SAMPLE-SIZE | | | 0.0002 | 0.0006 |
| | | | (0.0005) | (0.0002) |

| *Table 2 Continued* | | | | |
|---|---|---|---|---|
| **Variable** | **Estimate** | **Marginal Effects** | **Estimate** | **Marginal Effects** |
| YEAR | | | -0.011 | -0.004 |
| | | | (0.0253) | (0.009) |
| CONSTANT | 0.385 | | 21.177 | |
| | (0.333) | | (50.238) | |
| **Pseudo $R^2$** | 0.256 | | 0.2668 | |
| **LR Chi Squared Statistic** | 52.26** | | 60.33** | |

Table 2 notes: * Significant at the 0.10 level; ** Significant at the 0.05 level; [a] values in parentheses are robust standard errors.

Further, the presence of multiple observations from individual studies could potentially bias estimated standard errors. To address this concern, the probit probability models (1 and 2) presented in Table 2 correct for clustering bias and are estimated using robust standard errors. The clustering correction relaxes the assumption of independence between observations drawn from the same study while maintaining the assumption of independence for observations across studies.

Pseudo $R^2$ measures range from 0.256 in (Model 1) to 0.267 in (Model 2). Adding the four extra study/sample characteristics variables (POST-NOAA, OBS, SAMPLESIZE and YEAR) provides no significant change in explaining overall variation, and none of the estimated coefficients on these individual variables are statistically significant.

Across both model specifications (baseline and extended) the estimated coefficients on the variables INDUCED, CERTAINTY and CHEAP-TALK are negative and significant at the 0.05 level. INDUCED represents a particular experimental laboratory setting where values are induced to participants. Certainty corrections (CERTAINTY) use responses to follow-up (un)certainly level questions and various re-coding schemes on original valuation responses (e.g., to convert relatively uncertain yes responses in a dichotomous choice or referendum format to no responses). Cheap talk designs (CHEAP-TALK) use a variety of stylized scripts inserted prior to valuation questions in a survey to discourage potential hypothetical bias. All three of these variables significantly reduce the probability of observing a disparity between actual and hypothetical responses. In contrast, the estimated coefficient on the variable STUDENT is positive and significant (increasing the probability of observing a disparity), but only in the extended model specification at the 0.10 level.

Across both model specifications (baseline and extended), the estimated coefficients on the indicator variables for three different pairs of elicitation format-good type (PUB-REF, PUB-OE, and PUB-OTHER) are negative and statistically significant at either the 0.05 level (PUB-REF and PUB-OTHER) or 0.10 level (PUB-OE). None of the other estimated coefficients on pairings are significantly different from the base case of PRIV-DC. Of particular interest, relative to the prominent case of PRIV-DC (which would not be incentive compatible for truthful revelation of preferences), the PUB-REF case (which would be incentive compatible with truthful revelation of preferences) would reduce the probability of observing a disparity between actual and stated values. This is also clearly seen in the raw data, where 29 out of 45 observations (65%) for PRIV-DC were 1's (significant disparity identified), versus only 13 of 31 observations (42%) for PUB-REF. Finally, with a range from -0.31 to -0.47 across the significant variables PUB-REF, PUB-OE and PUB-OTHER, the observed marginal effects (against a sample mean probability of observing a disparity of 0.60) are clearly large enough to effect a transposition from whether or not there is an expectation of observing a disparity.

## Discussion and Conclusions

This preliminary analysis extends a probability of disparity meta-analysis approach first applied to the CV and stated preference literature by Little and Berrens (2004). Specifically, this study modifies that initial analysis by using the much larger set of comparisons now available, and adding a variety of new explanatory variables, including a set of dummy variables that represent combinations of elicitation format and the nature of the good (private or public). Preliminary econometric analysis from a set of probit probability models, with clustering corrections and robust standard errors, provide some intriguing new results.

First, controlling for any possible effect of more generically described "Lab" conditions for the study, induced value experimental studies (versus what are commonly referred to as "homegrown preference" studies") are significantly less likely to observe a disparity in hypothetical versus real response comparisons.

Second, while somewhat mixed, there is at least preliminary evidence supporting the argument (e.g., Loomis, 2011) that student samples may increase the probability of observing hypothetical bias. Split-sample treatments of student versus non-student samples, while controlling for other factors and the pairing of elicitation format and good type, are probably called for in future hypothetical bias comparison studies.

Third, combinations or pairings of chosen elicitation format and the nature of the good (public or private) clearly matter in the likelihood of observing hypothetical bias. This is consistent with the general Carson and Groves (2007) theoretical framework for differentiating between consequential and inconsequential valuation questions, as different pairings will have different incentive compatibility properties. More specifically, the preliminary meta-analysis results appear to support the theoretical prediction that public goods referenda will minimize the probability of disparity between hypothetical and actual responses relative to the baseline case of a private good dichotomous choice format (e.g., voluntary contribution). More generally, the evidence indicates that comparisons of several different combinations of elicitation formats (REF and OE and OTHER) with public goods are likely to reduce the probability of observing hypothetical bias against this baseline case. In the initial probability of disparity results, with a much smaller sample, Little and Berrens (2004) found no difference between public and private goods, using only the broad dummy variable indicator. (The initial List and Gallet [2001] calibration factor results found that private goods were less prone to hypothetical bias; but, this was not found in Little and Berrens [2004] calibration factor results). The probability of disparity results found here suggest that public and private goods are different, when paired with different elicitation formats. There is no evidence here that private goods are less prone to observing hypothetical bias. This raises doubts over any argument that results from studies with private goods (e.g., Cummings et al., 1995) should be used to make inferences involving public goods, even if the elicitation formats are the same. Certainly, using, say, student samples, in an experimental lab setting involving "homegrown values," when combined with a non-incentive compatible elicitation format and good type (e.g., PRIV-DC), is highly likely to generate "hypothetical bias."

Fourth, while often largely ad hoc in nature, commonly used (un)certainty corrections and cheap talk scripts are both estimated to significantly reduce the probability of observing a disparity in hypothetical versus real responses comparisons. Marginal effects estimates suggest that certainty corrections, in comparison to cheap talk scripts, have a larger incremental impact on reducing the probability of observing a disparity. But, there is still much that we do not understand about both of these types of corrections, and they both have different variations with apparently differential impacts (Carson and Groves, 2011; and Champ et al., 2009). Thus, it appears important to increase the number of split-sample comparisons of different types of each of these approaches in future hypothetical versus real comparisons.

Fifth, the statistical evidence does not support the argument that choice experiments (CE) are somehow less prone to hypothetical bias than more typical CV approaches and formats, as has been hypothesized by some (e.g. Arrow et al., 1993; Hanley et al., 2001).

Sixth, the statistical evidence does not support the argument that WTP studies are less prone to hypothetical bias than WTA studies (e.g., Arrow et al., 1993; List and Gallet, 2001).

In closing, these results are presented as preliminary. We hope that they help the applied researcher access an evolving literature, and where possible spur additional comparison studies of actual and hypothetical valuation responses. Some of the cells in the experimental design of this meta-analysis should be expanded with additional observations, and there could be refinements in the explanatory variables (e.g., altering some of the "lumping and splitting" choices, and adding interaction terms). Further, some interesting treatments in comparison studies, such as varying the degree of social isolation or context (e.g., List et al., 2004; Mozumder and Berrens, 2011) or controlling in some way the real or perceived consequentiality of the responses (e.g., Broadbent et al., 2010; Landry and List, 2007) are still too rare to create explanatory variables in a meta-analysis. Planned future research includes continuing to collect additional studies, and estimating both probability of disparity models, and updated calibration factor models, which make use of pairings of elicitation formats and the type of good. This could include further breakdowns than simply private versus public (e.g., quasi-private and quasi-public goods). Ideally, robust meta-data may allow rigorous hypothesis testing of the pattern of evidence of emerging theoretical frameworks (e.g., Carson and Groves, 2007 and 2011), which can never be fully tested in individual studies. While that remains to be seen, the emerging consequentiality perspective or paradigm (Poe and Vossler, 2011) may be shifting any "local, provisional methodology" (Randall, 1993) around CV and related approaches. Understanding patterns in stated preference survey results "is more complicated than previously thought" (Carson and Groves, 2011, p. 307). As further tests and comparisons emerge, improved meta-analyses will need to be rich enough to map these performance characteristics (e.g., controlling for factors that affect consequentiality, and known incentive compatibility properties).

## References


Aadland, D., B. Anatchkova, B. Simon, B. Grandjean, J. Shogren and P. Taylor. 2012.

"Valuing Access to US Public Lands: A Pricing Experiment to Inform Federal Policy." *Social Science Quarterly* 93(1): 248-269.

Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 15 January: 1402-1416.

Bateman, I., R. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee,G.Loomes, S. Mourato, E. Ozdemiroglu, D. Pearce, R. Sugden and J. Swanson. 2002. *Economic Valuation with Stated Preference Techniques: A Manual*. Cheltenham, UK: Edward Elgar.

Berrens, R., A. Bohara, H. Jenkins-Smith, C. Silva and D. Weimer. 2003. "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples." *Political Analysis*, 11(1): 1-23.

Boyle, K., 2003. "Contingent Valuation in Practice." Pp. 111-170. In: P. Champ, K. Boyle and T. Brown (Eds.) *A Primer on Nonmarket Valuation*. Boston, US: Kluwer Academic Publishers.

Broadbent, C., J. Grandy and R. Berrens. 2010. "Testing for Hypothetical Bias in a Choice Experiment Using a Local Public Good: Riparian Forest Restoration." *International Journal of Ecological Economics and Statistics* 19(F10): 1-19,

Carson, R. 2012. *Contingent Valuation: A Comprehensive Bibliography and History*. Cheltenham, UK: Edward Elgar Publishing.

Carson, R. and T. Groves. 2007. "Incentive and Informational Properties of PreferenceQuestions." *Environmental and Resource Economics* 37: 181-210.

Carson, R. and T. Groves. 2011. "Incentive and Informational Properties of PreferenceQuestions: Commentary and Extensions." Pp. 300-321. In: J. Bennett (Ed.) *The International Handbook on Non-Market Environmental Valuation*. Cheltenham, UK: Edward Elgar Publishing.

Champ, P. 2003. "Collecting Survey Data for Nonmarket Valuation." Pp. 59-98. In: P. Champ, K. Boyle and T. Brown (Eds.) *A Primer on Nonmarket Valuation*. Boston, US: Kluwer Academic Publishers.

Champ, P., K. Boyle and T. Brown (Eds.) 2003. *A Primer on Nonmarket Valuation*. Boston, MA: Kluwer Academic Publishers.

Champ, P., R. Moore, and R. Bishop. 2009. "A Comparison of Approaches to Mitigate Hypothetical Bias." *Agricultural and Resources Economics Review* 38(2): 166-180.

Cummings, R., D. Brookshire and W. Schulze (Eds.) 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman and Allanheld.

Cummings, R., G. Harrison, and E. Rutstrom. 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible? *American Economic Review* 85: 260-266.

Grijalva, T., R. Berrens, A. Bohara and W.D. Shaw. 2002. "Testing the Validity of Contingent Behavior Trip Responses." *American Journal of Agricultural Economics,* 84(2): 401-414.

Hanley, N., S. Mourato and R. Wright. 2001. "Choice Modeling Approaches: A Superior Alternative for Environmental Valuation?" *Journal of Economic Surveys* 15(3): 435-462.

Herriges, J., C. Kling, C. Liu and J. Tobias. 2010. "What are the Consequences of Consequentiality?" *Journal of Environmental Economics and Management* 59: 67-81.

Landry, C., and J. List. 2007. "Using Ex Ante Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field." *American Journal of Agricultural Economics* 89(2): 420-429.

Li, Z., and Y. Ho. 2008. "Use of Citation per Publication as an Indicator to Evaluate Contingent Valuation Research." *Scientometrics* 75(2): 97-110.

List, J., R. Berrens, A. Bohara, J. Kervliet. 2004. "Social Isolation and Stated Preferences." *American Economic Review* 94(3):7 41-752.

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

Little, J., K. Grimsrud, P. Champ, and R. Berrens. 2006. "Investigation of Stated and Revealed Preferences for an Elk Hunting Raffle." *Land Economics* 82(4): 623- 640.

Loomis, J., 2002 (2nd Ed.) *Integrated Public Lands Management: Principles and Applications to National Forests, Parks, Wildlife Refuges and BLM Lands*. New York, NY: Columbia University Press.

Loomis, J., 2011. "What's To Know About Hypothetical Bias in Stated Preference Valuation Studies?" *Journal of Economic Surveys* 25(2): 363-370.

Loomis, J., A. Gonzalez-Caban and J. Champ. 2011. "Testing the Robustness of Contingent Valuation Estimates of WTP to Survey Mode and Treatment of Protest Responses." Pp. 102-121. In: J. Bennett (Ed.) *The International Handbook on Non-Market Environmental Valuation*. Cheltenham, UK: Edward Elgar Publishing.

Loomis, J., T. Kroeger, L. Richardson and F. Casey. 2008. "A Benefit Transfer Toolkit for Fish, Wildlife, Wetlands and Open Space." *Western Economics Forum* Fall: 33-43.

Mozumder, P. and R. Berrens. 2011. "Social Context, Financial Stakes and Hypothetical Bias: An Induced Value Referendum Experiment." *Applied Economics* 43(29): 4487-4499.

Murphy, J.J., P.G. Allen, T.H. Stevens, and D. Weatherhead. 2005. "A Meta-analysis of Hypothetical Bias in Stated Preference Surveys." *Environmental and Resource Economics* 30: 313-325.

Nepal, M., R. Berrens, and A. Bohara. 2009. "Assessing Perceived Consequentiality: Evidence from a Contingent Valuation Survey on Global Climate Change." *International Journal of Ecological Economics and Statistics* 14(P09):1 4-29.

Poe, G., and C. Vossler. 2011. "Consequentiality and Contingent Values: An Emerging Paradigm." Pp. 122-141. In: J. Bennett (Ed.) *The International Handbook on Non-Market Environmental Valuation*. Cheltenham, UK: Edward Elgar Publishing.

President's Council of Advisors on Science and Technology (PCAST). 2011. Report to the President: Sustaining Environmental Capital: Protecting Society and the Economy. J. Holdren and E. Lander, Co-Chairs. Executive Office of the President, Washington DC, 20502.

Randall, A., 1993. "What Practicing Agricultural Economists Really Need to Know About Methodology." *American Journal of Agricultural Economics* 75: 48-59.

Randall, A., 1998. "Beyond the Crucial Experiment: Mapping the Performance Characteristics of Contingent Valuation." *Resource and Energy Economics* 20(2): 196-207.

Richardson, L., and J. Loomis, 2009. "The Total Economic Value of Threatened and Endangered Species: An Updated Meta-Analysis." *Ecological Economics* 68(5): 1535-1547.

Rose, J., S. Bain and M. Bliemer. 2011. "Experimental Design Strategies for Stated Preference Studies." Pp. 273-299. In: J. Bennett (Ed.) *The International Handbook on Non-Market Environmental Valuation*. Cheltenham, UK: Edward Elgar Publishing.

Talberth, J., R. Berrens, M. McKee, M. Jones. 2006. "Averting and Insurance Decisions in the Wildland Urban Interface: Implications of Survey and Experimental Data for Wildfire Risk Policy." *Contemporary Economic Policy*, 24(2): 203-223.

Taylor, L., M. McKee, S. Laury, and R. Cummings. 2001. "Induced Value Tests of the Referendum Voting Mechanism." *Economics Letters* 71: 61-65.

Vossler, C., and M. McKee. 2006. "Induced Value Tests of Contingent Valuation Elicitation Mechanisms." *Environmental and Resource Economics* 35: 137-168.